THE WORLD BANK

Research Observer

INSIDE

Decon on Rural Poverty

Symposium on Evaluation

Ravallion on Evaluation and Practice

•

King and Behrman on Timing and Exposure in Evaluations

Symposium on Financial Sector

Claessens on Competition in the Financial Sector

•

Beck on Access to Financial Services

ISSN 0257-3032 (Print)
ISSN 1564-6971 (Online)
www.wbro.oxfordjournals.org

OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

THE WORLD BANK

Research Observer

The World Bank Research Observer is intended for anyone who has a professional interest in development. *Observer* articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the *Observer* considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the Editor. Please direct all editorial correspondence to the Editor, *The World Bank Research Observer*, 1818 H Street, NW, Washington, DC 20433, USA.

For more information, please visit the Web sites of the *Research Observer* at www.wbro.oxfordjournals.org, the World Bank at www.worldbank.org, and Oxford University Press at www.oxfordjournals.org.

THE WORLD BANK

Research Observer

Volume 24 • Number 1 • February 2009

Rural Poverty: Old Challenges in New Contexts
Stefan Dercon 1

Symposium on Evaluation

Evaluation in the Practice of Development
Martin Ravallion 29

Timing and Duration of Exposure in Evaluations of Social Programs
Elizabeth M. King and Jere R. Behrman 55

Symposium on Financial Sector

Competition in the Financial Sector: Overview of Competition Policies
Stijn Claessens 83

Access to Financial Services: Measurement, Impact, and Policies
Thorsten Beck, Asli Demirgüç-Kunt, and Patrick Honohan 119

Subscriptions
A subscription to *The World Bank Research Observer* (ISSN 0257-3032) comprises 2 issues. Prices include postage; for subscribers outside the Americas, issues are sent air freight.
Annual Subscription Rate (Volume 24, 2 issues, 2009)
Academic libraries
Print edition and site-wide online access: US$151/£101/€151
Print edition only: US$143/£95/€143
Site-wide online access only: US$143/£95/€143
Corporate
Print edition and site-wide online access: US$227/£151/€227
Print edition only: US$216/£144/€216
Site-wide online access only: US$216/£144/€216
Personal
Print edition and individual online access: US$52/£35/€52
Please note: US$ rate applies to US & Canada, Euros€ applies to Europe, UK£ applies to UK and Rest of World.
Readers with mailing addresses in non-OECD countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request by writing to the UK address below.
There may be other subscription rates available; for a complete listing, please visit www.wbro.oxfordjournals.org/subscriptions.
Full pre-payment in the correct currency is required for all orders. Payment should be in US dollars for orders being delivered to the USA or Canada; Euros for orders being delivered within Europe (excluding the UK); GBP sterling for orders being delivered elsewhere (i.e. not being delivered to USA, Canada, or Europe). All orders should be accompanied by full payment and sent to your nearest Oxford Journals office. Subscriptions are accepted for complete volumes only. Orders are regarded as firm, and payments are not refundable. Our prices include Standard Air as postage outside of the UK. Claims must be notified within four months of despatch/order date (whichever is later). Subscriptions in the EEC may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Subscribers in Canada, please add GST to the prices quoted. Personal rate subscriptions are only available if payment is made by personal cheque or credit card, delivery is to a private address, and is for personal use only
Back issues: The current year and two previous years' issues are available from Oxford University Press. Previous volumes can be obtained from the Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. E-mail: psc@periodicals.com. Tel: (518) 537-4700. Fax: (518) 537-5899.
Contact information: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnls.cust.serv@oxfordjournals.org. Tel: +44 (0)1865 353907. Fax: +44 (0)1865 353485.
In the Americas, please contact: Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oxfordjournals.org. Tel: (800) 852-7323 (toll-free in USA/Canada) or (919) 677-0977. Fax: (919) 677-1714. **In Japan, please contact:** Journals Customer Service Department, Oxford University Press, 4-5-10-8F Shiba, Minato-ku, Tokyo, 108-8386, Japan. E-mail: custserv.jp@oxfordjournals.org. Tel: +81 3 5444 5858. Fax: +81 3 3454 2929.
Postal information: *The World Bank Research Observer* (ISSN 0257-3032) is published twice a year, in February and August, by Oxford University Press for the International Bank for Reconstruction and Development/THE WORLD BANK. Postmaster: send address changes to *The World Bank Research Observer*, Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009. Periodicals postage paid at Cary, NC and at additional mailing offices. Communications regarding original articles and editorial management should be addressed to The Editor, *The World Bank Research Observer*, The World Bank, 1818 H Street, NW, Washington, D.C. 20433, USA.
Oxford Journals Environmental and Ethical Policies: Oxford Journals is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see http://www.oxfordjournals.org/ethicalpolicies.html for further information on Oxford Journals' environmental and ethical policies.
Digital Object Identifiers: For information on dois and to resolve them, please visit www.doi.org.
Permissions: For information on how to request permissions to reproduce articles or information from this journal, please visit www.oxfordjournals.org/jnls/permissions.
Advertising: Advertising, inserts, and artwork enquiries should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Tel: +44 (0)1865 354767; Fax: +44(0)1865 353774; E-mail: jnlsadvertising@oxfordjournals.org.

Paper used: *The World Bank Research Observer* is printed on acid-free paper that meets the minimum requirements of ANSI Standard Z39.48-1984 (Permanence of Paper).
Indexing and abstracting: *The World Bank Research Observer* is indexed and/or abstracted by ABI/INFORM, CAB Abstracts, *Current Contents/Social and Behavioral Sciences, Journal of Economic Literature/EconLit, PAIS International, RePEc (Research in Economic Papers), Social Services Citation Index*, and *Wilson Business Abstracts*.

Typeset by Techset Composition Limited, Chennai, India; Printed by Edwards Brothers Incorporated, USA

Rural Poverty: Old Challenges in New Contexts

Stefan Dercon

Poverty is still a predominantly rural phenomenon. However, the context of rural poverty has been changing across the world, with high growth in some economies and stagnation in others. Furthermore, increased openness in many economies has affected the specific role of agricultural growth for rural poverty reduction. This paper revisits an 'old' question: how does growth and poverty reduction come about if most of the poor live in rural areas and are dependent on agriculture? What is the role of agricultural and rural development in this respect? Focusing on Sub-Saharan Africa, and using economic theory and the available evidence, the author comes to the conclusion that changing contexts has meant that agricultural growth is only crucial as an engine for growth in particular settings, more specifically in landlocked, resource-poor countries, which are often also characterized by relatively low potential for agriculture. However, extensive market failures in key factor markets and likely spatial effects give a remaining crucial role for rural development policies, including focusing on agriculture, to assist the inclusion of the rural poor in growth and development. How to overcome these market failures remains a key issue for further research. JEL codes: O41, Q10, O55

Even though poverty has been reducing in some parts of the world in recent years, not least in East Asia and China, and more recently in South Asia, its persistence in large parts of Africa and elsewhere keeps it high on the agenda. Poverty persistence is to a large extent related to a poor growth performance in national economies. Furthermore, in large parts of the world, poverty also remains in terms of sheer numbers a mainly rural and agricultural phenomenon, in that most of the poor depend on the rural sector for their livelihood.

In this paper I will revisit some well-rehearsed but nevertheless relevant questions. What is the place of rural development and agricultural growth in growth and poverty reduction? What are the main rural constraints on growth and

doi:10.1093/wbro/lkp003 Advance Access publication April 9, 2009 24:1–28

poverty? Has more recent economic theory and empirical research given more guidance as to what can be done? These issues will be discussed, keeping in mind some of the poorest parts of the world, most notably Sub-Saharan Africa. I will use some recent theoretical models, including building on Lewis (1954), as my guide to structure the discussion, including that of the empirical evidence, even though this is incomplete, to corroborate my conclusions.

These questions are well rehearsed and feature prominently in many general discussions on development. Textbooks typically engage with issues at various levels, for example Ray (1998) and Bardhan and Udry (1999). Longer treatments can be found in Timmer (2002) or de Janvry and others (2002). What is different in this paper is that I want to revisit some of the theory and evidence specifically with Sub-Saharan Africa in mind; as a result it is highly selective. Overall, the experience with agricultural growth, growth in general, and poverty reduction over the last decades has been disheartening in this region. It has provided the context for renewed calls for a strong focus on agriculture in Africa as a necessary condition for growth and poverty reduction. For example, Sachs has been most vocal in calling for a 'green revolution' in Africa as an essential part of its development strategy (Sachs 2005). More nuanced analyses, such as the recent *World Development Report 2008* on agriculture for development, emphasize the crucial role of agriculture as being "vital for stimulating growth in other parts of the economy" (World Bank 2007, p. xiii) in Sub-Saharan Africa, requiring strong productivity increases in smallholder agriculture.

In the context of the substantial heterogeneity in circumstances and opportunities across Sub-Saharan Africa, the next section will first briefly summarize the current evidence on the evolution of rural poverty, contrasting the African experience with other parts of the world. The evidence is consistent with a classic view that a move out of agriculture is correlated with overall poverty reduction. In the African context, neither a large scale poverty reduction nor a move out of the rural sector is occurring. Of course, this does not prove a causal link. Nor does it prove that a focus on agricultural growth is not warranted. I will ask how theory as well as the available evidence can inform us on what role agricultural and rural development has to play in both growth and poverty reduction in Africa.

In the next section I offer a discussion of the relevance of a macroview on the role of agriculture in growth and poverty reduction. This will require us to revisit some of the old and seemingly unfashionable questions related to sectoral and urban–rural linkages to understand better the role of the relevance of agricultural and rural development in the Africa setting. One of the key problems in this review is the relatively sparse evidence, so I rely largely on a simple but suggestive model of rural–urban linkages to make sense of this evidence. Combining this with recent approaches to the question of the scope for growth in Africa, emphasizing its heterogeneity in opportunities (Ndulu and others 2008), allows us to

identify those specific settings in which agricultural development is likely to be essential in stimulating growth and poverty reduction, as well as the nature and role of rural development in other settings. In particular, I will argue that the role of agriculture is likely to be very different in different settings, depending on whether a country can take advantage of manufacturing opportunities, whether it is dependent on others for its natural resources, or whether it is landlocked and with few natural resources of its own. I will argue that, especially in the latter case, a focus on agricultural growth may be an essential, if difficult, route out of poverty.

In the final section, the strong assumptions regarding the nature of markets implied in the macroview will be complemented by a microview which emphasizes different market failures and the possibility of poverty traps. I will focus on three examples of serious market failures—those related to credit, to risk, and to spatial effects—and I will review their theoretical impacts and the available evidence for them. These market failures, especially in circumstances where they may lead to poverty traps, bring rural and agricultural policies back to the fore, leading to suggestive policy conclusions on the role of agriculture and rural development.

Rural Poverty Patterns

Poverty is still predominantly a rural phenomenon. Pick a random poor person in the world and the odds are that this person will be living and working in the rural areas as a farmer or agricultural worker. Even though the data are not without problems, the most recent estimates suggest that about 76 percent of the poor in the world live in rural areas, well above the overall population share living in rural areas, which is 58 percent (Ravallion and others 2007).[1] Sub-Saharan Africa is no exception: while it has the highest poverty rate overall, rural poverty is about a quarter higher than urban poverty, with 65 percent of the population and 70 percent of the poor living in rural areas. At current patterns of growth, poverty reduction, and population growth, poverty is likely to remain a predominantly rural phenomenon for the next few decades (Ravallion and others 2007).

Is this changing? The figures in Ravallion and others (2007) also offer a suggestive insight into the recent patterns of urbanization of poverty over the period 1993 to 2002. Even though the urban poverty rate has marginally decreased in the world, the urban share of poverty has been increasing (from about 19 to 24 percent) as the urban population has grown faster than the rural population, largely due to in-migration.[2] There is considerable variation in this pattern as well. In Sub-Saharan Africa, a very marginal decrease in rural poverty and stagnating urban poverty rates, with a growing share of the urban population in the

Table 1. Growth of GDP and Agricultural GDP

	Growth GDP 1990–2000	*Growth GDP 2000–04*	*Growth Agricultural GDP 1990–2000*	*Growth GDP Agricultural 2000–04*
Low income	4.7	5.5	3.1	2.7
Middle income	3.8	4.7	2.0	3.4
East Asia & Pacific	8.5	8.1	3.4	3.4
South Asia	5.6	5.8	3.1	1.9
Sub-Saharan Africa	2.5	3.9	3.3	3.6

Source: World Development Indicators, World Bank.

total population, is responsible for a more substantial urbanization of poverty and little change in total poverty. The global overall urbanization of poverty in the context of relatively substantial poverty reduction, and with the larger part of the 'stock' of the poor living in rural areas, also implies that rural poverty has contributed most to overall declines in poverty: using a simple decomposition, Ravallion and others (2007) calculate that about 80 percent of aggregate poverty declines stem from rural poverty declines. But this obviously does not prove any causality between, say, urbanization and poverty, or indeed that what happens in the rural or agricultural economy is in itself the cause of poverty declines.

In particular, these patterns have to be seen in the broader context of the economy. 'Growth in GDP per capita and poverty reduction are well known to be coinciding with a gradual decrease of the share of GDP from agriculture as well as the share of the population engaged in agriculture. For example, across all low-income countries, growth stood at about 5 percent a year, with the share of agriculture declining from 32 to 23 percent of GDP between 1990 and 2004 (World Bank 2005b). Contrasting the experience in the 1990s of Sub-Saharan Africa with the relatively fast growing regions in East Asia and Pacific (including China) and South Asia, where substantial poverty reduction is taking place, is instructive (table 1). We observe growth of overall GDP far outpacing agricultural GDP growth in both East Asia and South Asia, but not so in Sub-Saharan Africa, where overall growth and growth in agricultural GDP were more similar at 2.5 and 3.3 percent a year respectively.[3] With population growth still at about 2.3 percent per year in Sub-Saharan Africa, much higher than in these other regions, per capita growth has been minimal. In other words, there is little sign of a structural transformation in Sub-Saharan Africa, despite signs of some urbanization of poverty, compared to the patterns observed in more successful regions.

For some, agricultural growth is seen as the key engine of growth (Timmer 2007). The higher recent growth rate in African agriculture is, then, a sign of hope, even if at current rates they can hardly be viewed as evidence of rapid agricultural transformation. In such a view, the key is to raise agricultural

productivity to allow growth to take off. A standard conceptualization of this process focuses on the existence of linkages in inputs and outputs between agriculture and other sectors (Johnston and Mellor 1961). Showing sympathy with this view, in the *World Development Report 2008* it is argued that, in currently agriculture-based economies, high growth in agriculture may well be the road to economic take off.[4] Even though the historical evidence on the role of agriculture in allowing the Industrial Revolution to start in Europe is suggestive, and that agricultural growth was also a significant element in the growth success of East Asia and China, it is much harder to argue that agricultural growth is *essential* to allow growth to take off, and the evidence for this is not clear-cut and is hard to come by.

Historians are still debating vigorously whether an agricultural revolution, in terms of a period of rapid increases in productivity during the 18th and 19th centuries, with its seeds earlier, was a key cause of the Industrial Revolution and subsequent growth, first in England and then elsewhere in Europe (Crafts 1985; Allen 1999). The timing of the growth in labor productivity is disputed, some dating it much earlier than commonly suggested (Allen 1999), while some even dispute that there was any significant productivity increase in agriculture in England in the period 1560 to 1850, so that its role as a precursor for overall growth increases is even more minimal (Clark 2002). Productivity increases in agriculture may have been a consequence of the Industrial Revolution, linked to competition over labor, rather than a precursor (Gantham 1989). Furthermore, recent comparisons of European agricultural productivity with levels in the Yangtze Delta in the 17th to early 19th centuries suggest that Chinese land and labor productivity in agriculture were close to the best performers in Europe in that period (England and the Low Countries), undermining the view that agricultural circumstances allowed industrialization to take off in these countries and not elsewhere. More recently, while the important role of policy changes leading to agricultural growth in the growth success of China is not disputed, this does not necessarily extend to other Asian success stories. For example, Korea appears not to have invested in agricultural productivity in the period leading up to its rapid industrialization (Amsden 1989).[5]

Growth linkages from agriculture in currently developing countries, including Africa, are commonly suggested to be considerable (for a review, see Staatz and Dembele 2007). While possible, the evidence is hard to compile and subject to considerable methodological problems. Simultaneity difficulties affect the econometric evidence from time series for particular countries, while panel data analysis produces ambiguous results (World Bank 2007). Most analysis depends on simulation models (such as that based on computed general equilibrium models or on input–output models), which inevitably have to impose strong and untested behavioral assumptions to derive results (Dorosh and Haggblade 2003).

There is stronger evidence in favor of the view that agricultural growth helps growth to be more pro-poor, albeit possibly dependent on context. For example, in China, it is estimated that growth from agricultural growth contributed up to four times more to poverty reduction than growth from industry or the service sector (Ravallion and Chen 2007). However, the rather favorable land distribution there may well have played a key role in this, a factor contributing to similar evidence in other East Asian settings, such as Vietnam. Evidence from India suggests a more subtle message: the impact of the growth in agriculture is matched by growth in services in terms of their poverty-reducing effects, although these effects, from non-farm growth, are larger in states with higher initial farm productivity (Ravallion and Datt 1996, 2002). Furthermore, evidence from Foster and Rosenzweig (2004) suggests that areas in India with the slowest growth in agricultural productivity had the largest growth in the rural non-farm tradable sector.

This discussion of patterns and change in rural poverty helps to motivate our further analysis on Africa. It is clear that poverty is highest in rural areas, but is that sufficient reason to focus on rural areas and agriculture? Successful poverty reduction is not simply equated with relatively high growth in agriculture. At best, during periods of rapid poverty reduction correlated with growth, rural growth is likely to be important for poverty reduction, but successful growth is associated with growth in the non-agricultural sector fast outpacing agricultural growth.

In any case, this implies that understanding rural poverty changes cannot naively focus only on what happens in the rural sector or in agriculture. Analysis will have to be done in the context of overall growth and changes, taking into account rural–urban linkages. Arguably, the first systematic treatment of these issues are in the Lewis model (Lewis 1954). This model is effectively part of a *theory of urban and rural interactions*, albeit with particular assumptions about the functioning of markets in urban settings and, especially, the nature of incentives and decisionmaking. Much work has refined this type of analysis: a shared underlying research question is how does growth and poverty reduction come about if much activity and labor is initially in the agricultural or rural sector?[6] These questions have somewhat been forgotten in much recent research.

However, the context of the rural sector has also rather dramatically changed in the developing world, when compared say to the 1970s and 1980s. In most of the developing world, not least in Sub-Saharan Africa, market-based reforms ("getting the prices right") has moved forward dramatically. In Africa, the context of farming and agricultural markets has substantially changed, and virtually everywhere important liberalization, mostly domestic but also increasingly of international trade, has been taking place. Globalization and increased openness, as well as investment in commercial agriculture and marketing, are changing the context slowly but in an irreversible way. High commodity prices, including for

cereals, provides new opportunities for agriculture. All these factors provide a different context to ask (again): Can and should agriculture be the engine for growth? Can and should it be an engine for poverty reduction in the context of growth?

A Role for Rural and Agricultural Growth in Growth and Poverty Reduction in Africa?

Much basic analysis of the importance of agricultural growth for poverty reduction is based on simplistic premises. For example, it tends to be stated that since the poor are employed in agriculture, agriculture must be the basis of poverty reduction efforts. Against this is one of the basic insights from most data on poverty: that a systematic increase in prosperity tends to be linked to having fewer people dependent on agriculture for their livelihoods. The key question in this respect is how to get them out of the agricultural sector in a sustainable way.

As it is one of the clearest analyses of this issue, I will build on the work by Eswaran and Kotwal (1993a, 1993b, 1994, 2002), which, though containing excellent economic theoretical insights, is bereft of even a single equation. The relevance of their work lies mainly in the clarity of the questions asked and answered; most of the points have been made by others, but rarely as crisply and convincingly. In this section, I will briefly summarize the key points they make and give a flavor of their analysis, which is done with India in mind. Then I will discuss how applicable these insights are for other parts of the world, most notably for Africa.

Eswaran and Kotwal's analysis can be thought of as a Lewis model within a proper general equilibrium framework. It also drops some of the most difficult assumptions underlying Lewis's original analysis, but which have remained present in many of the subsequent contributions. To put it bluntly, there is no assumption that rural labor markets work in a way equivalent to agricultural workers spending too much time sitting under trees (surplus labor that can be freely extracted). Furthermore, industrial workers do not choose to eat shirts during a critical phase of the growth process (that is, when total output in agriculture is going down, they are willing to shift to consuming less food and more manufacturing goods).[7]

Theoretical Framework

The Eswaran and Kotwal model assumes a two-sector economy, industry, and agriculture. There are two goods: shirts and food. Production in both sectors is

characterized by (different) constant returns to scale production technology, using labor as well as land in agriculture. There are landowners and workers in the rural economy, and workers in the industrial sector. A crucial assumption is that preferences are lexicographic: people will first need to have enough food before they will buy shirts. It captures an Engel effect that richer people will spend less on basic essentials, but, by making it more extreme, its relevance comes out more directly. An alternative way of looking at this is to state that there is no circumstance in which, for very poor people, lower prices for manufactured goods cannot induce them to cut back on essential basic commodities. Even though in reality it may not be as clear-cut, this view takes seriously that poverty is related to deprivation in essential food intake. As EW show, relaxing the assumptions does not fundamentally change the result, but it does make the dichotomy in implications under different processes of technological and other development marginally less striking. There is also an initial inequality in this economy. Some (the rich) have assets such as land; the poor only have labor. At first, the poor only eat food, since they do not have enough to satisfy their basic needs; once sated with food, they do not eat more. So there is a maximum level of spending on food, and a poor person is someone who only spends on food.

It is further assumed that there are clearing and integrated labor markets across rural and urban areas.[8] This means that people are indifferent as to whether they work in agriculture or industry: contrary to Lewis, these markets are not perfectly segmented, but integrated.[9] Clearing product markets are those where demand equals supply in each. All these assumptions about markets imply that poverty will go down if labor demand increases, resulting in an increase of real wages. In other words, real wage increases for the initially poor determine whether poverty declines. But how does this work? Using the assumptions, a generic model can be developed that can be used to contrast a number of alternative strategies to achieve this. Understanding the context and situation in which these different strategies are effective ways of reducing poverty will also help us to understand how important it is to focus on agricultural and rural development.

First, in a closed economy, the policy to be considered is (neutral) industrial progress via total factor productivity (TFP) growth. Under these assumptions, Eswaran and Kotwal show that this implies that more shirts are being produced for the same amount of labor. Prices for shirts will go down, but crucially the poor do not care for these cheaper shirts, since they still do not have enough food. The result is that there is no incentive for anyone to move out of agriculture, since total food supply would go down and demand for food would go up. In the end, the TFP growth only benefits the rich, who have enough food and are already consuming shirts, which they can consume more of. The marginal product of labor in industry goes up, but the price of shirts goes down. Employment is unchanged, and nominal wages, food prices, and therefore real

wages for the poor are the same as before. Poverty is simply unchanged, despite TFP growth in the industrial sector.

Next, consider again a closed economy and a policy of (neutral) technological progress in agriculture. More food is being produced for the same labor. This is obviously of interest to all the workers: there is more food for the same amount of work. Once there is more food consumed, some may cross the threshold and be sated with food, and become interested in buying shirts. The result is that shirt prices are being pushed up. This creates incentives for firms to expand production and attract more labor to deliver this increased production. Higher demand for labor will require increasing nominal wages. Rural wages will move up as well, while food prices will go down somewhat, due to the higher production and shift to shirts by some previously poor consumers. In equilibrium, labor will have moved out of agriculture into shirt production, and higher equilibrium real wages will imply a reduction in poverty.

The contrasting results are striking and lead to the conclusion that, in a closed economy, growth in agriculture may well be essential for poverty reduction, while industrial progress has no impact. The presence of demand linkages is the key factor, but, for poverty reduction, the relevant linkages are only via commodities consumed by the poor. Mellor has long emphasized this process as well (1999). But there is a difference here: the issue is not just growth linkages but also the link with poverty. Agriculture is then the central engine for poverty-reducing growth.

The results are nevertheless strongly affected by the assumptions about openness. In an open economy, the central demand and supply constraints do not matter anymore for traded commodities. Basic food staples can typically be imported, while shirts can be exported. Therefore, assume that both goods are tradable goods, so that only world prices matter.

We can now revisit both cases. First, consider the impact of industrial progress. More shirts are being produced for the same labor input, but prices of shirts remain the same, as world prices are not affected. Firms have an incentive to expand production, so the demand for labor and nominal wages will increase. Even though food supply goes down, in this case workers can move since food imports can go up. So the marginal product of labor goes up in agriculture as well, allowing rural and urban wages to increase in nominal and real terms. Food imports with higher real wages will mean that more food is being consumed and that some workers will start consuming shirts. The result is that poverty declines. Second, the impact of agricultural technological progress is now very similar to industrial progress. The demand linkages are not crucial anymore for the link between real wages and output, and real wages increase with more people buying shirts than before.

To put it simply, poverty reduction can then be achieved by any source of increased domestic competitiveness relative to the rest of the world, a

Relevance for Africa?

Are these results relevant for Sub-Saharan Africa? In the Eswaran and Kotwal model, if growth is driven by agricultural growth that is technologically neutral or labor intensive, then the scope for poverty reduction will be larger. Land is typically not highly unequally distributed in most African countries, where there is relatively low landlessness and where, in some countries, such as Ethiopia, there is a remarkably equal land distribution. Productivity increases will reward the poorer farmers as well, so that growth may have substantial poverty impacts. The changing circumstances in African agriculture, in terms of allowing many factor and product markets to work more freely, especially in terms of removing much of the urban bias, has improved the opportunities for this process to materialize. A "green revolution for Africa" may have substantial returns.

However, the *necessity* of agricultural growth to deliver both growth and poverty reduction is not so clear-cut. As in the model, opening up the economy has removed the crucial dependence on progress in agriculture for delivering poverty reduction as the crucial demand linkage with agriculture is removed: growth in other sectors, provided it is labor intensive, can similarly promote poverty reduction.12 To assess the case for the crucial role of agricultural growth, the growth opportunities for Africa have to be considered.

Recent work by Ndulu and others (2008) provides the foundation for a suggestive three-way description of the growth opportunities of Sub-African countries; a similar description is used in Collier (2007). These researchers make distinctions in terms of growth opportunities: first, there are resource-rich economies; second, there are coastal and other well-located countries; third, there are land-locked economies without natural resources. Each of these groups has very different problems at their core when trying to boost growth and to reduce poverty.

For resource-rich economies, the key issue is to manage wealth: how to translate the underlying wealth controlled by the nation into the basis for sustainable and shared prosperity. The key problems they tend to face are Dutch disease and governance problems, and they are more likely to be ravaged by violent conflict—think Nigeria, Angola, or Congo.

For coastal and other well-located countries the challenge is very different. They have no natural resources, and so no immediate source of wealth. Wealth needs to be generated. They have two production factors they can put to good use: they have people and they have their location to their advantage. Much of coastal Africa, not least Ghana, Côte d'Ivoire, Kenya, and South Africa, springs to mind. Their main challenge is how to take advantage of the opportunities offered by their location. They are countries that in principle should be able to take advantage of world trade opportunities, so their priorities are likely to have to include building up trade infrastructure, managing market institutions and

regulation, investing in skills, and supporting the formation of well-working labor markets. These are very different challenges, but globalization offers serious opportunities for them. Without working on their constraints, they are bound to be left behind; but the potential is there.

This leaves the landlocked, resource-poor economies without natural resources. They are suffering most from the agglomeration effects: they have little to offer, and they totally depend on their neighbors to overcome these effects. Matters are made worse if their better located or better endowed neighbors are mishandling their economies, or indeed if they are in conflict with these states. All these factors are creating further negative externalities. Examples are Burundi, Burkina Faso, and Ethiopia.

So when is agricultural growth essential? How does agricultural growth fit into this framework? First, take the resource-rich countries. Agriculture is unlikely to be an essential source of growth. Nevertheless, such an economy needs to find ways of diversifying and building up its productive capacity, and agriculture could play a role in this. In this context, the burden of agricultural growth to drive overall growth is not present, so that efforts for intensification or diversification can have a much more pro-poor bias. This could involve a focus on smallholder agriculture, for example by supporting new technologies and activities with higher labor productivity. But clearly, there are a variety of ways to encourage the distribution of the wealth of such a country, and it would be hard to argue that stimulating agricultural growth is essential. Moreover, investing in rural areas, including in basic services such as health, education, and infrastructure, could be an effective alternative form of redistribution, and it may have higher long-term returns in terms of transforming the economy than a narrow focus on agriculture. For example, there would be less of a burden to ensure that these investments were largely in high potential areas.

Second, consider the well-located economies, who are best placed to take advantage of world economic opportunities. Managing their comparative advantage, via labor markets, skills, regulation, and investment climate, is most essential. The role of agriculture is similar to the Eswaran and Kotwal model in an open economy: "industrial" progress is most likely the best route and a vehicle to take advantage of trade opportunities. The role of agriculture is then more subsidiary: it makes sense to encourage progress in agriculture as well, if only as a means of managing an exit from agriculture when trade-based growth takes off.[11] Skill creation via better health and education, also in the rural sector, is then most helpful as well, if only since it will facilitate the development of better skilled labor that can in due course be absorbed. The experience of Indonesia in the late 1970s and 1980s is rather reminiscent of this, with active rural policies but which ultimately led to more absorption in the urban economy. For African economies, the key challenge is to overcome their relative marginalization in the world

economy, as latecomers are put at a disadvantage when competing with countries that have already an established industrial base, such as many Asian economies. This may require specific support for manufacturing industrial development for this potential to materialize (Collier and Venables 2007), but this is still the best route for growth.

The landlocked, resource-poor countries are a rather different problem. In many cases—think of Ethiopia or Burkina Faso—the agricultural base is at best highly vulnerable. But their risk of total marginalization relative to the world economy is also highest. They are mainly dependent on the ability of their better located neighbors to pull them into trade-oriented opportunities, often involving migrant labor. In terms of active policies, the opportunities are limited: infrastructure and skill creation are sensible, but as locations for investment they are likely to remain down the pecking order for a long time, not least because many of their better located neighbors are still only barely integrating with the world economy at present. As a consequence, the best way to think of these economies is as if they were effectively closed economies, irrespective of active trade liberalization. As the model predicted, agricultural growth is then essential for both growth and poverty reduction—but don't expect any miracles.

Technological progress in agriculture has to be actively pursued, as well as other measures to raise rural productivities as the main way of delivering growth that also has clear poverty-reducing impacts. Rural and agricultural development is likely to be hardest here, as they are often agriculturally more vulnerable areas, but it is here where it has to be attempted with most vigor. In order to achieve poverty reduction, there are also likely to be important trade-offs between stimulating agriculture in the high potential areas (as likely to be required for growth) and promoting it in more marginal areas where poverty may be highest. With commodity prices high, as in recent years, this tension between growth and poverty is even more present, as promoting overall agricultural output becomes even more crucial.

The contrast between, say Ethiopia and India, is striking here. Ethiopia may have been trying to open up in recent years, but its neighbors that are better located have not been taking advantage of trade opportunities, and in general its relations with them are at best frosty (such as is the case with Somalia and Eritrea). This limits its options dramatically. At best it can create a basic infrastructure and create skills, but growth is likely to have to come via agriculture to encourage any systematic and persistent decline in rural poverty. Its population cannot take advantage of growth opportunities in neighboring countries for the purposes of trade or even of migration.

India has pockets and even states that face similar constraints on local natural resources, and other problems, as does Ethiopia (for example, Bihar). But its economy is broadly integrated, and with some states taking advantage of its

increased openness and location, growth externalities and employment opportunities provide options for these lagging states to at least take partial advantage of the overall change, with some beneficial impacts on local poverty reduction.

But even so, Ethiopia cannot expect miracles from a strategy of focusing on agriculture alone. Its relatively landlocked nature as well as its dependence on rainfall implies that some (unexpectedly) good harvests can push down temporarily food prices received by farmers to extremely low levels as export parity prices (the world price minus transactions costs to supply the world market from the farm) are bound to be low. In 2001/02 for example, after a brief period of yield increases and area expansion, good weather contributed to a bumper harvest for maize, but prices were pushed to levels at which farmers found it not profitable to harvest. Such events undermine agricultural transformation, making it imperative that demand growth takes place to avoid prices to collapse occasionally in such landlocked economies with high transactions costs.

Market Failures and Poverty Traps?

The previous section focused on a macro- or general equilibrium narrative on the role of agricultural and rural development in growth and poverty reduction. Changing contexts, not least that of increased openness of developing countries and the move toward more market-based economies, changes the role of agriculture and rural development in growth and poverty reduction, one reason being that urban–rural and sectoral interactions have to be properly integrated in the analysis. However, this analysis was conducted using the assumption of well-functioning factor and product markets. As much of the rural development research in recent decades has highlighted, market failures are prevalent, even despite the removal of many policy-induced market imperfections, such as in agricultural product markets.[12] Resources are then not allocated efficiently until marginal returns to different factors of production are equalized. Factor markets, including those for labor, serve different people differently, resulting in heterogeneity in the extent to which people with different initial circumstances can take advantage of opportunities, such as those offered by growth. Rural non-market institutions may have developed partially to substitute or compensate for market failures, but not completely. This has important implications, not just for the extent of growth that can be achieved, but most importantly, for our purposes, for the extent to which the poor can partake in growth. While the underlying taxonomy of general growth opportunities in different parts of Sub-Saharan Africa, which was developed in the previous section, is unlikely to be fundamentally affected, the ability of these growth narratives, to deliver poverty reduction and broad inclusion in the growth process, definitely is.

In this section, I will try to put some of the key lessons from the extensive microlevel research on rural households and institutions from the last few decades into the broader context of poverty reduction and growth. The key question is: What do we know about the key constraints of the rural poor to participate in or contribute to growth? Much of the recent academic literature on rural issues has explored many of these market failures in factor markets, such as land, labor, credit, and insurance. This typically forms the core of much of the teaching offered in graduate schools in the microeconomics of development.[13] However, the key issue for our purposes is to explore how microlevel issues can be put into the bigger picture of growth and poverty reduction.

Following this I will consider the question of what would cause some poor people to remain locked in rural poverty, even in those countries that manage to start growth processes that are intensive in terms of labor. Furthermore, for countries that are absolutely dependent on rural development efforts for poverty reduction via growth, such as landlocked, resource-poor countries with "poor" neighbors, the need to unlock rural potential means that we should be especially careful when examining what may cause growth in particular rural areas to lag behind. I will focus on three instances that illustrate a more general principle and finding in recent theoretical work, one suggested by empirical evidence: that initial poverty and market failures conspire to keep some poor people persistently poor or even in a poverty trap. In addition, I will focus on three problems linked to market failures which may induce the following processes: access to capital (credit market failure); risk (insurance market failure); and spatial externalities (the curse of geography).

Credit Market Failure and Poverty Traps

The most obviously observable market failure is that of credit markets to conform to the assumptions of perfectly competitive markets. Under perfect and complete markets, anyone with a profitable project should be able to get a loan at the current interest rate. If markets were perfect and efficient, no bank would ask for collateral to secure the loan.[14] In practice, without collateral, one typically would not get the loan. Collateral requirements can be understood as an important means by which credit markets handle the central problems that bedevil these markets: asymmetric information, such as moral hazard and adverse selection, and enforcement problems. Since imperfect information means that borrowers may not be able to know which projects are more risky among many risky projects, or whether lenders will implement other actions than initially committed to after the loans have been granted, collateral may be asked for to secure the loans. Collateral may also help to enforce the repayment of loans.

Starting from initial asset poverty for some, it is obvious that this may be a market failure that is particularly hurtful for the poor, by excluding them from profitable opportunities. A number of careful, suggestive models, most notably Eswaran and Kotwal (1986), show the key implications of this: the rich do not just earn more income because they have more assets, but they can also use assets more efficiently. Market failures force some of the poor to be inefficient, and they exacerbate any initial inequality—some of the poor may stay behind. There is much suggestive evidence that similar processes are common in agricultural settings, where they are often linked to credit market failures. A key prediction of this model is that the marginal return to bringing more land into production by the poor outweighs that of the rich, and that average output per hectare is larger for the poor than for the rich. This negative correlation between cultivated land area and output per hectare is commonly observed in developing countries. Binswanger and others (1995) provided a comprehensive overview of the evidence and looked into different explanations. Land quality heterogeneity is certainly part of the story, but factor market failures, including those related to credit, are likely to be relevant as well.

The model described is effectively a static model—but its potential dynamic implications are intuitively appealing. Starting from some inequality in assets, those with more wealth earn higher returns and plausibly can accumulate at a high rate, while the poor enter into technology or activity portfolios with lower returns and may not be able to start accumulating any wealth. This intuition is at the basis of a number of growth models leading to poverty traps for some and accumulation for others. Banerjee and Newman (1993) showed the adverse impact of asset inequality on growth, linked to credit market failures. When threshold levels of assets are needed to enter into different activities, then entry into profitable activities is closed off for those with limited assets, and they are trapped in poverty, while others can climb the occupational ladder. A poverty trap is an equilibrium outcome and a situation from which one cannot emerge without outside help, for example via a positive windfall to this group, such as by redistribution or aid, or via a fundamental change in the functioning of markets. Much other work suggests poverty traps and overall efficiency and growth losses, due to poverty and inequality combined with credit market failure, where some people are unable to exploit growth-promoting opportunities for investment (for example Galor and Zeira 1993; Bénabou 1996; Aghion and Bolton 1997). This model of credit market failure is also a central part of the narrative underlying the 2006 *World Development Report* (World Bank 2005a); it discusses also some suggestive evidence of its implications.

If poverty traps induced by the credit market are present in rural areas, or more generally, underinvestment is present, limiting growth. Furthermore, if growth picks up but access to profitable opportunities requires some minimal

investment (for example being able to invest in the migration of some household members, or sunk costs for a newly profitable activity), then credit market failures may result in the exclusion of some of the poor from the benefits of growth. Proving the existence of credit-induced poverty traps is difficult, but there is definitely evidence in rural Africa of entry costs (in particular activities) and assets (in the presence of limited access to investment capital) leading to poorer households holding less profitable portfolios (Dercon 1998; Barrett and others 2005).

What to do about them is less clear. Credit market interventions have long been a favored intervention, in recent times mainly via microfinance schemes, although other interventions may well be better and more useful for resolving credit market failures (Besley 1994). It is not altogether clear that the poorest are benefiting most from microfinance (Amendatriz de Aghion and Morduch 2005): more needs to be learned from careful evaluations of specific products that could have the highest impact (Karlan and Goldberg 2006). For Africa, the key question is also whether microcredit is not overrated as a means to fostering inclusion of the rural poor in the economy: often, if applied to rural settings, it is seen as a means of helping to get the rural poor into off-farm business activities, usually with restrictions on the eligible activities, via the support offered by the microfinance institution. Even though credit market restrictions may exclude them from certain profitable opportunities, it is unlikely to be the case that large scale poverty reduction is going to be achieved by making more and more people dependent on entrepreneurial activities. In most economies, the transformation toward a higher income and low poverty economy has been achieved by an increase in wage employment, via higher labor demand, with farmers becoming employees in industry and the service sector. It is likely that in the long run, the highest returns are to be obtained from health, education, and skills, not least in African economies that are resource rich or whose potential is in manufacturing exports. Of course, the transition is likely to be helped by allowing some people to take advantage of entrepreneurial opportunities in agriculture and the non-farm sector, but it is unlikely to be a large-scale successful route.

There is a dilemma here as well: as long as growth is not taking off, the poor are likely to be helped more by encouraging entrepreneurial activities, as labor demand is not picking up. In most African contexts, returns to education are convex: low for primary and, possibly, high for higher levels of education (Söderbom and others 2006), so that investing in education under the circumstances of the 1980s and 1990s, with low wage-labor demand, is unlikely to offer rapid routes out of poverty; though if growth were to pick up in a sustained way, this could change. Failing that, microcredit schemes could offer a solution for many of the poor, but without growth it does not offer real scope for large-scale poverty reduction. In any case, and at best, microcredit schemes with more flexibility, that do not try to tie people to particular entrepreneurial activities, but that

respond to the general financing needs of families, may be more effective (Karlan and Mullainathan 2007).

Insurance Market Failures and Risk-induced Poverty Traps

Another serious market failure impacting disproportionately on the poor is the lack of insurance and protection of the poor in the face of risk. The existence of complete insurance markets (or, to be technically more accurate, complete state-contingent markets) is another assumption for perfect markets that tends to be violated in practice. Problems with asymmetric information and enforcement issues, not dissimilar to those causing credit market failures, are again typically responsible for the limited spread of insurance mechanisms in developing countries. Even if they wanted to, the poor could not get any insurance for most of the risks they face.

Uninsured risk causes considerable hardship to the poor. Developing countries are still characterized by a high incidence of natural disasters, drought, conflict, and insecurity, as well as economic shocks, such as commodity price shocks and currency shocks. Health problems are widespread, as are pests in agriculture. It is commonplace to view these as "transitory" problems, requiring temporary solutions, such as some form of safety net, after which one should get back to the bigger issues of development. For the policymaker it also often means that it is just a social issue that should not distract the key (macroeconomic) policymakers from the bigger issue of how to stimulate growth in the economy and alleviate widespread poverty. However, this is misleading. There is increasing evidence that risk and shocks are a cause of lower growth, which results especially in the lower growth of the incomes of the poor and possibly in poverty traps. Focusing attention on the poor could then again be contributing to both growth and equity; in any case, it could be instrumental in ensuring that the poor can benefit from emerging growth.

Households in developing countries have developed sophisticated mechanisms to cope with risk. Typically, one could consider two types of responses: risk-management strategies and risk-coping strategies. Risk-management strategies involve trying to shape the risks faced by entering into activity portfolios that are more favorable in terms of risks. For example, entering into low risk activities or diversifying into portfolios of activities with differing risk profiles—growing more drought resistant crops, entering into petty trading or firewood collection, seasonal migration, and so on. Risk-coping strategies involve activities for coping with consequences of risk in income. Two types are commonly observed: self-insurance using savings, often in the form of cattle or small ruminants, to be sold off when the need arises; and informal mutual support mechanisms, where members of group or community provide transfers to each other in times of need, typically on a reciprocal basis (Fafchamps 1992).

These strategies are not without cost: income risk-management strategies result in a reduction in mean income and a variability in income, while adjusting asset portfolios to cope with risk typically involves investing in liquid assets with lower returns, rather than in productive illiquid investment. This affects their long-term income and their ability to move out of poverty. Indeed, there is growing evidence that these strategies imply substantial efficiency loss for the poor, which the rich—typically better protected via insurance, asset, and credit—do not have to endure (Dercon 2002). Morduch (1995) documented how more profitable technologies are not adopted because they are too risky in a particular setting in India. The same farmers have been found to hold livestock as a precaution against risk even when more productive investment opportunities exist (Rosenzweig and Wolpin 1993). Rosenzweig and Binswanger (1993) found that the loss in efficiency between the richest and poorest quintiles in their sample from India was more than 25 percent, attributable to portfolio adjustments in assets and activities due to risk exposure. In Ethiopia, Dercon and Christiaensen (2007) have found that modern input adoption is lowered due to the downside risk related to input loans, whose repayment is strictly enforced, even when rains and harvests fail. Over time, this results in substantial efficiency losses, which affect the poor disproportionately.

These risk-management strategies may trap the poor in poverty: to avoid further destitution, they are forced to forgo profitable but risky opportunities, and with it the opportunity to move out of poverty.[15] Even so, they cannot fully protect themselves: there is much evidence that although the strategies contribute to less variability in consumption and nutritional levels, they are still not able to cope with some serious, repeated shocks, not least those affecting whole communities, regions, or countries (Morduch 1999; Dercon 2002). These uninsured shocks typically wipe out assets, pushing the households down the asset distribution. They could be pushed below some critical threshold, trapping them into poverty from then on, for example due to the risk strategies they then need to follow to avoid further destitution, or due to other processes.

There is growing evidence that these processes are an important cause of poverty persistence and possibly permanent traps in developing countries. Jalan and Ravallion (2003) investigated the presence of poverty traps using data from China, and, although they did not find a pure poverty trap, they found that households took several years to recover from a single income shock, and that the recovery was much slower for the poor. There is also related evidence from Africa. Dercon (2004), using panel data from rural Ethiopia, found signs of poverty persistence linked to uninsured shocks, with the impact of rainfall for up to the previous four years affecting current growth rates, and the extent to which households had suffered in the famine of 1984–85 still an explanatory factor in growth rates in the 1990s. Furthermore, it took on average 10 years for livestock

holdings, a key form of savings in rural Ethiopia, to recover to the levels seen before the 1984–85 famine. In a careful study, Elbers and others (2007) use simulation-based econometric methods to calibrate a growth model that explicitly accounts for risk and risk responses applied to panel data from rural Zimbabwe. They found that risk substantially reduces growth, reducing the capital stock (in the steady state) by more than 40 percent. Two-thirds of this loss is due to *ex ante* strategies by which households try to minimize the impact of risk. Barrett (2005) has found suggestive evidence on poverty traps by looking at the livestock holdings of pastoralists in Kenya.

There is also increasing evidence of the long-term implications of uninsured shocks, focusing on health and education. For example, the permanent impact of drought on children is well documented—lower adult height, poor education outcomes, and therefore lower lifetime earnings. For example, the impact of drought and war in rural Zimbabwe in the early 1980s on a particularly vulnerable cohort of children was estimated at 7–12 percent of lifetime earnings or more (on this and other evidence, see Dercon and Hoddinott 2003).

All this evidence points to the important consequences of lack of insurance and protection in rural settings in developing countries, particularly as they affect the poor. Given that the root cause is again a market failure, exacerbated by poverty, there is a clear case for interventions that are potentially both poverty reducing and stimulating of efficiency and growth; in any case, they could ensure that the poor can more effectively take part in growth processes. In industrialized countries, not least in Europe, the failings of the insurance markets are largely resolved via some form of universal social insurance and substantial direct means-tested transfers. For developing countries, this is not likely to be cost-effective, involving high administrative costs and high informational requirements. To put it simply, the means for such systems are unlikely to be available.

Many responses can be considered, such as reducing the risks faced by rural households (for example by preventative health services, or better water management for agriculture), strengthening existing responses (such as investing in a better provision of savings products, or better functioning assets markets, such as that for live animals), and improving forms of insurance and broader social protection in the form of safety nets. While each have their problems and advantages, in recent years a number of specifically interesting initiatives have been taken with respect to insurance, even if their potential benefits are not quite substantiated yet.[16]

Spatial Effects

Another common cause for market failure is the presence of spatial externalities. Externalities are said to be present if economic or other interactions create social

gains or costs beyond those taken into account by those involved in the interaction. The standard example is environmental damage from production involving pollution not accounted for by the buyers and sellers of the commodity produced. A more general phenomenon in developing countries, that can be best understood in terms of externalities, involves geographically defined areas that appear to stay behind—poor neighborhoods, or even poor regions or countries. If one looks at the performance of the developing world, it has been striking over recent decades that some countries—largely in Africa—appear to have become increasingly marginalized, with low economic growth, persistent population growth, and, generally, persistent poverty. Less studied but at least as important is that even in countries where growth is high, there appears to be areas that systematically stay behind and do not benefit from the overall economic growth in terms of income growth and poverty reduction. Certain regions in China and India may well fit this bill. Much less documented but no less true, the geographical disparity in growth and poverty-reduction performance within African countries is similarly present.[17]

Such disparity may well be explained in terms of theories emphasizing agglomeration or location effects, predicting that firms will exploit increasing returns resulting from the presence of externalities to locating in the same geographical areas, implying that firms would locate in clusters (Fujita and others 1999). The corollary is that some less attractive locations may have missed the boat: not only would they not get the required investment, any capital present may well move out to capture the higher returns elsewhere. For those areas that missed the boat, there is a negative externality from the success of other areas. Clearly, this is a form of poverty trap: although initially these areas may not have been very different, once they missed the boat they can only escape by a serious exogenous shock or massive effort. They face a substantial threshold that they need to overcome to attract or retain capital for accumulation.

Other explanations similarly emphasize externalities related to the specific local context, for example low local endowments in terms of public goods, common property resources, and private asset holdings. If growth processes require a certain threshold of local endowments to take off, then poorly endowed areas may well find it hard to escape poverty. Proving this is difficult. There is some evidence for China and a few other countries.[18] For Africa, the role and effects of remoteness on growth and poverty in Africa are well discussed in Christiaensen and others (2005), with evidence from a number of countries.

Recall that these externalities are again market failures that specifically affect those at the lower end of the asset distribution—this time with assets broadly defined to include public and environmental goods. Given that poverty traps are identified, this empirical evidence would justify "poor areas" programs—massive investment programs in particular deprived areas to build up locational and community capital.

However, these empirical studies lack sufficient detail and a clear narrative about how these externalities come about. More evidence would be needed to guide and prioritize the type of interventions that would be most beneficial.

For example, most rural "poor areas" typically are characterized by remoteness—often linked to the lack of roads and communications infrastructure. One of the most common donor-policy responses is to build roads into poor areas. While undoubtedly bringing some benefits to remote communities, it is not necessarily the case that this is what is needed to unlock the growth potential of an area. In some countries, there is evidence that this may well be an appropriate response.[19] Still, historically much road building in developing countries has been in response to local economic growth or at least in recognition of some growth potential (such as cash crops or mining), and it was not the main cause of growth. Alternatives, such as irrigation, health, or educational schemes may be more important for unlocking their potential.

In any case, just doing a little is not going to help—in order for such areas to catch up, substantial levels of investment would be needed to lift them over the threshold. It may well be the case that creating opportunities for migration is a superior policy. This, however, also involves costs, and so may be difficult too. If these thresholds are substantial, then growth opportunities elsewhere may simply bypass many rural poor. But it is nevertheless the prime example of a potential rural poverty trap for which solutions have to be firmly considered in the context of rural–urban and other linkages, and a narrow focus on rural areas may be ineffective.

Conclusions

Contrary to most of the rest of the developing world, poverty levels have remained stagnant in Sub-Saharan Africa in the last few decades, in a context of slow growth. As most of the poor are living in rural areas, it may be concluded that agricultural growth and rural development policies have to be at the core of growth and poverty-reduction policies.

In this paper, I have used a "macro-" (intersectoral) and microperspective to discuss some of the key concerns and issues involved in an African context. Using a framework based on Eswaran and Kotwal (1993b), and evidence on growth opportunities in Africa based on Ndulu and others (2008), I conclude that agricultural growth is likely to be essential for landlocked and resource-poor economies in Africa, even though growth via agriculture is likely to be difficult. In other economies, in particular those with good locations for engagement into manufacturing exports, or resource-rich economies, agriculture is not the crucial constraint. Even if rural development policies are likely to be crucial for allowing

the gradual transformation of the economy, the pressure for agricultural growth to be an engine of growth is not present.

This analysis is however based on well-functioning factor and product markets, allowing specifically the poor to take advantage of opportunities irrespective of their endowments. Perfect markets in rural settings are not the appropriate assumption. There is considerable evidence that this implies that the poor may remain excluded of profitable opportunities to grow out of poverty, even if growth has picked up in the economy, potentially leading to poverty traps. Rural development policies targeted on the poor, and possibly including the stimulating of agricultural production by poor farmers, may then be part of the effort to make growth more inclusive. The mechanisms by which this exclusion may happen are well understood, and I have discussed credit and insurance market failures, as well as spatial effects in the form of "poor areas." The evidence on the appropriate responses is still only emerging, and clearly more experimentation and research is needed.

Notes

Stefan Dercon is Professor of Development Economics at Oxford University and affiliated to the European Development Network (EUDN), the Bureau of Economic Analysis and Research on Development (BREAD) and the Centre for Economic Policy Research (CEPR); email address: stefan.dercon@economics.ox.ac.uk. This paper was completed as part of research funded by the UK Department for International Development (DFID) in the context of a research program on improving institutions for pro-poor growth. An earlier version was presented at the World Bank's workshop entitled "Frontiers in Practice: Reducing Poverty through Better Diagnostics" in 2006. DFID and the World Bank are not responsible for the views expressed. I am grateful to Mark Koyama, Andrew Zeitlin, the editor, and three anonymous referees for very helpful comments on an earlier version of this paper.

1. This estimate is based on a poverty line of $1.08 a day in 1993 PPP, applied with a rural–urban correction based on cost-of-living differences between rural and urban areas.

2. As the definition of what constitutes an urban area is not standardized across the world's statistical offices, some caution is needed with these statements.

3. East Asia includes here the Pacific. Data from the World Development Indicators for 1990–2000 suggest 8.5 percent overall growth in East Asia and Pacific and 5.6 percent growth in South Asia, with respectively 3.4 and 3.1 percent growth in agricultural GDP. Growth rates post-2000 are not dissimilar, although overall growth in Sub-Saharan Africa appears to have picked up somewhat.

4. In the analysis of the World Development Report, the authors refer to agricultural-based economies as those with high shares of GDP and labor in agriculture. The geographical composition of this group in their report is such that 82 percent of the Sub-Saharan African rural population is in this group and that the share of Africans in the total population of this group is about 90 percent. In short, their analysis and prescriptions for agriculture-based countries are effectively about Africa.

5. The *World Development Report 2008* argues in this respect that rural investment did take place in Korea, but much earlier, in the first part of the 20th century.

6. Further 'classic' treatments are in Ranis and Fei (1964) and Jorgenson (1961).

7. For an excellent exposition of the Lewis model, see Ray (1998).

8. This assumption concerning perfect factor markets is crucial and will be challenged in the next section, qualifying some of the results in this section. Although offering a rather different

model from the one discussed here, Jorgensen (1961) can be seen as a predecessor, as he also brought in neoclassical market clearing assumptions and issues related to food consumption as a way of characterizing his 'dual economy' model. Ranis and Fei (1964) can be credited with offering some of the general equilibrium analysis presented in EW as well, including the impact of trade, although they used a dual economy setting rather than competitive factor markets.

9. This assumption can be questioned. However, as one purpose of this framework is to explain why in some contexts, it may not be crucial to focus on agriculture to get growth and poverty reduction, this assumption actually biases our arguments in favor of a focus on agriculture, as surplus labor implies that labor can be taken out of agriculture into the urban sector without any impact on total food supply.

10. The current high agricultural commodity prices across the world offer such an opportunity for agriculture, although the long-run level of incentives offered is harder to predict.

11. Of course, high-value agricultural activities, such as flowers, fruits, or vegetables in Kenya, are also effective means for taking advantage of locational and other advantages. Air transport is a (possibly increasingly) expensive means of transporting exports, making location sufficiently near to ports a continuing necessity for a trade-oriented growth model; which is not a straightforward option for landlocked countries such as Ethiopia.

12. This type of research is carefully discussed in detail in Bardhan and Udry (1999) and Mookherjee and Ray (2001).

13. Ray (1998) provides an excellent entry point to this literature.

14. In this discussion, I use the issue of collateral as a heuristic device to show the differential impact on the poor of credit market failures, without arguing that this is necessarily the key failure. Most other failures in credit markets can be shown to result in a specific disadvantage for poorer households. A useful discussion of different market failures in rural credit markets (and what to do about it) is in Besley (1994) and in Ray (1998).

15. See Banerjee (2003) for a formal poverty-trap model building on this idea.

16. For a helpful review of targeted transfers and safety nets, see Ravallion (2003). For a review of a broad set of means of offering "insurance" to the poor, see Dercon (2003) and the contributions therein. Of particular interest are initiatives to offer rainfall insurance, including in Africa, where even initial evaluations are yielding surprising results (Gine and Yang 2007).

17. For a review, see Kanbur and Venables (2005). Lack of convergence between rural and urban areas in a number of wealth indicators across 12 African countries is found in Sahn and Stifel (2003).

18. Jalan and Ravallion (2002) identified geographic poverty traps in rural China during the 1980s, finding that community characteristics affect the income growth performance of otherwise identical individuals, controlling for latent heterogeneity. These results show that in some areas living standards were falling while elsewhere *otherwise identical households* were enjoying rising living standards, an effect entirely due to externalities from the initial community characteristics.

19. Jalan and Ravallion (2002) presented evidence for China that roads are relevant for growth. Microdata from Ethiopia using a much smaller sample suggest similar effects, that is growth effects from levels of infrastructure (Dercon 2004).

References

Aghion, P., and P. Bolton. 1997. "A Trickle-Down Theory of Growth and Development with Debt Overhand." *Review of Economic Studies* 64(2):151–62.

Allen, R.C. 1999. "Tracking the Agricultural Revolution in England." *Economic History Review* 52(2):209–35.

Amsden, A. H. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.

Armendariz de Aghion, B., and J. Morduch. 2005. *The Economics of Microfinance*. Cambridge, Mass: MIT Press.

Banerjee, A. 2003. "The Two Poverties." In S. Dercon ed., *Insurance against Poverty*. Oxford: Oxford University Press.

Banerjee, A., and A. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101(2):274–98.

Bardhan, P., and C. Udry. 1999. *Development Microeconomics*. Oxford: Oxford University Press.

Barrett, C. 2005. "Rural Poverty Dynamics: Development Policy Implications." *Agricultural Economics* 32(1):45–60.

Barrett, C.B., M. Bezuneh, D.C. Clay, and T. Reardon. 2005. "Heterogeneous Constraints, Incentives and Income Diversification Strategies in Rural Africa." *Quarterly Journal of International Agriculture* 44(1):37–60.

Bénabou, R. 1996. "Inequality and Growth." In B. Bernanke, and J. Rotemberg eds., *National Bureau of Economic Research Macroeconomics Annual*. Cambridge: MIT Press, pp. 11–74.

Besley, T.. 1994. "How Do Market Failures Justify Intervention in Rural Credit Markets?" *World Bank Research Observer* 9(1):27–47.

Binswanger, H., K. Deininger, and G. Feder. 1995. "Power, Distortions, Revolt and Reform in Agricultural and Land Relations." In J. Behrman, and T.N. Srinivasan eds., *Handbook of Development Economics*, vol. 3. Amsterdam: North-Holland.

Christiaensen, L., L. Demery, and S. Paternostro. 2005. "Reforms, Remoteness and Risk in Africa: Understanding Inequality and Poverty During the 1990s." In R. Kanbur, and A.J. Venables eds., *Spatial Inequality and Development*. Oxford: Oxford University Press.

Clark, G. 2002. "The Agricultural Revolution and the Industrial Revolution, England 1500–1912." Mimeo, June, U.C. Davis Department of Economics.

Collier, P. 2007. *The Bottom Billion, Why the Poorest Countries are Failing and What Can Be Done About It*. New York: Oxford University Press.

Collier, P., and A.J. Venables. 2007. "Rethinking Trade Preferences: How Africa Can Diversify its Exports." *The World Economy* 30(8):1326–45.

Crafts, N. 1985. *British Economic Growth During the Industrial Revolution*. Oxford: Clarendon Press.

Datt, G., and M. Ravallion. 1998. "Why Have Some Indian States Done Better Than Others at Reducing Rural Poverty?" *Economica* 65(257):17–38.

———. 2002. "Is India's Economic Growth Leaving the Poor Behind?" *Journal of Economic Perspectives* 16(3):89–108.

Dercon, S. 1998. "Wealth, Risk and Activity Choice: Cattle in Western Tanzania." *Journal of Development Economics* 55(1):1–42.

———. 2002. "Income Risk, Coping Strategies and Safety Nets." *World Bank Research Observer* 17: 141–66.

———. 2003. *Insurance against Poverty*. Oxford: Oxford University Press and WIDER.

———. 2004. "Growth and Shocks: Evidence from Rural Ethiopia." *Journal of Development Economics* 74(2):309–29.

Dercon, S., and L. Christiaensen. 2007. "Consumption Risk, Technology Adoption and Poverty Traps: Evidence from Ethiopia." World Bank Policy Research Working Paper 4257.

Dercon, S., and J. Hoddinott. 2003. "Health, Shocks and Poverty Persistence." In S. Dercon ed., *Insurance against Poverty*. Oxford: Oxford University Press.

Dorosh, P., and S. Haggblade. 2003. "Growth Linkages, Price Effects and Income Distribution in Sub-Saharan Africa." *Journal of African Economies* 12(2):207–35.

Elbers, C., J.W. Gunning, and B. Kinsey. 2007. "Growth and Risk: Methodology and Micro Evidence." *The World Bank Economic Review* 21(1):1–20.

Eswaran, M., and A. Kotwal. 1986. "Access to Capital and Agrarian Production Organisation." *Economic Journal* 96:482–98.

———. 1993a. "Export Led Development: Primary vs. Industrial Exports." *Journal of Development Economics* 41, July:163–72.

———. 1993b. "A Theory of Real Wage Growth in LDCs." *Journal of Development Economics*(42, December:243–69.

———. 1994. *Why Poverty Persists in India*. New Delhi: Oxford University Press.

———. 2002. "The Role of Service Sector in the Process of Industrialization." *Journal of Development Economics*(68):401–20.

Fafchamps, M. 1992. "Solidarity Networks in Preindustrial Societies: Rational Peasants with a Moral Economy." *Economic Development and Cultural Change* 41 (October): 147–74.

Foster, A., and M.D. Rosenzweig. 2004. "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000." *Economic Development and Cultural Change* 52(3):509–42.

Fujita, M., P. Krugman, and A.J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, Mass.: MIT Press.

Galor, O., and J. Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Economic Studies* 60:35–52.

Gantham, G. 1989. "Agricultural Supply During the Industrial Revolution: French Evidence and European Implications." *The Journal of Economic History* 49(1):43–72.

Gine, X., and D. Yang. 2007. "Insurance, Credit, and Technology Adoption: Field Experimental Evidence from Malawi." World Bank Policy Research Working Paper 4425.

Jalan, J., and M. Ravallion. 2002. "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China." *Journal of Applied Econometrics* 17:329–46.

———. 2003. "Household Income Dynamics in Rural China." In S. Dercon ed., *Insurance against Poverty*. Oxford: Oxford University Press.

Janvry, A. de, E. Sadoulet, and R. Murgai. 2002. "Rural Development and Rural Policy." In B. GardnerG. Rausser (eds.), *Handbook of Agricultural Economics*, vol. 2, A, Amsterdam: North-Holland: 1593–658.

Johnston, B.E., and J.W. Mellor. 1961 "The Role of Agriculture in Economic Development." *American Economic Review* 51(4):566–93.

Jorgenson, D.W. 1961. "The Development of a Dual Economy." *Economic Journal* 71(282):309–34.

R. Kanbur, and A.J. Venables eds. 2005. *Spatial Inequality and Development*. Oxford: Oxford University Press.

Karlan, D., and N. Goldberg. 2006. "The Impact of Microfinance: A Review of Methodological Issues." Mimeo, Poverty Action Lab and Yale University.

Karlan, D., and S. Mullainathan. 2007. "Is Microfinance Too Rigid?" Mimeo, Poverty Action and Yale University.

Lewis, A.W. 1954. "Economic Development with Unlimited Supplies of Labor." *Manchester School* 139–91.

Mellor, J.. 1999. "Faster, More Equitable Growth: The Relation between Growth in Agriculture and Poverty Reduction." Agricultural Policy Development Project Research Report 4. Abt Associates Inc., Cambridge, Mass.

Mookherjee, D., and D. Ray. 2001. *Readings in the Theory of Economic Development*. Oxford: Blackwell.

Morduch, J. 1995. "Income Smoothing and Consumption Smoothing." *Journal of Economic Perspectives* 9 (Summer): 103–14.

———. 1999. "Between the State and the Market: Can Informal Insurance Patch the Safety Net?" *World Bank Research Observer* 14(2):187–207.

B.J. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier, and C.C. Soludo eds. 2008. *The Political Economy of Economic Growth in Africa, 1960–2000*. Cambridge: Cambridge University Press.

Ranis, G., and J.C.H. Fei. 1964. *Development of the Surplus Economy: Theory and Policy*. Homewood, Ill.: R.D. Irwin for the Economic Growth Center, Yale University.

Ravallion, M. 2003. "Targeted Transfers in Poor Countries: Revisiting the Trade-Offs and Policy Options." World Bank Policy Research Working Paper 3048.

Ravallion, M., and S. Chen. 2007. "China's (Uneven) Progress Against Poverty." *Journal of Development Economics* 82(1):1–42.

Ravallion, M., and G. Datt. 1996. "How Important to India's Poor is the Sectoral Composition of Economic Growth?" *World Bank Economic Review* 10:1–26.

———. 2002. "Why Has Economic Growth Been More Pro-poor in Some States of India than Others." *Journal of Development Economics* 68(2002):381–400.

Ravallion, M., S. Chen, and P. Sangraula. 2007. "New Evidence on the Urbanization of Global Poverty." World Bank Policy Research Paper 4199.

Ray, D. 1998. *Development Economics*. Princeton: Princeton University Press.

Rosenzweig, M., and H. Binswanger. 1993. "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments." *Economic Journal* 103:56–78.

Rosenzweig, M., and K. Wolpin. 1993. "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-income Countries: Investment in Bullocks in India." *Journal of Political Economy* 101(2):223–44.

Sachs, J. 2005. *The End of Poverty: Economic Possibilities for Our Time*. New York: Penguin Press.

Sahn, D., and D. Stifel. 2003. "Urban–Rural Inequality in Living Standards in Africa." *Journal of African Economies* 12(1):564–97.

Söderbom, M., F. Teal, A. Wambugu, and G. Kahyarara. 2006. "The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing." *Oxford Bulletin of Economics and Statistics* 68(3):261–88.

Staatz, J., and N.N. Dembele. 2007. "Agriculture for Development in Sub-Saharan Africa." Background paper for the World Development Report, World Bank.

Timmer, P.C. 2002. "Agriculture and Economic Development." In B. Gardner, and G. Rausser eds. *Handbook of Agricultural Economics*, vol. 2A, Amsterdam: North Holland.

———. 2007. "Agriculture and Economic Growth." In D.A. Clark ed., *The Elgar Companion to Development Studies*. Cheltenham: Edward Elgar.

World Bank. 2005a. *World Development Report 2006: Equity and Development*. Washington D.C.: Oxford University Press for the World Bank.

———. 2005b. *World Development Indicators*. Washington D.C.: Oxford University Press for the World Bank.

———. 2007. *World Development Report 2008: Agriculture for Development*. Washington D.C.: Oxford University Press for the World Bank.

Evaluation in the Practice of Development

Martin Ravallion

Standard methods of impact evaluation often leave significant gaps between what we know about development effectiveness and what we want to know—gaps that stem from distortions in the market for knowledge. The author discusses how evaluations might better address these knowledge gaps and so be more relevant to the needs of practitioners. It is argued that more attention needs to be given to identifying policy-relevant questions (including the case for intervention), that a broader approach should be taken to the problems of internal validity (including heterogeneity and spillover effects), and that the problems of external validity (including scaling up) merit more attention by researchers. JEL codes: H43, O22

Anyone who doubts the potential benefits to development practitioners from evaluation should study China's economic reforms. In 1978, the Communist Party's 11th Congress broke with its ideology-based view of policymaking in favor of a pragmatic approach, which Deng Xiaoping famously dubbed "feeling our way across the river." At its core was the idea that public action should be based on evaluations of experiences with different policies—"the intellectual approach of seeking truth from facts" (Du 2006, p. 2). In looking for facts, a high weight was put on demonstrable success in actual policy experiments on the ground. The first major application was to rural reform. While there had been much dissatisfaction with collectivized farming, there were competing ideas as to what needed to be done. The evidence from local experiments was eventually instrumental in persuading even the old guard of the Party's leadership (many of whom still favored collectivized farming) that household contracts could deliver higher food output. The evidence had to be credible. A new research group did field work studying the local experiments—though they were certainly not randomized experiments—in using contracts with individual farmers. The evidence might not be

doi:10.1093/wbro/lkp002 Advance Access publication March 25, 2009 24:29–53

conclusive by today's scientific standards, but it helped to convince skeptical policymakers (many still imbued in Maoist ideology) of the merits of scaling up the local initiatives (Luo 2007). The rural reforms that were then implemented nationally helped achieve what was probably the most dramatic reduction in the extent of poverty the world has yet seen.

Unfortunately we still have a long way to go before we will be able to say that this story from China is typical of development policymaking elsewhere. (And China still has much that it could do to enhance the credibility of its own efforts at evidence-based policymaking.) In this paper I argue that we underinvest in rigorous evaluations of development interventions *and* that the evaluations that are done currently are not as useful as they could be. Distortions in the "market for knowledge" about development effectiveness leave persistent gaps between what we know and what we want to know; and the learning process is often too weak to guide practice reliably. The outcome is almost certainly one of less overall impact on poverty.

I first try to understand how the gaps in our knowledge about development effectiveness come to exist and persist. I then identify a number of things that need to change in current approaches to evaluation if the potential to inform development practice is to be fulfilled. Examples are given from recent research, although a number of these issues remain under-researched. I hope that the discussion will help to change that.

Why Might We Underinvest in Rigorous Evaluations?

The focus of this paper is the problem of assessing the impact of a development project, where "impact" is measured against explicit counterfactual outcomes (such as in the absence of the project); the essential characteristic of a rigorous evaluation is that it includes a credible strategy for identifying the counterfactual. The topic embraces both *ex ante* and *ex post* evaluation (and possibly both for the same project). *Ex ante* evaluation is a key input to project appraisal. *Ex post* evaluation can sometimes provide useful insights into how a project might be modified along the way, and is certainly a key input to the accumulation of knowledge about development effectiveness, which guides future policymaking.

There are good reasons why not everything that is done in the name of development gets evaluated. Rigorous evaluations are rarely easy. Practical and logistical difficulties abound. Special-purpose data collection and close supervision are typically required. The analytic and computational demands for valid inferences can also be daunting and require specialized skills.

Knowledge-Market Failures

However, there are also reasons to doubt that the market for knowledge about development effectiveness works well. The outcome of these market failures is almost certainly that we underinvest in rigorous impact evaluations.

Suppliers and demanders of knowledge about development effectiveness do not typically have the same information about the quality of the evaluation—giving an example of what economists call "asymmetric information," which is a well-known source of market failure.[1] In the present context, development practitioners cannot easily assess the quality and expected benefits of an impact evaluation in order to weigh them against the costs. Short-cut methods promise quick results at low cost, though rarely are users well informed of the inferential dangers.[2] Since it is often hard for practitioners to know whether research is of good quality or not, there is a real risk that rigorous evaluations are driven out by nonrigorous ones of doubtful veracity.

Another important feature of this market is the degree of control that individual project "managers" (including staff in both aid agencies and governments) have over what gets evaluated and how much is spent on evaluation. This can be thought of as a noncompetitive feature of the market for knowledge about development effectiveness, in that the project manager more or less has the power to block the supply of knowledge. The decision about whether resources should be invested in data and research on a specific project or policy is often made by (or heavily influenced by) the individual practitioners involved, or by political stakeholders whose incentives need not be well-aligned with knowledge demands. The portfolio of evaluations is almost certainly biased towards programs that work well; managers of weak programs try to avoid rigorous evaluation, which threatens to expose the program's weaknesses. Lighter "evaluations" are often easier to manipulate for the purpose of showing seemingly positive results.

Decentralized decision-making about evaluation generates another source of market failure: the benefits from the rigorous evaluation of a development project spill over to other projects, which typically do not share in the cost of doing that evaluation. Development is a learning process, in which future practitioners benefit from current research. The individual project manager will typically not take account of these external benefits when deciding how much to spend on evaluation. This is what economists call an "externality." An implication of the externalities in the market for knowledge about development effectiveness is that we tend to underinvest in research that can draw useful lessons for other projects and settings besides that of the specific evaluation.

Certain types of evaluations are likely to be more prone to these sources of market failure. It is typically far easier to evaluate an intervention that yields all its likely impacts within 1 year (say) than an intervention that takes many years.

It can be no surprise that credible evaluations of the longer term impacts of (for example) infrastructure projects are rare. Similarly, we know very little about the long-term impacts of development projects that do deliver short-term gains; for example, we know much more about the short-term impacts of transfer payments on the current nutritional status of children in recipient families than about the possible gains in their longer term productivity from better nutrition in childhood. So future practitioners are often poorly informed about what works and what does not. There is a "myopia bias" in our knowledge, favoring development projects that yield quick results.

We probably also underinvest in evaluations of types of interventions that tend to have diffused, widespread benefits. Impacts for such interventions are often harder to identify than for cleanly assigned programs with well-defined beneficiaries, since one typically does not have the informational advantage of being able to observe nonparticipants (as the basis for inferring the counterfactual). It may also be hard to fund evaluations for such interventions, since they often lack a well-defined constituency of political support.

The implication of all this is that, without strong institutional support and encouragement, there will probably be too few evaluations, particularly of the long-term impacts of development interventions and of broader sectoral or economy-wide reforms. And the evaluations that do get done will focus too much on *internal validity* (whether valid inferences are drawn about the impact of that specific project in its specific setting) relative to *external validity* (whether valid inferences are drawn for other projects, either as scaled up versions of that project in the same setting or as similar projects in different settings). The fact that long-term evaluations are so rare (though it is widely agreed that development does not happen rapidly) and that we clearly know too little about external validity suggest that the available support is currently insufficient *or* it is misallocated.

Rising Support for Evaluations

Increasingly evaluations do receive support beyond what is demanded by the immediate practitioners. There has been substantial growth in donor support for impact evaluations in recent years. Donor governments are increasingly being pressed by their citizens to show the impact of development aid which has generated extra resources for financing impact evaluations. Unfortunately, the resources available are not always used for making rigorous evaluations. And it is not clear that the extra resources are having as much impact as they could on the incentives facing project managers and governments. Donor support needs to focus on increasing marginal private benefits from evaluation or reducing marginal costs. Nonetheless, there is now a broader awareness of the problems faced when trying to do evaluations, including the age-old problem of identifying "causal" impacts.

This has helped make donors less willing to fund weak proposals for evaluations that are unlikely to yield reliable knowledge about development effectiveness.

What does get evaluated, however, is still only a modest fraction of what gets done on the ground in the name of development. That may always be the case, given the costs of evaluation. But what is more worrying is that this fraction is a decidedly *nonrandom* one. Typically, a self-selected sample of practitioners approaches the funding sources, often with researchers already in tow. This process is likely to favor projects and policies that are expected to have benefits by their advocates.

All this makes it very important that new efforts by the development community to support impact evaluations of development policies—to address the market failures discussed above—should start from those knowledge gaps, not from a researcher's prior preference for one sort of data or method. That is not always the case. For example, while the recent enthusiasm for Randomized Control Trials (RCTs) (also called social experiments)—see, for example, Banerjee (2007) and Duflo and Kremer (2005)—has generated some interesting new research, it is not based on any clear strategic assessment of how this particular method would fill the knowledge gaps of highest priority. Nor is there any obvious reason why doing more social experiments would help correct for the distortions that generated those knowledge gaps. Randomization is clearly only feasible for a nonrandom subset of policies and settings; for example, it is rarely feasible to randomize the location of infrastructure projects and related programs, which are core activities in almost any poor country's development strategy. And even for the types of programs for which randomization is an option, it will be adopted more readily in some settings than others, given that social experiments raise ethical and political concerns—stemming from the fact that some of those to which a program is randomly assigned will almost certainly not need it, while some in the control group will. A better idea would be to randomize what gets evaluated rigorously and then choose a method appropriate to each sampled intervention, with randomization as one option.

The rest of this article will explore how we might assure that future work on impact evaluation is more relevant to the needs of development practitioners. While better approaches to evaluation will not, on their own, solve all the problems in the market for knowledge discussed above, recognizing those problems is the logical starting point for thinking about what constitutes better evaluation.

How Can We Do Better in Filling Key Knowledge Gaps?

The archetypal formulation of the evaluation problem aims to estimate the average impact on those to which a specific program is assigned (the participants)

by attempting to infer the counterfactual from those to which it is not assigned (nonparticipants). While this is an undeniably important and challenging problem, solving it is not sufficient for assuring that evaluation is relevant to development practice.

Questions for Evaluations

Evaluations should not take the intervention as predetermined, but must begin by probing the problem that a policy or project is addressing. Why is the intervention needed? How does it relate to overall development goals, such as poverty reduction? What are the market, or governmental, failures it addresses? What are its distributional goals? What are the trade-offs with alternative (including existing) policies or programs? As Devarajan and others (1997) argue, researchers can often play an important role in addressing these questions. This involves more precise identification of the policy *objectives* (properly weighing gains across different subgroups of a population and different generations); the relevant *constraints*, which include resources, information, incentives, and political economy constraints; and the *causal links* through which the specific intervention yields its expected outcomes.

This role in conceptualizing the case for intervention can be especially important when the capacity for development policymaking is weak or when it is captured by lobby groups advocating narrow sectoral interests. The *ex ante* evaluative role for research can also be crucial when practitioners have overly strong prior beliefs about what needs to be done. Over time, some practitioners become experts at specific types of intervention, and some may even lobby for those interventions. The key questions about whether the intervention is appropriate in the specific setting may not even get asked.

Evaluators themselves can also become lobbyists for their favorite methods. Too often it is not the question that is driving the evaluation agenda but a preference for certain types of data or certain methods; the question is then found that fits the methodology, not the other way around. Starting with the question, not the method, often points the evaluator toward types of data and methods outside the domain traditionally favored by his or her own disciplinary background. For example, some of the World Bank's research economists trying to understand persistent poverty and the impacts of antipoverty programs have been drawn to the theories and methods favored in other social sciences, such as anthropology, sociology, and social psychology; see, for example, the collection of papers in Rao and Walton (2004). Good researchers, like good detectives, assemble, and interpret diverse forms of evidence in testing empirical claims.

As already noted, rigorous impact evaluations require credible strategies for identifying the counterfactual—taking proper account of the likely sources of

bias, such as when outcomes are only compared over time for program participants, or when participants and nonparticipants are compared at only one date; see Ravallion (2008) for a survey of the (experimental and nonexperimental) methods available for this task. This is all about internal validity, which has been the main focus of researchers working on evaluations. In this discussion I will flag some issues that have received less attention yet matter greatly to the impact of an evaluation.

The choice of counterfactual is one such issue. The classic evaluation focuses on counterfactual outcomes in the absence of the program. This counterfactual may fall well short of addressing the concerns of policymakers. The alternative of doing nothing is rarely of interest to policymakers, who prefer instead to spend the same resources on some other program (possibly a different version of the same program). A specific program may appear to perform well against the option of doing nothing, but it is still performing poorly against some feasible alternative. For example, in an impact evaluation of a workfare program in India, Ravallion and Datt (1995) showed that the program substantially reduced poverty among the participants relative to the counterfactual of "no program," but that once the costs of the program were factored in (including the foregone income of workfare participants) the alternative counterfactual of a uniform (untargeted) allocation of the same budget outlay would have had more impact on poverty. Formally, the evaluation problem is essentially no different if some alternative program is the counterfactual; in principle we can repeat the analysis relative to the "do nothing counterfactual" for each possible alternative and compare them. But this is rare in practice.

Nor is it evident that the classic formulation of the impact evaluation problem yields the most relevant impact parameters. For example, there is often an interest in better understanding the *horizontal impacts* of a program, that is the differences in impacts at a given level of counterfactual outcomes, as revealed by the joint distribution of outcomes under treatment and outcomes under the counterfactual. We cannot know this from a standard impact evaluation, which only reveals net counterfactual mean outcomes for those treated. Instead of focusing solely on the net gains to the poor (say) we may ask how many losers there are among the poor, and how many gainers.

Counterfactual analysis of the joint distribution of outcomes over time is useful for understanding impacts on poverty dynamics. This approach is developed in Ravallion and others (1995) for the purpose of measuring the impacts of changes in social spending on the intertemporal joint distribution of income. Instead of only measuring the impact on poverty (the marginal distribution of income) the authors exploit panel data to distinguish impacts on the number of people who escape poverty over time (the "promotion" role of a safety net) from impacts on the number who fall into poverty (the "protection" role). (This is only possible if

one can identify how impacts vary with household characteristics; the discussion will return to this issue in discussing impact heterogeneity below.) Ravallion and others apply this approach to an assessment of the impact on poverty transitions of reforms in Hungary's social safety net.

Spillover Effects

A further way in which the classic impact evaluation problem often needs to be adapted to the needs of practitioners concerns its assumption that impacts for direct participants do not spill over to nonparticipants. Only under this assumption can we infer the counterfactual from an appropriate sample of the nonparticipants. Spillover effects are recognized as a concern in evaluating large public programs for which contamination of the control group can be hard to avoid due to the responses of markets and governments; spillover are also relevant in drawing lessons for scaling up based on an RCT. For further discussion, see Moffitt (2003, 2006).

An example of spillover effects can be found in the Miguel and Kremer (2004) study of treatments for intestinal worms in children. The authors argue that an evaluation design, in which some children are treated and some are retained as controls, would seriously underestimate the gains from treatment by ignoring the externalities between treated and "control" children. The design for the authors' own evaluation avoided this problem by using mass treatment at the school level instead of individual treatment (using control schools at sufficient distance from treatment schools).

Spillover effects can also arise from the way markets respond to an intervention. Consider the example of an Employment Guarantee Scheme (EGS) in which the government commits to give work to anyone who wants it at a stipulated wage rate; this was the aim of the famous EGS in the Indian state of Maharashtra; in 2006 the Government of India implemented a national version of this scheme. The attractions of an EGS as a safety net stem from the fact that access to the program is universal (anyone who wants help can get it) but that all participants must work to obtain benefits and at a wage rate that is considered low in the specific context. The universality of access means that the scheme can provide effective insurance against risk. The work requirement at a low wage rate is taken by proponents to imply that the scheme will be self-targeted to the income poor.

The EGS is an assigned program in that there are well-defined "participants" and "nonparticipants." And at first glance it might seem appropriate to collect data on both groups and compare their outcomes either by random assignment or after cleaning out observable heterogeneity. However, this classic evaluation design could give a severely biased result. The gains from such a program are

very likely to spill over into the private labor market. If the employment guarantee is effective then the scheme will establish a firm lower bound to the entire wage distribution—assuming that no able-bodied worker would accept non-EGS work at any wage rate below the EGS wage. So even if one picks a perfect comparison group, one will conclude that the scheme has no impact, since wages will be the same for participants and nonparticipants. But that would entirely miss the impact, which could be large for both groups.

Spillover effects can also arise from the behavior of governments. Chen and others (2009) find evidence of such spillover effects in their evaluation of a World Bank-supported poor-area development program in rural China. When the program selected certain villages to participate, the local government withdrew some of its own spending on development projects in those villages, in favor of nonprogram villages—the same set of villages from which the comparison group was drawn. Ignoring these spillover effects generated a nonnegligible underestimation of the impact of the program. Chen and others show how, under certain assumptions, one can estimate the maximum bias due to the specific type of spillover effects that arises from local government spending responses to external development aid. In the case of the poor-area program in China that Chen and others study, their results suggest that the spending responses of local governments to the external aid entail that the standard "difference-in-difference" method may well capture only two-thirds of the true impact.

Heterogeneity

Practitioners should never be happy with an evaluation that assumes common (homogeneous) impact. The impact of an assigned intervention can vary across those receiving it. Even with a constant benefit level, eligibility criteria entail differential costs to participants. For example, the foregone labor earnings incurred by participants in workfare or conditional cash transfer schemes (via the loss of earnings from child labor) will vary according to skills and local labor-market conditions.

By recognizing the scope for heterogeneity in impacts and the role of contextual factors, one can make evaluative research more relevant to good policymaking. For example, in the aforementioned evaluation of a poor-area development program in rural China, Chen and others (2009) find low overall impact but considerable heterogeneity, in that different types of households benefited more than others, with the relatively better educated amongst the poor achieving the highest returns to the project's investments. The policy implication is that choosing different beneficiaries would have greatly increased the project's overall impact; indeed, the study estimated that an alternative process of beneficiary selection that better exploited the heterogeneity in impacts could have led to a four-fold increase in the

project's overall rate of return. By developing a deeper understanding of such heterogeneity, evaluations can help develop better projects.

Heterogeneity of impacts in terms of observables is readily allowed for by adding interaction effects between the intervention and observables to one's model of outcomes. However, not all sources of heterogeneity are observable, and participants and stakeholders often react to factors unobserved by the researcher—confounding efforts to identify true impacts using standard methods, including experiments; this is what Heckman and others (2006) refer to as "essential heterogeneity." With some extra effort, one can also allow for latent heterogeneity in the impacts of an intervention (using a random coefficients estimator in which the impact estimate contains a stochastic component). Applying this approach to the evaluation data for PROGRESA (a conditional cash transfer program in Mexico), Djebbari and Smith (2008) found that they could convincingly reject the assumption of common (homogeneous) effects made by past evaluations of that program.

When there is such heterogeneity, it can be of interest to policymakers to distinguish marginal impacts (from small program expansions or contractions) from the average impacts that have received the bulk of attention. Following Björklund and Moffitt (1987), the marginal treatment effect can be defined as the mean gain to units that are indifferent between participating or not. This requires that we model explicitly the choice problem facing participants (Björklund and Moffitt 1987; Heckman and Navarro-Lozano 2004). We may also want to estimate the joint distribution of outcomes under treatment and outcomes under the counterfactual, and a method for doing so is outlined in Heckman and others (1997).

External Validity

Arguably the most important thing to learn from any evaluation relates to its lessons for future policies (including reforms to the interventions being evaluated). External validity is highly desirable, but it can be hard to achieve. We naturally want research findings to have a degree of generalizability, so they can provide useful knowledge to guide practice in other settings. Thus empirical researchers need to focus on *why* a policy or program has an impact; a question to which I will return. However, too often impact evaluations are a "black box"; under certain assumptions, they reveal average impacts among those who receive a program, but say little or nothing about the economic and social processes leading to that impact. And only by understanding those processes can we draw valid lessons for scaling up, or for taking the same project to other settings. Research that tests the theories that underlie the rationales for intervention can thus be useful in practice.

When the policy issue is whether to expand a given program at the margin, the classic estimator of mean impact is actually of rather limited interest. For example, we may want to know the marginal impact of a greater duration of exposure to the program. An example can be found in the study by Ravallion and others (2005) of the impacts on workfare participants of leaving the program relative to staying (recognizing that this entails a nonrandom selection process). Another example can be found in the study by Behrman and others (2004) of the impacts on children's cognitive skills and health status of longer exposure to a preschool program in Bolivia. The authors provided an estimate of the marginal impact of higher program duration by comparing the cumulative effects of different durations using a matching estimator. In such cases, selection into the program is not an issue, and we do not even need data on units who never participated.

Relatedly, one must recognize the importance of *context* since this can be key to drawing valid lessons for other settings. Relevant contextual factors may include the circumstances of participants, the economic, cultural, and political environment, and the administrative context. Unless we understand how such factors influence the outcomes of an intervention, the evaluation will have weak external validity. The next section returns to this issue.

Given that we can expect in general that any intervention will have heterogeneous impacts—some participants gain more than others—serious concerns can arise about the external validity of RCTs. The people who are normally attracted to a program, taking account of the expected benefits and costs to them personally, may differ systematically from the random sample of people who were included in the trial.[3] The RCT may well have evaluated a very different program to the one that is actually implemented on the basis of that RCT.

External validity concerns about impact evaluations can also arise when certain institutions need to be presented to even facilitate the evaluations. For example, when randomized trials are tied to the activities of specific non-governmental organizations (NGOs) as the facilitators, there is a concern that the same intervention at the national scale may have a very different impact in places where the NGO is not present. Making sure that the control group areas also have the NGO can help, but even then we cannot rule out interaction effects between the NGO's activities and the intervention. In other words, the effect of the NGO may not be "additive" but "multiplicative," such that the difference between measured outcomes for the treatment and control groups does not reveal the impact in the absence of the NGO. Furthermore, the very nature of the intervention may change when it is implemented by a government rather than an NGO. This may happen because of unavoidable differences in (among other things) the quality of supervision, the incentives facing service providers, and administrative capacity.

A further external validity concern is that, while partial equilibrium assumptions may be fine for a pilot, *general equilibrium effects* (sometimes called "feedback" or "macro" effects) can be important when the pilot is scaled up nationally. For example, an estimate of the impact on schooling of a tuition subsidy based on a randomized trial may be deceptive when scaled up, given that the structure of returns to schooling will alter. Heckman and others (1998) demonstrated that partial equilibrium analysis can greatly overestimate the impact of a tuition subsidy once relative wages adjust, although Lee (2005) found a much smaller difference between the general and partial equilibrium effects of a tuition subsidy in a slightly different model.

A special case of the general problem of external validity is *scaling up*. There are many things that can change when a pilot program is scaled up: the inputs to the intervention can change, the outcomes can change, and the intervention can change; Moffitt (2006) gave examples in the context of education programs. The realized impacts on scaling up can differ from the trial results (whether randomized or not) because the socio-economic composition of program participation varies with scale. Ravallion (2004) discussed how this can happen in theory and presented the results from a series of country case studies, all of which suggest that the incidence of program benefits becomes more pro-poor with scaling up. Trial results could over- or underestimate impacts on scaling up. Larger projects may be more susceptible to rent seeking or corruption (as Deaton [2006] suggests); alternatively, the political economy may entail that the initial benefits tend to be captured more by the nonpoor (as shown by Lanjouw and Ravallion 1999, using data for India).

Evaluative research should regularly test the assumptions made in operational work. Even field-hardened practitioners do what they do on the basis of some implicit model of how the world works, which rationalizes what they do, and how their development project is expected to have an impact. Existing methods of rapid *ex ante* impact assessment evidently also rely heavily on the models held by practitioners. Researchers can perform a valuable role in helping to make those models explicit and (where possible) helping to assess their veracity.

A case in point is the questionable assumption—routinely made by both project staff and evaluators—that the donor's money is actually financing what recipients claim it is financing. Research has pointed to a degree of fungibility in development aid, whereby the marginal use of public finds is unlikely to be the specific project that is being evaluated. Yet an assessment of "aid effectiveness" is (presumably) just that—an evaluation of the impact of the aid, not the project *per se*. These are different evaluation problems.

Assessments of aid effectiveness need to take a broader view of public spending, as advocated by Devarajan and others (1997). How broad it needs to be is unclear. There is some evidence that external aid sticks to its sector (quaintly

called a "flypaper effect" in economics); on this see van de Walle and Mu (2007). The existence of fungibility and flypaper effects points to the need for a sectoral approach in efforts to evaluate the impacts of development aid.

What Determines Impact?

The above discussion points to the need to supplement standard evaluations by information that can throw light on the factors influencing measured outcomes. That can be crucial for drawing useful policy lessons, including redesigning a program and scaling up. The relevant factors relate to both the participants (such as understanding program take-up decisions and how the outcomes are influenced by participants' characteristics) and program context (such as understanding how the quantity/quality of service provision affects outcomes and how the role of local institutions influences outcomes). This section elaborates some of the ways that we might learn more about how a program does, or does not, have an impact, so as to better address the issues raised above.

An obvious approach to understanding which factors influence a program's performance is to repeat it across different types of participants and in different contexts. Duflo and Kremer (2005) and Banerjee (2007) have argued that repeated RCTs across varying contexts and scales should be used to decide what works and what does not in development aid. Even putting aside the aforementioned problems encountered in social experiments, the feasibility of doing a sufficient number of trials—sufficient to span the relevant domain of variation found in reality for a given program, as well as across the range of policy options—is far from clear. The number of RCTs needed to test even one large national program could well be prohibitive. It is questionable whether this is a sound strategy for filling the existing gaps in our knowledge about development effectiveness.

Nonetheless, even if one cannot go as far as Banerjee (2007) would like, it can be agreed that evaluation designs should plan for contextual variation. Important clues can often be found in the geographic differences in impacts. These can stem from geographic differences in relevant population characteristics or from deeper location effects, such as agro-climatic differences and differences in local institutions (such as local "social capital" or the effectiveness of local public agencies). An example can be found in the study by Galasso and Ravallion (2005) in which the targeting performance of Bangladesh's Food-for-Education program was assessed across each of 100 villages in Bangladesh, with the results being correlated with the characteristics of those villages. The authors found that the revealed differences in performance were partly explicable in terms of observable village characteristics, such as the extent of intravillage inequality (with more unequal villages being less effective in reaching their poor through the program).

Failure to allow for such location differences has been identified as a serious weakness in past evaluations; see for example the comments by Moffitt (2003) on trials of welfare reforms in the United States.

The literature suggests that location is a key dimension of context. An implication is that it is less problematic to scale up from a pilot within the same geographic setting (with a given set of relevant institutions) than to extrapolate the trial to a different setting. In one of the few attempts to test how well evaluation results from one location can be extrapolated to another location, Attanasio and others (2003) divided the seven states of Mexico in which the PROGRESA evaluation was done into two groups. They found that results from one group had poor predictive power for assessing likely impacts in the other group.

Useful clues for understanding impacts can sometimes be found by studying impacts on what can be called "intermediate" or "structural" measures. The typical evaluation design identifies a small number of "final outcome" indicators, and it aims to assess the program's impact on those indicators. Instead of using only final outcome indicators, one may choose to also study impacts on certain intermediate indicators of behavior deemed relevant on theoretical grounds. For example, the intertemporal behavioral responses of participants in antipoverty programs are of obvious relevance to understanding their impacts. An impact evaluation of a program of compensatory cash transfers to Mexican farmers found that the transfers were partly invested, with second-round effects on future incomes (Sadoulet, de Janvry, and Davis 2001). Similarly, Ravallion and Chen (2005) found that participants in a poor-area development program in China saved a large share of the income gains from the program. Identifying responses through savings and investment provides a clue to understand the current impacts on living standards and the possible future welfare gains beyond the project's current life span. Instead of focusing solely on the agreed welfare indicator relevant to the program's goals, one collects and analyzes data on a potentially wide range of intermediate indicators relevant to understanding the processes determining impacts.

This also illustrates a common concern in evaluation studies, given behavioral responses, namely that the study period is rarely much longer than the period of the program's disbursements. However, a share of the impact on peoples' living standards will usually occur beyond the disbursement period. This does not necessarily mean that credible evaluations will need to track welfare impacts over much longer periods than is typically the case—raising concerns about feasibility. But it does suggest that evaluations need to look carefully at impacts on partial intermediate indicators of longer term impacts even when good measures of the welfare objective are available within the project cycle. The choice of such indicators will need to be informed by an understanding of participants' behavioral responses to the program. That understanding will be informed by both theory and data.

In learning from an evaluation, one often needs to draw on information external to the evaluation. Qualitative research (intensive interviews with participants and administrators) can be a useful source of information on the underlying processes determining outcomes; see the discussion on "mixed methods" in Rao and Woolcock (2003). One approach is to use such methods to test the assumptions made by an intervention; this has been called "theory-based evaluation," although that is hardly an ideal term given that identification strategies for mean impacts are often theory based. Weiss (2001) illustrated this approach in the abstract in the context of evaluating the impacts of community-based antipoverty programs. An example is found in a World Bank evaluation of social funds (SFs), as summarized in Carvalho and White (2004). While the overall aim of an SF is typically to reduce poverty, the study was interested in seeing whether SFs worked as intended by their designers. For example, did local communities participate? Who participated? Was there "capture" of the SF by local elites (as some critics have argued)? Building on Weiss (2001), the evaluation identified a series of key hypothesized links connecting the intervention to outcomes and tested whether each one worked. For example, in one of the country studies, Rao and Ibanez (2005) tested the assumption that an SF works by local communities collectively proposing the subprojects that they want; for an SF in Jamaica, the authors found that the process was often dominated by local elites.

In practice, it is very unlikely that all the relevant assumptions are testable (including alternative assumptions made by different theories that might yield similar impacts). Nor is it clear that the process determining the impact of a program can always be decomposed into a neat series of testable links within a unique causal chain; there may be more complex forms of interaction and simultaneity that do not lend themselves to this type of analysis. For these reasons, theory-based evaluation cannot be considered an alternative to assessing impacts on final outcomes by credible (experimental or nonexperimental) methods, although it can still be a useful complement to such evaluations for better understanding measured impacts.

Project monitoring databases are an important, underutilized, source of information for understanding how a program works. Too often, however, the project monitoring data collected and the information system used have negligible evaluative content. This is not inevitably the case. For example, Ravallion's (2000) method of combining spending maps with poverty maps can allow rapid assessments of the targeting performance of a decentralized antipoverty program. This illustrates how, at modest cost, standard monitoring data can be made more useful for providing information on how the program is working, and in a way that provides sufficiently rapid feedback to a project to allow corrections along the way.

The Proempleo experiment in Argentina provides an example of how information external to the evaluation can carry important insights. Proempleo was a pilot wage subsidy and training program for unemployed workers. The RCT by Galasso and others (2004) randomly assigned vouchers for a wage subsidy across (typically poor) people currently in a workfare program and tracked their subsequent success in getting regular work. A randomized control group located the counterfactual. The results indicated a significant impact of the wage-subsidy voucher on employment. But when cross-checks were made against central administrative data, supplemented by informal interviews with the hiring firms, it was found that there was very low take-up of the wage subsidy by firms. The scheme was highly cost effective; the government saved 5 percent of its workfare wage bill for an outlay on subsidies that represented only 10 percent of that saving. However, the cross-checks against these other data revealed that Proempleo did not work the way its design had intended. The bulk of the gain in employment for participants was not through higher demand for their labor induced by the wage subsidy. Rather the impact arose from supply-side effects; the voucher appeared to have had credential value to workers—it acted like a "letter of introduction" that few people had (and how it was allocated was a secret locally). This could not be revealed by the evaluation, but required supplementary data. The extra insight obtained about how Proempleo actually worked in the context of its trial setting also carried implications for scaling up, which put emphasis on providing better information for poor workers about how to get a job rather than providing wage subsidies.

Spillover effects also point to the importance of a deeper understanding of how a program operates. Indirect (or "second-round") impacts on nonparticipants are common. A workfare program may lead to higher earnings for nonparticipants; or a road improvement project in one area might improve accessibility elsewhere. Depending on how important these indirect effects are thought to be in the specific application, the "program" may need to be redefined to embrace the spillover effects. Or one might need to combine the type of evaluation discussed here with other tools, such as a model of the labor market, to pick up other benefits.

An extreme form of a spillover effect is an economy-wide program. The classic evaluation tools for assigned programs have little obvious role for economy-wide programs in which no explicit assignment process is evident, or, if it is, the spillover effects are likely to be pervasive. When some countries get the economy-wide program but some do not, cross-country comparative work (such as growth regressions) can reveal impacts. That identification task is often difficult, because there are typically latent factors at country level that simultaneously influence outcomes and whether a country adopts the policy in question. And even when the identification strategy is accepted, carrying the generalized lessons from cross-country regressions to inform policymaking in any one country can be highly

problematic. There are also a number of promising examples of how simulation tools for economy wide policies such as Computable General Equilibrium models can be combined with household-level survey data to assess impacts on poverty and inequality.[4] These simulation methods make it far easier to attribute impacts to the policy change, although this advantage comes at the cost of the need to make many more assumptions about how the economy works.

In both assessing impacts and understanding the reasons for those impacts, there is often scope for a "meso" level analysis in which theory is used to inform empirical analysis of what would appear to be the key mechanisms linking an intervention to its outcomes, and this is done in a way that identifies key structural parameters that can be taken as fixed when estimating counterfactual outcomes. This type of approach can provide deeper insights into the factors determining outcomes in *ex post* evaluations and can also help in simulating the likely impacts of changes in program or policy design *ex ante*.

Naturally, simulations require many more assumptions about how an economy works.[5] As far as possible one would like to see those assumptions anchored to past knowledge built up from rigorous *ex post* evaluations. For example, by combining a randomized evaluation design with a structural model of education choices and exploiting the randomized design for identification, one can greatly expand the set of policy-relevant questions about the design of a program that a conventional evaluation can answer; examples using the PROGRESA evaluation data can be found in Todd and Wolpin (2002), Attanasio and others (2004), and de Janvry and Sadoulet (2006). This strand of the literature has revealed that a budget-neutral switch of the enrolment subsidy in PROGRESA from primary to secondary school would have delivered a net gain in school attainments, by increasing the proportion of children who continue onto secondary school. While PROGRESA had an impact on schooling, it could have had greater impact. However, it should be recalled that this type of program has two objectives: increasing schooling (reducing future poverty) and reducing current poverty, through the targeted transfers. To the extent that refocusing the subsidies on secondary schooling would reduce the impact on current income poverty (by increasing the forgone income from children's employment), the case for this change in the program's design would need further analysis.

Many of these observations point to the important role played by theory in understanding why a program may or may not have an impact. However, the theoretical models found in the evaluation literature are not always the most relevant to developing country settings. The models have stemmed mainly from the literature on evaluating training and other programs in developed countries, in which selection is seen largely as a matter of individual choice amongst those eligible. This approach does not sit easily with what we know about many antipoverty programs in developing countries, in which the choices made by politicians

and administrators appear to be at least as important to the selection process as the choices made by those eligible to participate. We often need a richer theoretical characterization of the selection problem to assure relevance.

An example of one effort in this direction can be found in the Galasso and Ravallion (2005) model of a decentralized antipoverty program; their model focuses on the public-choice problem facing the government and the local collective action problem facing communities, with individual participation choices treated as a trivial subproblem. Such models can also point to instrumental variables for identifying impacts and studying their heterogeneity.

An example of the use of a more structural approach to assessing an *economy-wide* reform can be found in Ravallion and van de Walle (2008). Here the policy being studied was the decollectivization of agriculture in Vietnam and the subsequent efforts to develop a private market in land-use rights. These were huge reforms, affecting the livelihoods of the vast majority of the Vietnamese people. Ravallion and van de Walle developed models to explain how farmland was allocated to individual farmers at the time of decollectivization, how those allocations affected living standards, and how the subsequent reallocations of land amongst farmers (that were permitted by the subsequent market-oriented agrarian reforms) responded to the inefficiencies left by the initial administrative assignment of land at the time of decollectivization. Naturally, many more assumptions need to be made about how the economy works—essentially to make up for the fact that one cannot observe nonparticipants in these reforms as a clue to the counterfactual. Not all of those assumptions are testable. However, the principle of evaluation is the same, namely to infer the impacts of these reforms relative to explicit counterfactuals. For example, Ravallion and van de Walle assessed the welfare impacts of the privatization of land-use rights against both an efficiency counterfactual (the simulated competitive market allocation) and an equity counterfactual (an equal allocation of quality-adjusted land within communes). This type of approach can also throw light on the heterogeneity of the welfare impacts of large reforms; in the Vietnam case, the authors were able to assess both the overall impacts on poverty and identify the presence of both losers and gainers, including among the poor.

Does Published Knowledge Reliably Guide Development Practice?

The benefits from evaluations depend in part on their publication, which is the main way they feed into development knowledge. Development policymaking draws on accumulated knowledge built up in large part from published findings.

At the same time, publishing in refereed journals is important to a researcher's credibility and career prospects. Thus publication processes—notably the incentives facing journal editors and reviewers, researchers, and those who fund research—are relevant to our success in achieving development goals.

There are reasons for questioning how well the publication process performs in helping to realize the social benefits from rigorous evaluations. Three issues stand out. First, the cost of completing the publication stage in the cycle of research can be significant, and it is hard to reduce these costs; writing the paper the right way, documenting everything that was done, addressing the concerns of referees and editors, all take time. Practitioners are often unwilling to fund these costs, and they even question the need for publication. Again a large share of the benefits is external, to which individual project staff naturally attach low weight.

Second, received wisdom develops its own inertia through the publication process, with the result that it is often harder to publish a paper that reports unexpected or ambiguous impacts when judged against current theories, past evidence, or both. Reviewers and editors are likely to apply different standards according to whether they believe the results hold on *a priori* grounds.

In the context of evaluating development projects, the prior belief is often that the project will have positive impacts, for that is presumably the main reason why the project was funded in the first place. Then a preference for confirming prior beliefs will tend to bias our knowledge in favor of finding positive impacts. Negative or nonimpacts will not get reported as easily. When there is a history of research on a type of intervention, the results of the early studies will set the prior beliefs against which later work is judged. An initial bad draw from the true distribution of impacts may then distort knowledge for some time after.

A third source of bias is that the review process in scientific publishing (at least in economics) tends to put greater emphasis on the internal validity of an evaluative research paper than on its external validity. The bulk of the effort goes into establishing that valid inferences are being drawn about causal impacts within the sample of program participants. The authors may offer some concluding (and possibly highly cautious) thoughts on the broader implications for scaling up the program well beyond that sample. However, these claims will rarely be established with comparable rigor to the efforts put into establishing internal validity, and the claims are rarely challenged by reviewers.

These imperfections in the research publication industry undoubtedly have feedback effects on the production of evaluations. Researchers will tend to work harder to obtain positive findings, or at least results consistent with received wisdom, so as to improve their chances of getting their work published. No doubt, extreme biases (in either direction) will be eventually exposed. But this takes time.

Researchers have no shortage of instruments at their disposal to respond to the (often distorted) incentives generated by professional publication processes. Key

decisions on what to report, and indeed the topic of the research paper, naturally lie with the individual researcher, who must write the paper and get it published. In the case of impact evaluations of development projects, the survey data (often collected for the purpose of the evaluation) will typically include multiple indicators of "outcomes." If one collects 20 indicators (say) then there is a good chance that at least one of them shows statistically significant impacts of the project even when it had no impact in reality. A researcher keen to get published might be tempted to report results solely for the significant indicator. (Journal reviewers and editors rarely ask what other data were collected.) The dangers to knowledge generation are plain.

The threat of replication by another researcher can help assure better behavior. But in economics, replication studies tend to have low status and are actually quite rare. Thus, as Rodrik (2009) points out, there will be little or no incentive for researchers to carry out the great many repetitions that would probably be called for in the agenda for the mass RCTs proposed by Banerjee (2007) and Duflo and Kremer (2005), given that professional journals would have little interest in such replications of the same intervention and method in different settings.

Nor do researchers have a strong incentive to make their data publicly available for replication purposes. Some professional economics journals have adopted a policy that the datasets used in accepted papers should be made available this way, although enforcement is not uniformly strong.

In choosing how to respond to this environment, the individual researcher faces a trade-off between publishability and relevance. Thankfully, the fact of being policy relevant is not in itself an impediment to publishability in most journals, though any research paper that lacks originality, rigor, or depth will have a hard time getting published. It is by maintaining the highest standards that we assure that relevant research is publishable, as well as being credible when carried to policy dialogues. However, it must be acknowledged that the set of research questions that are most relevant to development policy overlap only partially with the set of questions that are seen to be in vogue by the editors of the professional journals at any given time. The dominance of academia in the respected publishing outlets is understandable, but it can sometimes make it harder for researchers doing work more relevant to development practitioners, even when that work meets academic standards. Academic research draws its motivation from academic concerns that overlap imperfectly with the issues that matter to development practitioners. Provided that scholarly rigor is maintained, the cost to a researcher's published output of doing policy relevant research might not be high, but it would be naïve to think that the cost is zero.

Communication and dissemination of the published findings on development effectiveness can also be deficient. Researchers sometimes lack the skills or personalities needed for effective communication with nontechnical audiences.

Having worked very hard to assure that the data and analysis are sound, and so pass muster by accepted scientific criteria, it does not come easily for all researchers to translate the results into just a few key policy messages, which do not seem to do justice to all the work involved. The externality problem can also arise here, whereby social returns from outreach exceed private returns. A research institution will often need to support its researchers with specialized staff who possess strong communication skills.

Conclusions

We underinvest in some of the most important tools for enhancing development effectiveness. Weak incentives facing key decision-makers—stemming from knowledge externalities, asymmetric information, and noncompetitive features of the market for knowledge—entail that too few rigorous impact evaluations of development interventions get done. This problem appears to be particularly severe for evaluations of projects that yield benefits over long periods and for efforts in rigorously understanding the lessons that can be drawn for other projects and settings. While donor support for evaluation is helping redress these problems, there is still a long way to go; greater support is needed, but existing support could also be made more effective if it were aimed at changing private incentives to evaluate, by either raising the marginal benefits or lowering the marginal costs facing project managers. The process of knowledge generation through evaluations is probably also affected by biases on the publication side, which distort the incentives facing individual researchers in doing evaluations.

None of this is helped by the fact that even the most rigorous methods found in practice often fall well short of delivering credible answers to the questions posed by practitioners. Those questions start at the outset of the project cycle and even embrace the rationale for the intervention. They include understanding why the intervention might have greater impact for some participants, and in some settings, than others. They include the lessons for both the intervention under study and (importantly) future interventions. The classic estimate of the mean impact on those treated is of strictly limited utility for addressing these issues.

Nor is the task helped by the fact that researchers have at times overstated what their favorite method can deliver for practitioners, and that they have often chosen what they evaluate according to whether their favorite method is feasible, rather than whether the question is important to development. Interventions are even being chosen, or designed, to fit certain preferred evaluation methods. At the same time, exaggerated claims are sometimes made by nonresearchers about what can be learnt about development effectiveness in a short time with little or no credible data.

Looking forward, greater effort is needed to develop approaches to evaluation that can throw more useful light on the external validity of findings on specific projects (including implications for scaling up) and that can provide a deeper understanding of what determines why an intervention does, or does not, have an impact. Fungibility and flypaper effects also point to the need for a broader sectoral approach to assessing aid effectiveness. There is still much to do if we want to realize the potential for evaluative research to inform development policy by "seeking truth from facts."

Notes

Martin Ravallion is Director, Development Research Group, World Bank; his email address is Mravallion@worldbank.org. For helpful comments on an earlier version of this article, and related discussions on this topic, the author is grateful to Francois Bourguignon, Asli Demirguc-Kunt, Gershon Feder, Jed Friedman, Emanuela Galasso, Markus Goldstein, Bernard Hoekman, Beth King, Danny Leipziger, David McKenzie, Luis Seven, Lyn Squire, Dominique van de Walle, Michael Woolcock, and the journal's reviewers. These are the views of the author and should not be attributed to the World Bank or any affiliated organization.

1. The classic account of this problem is given in Akerlof (1970).
2. For example, OECD (2007) outlines an approach to "*ex ante* poverty impact assessment" that claims to assess the "poverty outcomes and impacts" of a project in just 2–3 weeks at a cost of $10,000–40,000, which, as the authors point out, is appreciably less than standard impact evaluations. The OECD paper proposes that a consultant fills in a series of tables giving the project's "short-term and long-term outcomes" across a range of (economic and noneconomic) dimensions for each of the various groups of identified "stakeholders," as well as the project's "transmission channels," through induced changes in prices, employment, transfers, and so on. Many readers (including many practitioners) would not know just how hard it is to make such assessments in a credible way, and the paper offers no guidance to readers on what degree of confidence one can have in the results of such an exercise.
3. This is sometimes called "randomization bias"; see Heckman and Smith (1995). See also the discussion in Moffitt (2004).
4. See, for example, Bourguignon and Ferreira (2003) and Chen and Ravallion (2004).
5. For a useful overview of *ex ante* methods, see Bourguignon and Ferreira (2003).

References

Akerlof, George. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84:488–500.

Attanasio, Orazio, Costas Meghir, and Ana Santiago. 2004. *Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA*. Working Paper EWP04/04. London: Institute of Fiscal Studies.

Attanasio, Orazio, Costas Meghir, and Miguel Szekely. 2003. *Using Randomized Experiments and Structural Models for Scaling Up: Evidence from the PROGRESA Evaluation*. Working Paper EWP04/03. London: Institute of Fiscal Studies.

Banerjee, Abhijit. 2007. *Making Aid Work*. Cambridge, MA: MIT Press.

Behrman, Jere, Yingmei Cheng, and Petra Todd. 2004. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics* 86(1):108–32.

Björklund, Anders, and Robert Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-Selection." *Review of Economics and Statistics* 69(1):42–9.

Bourguignon, François, and Francisco Ferreira. 2003. "Ex-ante Evaluation of Policy Reforms Using Behavioural Models." In Francois F Bourguignon, and Luiz Pereira da Silva, eds., *The Impact of Economic Policies on Poverty and Income Distribution*. New York: Oxford University Press.

Carvalho, Soniya, and Howard White. 2004. "Theory-Based Evaluation: The Case of Social Funds." *American Journal of Evaluation* 25(2):141–60.

Chen, Shaohua, and Martin Ravallion. 2004. "Welfare Impacts of China's Accession to the World Trade Organization." *World Bank Economic Review* 18(1):29–58.

Chen, Shaohua, Ren Mu, and Martin Ravallion. 2009. "Are there Lasting Impacts of Aid to Poor Areas?" *Journal of Public Economics* 93(3–4):512–528.

Deaton, Angus. 2006. "Evidence-based Aid Must not Become the Latest in a Long String of Development Fads." *Boston Review* July. (http://bostonreview.net/BR31.4/deaton.html).

Devarajan, Shantayanan, Lyn Squire, and Sethaput Suthiwart-Narueput. 1997. "Beyond Rate of Return: Reorienting Project Appraisal." *World Bank Research Observer* 12(1):35–46.

Djebbari, Habiba, and Jeffrey Smith. 2008. "Heterogeneous Program Impacts of PROGRESA." *Journal of Econometrics* 145(1–2):64–80.

Du, Runsheng. 2006. *The Course of China's Rural Reform*. Washington, DC: International Food Policy Research Institute.

Duflo, Esther, and Michael Kremer. 2005. "Use of Randomization in the Evaluation of Development Effectiveness." In George Pitman, Osvaldo Feinstein,, and Gregory Ingram, eds., *Evaluating Development Effectiveness*. New Brunswick, NJ: Transaction Publishers.

Galasso, Emanuela, and Martin Ravallion. 2005. "Decentralized Targeting of an Anti-Poverty Program." *Journal of Public Economics* 89(4):705–27.

Galasso, Emanuela, Martin Ravallion, and Agustin Salvia. 2004. "Assisting the Transition from Workfare to Work: Argentina's *Proempleo* Experiment." *Industrial and Labor Relations Review* 57(5):128–42.

Heckman, James, and Salvador Navarro-Lozano. 2004. "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics* 86(1):30–57.

Heckman, James, and Jeffrey Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9(2):85–110.

Heckman, James, L. Lochner, and C. Taber. 1998. "General Equilibrium Treatment Effects." *American Economic Review Papers and Proceedings* 88(2):381–6.

Heckman, James, Jeffrey Smith, and Nancy Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies* 64(4):487–535.

Heckman, James, Serio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88(3):389–432.

de Janvry, Alain, and Elisabeth Sadoulet. 2006. "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality." *World Bank Economic Review* 20(1):1–29.

Lanjouw, Peter, and Martin Ravallion. 1999. "Benefit Incidence and the Timing of Program Capture," *World Bank Economic Review* 13(2):257–74.

Lee, Donghoon. 2005. "An Estimable Dynamic General Equilibrium Model of Work: Schooling, and Occupational Choice." *International Economic Review* 46(1):1–34.

Luo, Xiaopeng. 2007. "Collective Learning Capacity and the Choice of Reform Path." Paper presented at the IFPRI/Government of China Conference: Taking Action for the World's Poor and Hungry People, Beijing.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1):159–217.

Moffitt, Robert. 2003. *The Role of Randomized Field Trials in Social Science Research: A Perspective from Evaluations of Reforms of Social Welfare Programs.* Cemmap Working Paper, CWP23/02. Department of Economics, University College London.

———. 2004. "The Role of Randomized Field Trials in Social Science Research." *American Behavioral Scientist* 47(5):506–40.

———. 2006. "Forecasting the Effects of Scaling Up Social Programs: An Economics Perspective." In Barbara Schneider, and Sarah-Kathryn McDonald, eds., *Scale-Up in Education: Ideas in Principle.* Lanham: Rowman and Littlefield.

OECD (Organisation for Economic Co-operation and Development). 2007. *A Practical Guide to Ex Ante Poverty Impact Assessment.* Paris: Development Assistance Committee Guidelines and Reference Series, OECD.

Rao, Vijayendra, and Ana Maria Ibanez. 2005. "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* 41(5):788–838.

Rao, Vijayendra, and Michael Walton (eds.). 2004. *Culture and Public Action.* Stanford: Stanford University Press.

Rao, Vijayendra, and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In Francois J. Bourguignon, and Luiz Pereira da Silva, eds., *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools,* 165–90. New York: Oxford University Press.

Ravallion, Martin. 2000. "Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved." *World Bank Economic Review* 14(2):331–45.

———. 2004. "Who is Protected from Budget Cuts?" *Journal of Policy Reform* 7(2):109–22.

———. 2008. "Evaluating Anti-Poverty Programs." In Paul Schultz, and John Strauss, eds., *Handbook of Development Economics,* vol. 4. Amsterdam: North-Holland.

Ravallion, Martin, and Gaurav Datt. 1995. "Is Targeting through a Work Requirement Efficient? Some Evidence for Rural India." In Dominique van de Walle, and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore: Johns Hopkins University Press.

Ravallion, Martin, and Shaohua Chen. 2005. "Hidden Impact: Household Saving in Response to a Poor-Area Development Project." *Journal of Public Economics* 89(11–12):2183–204.

Ravallion, Martin, and Dominique van de Walle. 2008. *Land in Transition: Reform and Poverty in Rural Vietnam.* Basingstoke: Palgrave Macmillan and World Bank.

Ravallion, Martin, Dominique van de Walle, and Madhur Gaurtam. 1995. "Testing a Social Safety Net." *Journal of Public Economics* 57(2):175–99.

Ravallion, Martin, Emanuela Galasso, Teodoro Lazo, and Ernesto Philipp. 2005. "What Can Ex-Participants Reveal about a Program's Impact?" *Journal of Human Resources* 40(1):208–30.

Rodrik, Dani. 2009. "The New Development Economics: We Shall Experiment, but How Shall We Learn?" In Jessica Cohen, and William Easterly, eds., *What Works in Development? Thinking Big and Thinking Small*, Washington: Brookings Institution Press.

Sadoulet, Elisabeth, Alain de Janvry, and Benjamin Davis. 2001. "Cash Transfer Programs with Income Multipliers: PROCAMPO in Mexico." *World Development* 29(6):1043–56.

Todd, Petra, and Kenneth Wolpin. 2002. *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.* Penn Institute for Economic Research Working Paper 03-022, Department of Economics, University of Pennsylvania.

Van de Walle, Dominique, and Ren Mu. 2007. "Fungibility and the Flypaper Effect of Project Aid: Micro-Evidence for Vietnam." *Journal of Development Economics* 84(2):667–85.

Weiss, Carol. 2001. "Theory-Based Evaluation: Theories of Change for Poverty Reduction Programs." In Osvaldo Feinstein, and Robert Piccioto, eds., *Evaluation and Poverty Reduction.* New Brunswick, NJ: Transaction Publications.

Timing and Duration of Exposure in Evaluations of Social Programs

Elizabeth M. King • Jere R. Behrman

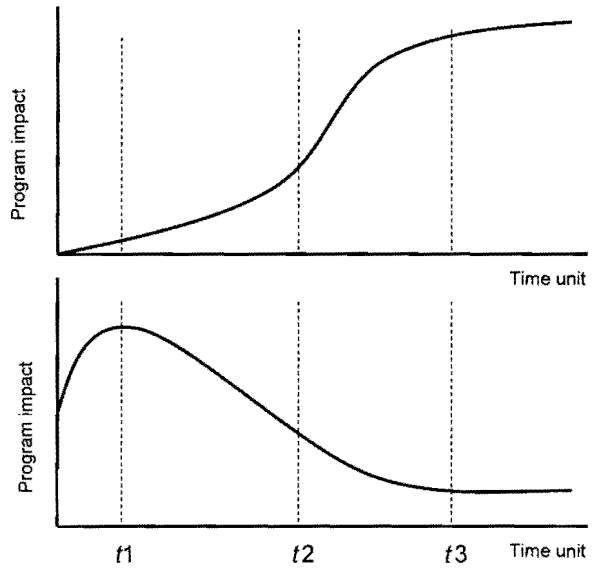
Impact evaluations aim to measure the outcomes that can be attributed to a specific policy or intervention. While there have been excellent reviews of the different methods for estimating impact, insufficient attention has been paid to questions related to timing: How long after a program has begun should it be evaluated? For how long should treatment groups be exposed to a program before they benefit from it? Are there time patterns in a program's impact? This paper examines the evaluation issues related to timing, and discusses the sources of variation in the duration of exposure within programs and their implications for impact estimates. It reviews the evidence from careful evaluations of programs (with a focus on developing countries) on the ways that duration affects impacts.

A critical risk that faces all development aid is that it will not pay off as expected—or that it will not be perceived as effective—in reaching development targets. Despite the billions of dollars spent on improving health, nutrition, learning, and household welfare, we know surprisingly little about the impact of many social programs in developing countries. One reason for this is that governments and the development community tend to expand programs quickly even in the absence of credible evidence, which reflects an extreme impatience towards adequately piloting and assessing new programs first. This impatience is understandable given the urgency of the problems being addressed, but it can result in costly but avoidable mistakes and failures; it can also result in really promising new programs being terminated too soon when a rapid assessment shows negative or no impact.

However, recent promises of substantially more aid from rich countries and large private foundations have intensified interest in assessing aid effectiveness. This interest is reflected in a call for more evaluations of the impact of donor-funded programs in order to understand what type of intervention works and

doi:10.1093/wbro/lkn009 Advance Access publication February 23, 2009 24:55–82

Figure 1. The Timing of Evaluations Can Affect Impact Estimates



what doesn't.[1] Researchers are responding enthusiastically to this call. There have been important developments in evaluation methods as they apply to social programs, especially on the question of how best to identify a group with which to compare intended program beneficiaries—that is, a group of people who would have had the same outcomes as the program group without the program.[2]

The timing question in evaluations, however, is arguably as important but relatively understudied. This question has many dimensions. For how long after a program has been launched should one wait before evaluating it? How long should treatment groups be exposed to a program before they can be expected to benefit from it, either partially or fully? How should one take account of the heterogeneity in impact that is related to the duration of exposure? This timing issue is relevant for all evaluations, but particularly so for the evaluation of social programs that require changes in the behaviors of both service providers and service users in order to bring about measurable outcomes. If one evaluates too early, there is a risk of finding only partial or no impact; too late, and there is a risk that the program might lose donor and public support or that a badly designed program might be expanded. Figure 1 illustrates this point by showing that the true impact of a program may not be immediate or constant over time, for reasons that we discuss in this paper. Comparing two hypothetical programs whose impact differs over time, we see that an evaluation undertaken at time $t1$ indicates that the case in the bottom panel has a higher impact than the case in the top panel, while an evaluation at time $t3$ suggests the opposite result.

This paper discusses key issues related to the timing of programs and the time path of their impact, and how these have been addressed in evaluations.[3] Many evaluations treat interventions as if they were instantaneous, predictable changes in conditions and equal across treatment groups. Many evaluations also implicitly assume that the effect on individuals is dichotomous (that is, that individuals are either exposed or not), as might be the case in a one-shot vaccination program that provides permanent immunization. There is no consideration of the possibility that the effects vary according to differences in program exposure.[4] Whether the treatment involves immunization or a more process-oriented program such as community organization, the unstated assumptions are often that the treatment occurs at a specified inception date, and that it is implemented completely and in precisely the same way across treatment groups.

There are several reasons why implementation is neither immediate nor perfect, why the duration of exposure to a treatment differs not only across program areas but also across ultimate beneficiaries, and why varying lengths of exposure might lead to different estimates of program impact. This paper discusses three broad sources of variation in duration of exposure, and reviews the literature related to those sources (see Appendix Table A-1 for a list of the studies reviewed). One source pertains to organizational factors that affect the leads and lags in program implementation, and to timing issues related to program design and the objectives of an evaluation. A second source refers to spillover effects, including variation that arises from the learning and adoption by beneficiaries and possible contamination of the control groups. Spillover effects are external (to the program) sources of variation in the treatment: while these may pertain more to compliance than timing, they can appear and intensify with time, and so affect estimates of program impact. A third source pertains to heterogeneous responses to treatment. Although there can be different sources of heterogeneity in impact, the focus here is on those associated with age or cohort, especially as these cohort effects interact with how long a program has been running.

Organizational Factors and Variation in Program Exposure

Program Design and the Timing of Evaluations

How long one should wait to evaluate a program depends on the nature of the intervention itself and the purpose of the evaluation. For example, in the case of HIV/AIDS or tuberculosis treatment programs, adherence to the treatment regime over a period of time is necessary for the drugs to be effective. While drug effectiveness in treating the disease is likely to be the outcome of interest, an evaluation of the program might also consider adherence rates as an intermediate

outcome of the program—and so the evaluation need not take place only at the end of the program but during the implementation itself. In the case of worker training programs, workers must first enroll for the training, and then some time passes during which the training occurs. If the training program has a specific duration, the evaluation should take place after the completion of the training program.

However, timing may not be so easy to pin down if the timing of the intervention itself is the product of a stochastic process. For example, a market downturn may cause workers to be unemployed, triggering their eligibility for worker training, or a market upturn may cause trainees to leave the program to start a job—as Ravallion and others (2005) observe in Argentina's Trabajar workfare program. In cases where the timing of entry into (or exit from) a program itself differs across potential beneficiaries, the outcomes of interest depend on an individual selection process and on the passage of time. An evaluation of these programs should consider selection bias. Randomized evaluations of trials with well-defined start and end dates do not address this issue.

In fact the timing of a program may be used for identification purposes. For example, some programs are implemented in phases. If the phasing is applied randomly, the random variation in duration can be used for identification purposes in estimating program impact (Rosenzweig and Wolpin 1986 is a seminal article on this point). One instance is Mexico's PROGRESA (Programa de Educación, Salud y Alimentación) which was targeted at the poorest rural communities when it began. Using administrative and census data on measures of poverty, the program identified the potential beneficiaries. Of the 506 communities chosen for the evaluation sample, about two-thirds were randomly selected to receive the program activities during the first two years of the program, starting in mid-1998, while the remaining one-third received the program in the third year, starting in the fall of 2000. The group that received the intervention later has been used as a control group in evaluations of PROGRESA (see, for example, Schultz 2004; Behrman, Sengupta, and Todd 2005).

One way to regard duration effects is that, given constant dosage or intensity of a treatment, lengthening duration of exposure is akin to increasing intensity, and thus the likelihood of greater impact. Two cases show that impact is likely to be underestimated if the evaluation coverage is too short. First, skill development programs are an obvious example of the importance of the duration of program exposure: beneficiaries who attend only part of a training course are less likely to benefit from the course and attain the program goals than those who complete the course. In evaluating the impact of a training course attended by students, Rouse and Krueger (2004) distinguish between students who completed the computer instruction offered through the Fast ForWord program and those who did not. The authors define completion as a function of the amount of training

attended and the actual progress of students toward the next stage of the program, as reflected in the percentage of exercises at the current level mastered at a prespecified level of proficiency.[5] The authors find that, among students who received more comprehensive treatment—as reflected by the total number of completed days of training and the level of achievement of the completion criteria—performance improved more quickly on one of the reading tests (but not all) that the authors use.

Banerjee and others (2007) evaluate two randomly assigned programs in urban India: a remedial training program that hired young women to teach children with low literacy and numeracy skills, and a computer-assisted learning program. Illustrating the point that a longer duration of exposure intensifies treatment, the remedial program raised average test scores by 0.14 of a standard deviation in the first year and 0.28 of a standard deviation in the second year of the program, while computer-assisted learning increased math scores by 0.35 of a standard deviation in the first year and 0.47 of a standard deviation in the second year. The authors interpret the larger estimate in the second year as an indication that the first year laid the foundation for the program to help the children benefit from its second year.

Lags in Implementation

One assumption that impact evaluations often make is that, once a program starts, its implementation occurs at a specific and knowable time that is usually determined at a central program office. Program documents, such as World Bank project loan documents, typically contain official project launch dates, but these dates often differ from the date of actual implementation in a project area. When a program actually begins depends on supply- and demand-related realities in the field. For example, a program requiring material inputs (such as textbooks or medicines) relies on the arrival of those inputs in the program areas: the timing of the procurement of the inputs by the central program office may not indicate accurately when those inputs arrive at their intended destinations.[6] In a large early childhood development program in the Philippines, administrative data indicate that the timing of the implementation differed substantially across program areas: because of lags in central procurement, three years after project launch not all providers in the program areas had received the required training (Armecin and others 2006). Besides supply lags, snags in information flows and project finances can also delay implementation. In conditional cash transfer programs in Mexico and Ecuador, delays in providing the information about intended household beneficiaries prevented program operators in some sites from making punctual transfers to households (Rawlings and Rubio 2005; Schady and Araujo 2008).[7] In Argentina poor municipalities found it more difficult to raise the

cofinancing required for the subprojects of the country's Trabajar program, which weakened the program's targeting performance (Ravallion 2002).

It is possible to address the problem of implementation lags in part if careful and complete administrative data on timing are available for the program: cross-referencing such data with information from public officials or community leaders in treatment areas could reveal the institutional reasons for variation in implementation. For example, if there is an average gap of one year between program launch and actual implementation, then it is reasonable for the evaluation to make an allowance of one year after program launch before estimating program impact.[8] However, reliable information on dates is often not readily available, so studies have tended to allot an arbitrary grace period to account for lags.

Assuming a constant allowance for delays, moreover, may not be an adequate solution if there is wide variation in the timing of implementation across treatment areas. This is likely to be the case if the program involves a large number of geographical regions or a large number of components and actors. In programs that cover several states or provinces, region or state fixed-effects might control for duration differences if the differences are homogeneous within a region or state. If the delays are not independent of unobservable characteristics in the program areas, that may also influence program impact. An evaluation of Madagascar's SEECALINE program provides an example of how to define area-specific starting dates. It defined the start of the program in each treatment site as the date of the first child-weighing session in that site. The area-specific date takes into account the program's approach of gradual and sequential expansion, and the expected delays between the signing of the contract with the implementing NGO and the point when a treatment site is actually open and operational (Galasso and Yau 2006). This method requires detailed program-monitoring data.

If a program has many components, the solution may hinge on the evaluator's understanding of the technical production function and thus on identifying the elements that must be present for the program to be effective. For example, in a school improvement program that requires additional teacher training and instructional materials, the materials might arrive in schools at about the same time, but the additional teacher training might be achieved only over a period of several months, perhaps because of differences in teacher availability. The evaluator, when considering the timing of the evaluation, must decide whether the effective program start should be defined according to the date when the materials arrive in schools or when all (or most?) of the teachers have completed their training. In the Madagascar example above, although the program has several components (for example, growth monitoring, micronutrient supplementation, deworming), the inception date of each site was fixed according to a growth-monitoring activity, that of the first weighing session (Galasso and Yau 2006).

Although the primary objective of evaluations is usually to measure the impact of programs, often they also monitor progress during the course of implementation and thus help to identify problems that need correction. An evaluation of the Bolivia Social Investment Fund illustrates this point clearly (Newman and others 2002). One of the program components was to improve the drinking water supply through investments in small-scale water systems. However, the first laboratory analysis of water quality showed little improvement in program areas. Interviews with local beneficiaries explained why: contrary to plan, people designated to maintain water quality lacked training; inappropriate materials were used for tubes and the water tanks; and the lack of water meters made it difficult to collect fees needed to finance maintenance work. After training was provided in all the program communities, a second analysis of water supply indicated significantly less fecal contamination in the water in those areas.

How are estimates of impact affected by variation in program start and by lags in implementation? Variation in program exposure that is not incorporated into program evaluation is very likely to bias downward the intent-to-treat (ITT) estimates of the program's impact, especially if such impact increases with the exposure of the beneficiaries who are actually treated. But the size of this underestimation, for a given average lag across communities, depends on the nature of the lags. If the program implementation delays are not random, it matters if they are inversely or directly correlated with unobserved attributes of the treated groups that may positively affect program success. If the implementation lags are *directly* correlated with unobserved local attributes, then the true ITT effects are underestimated to a larger extent; for example, central administrators may put less effort into starting the programs in areas that have worse unobserved determinants of the outcomes of interest, such as a weaker management capability of local officials to implement a program in these areas. If implementation delays are instead *inversely* associated with the unobserved local attributes (that is, the central administrators put more effort into starting the program in those same areas), then the ITT effects are underestimated to a lesser extent. If instead the program delays are random, the extent of the underestimation depends on the variance in the implementation lags (still given the same mean lag). All else being equal, greater random variance in the lags results in greater underestimation of the ITT effects. This is because a larger classical random measurement error in a right-side variable biases the estimated coefficient more towards zero.

If the start of the treatment for individual beneficiaries has been identified correctly, implementation delays in themselves do not necessarily affect estimates of treatment-on-the-treated (TOT) effects. In some cases this date of entry can be relatively easy to identify: for example, the dates on which beneficiaries enroll in a program may be established through a household or facility survey or

administrative records (for example, school enrollment rosters or clinic logbooks). In other cases, however, the identification may be more difficult: for example, beneficiaries may be unable to distinguish among alternative, contemporaneous programs or to recall their enrollment dates, or the facility or central program office may not monitor beneficiary program enrollments.[9] Nonetheless, even if the variation in treatment dates within program areas is handled adequately and enrollment dates are identified fairly accurately at the beneficiary level, nonrandom implementation delays bias TOT estimates. Even a well-specified facility or household survey may be adversely affected by unobservables that may be related to the direction and size of the program impact. The duration of exposure, like program take-up, has to be treated as endogenous. The problem of selection bias motivates the choice of random assignment to estimate treatment effects in social programs.

Learning by Providers

A different implementation lag is associated with the fact that program operators (or providers of services) themselves face a learning curve that depends on time in training and on-the-job experience. This most likely produces some variation in the quality of program implementation that is independent of whether there has been a lag in the procurement of the training. This too is an aspect of program operation that is often not captured in impact evaluations. Although the evaluation of Madagascar's SEECALINE program allotted a grace period of two to four months for the training of service providers, it is likely that much of the learning by providers happened on the job after the formal training.

While the learning process of program operators may delay full program effectiveness, another effect could be working in the opposite direction. The "pioneering effect" means that implementers may exhibit extra dedication, enthusiasm, and effort during the first stages, because the program may represent an innovative endeavor to attain an especially important goal. (A simplistic diagram of this effect is shown in Figure 1, bottom panel.) Jimenez and Sawada (1999) find that newer EDUCO schools in El Salvador had better outcomes than older schools (with school characteristics held constant). They interpret this as evidence of a Hawthorne effect—that is, newer schools were more motivated and willing to undertake reforms than were the older schools. If such a phenomenon exists, it would exert an opposite pull on the estimated impacts and, if sufficiently strong, might offset the learning effect, at least in the early phases of a new program. Over time, however, this extra dedication, enthusiasm, and effort are likely to wane.[10] If there are heterogeneities in this unobserved pioneering effect across program sites that are correlated with observed characteristics (for example, schooling of program staff), the result will be biased estimates of the impact of such characteristics on initial program success.

Spillover Effects

The observable gains from a social program during its entire existence, much less after only a few years of implementation, may be an underestimate of its full potential impact for several reasons that are external to the program design. First, evaluations are typically designed to measure outcomes at the completion of a program, and yet the program might yield additional and unintended outcomes in the longer run. Second, while the assignment of individuals or groups of individuals to a treatment can be defined, program beneficiaries may not actually take up an intervention—or may not do so until after they have learned more about the program. Third, with time, control groups or groups other than the intended beneficiaries might find a way of obtaining the treatment, or they may be affected simply by learning about the existence of the program—possibly because of expectations that the program will be expanded to their area. If noncompliance is correlated with the outcome of interest, then the difference in the average outcomes between the treatment and the control groups is a biased estimate of the average effect of the intervention. We discuss these three examples below.

Short-Run and Long-Run Outcomes

Programs that invest in cumulative processes, such as a child's physiological growth and accumulation of knowledge, require the passage of time. This implies that longer program exposure would yield greater gains, though probably with diminishing marginal returns. Also, such cumulative processes could lead to outcomes beyond those originally intended—and possibly beyond those of immediate interest to policymakers. Early childhood development (ECD) programs are an excellent example of short-run outcomes that could lead to long-run outcomes beyond those envisioned by the original design. These programs aim to mitigate the multiple risks facing very young children, and to promote their physical and mental development by improving nutritional intake and/or cognitive stimulation. The literature review by Grantham-McGregor and others (2007) identifies studies that use longitudinal data from Brazil, Guatemala, Jamaica, the Philippines, and South Africa that establish causality between preschool cognitive development and subsequent schooling outcomes. The studies suggest that a one standard deviation increase in early cognitive development predicts substantially improved school outcomes in adolescence, as measured by test scores, grades attained, and dropout behavior (for example, 0.71 additional grade by age 18 in Brazil).

Looking beyond childhood, Garces, Thomas, and Currie (2002) find evidence from the U.S. Head Start program that links preschool attendance not only to higher educational attainment but also to higher earnings and better adult social

outcomes. Using longitudinal data from the Panel Study of Income Dynamics and controlling for the participants' disadvantaged background, they conclude that exposure to Head Start for whites is associated in the short run with significantly lower dropout rates, and in the long run with 30 percent greater probability of high school completion, 28 percent higher likelihood of attending college, and higher earnings in their early twenties. For African-Americans participation in Head Start is associated with a 12-percentage-point lower probability of being booked for or charged with a crime.

Another example of an unintended long-run outcome is provided by an evaluation (Angrist, Bettinger, and Kremer 2004) of Colombia's school voucher program at the secondary level (PACES or Programa de Ampliación de Cobertura de la Educación Secundaria). This finds longer-run outcomes beyond the original program goal of increasing the secondary school enrollment rate of the poorest youths in urban areas. Using administrative records, the follow-up study finds that the program increased high-school graduation rates of voucher students in Bogota by 5–7 percentage points, which is consistent with the earlier outcome of a 10-percentage-point increase in eighth-grade completion rates (Angrist and others 2002). Correcting for the greater percentage of lottery winners taking college admissions tests, the program increased test scores by two-tenths of a standard deviation in the distribution of potential test scores.

In their evaluation of a rural roads project in Vietnam, Mu and van de Walle (2007) find that, because of developments external to the program, rural road construction and rehabilitation produced larger gains as more time elapsed after project completion. The impacts of roads depend on people using them, so for the benefits of the project to be apparent, more bicycles or motorized vehicles must be made available to rural populations connected by the roads. But the impacts of the new roads also include other developments that arose more slowly, such as a switch from agriculture to non-agricultural income-earning activities, and an increase in secondary schooling following a rise in primary school completion. These impacts grew at an increasing rate as more months passed, taking two years more on average to emerge.

In the long run, however, impacts can also vanish. Short-term estimates are not likely to be informative about such issues as the extent of diminishing marginal returns to exposure, which would be an important part of the information basis of policies. In Vietnam the impact of the rural roads project on the availability of foods and on employment opportunities for unskilled jobs emerged quite rapidly: it then waned as the control areas caught up with the program areas, an effect we return to below (Mu and van de Walle 2007). In Jamaica a nutritional-supplementation-cum-psychological-stimulation program for children under two yielded mixed effects on cognition and education years later (Walker and others 2005). While the interventions benefited child development—even at age 11,

stunted children who received stimulation continued to show cognition benefits—small improvements from supplementation noted at age 7 were no longer present at age 11. In fact, impact can vanish much sooner after a treatment ends. In the example of two randomized trials in India, although impact rose in the second year of the program, one year after the programs had ended, impact dropped. For the remedial program, the gain fell to 0.1 of a standard deviation and was no longer statistically significant; for the computer learning program, the gain dropped to 0.09 of a standard deviation, though it was still significant (Banerjee and others 2007).

Chen, Mu, and Ravallion (2008) point to how longer-term effects might be invisible to evaluators of the long-term impact of the Southwest China Project, which gave selected poor villages in three provinces funding for a range of infrastructure investments and social services. The authors find only small and statistically insignificant average income gains in the project villages four years after the disbursement period. They attribute this partly to significant displacement effects caused by the government cutting the funding for nonproject activities in the project villages and reallocating resources to the nonproject villages. Because of these displacement effects, the estimated impacts of the project are likely to be underestimated. To estimate an upper bound on the size of this bias, the increase in spending in the comparison villages is assumed to be equal to the displaced spending in the project villages. Under this assumption, the upper bound of the bias could be as high as 50 percent—and it could be even larger if the project actually has positive long-term benefits.

Long-term benefits, however, are often not a powerful incentive to support a program or policy. The impatience of many policymakers with a pilot–evaluate–learn approach to policymaking and action is usually coupled with a high discount rate. This results in little appetite to invest in programs for which benefits are mostly enjoyed in the future. Even aid agencies exhibit this impatience, and yet programs that are expected to have long-run benefits would be just the sort of intervention that development aid agencies should support because local politicians are likely to dismiss them.

Learning and Adoption by Beneficiaries

Programs do not necessarily attain full steady-state effectiveness after implementation commences. Learning by providers and beneficiaries may take time, a necessary transformation of accountability relationships may not happen immediately, or the behavioral responses of providers and consumers may be slow in becoming apparent.

The success of a new child-immunization or nutrition program depends on parents learning about the program and bringing their children to the providers,

and the providers giving treatment. In Mexico's PROGRESA the interventions were randomly assigned at the community level. If program uptake were perfect, a simple comparison between eligible children in the control and treatment localities would have been sufficient to estimate the program TOT effect (Behrman and Hoddinott 2005). However, not all potential beneficiaries sought services: only 61–64 percent of the eligible children aged 4 to 24 months and only half of those aged 2 to 4 years actually received the program's nutritional supplements. The evaluation found no significant ITT effects, but did find that the TOT effects were significant, despite individual and household controls.

In Colombia's secondary-education voucher program too, information played a role at both the local government level and the student level (King, Orazem, and Wohlgemuth 1999). Since the program was cofunded by the central and municipal governments, information given to the municipal governments was critical to securing their collaboration. At the beginning of the program, the central government met with the heads of the departmental governments to announce the program and solicit their participation; in turn the departmental governors invited municipal governments to participate. Dissemination of information to families was particularly important, because participation was voluntary and the program targeted only certain students (specifically those living in neighborhoods classified among the two lowest socioeconomic strata in the country) on the basis of specific eligibility criteria. Some local governments used newspapers to disseminate information about the program.

In decentralization reforms, the learning and adoption processes are arguably more complex because the decision to participate and the success of implementation depend on many more actors. Even the simplest form of this type of change in governance entails a shift in the accountability relationships between levels of government and between governments and providers—for example, the transfer of the supervision and funding of public hospitals from the national government to a subnational government. In Nicaragua's autonomous schools program in the 1990s, for example, the date a school signed the contract with the government was considered to be the date the school officially became autonomous. In fact, the signing of the contract was merely the first step toward school autonomy: it would have been followed by training activities, the election of the school management council, the development of a school improvement plan, and so on. Hence, the reform's full impact on outcomes would have been felt only after a period of time, and the size of this impact might have increased gradually as the elements of the reform were put in place. However, it is not easy to determine the length of the learning period. Among teachers, school directors, and parents in the so-called autonomous schools, the evaluation finds a lack of agreement on whether their schools had become autonomous and the extent to which this had been achieved (King and Özler 1998). An in-depth qualitative analysis in

a dozen randomly selected schools confirms that school personnel had different interpretations of what had been achieved (Rivarola and Fuller 1999).

Studies of the diffusion of the Green Revolution in Asia in the mid-1960s highlight the role of social learning among beneficiaries. Before adopting the new technology, individuals seem to have learned about it from the experiences of their neighbors (their previous decisions and outcomes). This wait-and-see process accounted for some of the observed lags in the use of high-yielding seed varieties in India at the time (Foster and Rosenzweig 1995; Munshi 2004). In rice villages the proportion of farmers who adopted the new seed varieties rose from 26 percent in the first year following the introduction of the technology to 31 percent in the third year; in wheat villages, the proportion of adopters increased from 29 percent to 49 percent. Farmers who did not have neighbors with comparable attributes (such as farm size or characteristics unobserved in available data such as soil quality) may have had to carry out more of their own experimentation. This would probably have been a more costly form of learning because the farmers bore all the risk of the choices they made (Munshi 2004).

The learning process at work during the Green Revolution is similar to that described by Miguel and Kremer (2003, 2004) about the importance of social networks in the adoption of new health technology, in this case deworming drugs. Survey data on individual social networks of the treatment group in rural Kenya reveal that social links provided nontreatment groups better information about the deworming drugs, and thus led to higher program take-up. Two years after the start of the deworming program, school absenteeism among the treatment group had fallen by about one-quarter on average. There were significant gains in several measures of health status—including reductions in worm infection, child growth stunting, and anemia—and gains in self-reported health. But children whose parents had more social links to early treatment schools were significantly less likely to take deworming drugs. The authors speculate that this disappointing finding could be due to overly optimistic expectations about the impact of the drugs, or to the fact that the health gains from deworming take time to be realized, while the side effects of the drugs are immediately felt.

Providing information about a program, however, is no guarantee of higher program uptake. One striking example of this is given by a program in Uttar Pradesh, India, which aimed to strengthen community participation in public schools by providing information to village members (Banerjee and others 2008). More information apparently did not lead to higher participation by the Village Education Committee (VEC), by parents, or by teachers. The evaluators attribute this poor result to more deep-seated information blockages: village members were unaware of the roles and responsibilities of the VEC, despite the existences of these committees since 2001, and a large proportion of the VEC members were not even aware of their membership.

The nutritional component in PROGRESA was undersubscribed (because parents lacked information about the program and its benefits), and the community mobilization in Uttar Pradesh was found wanting (because basic information about the roles and powers of village organizations is difficult to convey). Impact evaluations that do not take information diffusion and learning by beneficiaries into account obtain downward-biased ITT and TOT impact estimates. The learning process might be implicit—for example when program information diffuses to potential beneficiaries during the course of implementation, perhaps primarily by word of mouth—or it could be explicit, for example when a program aims an information campaign at potential beneficiaries during a well-defined time period.

Two points are worth noting about the role of learning in impact evaluation. One is the simple point discussed above that learning takes time. A steady-state level of effective demand among potential beneficiaries (effective in the sense that the beneficiaries actually act to enroll in or use program services) is related to the process of expanding effective demand for a program.[11] This implies that ITT estimates of program impact are biased downward if the estimates are based on data obtained prior to the attainment of this steady-state effective demand. The extent of the bias depends on whether learning (or the expansion of effective demand) is correlated with unobserved program attributes; specifically, there is less downward bias if this correlation is positive. There may be heterogeneity in this learning process: those programs that have better unobserved management capabilities may promote more rapid learning, while those that have worse management capabilities may face slower learning. Heterogeneity in learning would affect the extent to which the ITT and TOT impacts that are estimated before a program has approached effectiveness are downward-biased—but to a lesser degree if the heterogeneity in learning is random.

The second point is that the learning process itself may be a program component, and thus an outcome of interest in an impact evaluation. How beneficiaries learn and decide to participate is often external to a program, since the typical assumption is that beneficiaries will take up a program if the program exists. In fact, the exposure of beneficiaries to specific communication interventions about a program may be necessary to encourage uptake. There is a large literature, for example, that shows a strong association between exposure to mass-media information campaigns and the use of contraceptive methods and family planning services. The aims of such campaigns have been to make potential beneficiaries aware of these services, and to break down sociocultural resistance to them (Cleland and others 2006). This "social marketing" approach has been used also to stimulate the demand for insecticide-treated mosquito nets for malaria control, and has increased demand, especially among the poorest and most remote households (Rowland and others 2002; Kikumbih and others 2005). To understand how learning takes place is to begin to understand the "black box" that lies between program design and outcomes—and if this learning were

promoted in a random fashion, it could serve as an exogenous instrument for the estimation of program impact.

Peer Effects

The longer a program has been in operation, the more likely it is that specific interventions will spill over to populations beyond the treatment group and thus affect impact estimates. Peer effects increase impact, as in the case of the Head Start example already mentioned. Garces, Thomas and Currie (2002) find strong spillover effects within the family—higher birth-order children (that is, younger siblings) seem to benefit more than their older siblings, especially among African-Americans, because older siblings are able to teach younger ones. Hence, expanding the definition of impact to include peer effects adds to impact estimates.

Peer effects also arise when specific program messages (either directly from communications interventions or from observing treatment groups) diffuse to control groups and alter their behavior in the same direction as in the treatment group. While this contagion is probably desirable from the point of view of policymakers, it likely depresses impact estimates since differences between the control and treatment groups are diminished. Another form of leakage that grows with time may not be so harmless from the point of view of program objectives. For programs that target only specific populations, time allows political pressure to build for the program to be more inclusive and even for nontargeted groups to find ways of obtaining treatment (for example through migration into program sites). Because of the demand-driven nature of the Bolivia Social Investment Fund, for instance, not all communities selected for active promotion applied for and received a SIF-financed education project, but some communities not selected for active promotion nevertheless applied for promotion and obtained an education project (Newman and others 2002).

Heterogeneity of impact

An examination of how program impact varies according to the observable characteristics of the beneficiaries can teach us important lessons on policy and program design. Our focus here is on occasions when duration or timing differences interact with the sources of heterogeneity in impact. One important source of heterogeneity in some programs is cohort membership.

Cohort Effects

The age of beneficiaries may be one reason why duration of exposure to a program matters, and the estimates of ITT and TOT impacts can be affected

substantially by whether the timing is targeted toward critical age ranges. Take the case of ECD programs, such as infant feeding and preschool education, which target children for just a few years after birth. This age targeting is based on the evidence that a significant portion of a child's physical and cognitive development occurs at a very young age, and that the returns to improvements in the living or learning conditions of the child are highest at those ages. The epidemiological and nutritional literatures emphasize that children under three years of age are especially vulnerable to malnutrition and neglect (see Engle and others 2007 for a review). Finding that a nutritional supplementation program in Jamaica did not produce long-term benefits for children, Walker and others (2005) suggest that prolonging the supplementation—or supplementing at an earlier age, during pregnancy, and soon after birth—might have benefited later cognition. It might have been more effective than the attempt to reverse the effects of undernutrition through supplementation at an older age. Applying evaluation methods to drought shocks, Hoddinott and Kinsey (2001) also conclude that in rural Zimbabwe children in the age range of 12 to 24 months are the most vulnerable to such events: these children lose 1.5–2 centimeters of physical growth, while older children 2 to 5 years of age do not seem to experience a slowdown in growth.[12] In a follow-up study Alderman, Hoddinott, and Kinsey (2006) conclude that the longer the exposure of young children to civil war and drought, the larger the negative effect of these shocks on child height; moreover, older children suffer less than younger children in terms of growth.[13]

Interaction of Cohort Effects and Duration of Exposure

As discussed above, the impacts of some programs crucially depend on whether or not an intended beneficiary is exposed to an intervention at a particularly critical age range, such as during the first few years of life. Other studies illustrate that the duration of exposure during the critical age range also matters. The evaluation by Frankenberg, Suriastini, and Thomas (2005) of Indonesia's Midwife in the Village program shows just this. The program was intended to expand the availability of health services to mothers and thus improve children's health outcomes. By exploiting the timing of the (nonrandom) introduction of a midwife to a community, the authors distinguish between the children, living in the same community, who were exposed and those who were not exposed to a midwife. The authors group the sample of children into three birth cohorts. For each group, the extent of exposure to a village midwife during the vulnerable period of early childhood varied as a function of whether the village had a midwife and, if so, when she had arrived. In communities that had a midwife from 1993 onward, children in the younger cohort had been fully exposed to the program when data were collected, whereas children in the middle cohort had been only partially

exposed. The authors conclude that partial exposure to the village midwife program conferred no benefits in improved child nutrition, while full exposure from birth yielded an increase in the height-for-age z-score of 0.35 to 0.44 of a standard deviation among children aged 1 to 4 years.

Three other studies test the extent to which ECD program impacts are sensitive to the duration of program exposure and the ages of the children during the program. Behrman, Cheng, and Todd (2004) evaluated the impact of a preschool program in Bolivia, the Proyecto Integral de Desarrollo Infantil. Their analysis explicitly takes into account the dates of program enrollment of individual children. In their comparison of treated and untreated children, they find evidence of positive program impacts on motor skills, psychosocial skills, and language acquisition that are concentrated among children 37 months of age and older at the time of the evaluation. When they disaggregated their results by the duration of program exposure, the effects were most clearly observed among children who had been involved in the program for more than a year.

Like the Bolivia evaluation, the evaluation of the early childhood development program in the Philippines mentioned above finds that the program impacts vary according to the duration of exposure of children, although this variation is not as dramatic as the variation associated with children's ages (Armecin and others 2006). Administrative delays and the different ages of children at the start of the program resulted in the length of exposure of eligible children varying from 0 to 30 months, with a mean duration of 14 months and a substantial standard deviation of 6 months. Duration of exposure varied widely, even when a child's age was controlled for. The study finds that, for motor and language development, two- and three-year-old children exposed to the program had z-scores 0.5 to 1.8 standard deviations higher, depending on length of exposure, than children in the control areas, and that these gains were much lower among older children.

Gertler (2004) also estimates how duration of exposure to health interventions in Mexico's PROGRESA affected the probability of child illness, using two models—one assumes that program impact is independent of duration, and the other allows impact to vary according to the length of exposure. The interventions required that children under 2 years be immunized, visit nutrition monitoring clinics, and obtain nutritional supplements, and that their parents receive training on nutrition, health, and hygiene; children between 2 and 5 years of age were expected to have been immunized already, but were to obtain the other services. Gertler finds no program impact after a mere 6 months of program exposure for children under 3 years of age, but with 24 months of program exposure the illness rate among the treatment group was about 40 percent lower than the rate among the control group, a difference that is significant at the 1 percent level.

The interaction of age effects and the duration of exposure has been examined also by Pitt, Rosenzweig, and Gibbons (1993) and by Duflo (2001) in Indonesia

and by Chin (2005) in India in their evaluations of schooling programs. These studies use information on the region and year of birth of children, combined with administrative data on the year and placement of programs, to measure duration of program exposure. Duflo (2001), for example, estimates the impact of a massive school construction program on subsequent schooling attainment and on the wages of the birth cohorts affected by the program in Indonesia. From 1973 to 1978 more than 61,000 primary schools were built throughout the country, and the enrollment rate among children aged 7–12 rose from 69 percent to 83 percent. By linking district-level data on the number of new schools by year and matching these data with intercensal survey data on men born between 1950 and 1972, Duflo defines how long an individual was exposed to the program. The impact estimates indicate that each new school per 1,000 children increased years of education by 0.12–0.19 percent among the first cohort *fully* exposed to the program.

Chin (2005) uses a similar approach in estimating the impact of India's Operation Blackboard. Taking grades 1–5 as the primary school grades, ages 6–10 as the corresponding primary school ages, and 1988 as the first year that schools would have received program resources, Chin supposes that only students born in 1978 or later would have been of primary school age for at least one year in the program regime, and therefore were potentially exposed to the program for most of their schooling. The evaluation compares two birth cohorts: a younger cohort born between 1978 and 1983, and therefore potentially exposed to the program, and an older cohort. The impact estimates suggest that accounting for duration somewhat lowers the impact as measured, but it remains statistically significant, though only for girls.

Conclusions

This paper has focused on the dimensions of timing and duration of exposure that relate to program or policy implementation. Impact evaluations of social programs or policies typically ignore these dimensions; they assume that interventions occur at a specified date and produce intended or predictable changes in conditions among the beneficiary groups. This is perhaps a reasonable assumption when the intervention itself occurs within a very short time period and has an immediate effect, such as some immunization programs, or is completely under the direction and control of the evaluator, as in small pilot programs. In the examples we have cited (India's Green Revolution, Mexico's PROGRESA conditional cash transfer program, Madagascar's child nutrition SEECALINE program, and an early childhood development program in the Philippines, among others), this is far from true. Indeed, initial operational fits and starts in most programs, and a learning process for program operators and beneficiaries, can delay

full program effectiveness; also, there are many reasons why these delays are not likely to be the same across program sites.

We have catalogued sources of the variation in the duration of program exposure across treatment areas and beneficiaries, including program design features that have built-in waiting periods, lags in implementation due to administrative or bureaucratic procedures, spillover effects, and the interaction between sources of heterogeneity in impact and duration of exposure. Some evaluations demonstrate that accounting for these variations in length of program exposure alters impact estimates significantly, so ignoring these variations can generate misleading conclusions about an intervention. Appendix Table A-1 indicates that a number of impact evaluation studies do incorporate one or more of these timing and duration effects. The most commonly addressed source of duration effects is cohort affiliation. This is not surprising, since many interventions, such as education and nutrition programs, are allocated on the basis of age, in terms of both timing of entry into and exit from the program. On the other hand, implementation lags are recognized but often not explicitly addressed.

What can be done to capture timing and the variation in length of program exposure? First, *the quality of program data should be improved.* Such data could come from administrative records on the design and implementation details of a program, in combination with survey data on program take-up by beneficiaries. Program data on the timing of implementation are likely to be available from program management units, but these data may not be available at the desired level of disaggregation—this might be the district, community, providers, or individual, depending on where the variation in timing is thought to be the greatest. Compiling such data on large programs that decentralize to numerous local offices could be costly. There is obviously a difference in the primary concern of the high-level program manager and of the evaluator. The program manager's concern is the disbursement of project funds and the procurement of major expenditure items, whereas the evaluator's concern would be to ascertain when the funds and inputs reach treatment areas or beneficiaries.

Second, *the timing of the evaluation should take into account the time path of program impacts.* Figure 1 illustrates that program impact, however measured, can change over time, for various reasons discussed in the paper, so there are risks of not finding significant impact when a program is evaluated too early or too late. The learning process by program operators or by beneficiaries could produce a curve showing increasing impact over time, while a pioneering effect could show a very early steep rise in program impact that is not sustainable. Figure 1 thus suggests that early rapid assessments to judge the success of a program could be misleading, and also that repeated observations may be necessary to estimate true impact. Several studies that we reviewed measured their outcomes of interest more than once after the start of the treatment, and some compared short-run

and long-run effects to examine whether the short-run impact had persisted. Possible changes in impact over time imply that evaluations should not be a once-off activity for any long-lasting program or policy. In fact, as discussed above, examining long-term impacts could point to valuable lessons about the diffusion of good practices over time (Foster and Rosenzweig 1995) or, sadly, how governments can reduce impact by implementing other policies that (perhaps unintentionally) disadvantage the program areas (Chen, Mu, and Ravallion 2008).

Third, *the appropriate evaluation method applied should take into account the source of variation in duration of program exposure.* Impact estimates are affected by the length of program exposure, depending on whether or not the source of variation in duration is common within a treatment area and whether or not this source is a random phenomenon. Some pointers are: If the length of implementation lags is about equal across treatment sites, then a simple comparison between the beneficiaries in the treatment and control areas would be sufficient to estimate the average impact of the program or the ITT effects under many conditions—though not if there are significant learning or pioneering effects that differ across them. If the delays vary across treatment areas but not within those areas—and if the variation is random or independent of unobservable characteristics in the program areas that may also affect program effectiveness—then it is also possible to estimate the ITT effects with appropriate controls for the area, or with fixed effects for different exposure categories. In cases where the intervention and its evaluation are designed together, such as pilot programs, it is possible and desirable to explore the time path of program impact by allocating treatment groups to different lengths of exposure in a randomized way. This treatment allocation on the basis of duration differences can yield useful operational lessons about program design, so it deserves more experimentation in the future.

Appendix

Table A-1. Examples of Evaluations That Consider Timing Issues and Duration of Program Exposure in Estimating Program Impact

			Sources of variation in timing and duration of program exposure					
Studies	*Country*	*Intervention*	*Implementation lags*	*Short-run and long-run outcomes*	*Learning by beneficiaries*	*Learning and use by beneficiaries*	*Cohort effects*	*Cohort interacted with duration of exposure*
Angrist and others (2002)	Colombia	School voucher program for secondary level		x		x		
Angrist, Bettinger, and Kremer (2004)								
Armecin and others (2006)	Philippines	Comprehensive early childhood development program (ECD)	x				x	x
Banerjee and others (2007)	India	Balsakhi school remedial and computer-assisted learning programs		x				x
Behrman and Hoddinott (2005)	Mexico	PROGRESA nutrition intervention				x		
Behrman, Cheng, and Todd (2004)	Bolivia	PIDI preschool program					x	x
Behrman, Sengupta, and Todd (2005)	Mexico	PROGRESA education intervention					x	
Schultz (2004)								
Chin (2005)	India	Operation Blackboard: additional teachers per school	x				x	x
Duflo (2001)	Indonesia	School construction program		x			x	x

Continued

Table A-1. *Continued*

			Sources of variation in timing and duration of program exposure					
Studies	*Country*	*Intervention*	*Implementation lags*	*Short-run and long-run outcomes*	*Learning by beneficiaries*	*Learning and use by beneficiaries*	*Cohort effects*	*Cohort interacted with duration of exposure*
Foster and Rosenzweig (1995)	India	Green Revolution: new seed varieties				x		
Frankenberg, Suriastini, and Thomas (2005)	Indonesia	Midwife in the Village program					x	x
Galasso and Yau (2006)	Madagascar	SEECALINE child nutrition program	x		x	x		x
Garces, Thomas, and Currie (2002)	United States	Head Start program: ECD		x				
Hoddinott and Kinsey (2001)	Zimbabwe	Drought shocks; civil war		x			x	x
Alderman, Hoddinott, and Kinsey (2006)								
Jimenez and Sawada (1999)	El Salvador	EDUCO schools: community participation			x			

Gertler (2004)	Mexico	PROGRESA health and nutrition services				x	x
King and Özler (1998)	Nicaragua	School autonomy reform		x			
Rivarola and Fuller (1999)							
Miguel and Kremer (2003, 2004)	Kenya	School-based deworming program			x		
Mu and van de Walle (2007)	Vietnam	Rural roads rehabilitation project	x		x		
Munshi (2004)	India	Green Revolution: new seed varieties			x		
Rouse and Krueger (2004)	United States	Fast ForWord program: computer-assisted learning					x
Walker and others (2005)	Jamaica	Nutrition supplementation	x			x	

Note: Review articles on early childhood development programs—for example, Engle and others (2007) and Grantham-McGregor and others (2007)—cover a long list of studies that we mention in the text but are not listed in this table; many of those studies examine age-specific effects, and some examine short- and long-run impacts.

Notes

Elizabeth M. King (corresponding author) is Research Manager, Development Research Group, at the World Bank; her address for correspondence is eking@worldbank.org. Jere R. Behrman is Professor, Department of Economics, at the University of Pennsylvania. The authors are grateful to Laura Chioda and to three anonymous referees for helpful comments on a previous draft. All remaining errors are ours.

1. For instance, the International Initiative for Impact Evaluation (3IE) has been set up by governments of several countries, donor agencies, and private foundations to address the desire of the development community to build up systematically more evidence about effective interventions.

2. There have been excellent reviews of the choice of methods as applied to social programs. See, for example, Grossman (1994), Heckman and Smith (1995), Ravallion (2001), Cobb-Clark and Crossley (2003), and Duflo (2004).

3. To keep the discussion focused on the timing issue and the duration of exposure, we avoid discussing the specific evaluation method (or methods) used by the empirical studies that we cite. However, we restrict our selection of studies to review to those that have a sound evaluation design, whether experimental or using econometric techniques. Nor do we discuss estimation issues such as sample attrition bias, which is one of the ways in which a duration issue has been taken into account in the evaluation literature.

4. See Heckman, Lalonde, and Smith (1999) for a review.

5. Because Rouse and Krueger (2004) define the treatment group more stringently, however, the counterfactual treatment received by the control students becomes more mixed, and a share of these students is contaminated by partial participation in the program.

6. In their assessment of the returns to World Bank investment projects, Pohl and Mihaljek (1992) cite construction delays among the risks that account for a wedge between ex ante (appraisal) estimates and ex post estimates of rates of returns. They estimate that, on average, projects take considerably more time to implement than expected at appraisal: six years rather than four years.

7. In Mexico's well-known PROGRESA program, payment records from an evaluation sample showed that 27 percent of the eligible population had not received benefits after almost two years of program operation, possibly as a result of delays in setting up the program's management information system (Rawlings and Rubio 2005). In Ecuador's Bono de Desarrollo Humano, the lists of the beneficiaries who had been allocated the transfer through a lottery did not reach program operators, and so about 30 percent of them did not take up the program (Schady and Araujo 2008).

8. Chin (2005) makes a one-year adjustment in her evaluation of Operation Blackboard in India. Although the Indian government allocated and disbursed funds for the program for the first time in fiscal 1987, not all schools received program resources until the following school year. In addition to the delay in implementation, Chin also finds that only one-quarter and one-half of the project teachers were sent to one-teacher schools, while the remaining project teachers were used in ways the central government had not intended. Apparently, the state and local governments had exercised their discretion in the use of the OB teachers.

9. In two programs that we know, administrative records at the individual level were maintained at local program offices, not at a central program office, and local record-keeping varied in quality and form (for example, some records were computerized and some were not), so that a major effort was required to collect and check records during the evaluations.

10. Leonard (2008) provides an example of such an effect. He uses the presence of a Hawthorne effect, produced by the unexpected arrival of a research team to observe a physician, in order to measure an exogenous, short-term change in the quality of service provided by a physician. Indeed, there was a significant jump in quality upon the arrival of observers, but quality returned to previsit levels after some time.

11. Information campaigns for programs that attempt to improve primary-school quality or to enhance child nutrition through primary-school feeding programs in a context in which virtually

all primary-school-age children are already enrolled would seem less relevant than such campaigns as part of a new program to improve preschool child development where there had previously been no preschool programs.

12. The authors estimate the impact of atypically low rainfall levels by including a year's delay because the food shortages would be apparent only one year after the drought, but before the next harvest was ready.

13. To estimate these longer-run impacts, Alderman, Hoddinott, and Kinsey (2006) combine data on children's ages with information on the duration of the civil war and the episodes of drought used in their analysis. They undertook a new household survey to trace children measured in earlier surveys.

References

Alderman, Harold, John Hoddinott, and William Kinsey. 2006. "Long Term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers* 58(3):450–74.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2004. "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." National Bureau of Economic Research Working Paper No. 10713, August.

Angrist, Joshua D., Eric Bettinger, Erik Bloom, Elizabeth M. King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92(5):1535–58.

Armecin, Graeme, Jere R. Behrman, Paulita Duazo, Sharon Ghuman, Socorro Gultiano, Elizabeth M. King, and Nannette Lee. 2006. *Early Childhood Development through an Integrated Program: Evidence from the Philippines.* Policy Research Working Paper 3922. Washington, DC: World Bank.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122(3): 1235–64.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2008. *Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India.* Policy Research Working Paper 4584. Washington, DC: World Bank.

Behrman, Jere R., and John Hoddinott. 2005. "Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican 'Progresa' Impact on Child Nutrition." *Oxford Bulletin of Economics and Statistics* 67(4):547–69.

Behrman, Jere R., Yingmei Cheng, and Petra E. Todd. 2004. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics* 86(1):108–32.

Behrman, Jere R., Piyali Sengupta, and Petra Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of Mexico's School Subsidy Experiment." *Economic Development and Cultural Change* 54(1):237–75.

Chen, Shaohua, Ren Mu, and Martin Ravallion. 2008. *Are There Lasting Impacts of Aid to Poor Areas?* Policy Research Working Paper 4084. Washington, DC: World Bank.

Chin, Aimee. 2005. "Can Redistributing Teachers across Schools Raise Educational Attainment? Evidence from Operation Blackboard in India." *Journal of Development Economics* 78(2):384–405.

Cleland, John, Stan Bernstein, Alex Ezeh, Anibal Faundes, Anna Glasier, and Jolene Innis. 2006. "Family Planning: The Unfinished Agenda." *Lancet* 368(November):1810–27.

Cobb-Clark, Deborah A., and Thomas Crossley. 2003. "Econometrics for Evaluations: An Introduction to Recent Developments." *Economic Record* 79(247):491–511.

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.

———. 2004. "Scaling Up and Evaluation." InF. Bourguignon, and B. Pleskoviceds., *Annual World Bank Conference on Development Economics: Accelerating Development*. Washington, DC: World Bank.

Engle, Patrice L., Maureen M. Black, Jere R. Behrman, Meena Cabral de Mello, Paul J. Gertler, Lydia Kapiriri, and Reynaldo Martorell, and Mary Eming Young, and the International Child Development Steering Group. 2007. "Strategies to Avoid the Loss of Developmental Potential in More Than 200 Million Children in the Developing World." *Lancet* 369(January):229–42.

Foster, Andrew D., and Mark R. Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103(6): 1176–1209.

Frankenberg, Elizabeth, Wayan Suriastini, and Duncan Thomas. 2005. "Can Expanding Access to Basic Health Care Improve Children's Health Status? Lessons from Indonesia's 'Midwife in the Village' Programme." *Population Studies* 59(1):5–19.

Galasso, Emanuela, and Jeffrey Yau. 2006. *Learning through Monitoring: Lessons from a Large-Scale Nutrition Program in Madagascar*. Policy Research Working Paper 4058. Washington, DC: World Bank.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92(4):999–1012.

Gertler, Paul J. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from Progresa's Control Randomized Experiment." *American Economic Review* 94(2):336–41.

Grantham-McGregor, Sally, Yin Bun Cheung, Santiago Cueto, Paul Glewwe, Linda Richter, and Barbara Strupp, and the International Child Development Steering Group. 2007. "Developmental Potential in the First 5 Years for Children in Developing Countries." *Lancet* 369(9555):60–70.

Grossman, Jean Baldwin.. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *World Bank Research Observer* 9(2):159–80.

Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9(2):85–110.

Heckman, James J., R. J. Lalonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In Orley Ashenfelter, and David Card, eds., *Handbook of Labor Economics*. Vol. 1. Amsterdam: North-Holland.

Hoddinott, John, and William Kinsey. 2001. "Child Growth in the Time of Drought." *Oxford Bulletin of Economics and Statistics* 63(4):409–36.

Jimenez, Emmanuel, and Yasuyuki Sawada. 1999. "Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program." *World Bank Economic Review* 13(3):415–41.

Kikumbih, Nassor, Kara Hanson, Anne Mills, Hadji Mponda, and Joanna Armstrong Schellenberg. 2005. "The Economics of Social Marketing: The Case of Mosquito Nets in Tanzania." *Social Science and Medicine* 60(2):369–81.

King, Elizabeth M., and Berk Özler. 1998. *What's Decentralization Got to Do with Learning? The Case of Nicaragua's School Autonomy Reform*. Working Paper on Impact Evaluation of Education Reforms 9 (June). Washington, DC: Development Research Group, World Bank.

King, Elizabeth M., Peter F. Orazem, and Darin Wohlgemuth. 1999. "Central Mandates and Local Incentives: Colombia's Targeted Voucher Program." *World Bank Economic Review* 13(3):467–91.

Leonard, Kenneth L.. 2008. "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect." *Journal of Health Economics* 27(2):444–59.

Miguel, Edward, and Michael Kremer. 2003. "Social Networks and Learning about Health in Kenya." National Bureau of Economic Research Working Paper and Center for Global Development, manuscript, July 2003.

_____. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1):159–217.

Mu, Ren, and Dominique van de Walle. 2007. *Rural Roads and Poor Area Development in Vietnam.* Policy Research Working Paper Series 4340. World Bank.

Munshi, Kaivan. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics* 73(1):185–213.

Newman, John L., Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16(2):241–74.

Pitt, Mark M., Mark R. Rosenzweig, and Donna M. Gibbons. 1993. "The Determinants and Consequences of the Placement of Government Programs in Indonesia." *World Bank Economic Review* 7(3):319–48.

Pohl, Gerhard, and Dubravko Mihaljek. 1992. "Project Evaluation and Uncertainty in Practice: A Statistical Analysis of Rate-of-Return Divergences of 1,015 World Bank Projects." *World Bank Economic Review* 6(2):255–77.

Ravallion, Martin. 2001. "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation." *World Bank Economic Review* 150(1):115–40.

Ravallion, Martin.. 2002. "Are the Poor Protected from Budget Cuts? Evidence for Argentina." *Journal of Applied Economics* 5(1):95–121.

Ravallion, Martin, Emanuela Galasso, Teodoro Lazo, and Ernesto Philipp. 2005. "What Can Ex-Participants Reveal about a Program's Impact?" *Journal of Human Resources* 40(1):208–30.

Rawlings, Laura B., and Gloria M. Rubio. 2005. "Evaluating the Impact of Conditional Cash Transfer Programs." *World Bank Research Observer* 20(1):29–55.

Rivarola, Magdalena, and Bruce Fuller. 1999. "Nicaragua's Experiment to Decentralize Schools: Contrasting Views of Parents, Teachers, and Directors." *Comparative Education Review* 43(4): 489–521.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 1986. "Evaluating the Effects of Optimally Distributed Public Programs: Child Health and Family Planning Interventions." *American Economic Review* 76(3):470–82.

Rouse, Cecilia Elena, and Alan B. Krueger. 2004. "Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically Based' Reading Program." *Economics of Education Review* 23(4):323–38.

Rowland, Mark, Jayne Webster, Padshah Saleh, Daniel Chandramohan, Tim Freeman, Barbara Pearcy, Naeem Durrani, Abdur Rab, and Nasir Mohammed. 2002. "Prevention of Malaria in Afghanistan through Social Marketing of Insecticide-Treated Nets: Evaluation of Coverage and Effectiveness by Cross-sectional Surveys and Passive Surveillance." *Tropical Medicine and International Health* 7(10):813–22.

Schady, Norbert, and Maria Caridad Araujo. 2008. "Cash Transfers, Conditions, and School Enrollment in Ecuador." *Economía* 8(2):43–70.

Schultz, T. Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74(2):199–250.

Walker, Susan P., Susan M. Chang, Christine A. Powell, and Sally M. Grantham-McGregor. 2005. "Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-Stunted Jamaican Children: Prospective Cohort Study." *Lancet* 366(9499):1804–07.

Competition in the Financial Sector: Overview of Competition Policies

Stijn Claessens

Competition in the financial sector, as in other sectors, matters for allocative, productive, and dynamic efficiency. Theory suggests, however, that unfettered competition is not necessarily best given the special features of financial services. The author discusses these analytical complications before reviewing how to assess competition in the financial sector and its determinants. It is shown that competitiveness varies greatly across countries, in perhaps surprising ways, and that it is not driven by financial system concentration. Rather, systems with greater foreign entry and fewer entry and activity restrictions tend to be more competitive, confirming that contestability—the lack of barriers to entry and exit—determines effective competition. The author then analyzes how competition policy in the financial sector has generally been conducted and how changes in competition in the financial services industries should affect competition policy going forward. In part based on comparison with other industries, the author provides some suggestions on how competition policy in the financial sector could be better approached as well as what institutional arrangements best fit a modern view of competition policy in the sector. The specific competition challenges for developing countries is also highlighted. The author concludes that practices today fall far short of the need for better competition policy in the financial sector. JEL codes: G10, G18, G28, L1, L5

Competition in the financial sector matters for a number of reasons. As in other industries, the degree of competition in the financial sector matters for the efficiency of the production of financial services, the quality of financial products, and the degree of innovation in the sector. The view that competition in financial services is unambiguously good, however, is more naive than is the case in other industries: vigorous rivalry here may not be the best approach. Specific to the financial sector is the effect of excessive competition on financial stability, long recognized in theoretical and empirical research and, most importantly, in the

doi:10.1093/wbro/lkp004 Advance Access publication April 9, 2009 24:83–118

actual conduct of (prudential) policy toward banks. There are other complications, however. It has been shown, theoretically and empirically, that the degree of competition in the financial sector can matter (negatively or positively) for the access of firms and households to financial services, which in turn affects overall economic growth.

In terms of the factors driving competition in the financial sector and the empirical measurement of competition, one needs to consider the standard industrial organization factors, such as entry/exit and contestability. But financial services provision also has many network properties: in their production (for example the use of information networks), in their distribution (for example the use of ATMs), and in their consumption (for example the large externalities of stock exchanges and the agglomeration effects in liquidity). This makes for complex competition structures since aspects such as the availability of networks used or the first mover advantage in introducing financial contracts become important.

Not only are many of the relationships and trade-offs among competition, financial system performance, access to financing, stability, and growth complex from a theoretical perspective, but empirical evidence on competition in the financial sector has been scarce, and to the extent that it is available it is often not (yet) clear. What is evident from theory and empirics, however, is that these trade-offs mean that it is not sufficient to analyze competitiveness from a narrow concept alone or to focus on one effect only. One has to consider competition as part of a broad set of objectives—including financial sector efficiency, access to financial services for various segments of users, and systemic financial sector stability—and to consider possible trade-offs among these objectives. And since competition depends on several factors, one has to give thought to a broad set of policy tools when trying to increase competition in the financial sector.

In all, this means that competition policy in the financial sector is complex and can be hard to analyze. Empirical research on competition in the sector is also still at an early stage. The evidence nevertheless shows that competition and factors driving competition have been important aspects of recent improvements. To date, greater competition has been achieved by traditional means: removing entry barriers, liberalizing product restrictions, abolishing restrictive market definitions, eliminating intrasectoral restrictions, and so on. Making in this way financial systems more open and contestable, that is having low barriers to entry and exit, has generally led to greater product differentiation, lower cost of financial intermediation, more access to financial services, and enhanced stability. The evidence for these effects is fairly universal, from the United States, the European Union, and other industrialized countries to many developing countries.

As globalization, technological improvements, and deregulation advance, the gains of competition can be expected to become even more widespread across and

within countries. At the same time, once the easier steps have been taken, policies to achieve effective competition in all dimensions, and balancing the trade-offs between competition and other concerns, become more challenging. The rapid competitive gains due to the first rounds of liberalization over the past few decades will be hard to sustain in the future. Complexity will also become greater as financial services industries evolve, as financial markets and products become more complex and global, and as new regulatory and competition policy issues arise. The many new forms of financial services provision means all the more that approaches to competition issues need to be adjusted.

This is the more important since competition policy in the sector is often already behind that of many other sectors in many countries. Too often, competition is seen as an afterthought, rather than being considered an essential ingredient in the having of a strategy for the development of the sector. To assure markets remain and become even more competitive will require taking into account the special properties of financial markets, including the existence of many networks in finance. But here the theoretical and empirical literature is just catching up with the changes. And competition policy will become more difficult institutionally to organize, both within and across countries, yet it remains necessary given the global dimensions of many financial markets these days. Furthermore, financial systems are often entrenched, in developing countries especially, including through links between the financial and real sectors, and in certain odious relationships with the political sector, all of which can make achieving effective competition a complex task.

Recent events in global financial markets, while too recent to draw firm conclusions, highlight some issues on which it will be necessary to reflect. The global dimensions of the financial crisis clearly confirm the need for many policies, whether aimed at stability or at improving market functioning, to operate in a consistent manner across jurisdictions, especially for systemically important financial institutions and activities. The crisis also makes clear the need for a more holistic approach to prudential regulation, at both the institutional and macroeconomic level, to address wider systemic risks. This objective will likely require measures aimed at strengthening capital and liquidity requirements for individual institutions, avoiding the build-up of systemic risk across institutions and the economy over time, and improving national and international resources and financial sector responses to distress. The financial crisis has also shown the need to strengthen market discipline, address key information gaps, and encourage more robust private governance and risk management systems. One other lesson the financial crisis calls for is to revisit the institutional infrastructure for financial services provision, including the role of rating agencies and the need for derivatives trading to move to more regulated markets. The crisis also confirms the need for competition policy to adjust and adapt to developments in the

markets. As some have suggested, in the context of regulatory failures and weaknesses in private market discipline, increased competition can lead to excessive risk-taking, implying the need for competition policy to consider broader aspects. When considering these and other changes, the new architecture will need to take into account the inherent limitations of regulation and supervision, and to guard against over-regulation.

I will review the state of knowledge on these issues and how competition policy is and should be organized. In the next section I provide a review of the literature, concerning both the nature and effects of competition in the financial sector as well as how to go about measuring competition in general and in the financial sector specifically. I will discuss the approaches and methodology used to test for the degree of competition in a particular country or market, present some data on measures that are starting to be used for assessing competitiveness, show how these measures relate to structural and policy variables, and suggest what tools might be used to measure competition. I will also discuss the current state of affairs in competition policy and how changes in the financial services industries that are underway affect the nature of competition. In the following section I will assess the implications for competition policy, how to approach it, and how to organize it. In the last section, I will present my conclusions, although many are not definitive. I will stress, however, that practices in many countries fall far short of the important need for better competition policy in the financial sector.

Nature and Status of Financial Sector Competition

What is special about competition in the financial sector? And how does competition matter? The two questions are closely related and depend in turn on what dimensions one analyzes. For the purpose of this paper, I consider the links between competition and the following three dimensions: financial sector development (including the efficiency of financial services provision); access to financial services for households and firms (that is the availability, or lack thereof, of financial services at reasonable cost and convenience); and financial sector stability (that is the absence of systemic disturbances that have major real sector impact).

Regarding the first link, development and efficiency, one can consider questions like: With greater competition, is the system more developed? For example is it larger, does it provide better quality financial products/services in a static and dynamic way? Is it more efficient, that is exhibits a lower cost of financial intermediation? Is it less profitable? Is it closer to some competitive benchmark? Concerning the second link, one can consider whether access to financing—particularly for smaller firms and poorer individuals, but also in general for households, large firms, and other agents—is improved, in terms of volume and

costs, with greater competition. In terms of the stability link, one can consider whether the banking system has less volatility, fewer financial crises, and is generally more robust and its financial integrity higher with more competition.

I will consider what theory predicts for each of these three dimensions. Since all are important, and that there can be relationships among them, individual analyses will be incomplete.[1] I will next review what both theory and empirics predict on what drives competition in the financial sector. I analyze specifically the internationalization of financial services, which is growing rapidly and which has had an especially large impact on competition in the sector in many developing countries. Lastly, I will suggest what these theoretical and empirical findings suggest in terms of what tools regulators should use for the application of competition policy.

Effects of Competition in the Financial Sector: Theory

As in other sectors, one can expect competition policy to have effects on the development and efficiency of the financial sector, in both static and dynamic ways. Special to the financial sector is that the degree of competition can affect the access to financial services of borrows and other types of consumers of financial services. And, as has been long recognized, and recently confirmed with the global financial crisis, the degree of competition can affect the stability of a financial system.

Development and efficiency, static and dynamic. As a first-order effect, one expects increased competition in the financial sector to lead to lower costs and enhanced efficiency of financial intermediation, to greater product innovation, and to improved quality. Even though the financial services have some special properties, the channels are similar to other industries. In a theoretical model, Besanko and Thakor (1992), for example, allowing for the fact that financial products are heterogeneous, analyzed the allocational consequences of relaxing entry barriers. They found that equilibrium loan rates decline and deposit interest rates increase, even when allowing for differentiated competition. In turn, by lowering the costs of financial intermediation, and thus lowering the cost of capital for non-financial firms, more competitive banking systems lead to higher growth rates. Of course, it is not just efficiency and costs, but also the incentives of institutions and markets to innovate, that are likely to be affected by the degree of competition.

Access to financial services. As a first-order effect, greater development, lower costs, enhanced efficiency, and a greater and wider supply resulting from competition will lead to greater access. The relationships between competition and banking system performance in terms of access to financing are more complex, however.

In the theoretical literature it has been analyzed how access can depend on the franchise value of financial institutions and how the general degree of competition can negatively or positively affect access. Market power in banking, for example, may, to a degree, be beneficial for access to financing (Petersen and Rajan 1995). With too much competition, banks may be less inclined to invest in relationship lending (Rajan 1992). At the same time, because of hold-up problems, too little competition may tie borrowers too much to an individual institution, making the borrower less willing to enter a relationship (Petersen and Rajan 1994; Boot and Thakor 2000). More competition can then, even with relationship lending, lead to more access.

The quality of information can interact with the size and structure of the financial system to affect the degree of access to financial services. Financial system consolidation can lead to a greater distance and thereby to less lending to more opaque firms such as small and medium-sized enterprises. Improvements in technology and better information that spur consolidation can be offsetting factors, however. Theory has shown some other complications. Some researchers have highlighted that competition is partly endogenous as financial institutions invest in technology and relationships (for example Hauswald and Marquez 2003). Theory has also shown that technological progress lowering production or distribution costs for financial services providers does not necessarily lead to more or better access to finance. Models often end up with ambiguous effects of technological innovations, access to information, and the dynamic pattern of entry and exit on competition, access, stability, and efficiency (for example Marquez 2002; Dell'Ariccia and Marquez 2004). Increased competition can, for example, lead to more access, but also to weaker lending standards, as observed recently in the subprime lending market in the US (Dell'Ariccia and others 2008) but also in other episodes.

These effects are further complicated by the fact that network effects exist in many aspects of the supply, demand, or distribution of financial services. In financial services production, much use is made of information networks (for example, credit bureaus). In distribution, networks are also extensively used (for example, the use of ATMs). Furthermore, in their consumption, many financial services display network properties (for example, liquidity in stock exchanges). As for other network industries, this makes competition complex (see Ausubel 1991; Claessens and others 2003).

Stability. The relationships between competition and stability are also not obvious. Many academics and especially policymakers have stressed the importance of franchise value for banks in maintaining incentives for prudent behavior. This in turn has led banking regulators to balance carefully entry and exit. Licensing, for example, is in part used as a prudential policy, but often with little

regard for its impact on competition. This has often been a static view, however. Perotti and Suarez (2002) showed in a formal model that the behavior of banks today will be affected by both current and future market structure and the degree to which authorities will allow for a contestable (open) system in the future. In such a dynamic model, current concentration does not necessarily reduce risky lending, but an expected increase in future market concentration can make banks choose to pursue safer lending today. More generally, there may not be a trade-off between stability and increased competition, as shown among others by Allen and Gale (2004) and Boyd and De Nicolò (2005) and reviewed recently by Allen and Gale (2007). Allen and Gale (2004), furthermore, showed that financial crises, possibly related to the degree of competition, are not necessarily harmful for growth.

The Determinants of Competition and Assessing Competition: Theory and Empirics

I will first review what theory predicts is the driver of competition, in general and in the financial sector specifically, and then what theory suggests as to how best to measure competition and the tools that can be used.

Theory of the determinants of competition. In terms of empirical measurement and associated factors driving competition one can consider three types of approaches: market structure and associated indicators; contestability and regulatory indicators to gauge contestability; and formal competition measures. Much attention in policy context and empirical tests is given to market structure and the actual degree of entry and exit in particular markets as determining the degree of competition. The general Structure-Conduct-Performance (SCP) paradigm, the dominant paradigm in industrial organization from 1950 until the 1970s, made links between structure and performance. "Structure" refers to market structure defined mainly by the concentration in the market. "Conduct" refers to the behavior of firms—competitive or collusive—in various dimensions (pricing, R&D, advertising, production, choice of technology, entry barriers, predation, and so on). "Performance" refers to (social) efficiency, mainly defined by extent of market power, with greater market power implying lower efficiency. The paradigm was based on the hypotheses that i) structure influences conduct (for example, lower concentration leads to more competitive behavior of firms); ii) conduct influences performance (for example, more competitive behavior leads to less market power and greater social efficiency); iii) structure therefore influences performance (for example, lower concentration leads to lower market power).[2]

Theoretically and empirically there are a number of problems with the SCP paradigm and its implications that, directly and indirectly, structure determines

performance. For one, structure is not (necessarily) exogenous since market structure itself is affected by firms' conduct and hence by performance. Another conceptual problem is that industries with rapid technological innovation and much creative destruction, like the financial sector, may have high concentration and market power, but this is necessary to compensate these firms for their innovation and investment and does not mean reduced social welfare. Most importantly, and different from the SCP paradigm, the more general competition and contestability theory suggests that market structure and actual degree of entry or exit are not necessarily the most important factors in determining competition. The degree of contestability, that is, the degree of absence of entry and exit barriers, rather than actual entry, matters for competitiveness (Baumol and others 1982). Contestable markets are characterized by operating under the threat of entry. If a firm in a market with no entry or exit barriers raises its prices above the marginal cost and begins to earn abnormal profits, potential rivals will enter the market to take advantage of these profits. When the incumbent firm(s) respond(s) by returning prices to levels consistent with normal profits, the new firms will exit. In this manner, even a single-firm market can show highly competitive behavior.

The theory of contestable markets has also drawn attention to the fact that there are several sets of conditions that can yield competitive outcomes, with these possible even in concentrated systems since it does not mean that the firm is harming consumers by earning super-normal profits. On the other hand, collusive actions can be sustained even in the presence of many firms. The applicability of the contestability theory to specific situations can vary, however, particularly as there are very few markets which are completely free of sunk costs and entry and exit barriers. Theory specific to the financial sector adds to this some specific considerations. While the threat of entry or exit can also be an important determinant of the behavior of financial market participants, issues such as information asymmetries, investment in relationships, the role of technology, networks, prudential concerns, and other factors can matter as well for determining the effective degree of competition (Bikker and Spierdijk 2008).

Empirical approaches for measuring competition. There are three approaches that have been proposed for measuring competition. The first empirical approach considers factors such as financial system concentration, the number of banks, or Herfindahl indices. It relies on the SCP paradigm, that there are relationships between structure and conduct and performance, but it does not directly gauge banks' behavior. The second approach considers regulatory indicators to gauge the degree of contestability. It takes into account entry requirements, formal and informal barriers to entry for domestic and foreign banks, activity restrictions, and so on. It also considers changes over time in financial instruments, innovations, and so on, as these can lead to changes in the competitive landscape.

The third approach uses formal competition measures, such as the so-called H-statistics, that proxy the reaction of output to input prices. These formal competition measures are theoretically well-motivated and have often been used in other industries, but they do impose assumptions on (financial intermediaries') costs and production functions.

In terms of the first two approaches, the theory has made clear that documenting an industry's structure, the degree of competition, its determinants, and its impacts can be complicated. For one, the competitiveness of an industry cannot be measured by market structure indicators alone, such as the number of institutions or concentration indexes. And no specific market concentration measure is best: neither the number of firms, nor the market share of the top three or five, or the often used Herfindahl index is necessarily the best. Second, traditional performance measures used in finance, such as the size of banks' net interest margins or profitability or transaction costs in stock markets, do not necessarily indicate the competitiveness of a financial system. These performance measures are also influenced by a number of factors, such as a country's macroperformance and stability, the form and degree of taxation of financial intermediation, the quality of a country's information and judicial systems, and financial institution specific factors, such as leverage, the scale of operations, and risk preferences. As such, these measures can be poor indicators of the degree of competition. Yet, they have often been so used as such in spite of these weaknesses. Fortunately, general structure and performance measures have declined in empirical studies in favor of more specific tests.

Indeed, the third approach, emphasizing that documenting the degree of competition requires specific measures and techniques, has become more used. It points out that one needs to study actual behavior—in terms of marginal revenue, pass-through cost pricing, and so on.—using a model and develop from there a specific measure of competitiveness. While such theoretical well-founded tests have been conducted for many industries, they were at an early stage a decade or so ago for the financial sector, particularly cross-country (Cetorelli 1999). More and more, however, formal empirical tests for competition are being applied to the financial sector, mostly to banking systems in individual countries (see Bikker and Spierdjik 2008 for a review). Data problems were previously a hindrance for the cross-country research—since little bank-level data were available outside of the main developed countries—but recently established databases have allowed for better empirical work in comparing countries.

The Pazar and Rosse methodology. Generally, as in other sectors, the degree of competition is measured with respect to the actual behavior of (marginal) bank conduct. Broad cross-country studies using formal methodologies have been undertaken by Claessens and Laeven (2004) and Bikker and Spierdijk (2007).

Using bank-level data and applying the Panzar and Rosse (1987; PR) methodology, the first study estimates the degree of competition in 50 countries' banking systems. Specifically, it investigates the extent to which a change in factor input prices is reflected in (equilibrium) revenues earned by a specific bank. The PR model is, as is typical, estimated using pooled samples for each country.[3]

Under perfect competition, an increase in input prices raises both marginal costs and total revenues by the same amount as the rise in costs. Under a monopoly, an increase in input prices will increase marginal costs, reduce equilibrium output, and consequently reduce total revenues. The PR model provides a measure (H-statistic) of the degree of competitiveness of the industry, which is calculated from reduced-form bank-revenue equations as the sum of the elasticities of the total revenue of the banks with respect to the bank's input prices. The H-measure is between 0 and 1, with less than 0 being a collusive (joint monopoly) competition, less than 1 being monopolistic competition, and 1 being perfect competition. It can be shown, if the bank faces a demand with constant elasticity and a Cobb-Douglas technology, that the magnitude of H can be interpreted as an inverse measure of the degree of monopoly power or, alternatively, as a measure of the degree of competition.

The second study, Bikker and Spierdijk (2007), also uses the PR methodology but allows the degree of competition to vary over time; it covers 101 countries. Table 1 documents by individual country the measures of the two studies. The H-statistic of Claessens and Laeven varies generally between 0.50 and 0.85, suggesting that monopolistic competition sometimes approaching full competition is the best description of the degree of competition. The Bikker and Spierdijk data show an even larger variation in the degree of competitiveness across the larger sample of countries, possibly due to their estimation technique allowing for time variation.

While there does not appear to be any strong pattern among types of countries, it is interesting that some of the largest countries (in terms of the number of banks and the general size of their economy) have relatively low values for the H-statistics. In both studies, Japan and the United States, for example, have H-statistics less than 0.5. This may in part be due to the more fragmented banking markets in these countries, where small banks operate in local markets that are less competitive. Since studies find differences between types of banks, especially in countries with a large number of banks, such as the United States, studying all banks may lead to a distorted measure of the overall competitiveness of a banking system.[4] However, even if one computes H-statistics using data on large banks, rather than all banks for countries with many banks, results remain similar.

Other papers that use this methodology mostly also reject both perfect collusion as well as perfect competition, that is, they find mostly evidence of monopolistic competition (for example, Wong and others 2006 for Hong Kong; Gutiérrez de

Table 1. PR Measures (H-statistics) of Competitiveness of Banking Systems around the World

Country	*Bikker and Spierdijk H (at end of the period)*	*Claessens and Laeven H average*
Algeria	0.34	—
Andorra	0.88	—
Argentina	0.55	0.73
Armenia	0.43	—
Australia	0.29	0.80
Austria	−0.05	0.66
Azerbaijan	0.00	—
Bahamas	0.60	—
Bahrain	0.41	—
Bangladesh	0.87	0.69
Belgium	0.73	0.73
Bermuda	0.87	—
Bolivia	0.99	—
Botswana	0.23	—
Brazil	0.55	0.83
Canada	0.20	0.67
Cayman Islands	0.96	—
Chile	0.93	0.66
China PR	1.57	—
Colombia	0.78	0.66
Costa Rica	0.78	0.92
Croatia	0.04	0.56
Cyprus	−0.09	—
Czech Republic	0.82	0.73
Denmark	0.27	0.50
Dominican Republic	0.23	0.72
Ecuador	0.67	0.68
El Salvador	0.45	—
Estonia	0.11	—
Finland	−0.07	—
France	0.82	0.69
Germany	0.80	0.58
Ghana	0.61	—
Greece	0.47	0.76
Hong Kong	−0.04	0.70
Honduras	—	0.81
Hungary	0.79	0.75
Iceland	0.55	—
India	0.49	0.53
Indonesia	−0.06	0.62
Ireland	1.12	—
Israel	0.15	—

(Continued)

Table 1. *Continued*

Country	*Bikker and Spierdijk H (at end of the period)*	*Claessens and Laeven H average*
Italy	0.08	0.60
Ivory Coast	−0.04	—
Japan	0.44	0.47
Jordan	0.33	—
Kazakhstan	0.28	—
Kenya	0.62	0.58
Korea	1.03	—
Kuwait	0.36	—
Latvia	0.52	0.66
Lithuania	0.40	—
Luxembourg	0.37	0.82
Macau	0.23	—
Macedonia	−0.01	—
Malaysia	0.70	0.68
Malta	0.30	—
Mauritius	0.58	—
Mexico	0.37	0.78
Moldova	0.58	—
Monaco	0.41	—
Morocco	0.32	—
Mozambique	0.61	—
Nepal	0.90	—
Netherlands	0.92	0.86
New Zealand	−0.25	—
Nigeria	0.74	0.67
Norway	0.50	0.57
Oman	0.35	—
Pakistan	0.54	0.48
Panama	0.56	0.74
Paraguay	0.75	0.60
Peru	1.37	0.72
Philippines	0.28	0.66
Poland	0.03	0.77
Portugal	−0.02	0.67
Romania	0.59	—
Russian Federation	0.41	0.54
Saudi Arabia	0.51	—
Senegal	0.18	—
Singapore	0.51	—
Slovakia	0.16	—
Slovenia	0.29	—
South Africa	2.03	0.85

(Continued)

Table 1. *Continued*

Country	*Bikker and Spierdijk H (at end of the period)*	*Claessens and Laeven H average*
Spain	0.52	0.53
Sri Lanka	0.67	—
Sweden	−0.08	—
Switzerland	0.74	0.67
Taiwan	0.94	—
Thailand	0.63	—
Trinidad Tobago	0.21	—
Turkey	0.43	0.46
Ukraine	0.44	0.68
United Arab Emirates	0.46	—
United Kingdom	0.76	0.74
United States	0.46	0.41
Uruguay	0.53	—
Venezuela	0.74	0.74
Vietnam	0.74	—
Zambia	0.53	—

— Not available.

Notes: The table displays two measures. The Bikker and Spierdijk measure allows for variation over time and the reported H-statistic for each country is the one estimated for the end of the sample period. The samples used vary considerably across countries. The Claessens and Laeven measure is the estimated average H-statistic for each country in their sample calculated for the years 1994–2001 using the Panzar-Rosse (1987) approach. In this latter case, the H-statistics are based on a sample that includes observations from countries with a total number of at least 50 bank-year observations and observations on at least 20 banks.

Source: Claessens and Laeven (2004) and Bikker and Spierdijk (2007).

Rozas 2007 for Spain; Hempell 2002 for Germany; Bikker and Haaf 2001 summarized the results of some 10 studies; Berger 2000 provided further reviews). Tests for emerging markets are rarer, but those done (for example, Nakane 2001 for Brazil; Prasad and Ghosh 2005 for India; Yildirim and Philippatos 2007 for a sample of Latin America countries) also find evidence of monopolistic competition. There remain large variations across countries, however, and the ability to capture the degree of competition is still imperfect, as estimates vary considerably among studies for the same banking systems (Bikker and others 2006 review a number of studies). This is also clear from table 1 since there can be large differences between the two measures reported for individual countries (the correlation is only 0.38, and the rank correlation only 0.29), showing some of the difficulty in measuring competitiveness.

Empirical approaches for explaining competition. A lesser number of studies have tried to explain the degree of competition in particular markets. Claessens and Laeven (2004) related competitiveness (the H-measure) to indicators of countries'

banking system structures and regulatory regimes. Importantly, and consistent with some other studies, they found no evidence that their measure of competitiveness relates negatively to banking system concentration or the number of banks in the market. They did find systems with greater foreign bank entry, and fewer entry and activity restrictions, to be more competitive. Their findings suggest that measures of market structures do not necessarily translate into effective competition, consistent with the theory that contestability rather than market structure determines effective competition. Others have studied the impact of financial liberalization on the degree of competitiveness and find generally that liberalized systems are more competitive, in the sense of having a higher H-measure.

There are some identification issues here, of course. Just as some studies have found that trade openness raises efficiency in sectors open to foreign competition, it could be that more efficient banking sectors are more likely to allow (external) competition, so that efficiency is the cause, rather than just an effect, of contestability. And there can be omitted variables, as when financial deregulation is adopted along with other efficiency-enhancing measures. Also, there can be political economy arguments creating reverse causality or omitted factors—for example, when insiders prefer closed, but inefficient, financial systems. This means that studies assessing the impact of openness in financial systems and aggregate economic performance may have a hard time identifying the direction of causality and disentangling the effects of financial reforms from those of other measures. These caveats apply to most of the financial liberalization and reform analyses, including those referred to here.

Other empirical regularities. There is a broad literature that has documented many empirical regularities between financial system performance and structural factors within and across countries. This literature has related the actual behavior of financial markets to factors deemed to be related to competition, including not only structure but also entry barriers, as well as foreign ownership and the severity of activity restrictions, since these can limit intra-industry competition. Especially for banking systems, a number of empirical studies have found that the ownership of the entrants and incumbents, and the size and the degree of financial conglomeration (that is, the mixture of banking and other forms of financial services, such as insurance and investment banking), matter in a number of ways.

Many of these studies, however, do not use a structural, contestability approach to measure the actual degree of competitive conduct, and as such cannot indicate whether the underlying behavior is based on competitive or, say, oligopolistic behavior. Furthermore, often the focus has been on the banking system only, neglecting other forms of financial intermediation that have become more directly important in financial intermediation (capital markets, non-bank financial

institutions, insurance companies) or that play a role in determining the competitiveness of the banking system. Nevertheless, this literature suggests that the degree of competition has consequences for financial sector performance and sheds some light on competition issues. The literature can be classified by individual country and regional studies, cross-country studies, and studies on the specific effects of internationalization.

Country and regional studies. Evidence that competition matters is most convincingly available from liberalization steps, that is, when reform "unambiguously" introduces more competition. This has notably been the case in the United States with the abolishment of restrictions on intra- and interstate banking. Strahan (2003), a major contributor himself, reviewed this large literature and noted how it had documented large real effects of U.S. branching deregulation.[5] Besides documenting cross-sectional regularities, many studies have investigated the effects of changes in structure and especially consolidation in financial systems. (For an early review, see Berger and others 1999.) Similar experiences have been documented for the European Union following the Single Banking Directives and other measures aimed at creating more integrated and competitive financial markets (see for example, CEPR 2005, for a review). And for most emerging markets similar effects have been documented (see BIS 2006, for a collection of country experiences).

These experiences have also highlighted the symbiotic relationships between increased competition and changes in regulation: as competition intensifies, regulators are forced to evaluate the remaining rules by market participants, leading to more focused rules. Foreign banks entering local markets have especially been found to stimulate improvements in the quality of regulation and supervision (Levine 1996). As such, foreign bank entry and other forms of international financial liberalization need not necessarily wait until the local institutional environment is fully developed. And greater competition can highlight deficiencies that raise the costs of financial intermediation or hinder access to financial services, and as such it can be an impetus to reform.

There have thus been large and rapid competitive gains in industrialized countries due to intracountry and regional deregulation, and much progress has been made in the financial systems of those developing countries that opened up and experienced large entry, especially those in Central and Eastern Europe and Latin America. Experience has also shown, however, that gains will be hard to maintain in the future, largely as it involved the easy steps of liberalization and opening up. The tasks now are to deepen the competitive impact of liberalization. In most countries, major gains have come first and foremost to the wholesale capital and corporate finance markets. And even this has been somewhat limited since the competitive effects, even after eliminating barriers, can remain restricted

to certain wholesale markets, regions, or segments. Even without (many) formal barriers, competition in many other markets remains imperfect, and the gains from competition can be limited to certain financial services (segments).

Extending the gains to other types of consumers of financial services has not proven easy. Even in the most industrialized countries, with good financial institutions and solid institutional infrastructures, the degree of effective competition in consumer and retail services still lags behind that in other financial service segments. In the European Monetary Union, for example, following the introduction of the euro, retail deposit and mortgage interest rates have converged—beyond what was due to the elimination of exchange rate risk—yet other financial services still show large price and cost differences (CEPR 2005). Indeed, in the European Union, there remain very large differences in the cost of a typical basket of retail banking services. The World Retail Banking Report (2005), for example, estimated that for 19 countries in Europe, North America, Eastern Europe, and the Far East the cost of a basket varied from €34 to €252, a range of 1.0 to 7.4, with the high and the low being two European Monetary Union countries. And beyond the traditional loan and deposit services, many wholesale products still show large price differences, possibly due to imperfect competition.[6]

Similar analysis for developing countries done by Beck and others (2008) shows even larger differences, especially when scaled by income level. They show, for example, that over $700 are required to open a bank account in Cameroon, an amount higher than the GDP per capita of that country. And they highlight that the annual costs to maintain an account in developing countries can amount to as much as 7 percent of GDP per capita. While not due to lack of competition alone, they do find that barriers are higher in countries where there are more stringent restrictions on bank activities and entry, and when banking systems are predominantly government-owned, whereas more foreign bank participation is associated with lower barriers.

Many countries, therefore, have given improving competition in these segments a greater priority (for example, in the European Union in 2005 following the Financial Sector Action Plan 2005–2010 of the European Commission; in the United Kingdom following the 2001 Cruickshank Report). This shows that to further facilitate competition it is not just a matter of opening up or liberalizing more, but rather that many (subtle) barriers still need to be removed. The Cruickshank report in the United Kingdom showed that the barriers to lowering costs for consumers and SMEs are often subtle, involving combinations of high costs of switching bank accounts, hidden fees, and limited transparency. The 2007 European Union extensive competition inquiry into retail banking found a number of competition concerns in the markets for payment cards, payment systems, and retail banking products, with barriers not easy to eliminate. The recent poor experience with subprime lending in the United States, although

mostly an issue of lack of consumer protection and failure to conduct proper regulation and supervision, has brought to the fore, as well, how unfettered competition does not prevent misuse and poor outcomes.

Experience has also brought some risks to light. Increased competition can have adverse effects on access to financial services, as it can undermine the incentives of banks to invest in information acquisition and thereby lower their lending to information-intensive borrowers. More generally, more formal lending arrangements, often associated with consolidation, increase distance between lenders and borrowers, increase foreign bank entry, and increase use of technology and more competition—as has happened in many markets—which may in theory have an adverse impact on access for some classes of borrowers. There is indeed the evidence of the United States, the European Union, and some emerging markets: that consolidation has led to a greater distance and thereby to less lending to more opaque SMEs (Sapienza 2002; Carow and others 2004; Berger and others 2005; Karceski and others 2005; Degryse and others 2005; Boot and Schmeits 2005 review this literature).

Evidence of this happening on a large scale in industrialized countries has been limited, though there have been clearly offsetting trends as well.[7] But for developing countries and emerging markets especially, these risks may be higher. Due to institutional weaknesses, including poor information and institutional infrastructure, and weak contracting environments, and more generally higher degrees of inequality, the access of SMEs and households in developing countries is often already less than desirable. In these markets, the possibility exists of bifurcated markets: large (international) banks may (i) concentrate on large corporations, serving them by using domestic and international platforms with a wide variety of products; and (ii) concentrate on consumers, providing them with financial services based on advanced scoring techniques and the like. The left out, middle segment under such a scenario could be the SMEs. As competition intensifies, profitability may go down and banks may have little incentive to invest in longer-term, relationship-based lending and the information collection necessary for lending to this segment (for a model along these lines, see Sengupta 2007).

Improving access and promoting financial inclusion can therefore require some specific measures, not just complementary to those increasing competition, but partly to offset possible negative effects of competition. Whether improvements in the quality and availability of information and the contracting environment can compensate fast enough, such that SMEs and the like (still) have sufficient access to financial services, is an open question. Empirical evidence on this has been limited to date, but early evidence is positive (see de la Torre and others 2008).

Cross-country studies. A number of recent papers have investigated across countries the effects on financial sector performance of (changes in) financial regulations

and specific structural or other factors relating to how competitive the environment is. Factors analyzed include entry and exit barriers, activity restrictions, limits on information sharing, and other barriers. Here the empirical findings are fairly clear. In terms of development and efficiency, deregulation leading to increased competition has led to lower costs of capital for borrowers and higher rates of return for lenders, that is, lower margins and lower costs of financial intermediation, spurring growth. Barth and others (2001) documented, for 107 countries, various regulations on commercial banks that were in place in 1999, including various entry and exit restrictions and practices. Using this data, Barth and others (2004) showed, among others, that tighter entry requirements are negatively linked with bank efficiency, leading to higher interest rate margins and overhead expenditures. These results are consistent with both tighter entry restrictions limiting competition and the contestability of a market determining bank efficiency.

Using bank-level data for 77 countries, Demirgüç-Kunt and others (2004) found that bank concentration, which as noted needs not proxy for the degree of competition, has a negative effect on banking system efficiency, except in rich countries with well-developed financial systems and more economic freedoms. Furthermore, limiting entry of new banks and implicit and explicit restrictions on bank activities are associated with higher bank margins. The fact that too much competition can undermine stability and lead to financial crises has been often argued (Allen and Gale 2004); however, this has been difficult to document systematically (Beck and others 2006).[8] Finally, since overall growth combines a number of aspects—efficiency, access, and stability—the relationship between competition in the financial sector and growth can be insightful. Claessens and Laeven (2005) related their competition measure to industrial growth in 16 banking systems. They found that greater competition in countries' banking systems allows financially dependent industries to grow faster, thus providing comprehensive evidence that more competition in the financial sector serves the broader economy well.

Internationalization. There is also much evidence on the competitive effects of international openness and capital account liberalization, particularly relevant for developing countries. Overall, the competitive effects of cross-border capital flows have been found to be generally favorable. In terms of development and efficiency, competition through cross-border capital flows has been shown to lead to lower costs of capital for borrowers and higher (risk-adjusted) rates of return for lenders. Evidence has shown that opening up internationally can spur growth, including through improved financial intermediation (Bekaert and others 2005; Henry 2006 reviews this literature; Claessens 2006a reviews the competitive effects of cross-border banking).

Greater cross-border capital flows have been found, though, to increase access more for selected groups of borrowers, for example, large corporations that already had preferential access, especially in developing countries. Tressel and Verdier (2007) find that in countries with weaker overall governance, politically connected firms benefit relatively more from international financial integration than other firms do. The growth benefits are consequently not always there. Kose and others (2008) highlight the need for a minimum set of initial conditions—good macroeconomic policies, financial and institutional development—for countries to benefit from financial globalization. And while the effects on stability have generally been found to be favorable—as international financial integration allows for greater international specialization and diversification—international capital flows can add to risks, among others, through contagion and greater risk of financial crises (IMF 2007).

Foreign bank entry can be an alternative way for cross-border capital flows to reach a market. The entry of foreign banks has been found to have generally favorable competitive effects on the development and efficiency of domestic, host banking systems (Chopra 2007 provides an extensive review of the literature). These generally positive results have occurred through various channels, resulting from both direct financial intermediation and from competitive pressures being put on existing banks. There is little evidence of increased volatility associated with foreign bank entry; rather risks seem to be diversified better. Barth and others (2004) showed, for example, that allowing foreign bank participation tends to reduce bank fragility. The qualitative aspects of competitive foreign bank entry—new and better products—have by nature been harder to document, but have possibly been the most important.

The effects of entry of foreign banks have been found, though, to depend on some conditions, and in some cases there can be negative consequences. Detragiache and others (2008) find that foreign bank entry in poor countries is associated with lower (growth in) private credit. Beck and Soledad Martinez Peria (2007) find contrasting patterns for different classes of borrowers for Mexico (see also Schulz 2006, who reviews foreign bank entry in Mexico). More generally, while evidence of immiserizing effects of internationalization is limited, achieving the full gains from entry often requires some (minimal) adjustment of regulations, legal and other institutional infrastructure to international norms. Furthermore, interactions between capital account liberalization, financial services liberalization, and domestic deregulation can affect the gains.

Complexity is, however, increasingly due to the changing nature of financial services provision. Financial services industries are continuously changing, not just due to the removal of barriers and the increased role of non-bank financial institutions, but also due to increased globalization and technological progress, which are all affecting the degree and type of competition. Even in market

segments where competition has been intense and benefits in terms in access and costs have been very favorable, such as wholesale and capital markets, new competition policy challenges have arisen, nationally and internationally. The consolidation of financial services industries, the emergence of large, global players, the large investments in information technology and brand names necessary to operate effectively and to gain scale, and the presence of large sunk costs make it difficult to assure full competition, even abstracting from the special characteristics of financial services. The increased importance of networks is also affecting the nature and degree of competition. As such, the positive experiences documented here may not prevail unless policies adjust as well.

Tools to Use

The literature reviewed makes clear that the tools used by policymakers for identifying and addressing competition issues in the financial sector need to be specific. It is also clear that tools will need to be adjusted in light of the changes in financial services industries. Measures typically used to date for measuring lack of competition (for example, Herfindahl or concentration indexes of banks or branches within a geographic area) are clearly quite limited, and were even a few decades ago, and are now even more so given the changes that have occurred. For example, rigid rules and guidelines—such as certain cut-offs for Herfindahl or concentration indexes—will not suffice. Rather, it is necessary to rely on solid analyses of the degree of competitive behavior. Yet, this ends up being especially complicated since the more sophisticated analytical and empirical tools developed for measuring competition in other industries are hard to apply to financial services industries. The unclear production function for financial services, the tendency to produce and sell bundles of services, the weaker and more volatile data, the presence of network properties, and so on, make assessing the degree of competitive behavior complex. A few examples illustrate the difficulties.

To measure effectively, using the tools from the traditional industrial organization literature (such as pass-through coefficients), (changes in) competition in banking (the most traditional financial service for which much data is available) is already a complex process. Data are often limited and span few observations. Most tests require at least 50 bank-year observations. Since in many developing countries the number of banks with good financial information is low, one cannot conduct year-by-year estimates of the degree of competition, or only do so subject to large confidence intervals, making comparisons over time hard. Using data from a larger sample of countries, for example, all Latin American countries or the EU-15, creates other difficulties, such as comparability.

Networks are another complication that can give rise to special competition measurement issues. Financial services provision involves the use of an ever greater

number of networks, such as payments, distribution, and information systems. This means barriers to entry can arise due to a lack of access for some financial services providers to essential services. In banking, for example, network barriers can be closely related to financial institutions which have access to the payments system, typically banks. ATM and other distribution networks can further be limited to banks, or only be available at higher costs to non-bank financial institutions. Obvious network effects arise when some banks have large nationwide coverage in branches or ATMs, as it can allow them to service customers more cheaply.

A recent advance in developing countries especially is banking through networks of agents, where, say, retail chains with a large network of stores serve as correspondent agents for banks (examples include Bolivia, Brazil, Colombia, India, Mexico, Pakistan, Peru, and South Africa; see Mas and Siedek 2008). This links competition in the retail sector with that in the financial sector. Also, two-sided network effects exist in payment card markets, since larger point-of-sale (POS) networks are more valuable to both cardholders and merchants. This leads to complex measurement issues, for which the credit card industries provide an interesting example (Pindyck 2007).

Another example of a network is that access to credit and other information on borrowers and other clients is often limited to (a subset of) incumbent banks. In addition, network externalities—especially in capital markets, for example, the agglomeration effects of liquidity—can complicate the application of competition policy. Ownership and governance structures can play a role as well. In many stock exchanges, derivatives, and other formal trading markets, ownership and governance structures are changing from mutual to for-profit ownership and with fewer owners. The traditional means of ensuring competition can thus be impaired, or at least changed,[9] and new competition approaches can be required. There are also forces toward vertical integration, especially in capital markets (for example, the integration of trading systems with clearing and settlement), while other forces push toward more separation in other aspects (for example, clarity in functions) or toward horizontal consolidation (for example, economies of scale). Each of these raises (new) forms of possible anti-competitive behavior.

An example of the complexity in defining a market is payments services. In payments services, as in many other financial services these days, the issue arises as to what constitutes the relevant product market. Payment cards include credit cards, debit cards, and charge (or stored value) cards. While different in terms of underlying technology, pricing schemes, and some auxiliary services, these cards are similar in their cash substitute function, in which case competition analysis should cover all types of cards. Alternatively, however, payment cards can be seen as part of a bundle of different services, like ATM cash withdrawal, payment service at POS terminals, and so on. In which case, payment cards should be analyzed as part of the competition in bundles of household financial services,

including deposit services. This issue is relevant not just in industrialized countries, but also in developing countries where cards are rapidly becoming a substitute for traditional banking services. For example, in some markets cards with pre-paid balances have been introduced that can be used for (small) payments without the need for a bank account. These and others forms of branchless banking (for example, mobile banking using phones; see Ivatury and Mas 2008) introduce new competition issues.

An example from the capital markets is the increasing trend toward internalization of trading within financial institutions, institutional investors, and other financial intermediaries. While this can save transactions costs for the final consumer, it makes for less overall transparency and can lead to anti-competitive outcomes. Yet, data are more difficult to obtain, and analysis to establish anti-competitive behavior is more complex.

In addition to these complications, market and product definitions have become (more) difficult. It is somewhat trite, but nevertheless very important from a competition policy point of view, to state that many financial markets today are global in nature, making any application of competition policy to national markets only of lesser value than in the past. In addition, the definition of a specific financial service (and its market) has become more difficult. Today, for example, there are fewer differences than in the past between the markets for pension services and for asset management services; after all, many people can save in both ways and, provided tax rules are largely harmonized between the two, will do so. And with many non-financial institutions providing (near) banking and other financial services, the boundary between banks and non-bank financial institutions has become more blurred.

The upshot of these complications and changes in financial services industries is that, for the same degree of liberalization, competition may be less assured today than in the past. The increasing presence of high fixed costs and large sunk costs in the production of wholesale financial services can mean significant first mover and scale advantages, possibly leading to natural monopoly and market power. Externalities, say in e-finance in the adoption of payments using mobile phones, can make the adoption of new technologies exhibit critical mass properties. In consumer finance, switching costs may have increased—for example, because automatic payments are increasingly linked to one's specific bank account number. This means in turn that customers cannot and do not easily change provider, leading to more complex competition. There is indeed some evidence from studies on banking systems that the progress in increasing competition may have slowed down from the early 2000s on, with even some indications of a decline in competition in some markets (Bikker and Spierdijk 2007). The robustness of this finding is yet to be confirmed, and the exact causes are unclear; in any case they are likely to be multiple.

The difficulty in documenting (changes in) competition with current tools shows that the tools to measure the (lack of) competition policy in financial service industries need to be enhanced or even be newly designed to address new issues and changes. This will often require developing and applying new, economically fully justified, models, which will take time to do and which can be complex. Short of doing so, however, much can be done. Much information on the competitive structure can still be discerned by focusing on price setting for specific products or financial functions, for example, what fees are charged for consumer retail products or for processing individual pension premiums or payments. In addition, more focus can be given to the pricing and availability of inputs necessary to produce financial services, for example, do all types of financial institutions have access on the same basis as the retail payments system? This type of information can also be better disclosed such that users can act on it.

Implications for Competition Policy in the Financial Sector

There had been little analysis of the design and conduct of competition policy in the financial sector. And the "special nature" of finance, with its emphasis on stability, always meant that competition policy was considered more complicated in the financial services industry. Changes in these industries create their own set of new competition issues, and there are surely no easy answers as to how to reflect these in policy. As such it is hard to be definitive on how competition policy should be conducted in the financial sector. What is clear is that the two aspects that have to be considered afresh in competition policy include the *approaches* and the *institutional arrangements*.

Approaches

One can think of three possible, and largely complementary, approaches to conducting competition policy. One is assuring that entry/exit rules allow for contestable markets in terms of financial institutions and products. Two is leveling the playing field across financial services providers and financial products, such that there is effective intrasectoral competition. Three is assuring that the institutional environment (payments system, credit bureaus, and so on) is contestable. The first has been the traditional approach and the norm. As analyzed above, it has been quite effective, and it will have to remain the essential cornerstone of competition policy in the financial sector, as in other sectors. But, as noted, on its own it may have reached its limits.

The second, leveling the playing field, means harmonization (or convergence), both among financial services providers (banks, insurance companies, pension

funds, asset management, and so on), markets—national, regional, and global—and between different, but functionally equivalent, types of products, whether called banking, insurance, or capital market products. The goal of harmonization should be that, within particular markets, products are not regulated differently depending on what type of financial institution provides the service. And products that offer the same functionality of service, but may be "labeled" differently, that is, fall under different regulatory approaches, need to be treated similarly. Harmonization (or convergence) includes addressing differences in taxes, capital adequacy requirements, transparency/disclosure, and so on across sectors and products. This will be useful not just to increase competition, but also to avoid regulatory arbitrage and to reduce differences in the net overall regulatory burden of products. The increased creation of complex financial products that straddle various markets and institutions makes the need for a common regulatory approach all the more necessary.

Harmonization across financial service sectors and products is a long-standing issue. The big barriers across financial service sectors have been removed: only in some countries, but increasingly less so, are there still (large) regulatory barriers between commercial banks, investment banks, insurance companies, and other financial institutions. The fact that these large barriers have been removed, however, does not make the issue of harmonization moot, since often many smaller barriers remain. Some differences will be due to some "path dependence"; for example, some products emerged as insurance products but migrated to becoming savings products. Others arise from the existence of subtle barriers, for example, some products may be linked to the payments system for which access is limited. And others again exist because of linkages with other economic policies, for example, tax preferences may be linked to pension products but not to savings. Furthermore, many financial products come bundled (for example, a checking account has savings, payments, and often as well credit—overdraft—functions linked to it), making it hard to compare regulatory burdens of individual products with each other (for example, the costs of complying with Anti-Money Laundering and Combating the Financing of Terrorism (AML/CFT) may be assigned to a checking account or may be spread over various products).

In all cases, there is a need to go deeper. Yet, designing an *ex ante* approach to level perfectly the playing field is conceptually and in practice very difficult. The current approach, which is largely reactive—as producers and consumers are faced with differences, they may approach the various regulators and appeal for harmonization—has therefore benefits. It has also risks, however. There can be a race to the bottom, as the lowest treatment becomes the norm for all products. It also opens up the possibility of lobbying for favorable treatments. This can go counter to the valid reasons for differences in regulatory treatments based on, say,

prudential concerns or consumer protection. A proactive approach by authorities and competition agencies can therefore still be useful.

This can be complemented by giving consumers more information. Better price information and more disclosure on the costs of various financial services can help consumers identify uncompetitive products, start formal complaints, or both. Many countries have centralized places where, say, interest rates on deposit and standard loan contracts can be found. Experience shows though that this remains of limited effectiveness when done alone. Market solutions can greatly and often more effectively foster competition than government initiatives alone: witness the many firms offering price comparators. Regardless, and similar to what is needed for assessing the degree of competition, agencies could require better data on prices and costs at the level of individual products and make this data available. This would be a very important starting point for users of financial services that often lack empirical bases.

The third approach, assuring that the institutional environment is contestable, is complex as well. This would mean that the various inputs required for the production and distribution of financial services, including network services (for example, payments and check system, credit bureaus, other networks, and so on), need to be: available to all interested in using them, be fairly and uniformly priced, and be efficiently provided. For no part of a specific financial service production and distribution chain should there be any undue barriers or unfair pricing. These steps are considered basic requirements in most other network industries, where (private) firms are producing and delivering services (for example, phone, other telecommunications, energy, and water), and using common networks (for example, telecommunication lines, power lines, railroads, pipelines, and so on).

With the often subtle barriers in financial services industries, however, these steps and policy recommendations to foster more effective competition are not easy. In many markets, policy actions and recommendations have largely been in the form of putting more pressure on the financial industry, including by relying on codes of conduct, reducing extensive barriers, converging standards, limiting collusive practices, and encouraging consumer mobility by lowering switching costs. Some strong general policy intervention can at times be necessary, however, to force more rapid adjustments, create standardization, or remove barriers. Some examples will illustrate the benefits of strong actions.

Over the past decade many governments have required various retail payments systems initially developed by (groups of) individual banks within a nation to be integrated and available to all consumers. This greatly increased not only the quality of payments services, but also often lowered costs. The European Union recently required charges for financial transfers among eurozone countries to be equal to those for domestic transfers (subject to some conditions), which

illustrates the benefits of strong actions. Another option is mandating easy portability of one's bank account number, which is being introduced in some European Union countries. Mandating by government rule a level playing field can be equally necessary in capital markets to assure fair trading and pricing for small as well as large investors. In many markets, traders are required to always use the best price.[10]

Important to assuring a contestable institutional infrastructure in finance will be the formulation and application of standards; but here policymakers will face trade-offs. As the payments system examples show, in networks compatibility of systems is mostly based on standards. Standards can also help avoid coordination problems in the technology choices made by firms, and can help consumers forecast whether the specific technology will be widespread, leading to reduced uncertainty and less risk of consumer lock-in, and thereby avoidance of non-adoption (waiting). In several cases after the industry agreed on a common standard, the adoption of the good or service did indeed increase sharply. In financial services, one good example has been the Society for Worldwide Interbank Financial Telecommunication (SWIFT) protocol for transacting international payments, introduced in 1977. At the same time, with standards, users can be forced to make a choice. Furthermore, joining more than one network is often ruled out by contract. Exclusivity arrangements can lead to the predominance of a large network, even when more differentiated networks with more consumer choice could proliferate. Anti-competitive behavior can then easily follow. Policymakers then face a trade-off between, on the one hand, encouraging market development by supporting (a particular) standard(s) and achieving critical mass with the best technology and, on the other hand, stimulating competition at the same time as not favoring incumbents.

In these and other areas, approaches to competition policy in the financial sector can perhaps learn from those used in other network industries, many of which have adopted relatively sophisticated competition policies. For example, in many infrastructure industries, the ownership, management, or both of the network has been separated in recent years from the provision of services to assure fairer competition. Access policies and pricing of network services are often subject to government regulatory review. In these other industries, some rules for operating on the network may be standardized through direct government actions or through self-regulatory agencies assigned with this task, and not left to the (private sector) operators or owners (alone).

Some of these other network industries have also come to grips with the issue of assuring access to basic services for a wide class of consumers. Through mechanisms such as "universal service obligations," uniform price rules for essential inputs in producing services or key outputs, selected subsidies, and other (tax) incentives policymakers have been able to assure (near) universal access in these

other network industries, at least in the most industrialized countries. These models are also being applied in developing countries. These approaches may equally apply to those financial services with large network properties. For example, in payment services, standard uniform pricing rules could be imposed, similar to the uniform rates applied in certain basic postal, phone, telecommunications, water, or electricity services.

Institutional Arrangements

The institutional arrangements for competition policy will often need to change as well. For one, competition policy needs to be separated more clearly from prudential oversight. Some countries have already taken competition policy out of the central bank or supervisory authority, but in many countries the responsibility for competition policy still lies with the prudential authority. This creates a conflict of interests (for a review of the arguments, see Carletti and Hartmann 2002). Separation does not mean that the prudential authority would have no say in competition: it could have some (veto) rights in any specific decisions or general policy changes. Furthermore, the competition authority could still rely on analyses by the prudential authority when, say, in the case where technical expertise is scarce in the competition authority. But clearer separation does address the conflict of interest issue that has hindered effective competition policy in the financial sector.

Second, there is much more need to coordinate better, and preferably bring together, competition policy functions presently dispersed among various agencies within a country (for example, separation of banking and non-bank financial institutions, or prudential regulators, or both specialized and general competition policy agencies). Reducing this dispersion will avoid the inconsistent application of competition policy across financial institutions and products that are functionally equivalent. It will also allow for the build-up of skills necessary for proper competition policy analysis. Of course, in many countries, there is also a need to improve the skill base in the judicial system where competition cases may be finally settled or arbitrated.

It will also be important to consider the interactions between competition policy and consumer protection policies specific to the financial sector. This concerns three sets of issues: assuring markets work better for all final consumers (what is sometimes called assuring a proper business conduct); protecting individual consumers (which can be considered a narrow version of consumer protection); and assuring consumers obtain the greatest benefits from financial services provision (for example, through proper information and education—which makes for an even wider concept of consumer protection). Competition policy is relevant for all these issues, as both too little and too much competition can hurt consumers through each of these channels.

The costs and benefits of single versus multiple supervisory agencies have been debated for some time[11] and no simple answers exist here on the best balance, from the point of view of financial stability or from the perspective of efficiency of financial services provision. It does relate, however, to the issues of competition and harmonization across financial services and financial services providers. In the design of competition policy in the financial sector the organization of the supervisory agencies thus has to be considered. The move toward single supervisory authorities across the world—countries as diverse as Estonia, Kazakhstan, South Korea, Nicaragua, and the United Kingdom have adopted it in the last decade—presumably could help with reducing unnecessary differences arising from multiple regulatory regimes.[12]

Superficially, differences in the degree of *de jure* or de facto harmonization (or lack thereof) among financial instruments are not obvious between various supervisory regimes. Even where there is a single supervisory authority, it has not done away with all (or even many) of the regulatory harmonization issues across sectors or products. Presumably, competitive pressures from producers and users and the lobbying strength of these constituencies relative to regulators will be the most important factors driving the (de facto) reduction in barriers. In that respect, a more fragmented structure of regulation and supervision may well lead to more de facto harmonization and convergence as financial services industries are more strongly positioned to argue for regulatory changes and agencies "compete" with each other.[13] Nevertheless, whether any of these institutional arrangements are superior from the point of view of efficient financial services provision has not been researched in depth and may remain unclear in any case given the difficulty of attribution. And the organization of a supervisory authority in a single country may be of little relevance when competition for some financial services already is on a global basis.

The changing nature of financial services provision also means that other aspects affect the competitive environment. For example, the competitive structure in telecommunication markets may affect the market for electronic (or remote) finance, as in the case of mobile payments. And, obviously, there is a much greater need today for international cooperation among various national agencies in the application of competition policy. Harmonization and convergence across markets, already a very complex undertaking within countries, will be compounded regionally or globally. The European Union experience, which has been engaged for quite some time now in a process of financial integration and convergence, shows the tenacity needed to create a single market for financial services. It shows that requiring some uniformity in minimal regulations is not sufficient, since inconsistencies with national rules and laws still arise, as other policy areas need to be adjusted, which take much time and effort.

These national, regional, and global experiences also show how many conceptually difficult questions can arise with convergence. For example, while many banks operate across borders without barriers, liquidity support and lender of last resort facilities are still organized nationally. This creates inconsistencies with policies for dealing with financial insolvency.[14] While this topic largely concerns financial stability, and is beyond this paper (there is a large literature here: see the papers collected in Caprio and others 2006), these differences can also have competitive implications. For example, banks from some countries may have more generous access to the local safety net than banks from other countries do. Of course, these issues also arise within countries, as when state-owned banks attract deposits at a low interest rate because they are (perceived to be) covered more generously by the safety net, as has happened often in developing countries. And they arise both *ex ante* and *ex post*, as when weak banks receive liquidity, solvency support, or both.[15] These and other issues mean that competition agencies will have to be both reactive and proactive in their investigations. Today, agencies often only respond to events, such as large scale mergers and acquisitions, but undertake little analysis of competitive conditions in existing markets. An approach targeted at key areas of concern of possible anti-competitive behavior would be useful.

Lastly, harmonization and convergence depend these days to a great extent on international standards, of which the ones developed by the Basel Committee on Banking Supervision, the International Organization of Securities Commissions, the International Association of Insurance Supervisors, and the Committee on Payment and Settlement Systems are the most visible. This has become a large body of "soft law."[16] The ambition levels of these standards vary, from suggestions to achieve a minimum common denominator among existing national requirements, which is most often the case, to going beyond existing national requirements. Although the standards are voluntary in nature and implementation is left to countries themselves, some of the standards can be quite intrusive. In most cases, functional convergence and arbitrage would make remaining cross-border regulatory differences of little consequence in hindering competition. Some major initiatives, however, like Basel II and other rules affecting (cross-border) banking, may end up hindering effective competition in some respects.[17]

Conclusions

I have reviewed the state of knowledge on competition in the financial sector and how competition policy is and should be organized. I have shown that competition matters as in other industries, but that there are some specific analytical issues. Notably, there is the effect of excessive competition on financial stability;

but also that the degree of competition matters for the access of firms and households to financial services. As a consequence, the view that competition in financial services is unambiguously good is more naive than in other industries. And it is not sufficient to analyze competitiveness from a narrow concept alone or to focus on one effect only. One has to consider a broader set of objectives, including efficiency, access to services to various segments of users, and systemic financial sector stability, as well as possible trade-offs among these objectives. In terms of the factors driving competition in the financial sector, and empirical measurement of competition, I have highlighted that one needs to consider standard industrial organization factors, such as entry and exit and contestability, but also that financial services provision has many network properties in production, distribution, and consumption, making for complex competition structures.

Besides the theoretical complexity, empirical evidence on competition in the financial sector is scarce and often not (yet) clear. Much of the current literature relates performance indicators to the financial system structures and regulatory regimes of countries without formal measures of competitiveness. And the contestability view of competition is not the one typically applied. Rather, the market Structure-Conduct-Performance paradigm is at best used. What is available, however, suggests that competition has spurred improvements, including greater product differentiation, lower cost of financial intermediation, more access to services, and enhanced stability. This evidence is fairly universal, from industrialized countries to many developing countries. In terms of factors driving competition, to date, the latter has been achieved by traditional means, that is making systems more open and contestable—having low barriers to entry and exit—in developing countries; this has meant that the internationalization of financial services has often driven changes. As globalization, technological improvements and, deregulation further progress, the gains of competition can be expected to become even more widespread across and within countries.

At the same time, the review shows that once the easier steps have been taken, policies to achieve effective competition in all dimensions, and balancing the trade-offs between competition and other concerns, become more challenging. As financial services industries evolve, and as financial markets and products become more complex and global, new regulatory and competition policy issues arise. This means that approaches to competition issues need to adjust. This is important since competition policy in the financial sector is often already lagging behind, though the theoretical and empirical literature is just catching up with the special issues and changes in financial services industries.

To move forward, therefore, besides improving the measurement of competition, much can be learned from policies already standard in many other industries, especially network industries. I have made some suggestions as to what approaches, as well as institutional arrangements and tools, best fit a modern

view of competition policy in the financial sector. I have also suggested that policymakers can greatly enhance the available data so that users will have the information needed to assess the costs of different financial services. Finally, with rapid changing financial services industries, there is a need to remain agile and adjust competition policies and procedures.

Notes

Paper prepared for the G-20 meeting on *Competition in the Financial Sector*, Bali, February 16–17, 2008. Stijn Claessens is Assistant Director in the Research Department of the IMF, a Professor of International Finance Policy at the University of Amsterdam, and a Research Fellow at the CEPR. His email address is SClaessens@imf.org. This paper's findings, interpretations and conclusions are entirely those of the author and do not necessarily represent the views of the IMF, its Executive Directors or the countries they represent. I would like to thank Thorsten Beck, Jaap Bikker, Giovanni Dell'Ariccia and Luc Laeven for useful comments and discussions, the discussant, Mario Nakani, participants at the G-20 meeting, participants at seminars at the IMF Regional Office in Tokyo, the World Trade Organization (Geneva) and Bruegel (Brussels), and the referees and the editor for useful comments.

1. For a recent review of the theoretical literature on competition and banking, see Vives (2001).

2. Within this general paradigm, many aspects have been investigated. For example, there exist studies of the degree to which firms deviate from a production-efficient frontier, so-called x-inefficiency (see Berger and Humphrey 1997 for an international survey of x-inefficiency studies for financial institutions).

3. Specifically, the model to estimate the H-statistics for banking is:

$$\begin{aligned}\ln(P_{it}) = &\alpha + \beta_1 \ln(W_{1,it}) + \beta_2 \ln(W_{2,it}) + \beta_3 \ln(W_{3,it}) \\ &+ \gamma_1 \ln(Y_{1,it}) + \gamma_2 \ln(Y_{2,it}) + \gamma_3 \ln(Y_{3,it}) + \delta D + \varepsilon_{it}\end{aligned}$$

where P_{it} is the ratio of gross interest revenue to total assets (proxy for output price of loans), $W_{1,it}$ is the ratio of interest expenses to total deposits and money market funding (proxy for input price of deposits), $W_{2,it}$ is the ratio of personnel expense to total assets (proxy for input price of labor), $W_{3,it}$ is the ratio of other operating and administrative expenses to total assets (proxy for input price of equipment/fixed capital). The subscript i denotes bank i, and the subscript t denotes year t.

4. For example, De Bandt and Davis (2000) found monopoly behavior in small banks in France and Germany, and monopolistic competition in small banks in Italy and large banks in all three countries. This suggests that in these countries, small banks have more market power, perhaps as they cater more to local markets.

5. He summarized it as follows: "This paper focuses on how one dimension of this broad-based deregulation—the removal of limits on bank entry and expansion—affected economic performance. In a nutshell, the results suggest that this regulatory change was followed by better performance of the real economy. State economies grew faster and had higher rates of new business formation after this deregulation. At the same time, macroeconomic stability improved. By opening up markets and allowing the banking system to integrate across the nation, deregulation made local economies less sensitive to the fortunes of their local banks."

6. Even within fully integrated wholesale markets (no currency risks, limited legal and regulatory differences, good information, and so on), such as the United States and increasingly the European Union/European Monetary Union, there still is, for example, a familiarity bias, for example, more investment and entry closer to the home of the investor.

7. In part due to technology, banks are better able today to combine soft and hard information

in efficient ways, and some banks have become very profitable when specializing in SME lending. Also, larger multiple-service banks can have a comparative advantage in offering a wide range of products and services on a large scale, through the use of new technologies, business models, and risk management systems, making them effective in the SME markets.

8. Boyd and others (2006) find for the United States no trade-off between bank competition and stability, and that bank competition fosters the willingness of banks to lend. See also Čihák and others (2006) and Čihák and Schaeck (2007).

9. For example, a private provider of an essential service will have different incentives to serve all in need than a mutual-owned provider, where all uses are also members/owners.

10. In the United States, this is embodied in the Securities and Exchange Commission's (SEC) "order protection rule": no matter where a customer order is routed, he or she should receive the best price that is immediately and automatically available anywhere in the national market system. This principle promotes competition among individual market centers by ensuring that dominant markets cannot ignore smaller markets displaying the best price.

11. The issue of consolidated supervision is less debated.

12. Although there is this trend, it is not general. Some countries have recently adopted the model of integrating systemic stability with the prudential oversight of all—banking, insurance, and pension—individual financial institutions into one agency, but keeping this separate from the agency for market conduct supervision. Others have left systemic stability with the central bank, but organized prudential and market conduct under two separate agencies. Yet others have made no changes and still have separate prudential banking, securities markets, and insurance supervisors operating in one country (and sometimes multiples of each, for example, the United States).

13. Obviously, this is highly context and country dependent, and ignores many other dimensions. For example, with strong financial institutions and weak regulators, a greater influence of private interests could lead in some countries to lax and low-cost standards, with perhaps greater competitiveness, but with more risk of financial instability. In other environments, capture of the regulator may lead to rent-seeking by (selected) financial institutions, but with limited risks.

14. Although liquidity management may be done centrally by the foreign bank in its home country, branches of foreign banks are typically eligible to receive liquidity support from the local host central bank. In case of insolvency of the head bank, however, the home country authorities are responsible, which can involve home government resources in case the whole bank fails. In single currency regions, like the European Monetary Union, there is an additional need for coordination between member countries' liquidity support and the European Central Bank's monetary policy.

15. This has happened in many financial crises (see Claessens and others 2003 for an overview of measures used in restructuring), but in the past it has not led to competitive questions. The recent cases of the liquidity support for Northern Rock in the United Kingdom, the solvency support for IKB in Germany, and the "bailout" of Bear Stearns, however, have attracted some attention for their potential anti-competitive implications. Also, the (on-going) large scale liquidity support during the recent financial crisis from the U.S. Federal Reserve Board, the European Central Bank, and the Bank of England could raise such questions.

16. There are issues of the legitimacy and governance of the standards setting bodies, which are not discussed here.

17. The Basel II rules, for example, encourage international banks to use the same risk management approaches across national jurisdictions, which creates a level playing field and can help with competition. At the same time, too uniform application could lead credit risks to be priced too rich in some countries (for example, emerging markets) and too thin in other countries. Adapting the approaches to capture the risks in various markets appropriately is necessary, but would negate some of the gains of uniformity.

References

Allen, Franklin, and Douglas Gale. 2004. "Competition and Financial Stability." *Journal of Money, Credit and Banking* 363(2):453–80.

———. 2007. *Understanding Financial Crises*. Oxford: Oxford University Press.

Ausubel, Lawrence. 1991. "The Failure of Competition in the Credit Card Market." *American Economic Review* 81(1):50–81.

Barth, James R., Gerard Caprio Jr., and Ross Levine. 2001. "The Regulation and Supervision of Banks around the World: A New Database." In Robert E. Litan, and Richard Herring, eds., *Integrating Emerging Market Countries into the Global Financial System*. Washington, DC: Brookings Institution Press, pp. 183–241.

———. 2004. "Bank Regulation and Supervision: What Works Best?" *Journal of Financial Intermediation* 132:205–48.

———. 2006. *Rethinking Bank Regulation; When Angels Govern*. New York, NY: Cambridge University Press.

Baumol, William J., John C. Panzar, and Robert D. Willig. 1982. *Contestable Markets and the Theory of Industry Structure*. San Diego, CA.: Harcourt Brace Jovanovich.

Beck, Thorsten, and Maria Soledad Martinez Peria. 2007. "Foreign Bank Acquisitions and Outreach: Evidence from Mexico." World Bank Policy Research Working Paper 4467, Washington, DC.

Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine. 2006. "Bank Concentration, Competition, and Crises: First Results." *Journal of Banking and Finance* 30:1581–603.

Beck, Thorsten, Asli Demirguç-Kunt, and Maria Soledad Martinez Peria. 2008. "Banking Services for Everyone? Barriers to Bank Access and Use Around the World." *World Bank Economic Review* 22:397–43.

Bekaert, Geert, Campbell Harvey, and Christian Lundblad. 2005. "Does Financial Liberalization Spur Growth." *Journal of Financial Economics* 77:3–56.

Berger, Allen N. 2000. "Efficiency in Banking: Professional Perspectives." In A. Saunders, ed., *Financial Institutions Management*. Columbus, OH: McGraw-Hill, pp. 300–1.

Berger, Allen N., and B.C. Humphrey. 1997. "Efficiency in Financial Institutions: International Survey and Directions for Future Research." *European Journal of Operational Research* 98: 175–212.

Berger, Allen N., Rebecca S. Demsetz, and Philip E. Strahan. 1999. "The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future." *Journal of Banking and Finance* 23:135–94.

Berger, Allen N., Nathan H. Miller, Mitchell A. Petersen, Raghuram Rajan, and Jeremy Stein. 2005. "Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks." *Journal of Financial Economics* 76:237–69.

Besanko, David A., and Anjan V. Thakor. 1992. "Banking Deregulation: Allocational Consequences of Relaxing Entry Barriers." *Journal of Banking and Finance* 16:909–32.

Bikker, Jacob A., and Katharina Haaf. 2001. "Competition, Concentration and Their Relationship: An Empirical Analysis of the Banking Industry." DNB Staff Report 68, De Nederlandsche Bank, July.

Bikker, Jacob A., and Laura Spierdijk. 2007. "How Banking Competition Changed over Time." Mimeo, De Nederlandsche Bank, December, Amsterdam.

———. 2008. "Measuring and Explaining Competition in the Financial Sector." Mimeo, De Nederlandsche Bank.

Bikker, Jacob A., Laura Spierdijk, and P. Finnie. 2006. "Misspecification in the Panzar-Rosse Model: Assessing Competition in the Banking Industry." DNB Working Paper 114, De Nederlandsche Bank, Amsterdam.

BIS (Bank for International Settlement). 2006. "The Banking System in Emerging Economies: How Much Progress Has Been Made?" BIS Papers no. 28, Basle.

Boot, Arnoud W.A., and Anjolein Schmeits. 2005. "The Competitive Challenge in Banking." Amsterdam Center for Law & Economics ACLE, Working Paper 2005-08.

Boot, Arnoud, and Anjan V. Thakor. 2000. "Can Relationship Banking Survive Competition?" *Journal of Finance* 55(2):679–713.

Boyd, John H., and Gianni De Nicolò. 2005. "The Theory of Bank Risk-Taking and Competition Revisited." *Journal of Finance* 60(3):1329–43.

Boyd, John H., Gianni De Nicolò, and Abu M. Jalal. 2006. "Bank Risk-Taking and Competition Revisited: New Theory and New Evidence." IMF Working Paper WP/06/297.

Caprio, Gerard, Jr., Douglas Darrell Evanoff, and George G. Kaufman, eds. 2006. *Cross-Border Banking: Regulatory Challenges*. Blackensack, NJ: World Scientific Publishing. .

Carletti, Elena, and Philipp Hartmann. 2002. "Competition and Stability: What's Special About Banking?" ECB Working Paper 146, May.

Carow, K.A., Edward Kane, and R. Narayanan. 2004. "How Have Borrowers Fared in Banking Mega-Mergers?" Unpublished manuscript, Boston College.

CEPR (Centre for Economic Policy Research). 2005. "Barros, Pedro Pita, Erik Berglof, Paolo Fulghieri, Jordi Gual, Colin Mayer, and Xavier Vives." *Integration of European Banking: The Way Forward. Monitoring European Deregulation*, vol. 3. London: Centre for Economic Policy Research CEPR and Fundacion BBVA.

Cetorelli, Nicola. 1999. "Competitive Analysis in Banking: Appraisal of the Methodologies." *Economic Perspectives* 1:2–15.

Chopra, Ajai. 2007. "Opening up to Foreign Banks: Harnessing the Benefits and Avoiding the Pitfalls." Mimeo, IMF.

Čihák, Martin, and Klaus Schaeck. 2007. "Banking Competition and Capital Ratios." IMF Working Paper 07/216, Washington, DC.

Čihák, Martin, Klaus Schaeck, and Simon Wolfe. 2006. "Are More Competitive Banking Systems More Stable?" IMF Working Paper 06/143, Washington, DC.

Claessens, Stijn. 2006a. "Competitive Implications of Cross-Border Banking." In Gerard Caprio, Douglas D. Evanoff, and George G. Kaufman, eds., *Cross-Border Banking: Regulatory Challenges*. Blackensack NJ.: World Scientific Publishing, pp. 151–82.

Claessens, Stijn.. 2006b. "Current Challenges in Financial Regulation." Policy Research Working Paper 4103, World Bank, Washington, DC.

Claessens, Stijn, and Luc Laeven. 2004. "What Drives Banking Competition? Some International Evidence." *Journal of Money Credit and Banking* 36(3):563–83.

———. 2005. "Financial Dependence, Banking Sector Competition, and Economic Growth." *Journal of the European Economic Association* 3: 1, 179–201.

Claessens, Stijn, Daniela Klingebiel, and Luc Laeven. 2003. "Financial Restructuring in Banking and Corporate Sector Crises, What Policies to Pursue?" In Michael Dooley, and Jeffrey Frankel, eds., *Currency Crises Management*. Chicago, IL: NBER/University of Chicago Press, pp. 147–80.

Claessens, Stijn, Gergely Dobos, Daniela Klingebiel, and Luc Laeven. 2003. "The Growing Importance of Networks in Finance and Its Effects on Competition." In Anna Nagurney, ed., *Innovations in Financial and Economic Networks*. Northampton, MA: Edward Elgar, pp. 110–35.

De Bandt, Olivier, and E. Philip Davis. 2000. "Competition, Contestability and Market Structure in European Banking Sectors on the Eve of EMU." *Journal of Banking and Finance* 24:1045–66.

de la Torre, Augusto, María Soledad Martínez Pería, and Sergio Schmukler. 2008. "Bank Involvement With SMEs: Beyond Relationship Lending." World Bank, Washington, DC.

Degryse, Hans, Nancy Masschelein, and Janet Mitchell. 2005. "SMEs and Bank Lending Relationships: The Impact of Mergers." CEPR Working Paper 5061.

Dell'Ariccia, Giovanni, and Robert Marquez. 2004. "Information and Bank Credit Allocation." *Journal of Financial Economics* 71:185–214.

Dell'Ariccia, Giovanni, Luc Laeven, and Deniz Igan. 2008. "Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market." IMF Working Paper 08/106.

Demirgüç-Kunt, Asli, Luc Laeven, and Ross Levine. 2004. "Regulations, Market Structure, Institutions, and the Cost of Financial Intermediation." *Journal of Money, Credit, and Banking* 36(3):593–622.

Detragiache, Enrica, Poonam Gupta, and Thierry Tressel. 2008. "Foreign Banks in Poor Countries: Theory and Evidence." *Journal of Finance*, 63: 5, 2123–60.

European Central Bank. 2005. *Indicators of Financial Integration in the Euro area*, September.

European Commission. 2005. "White Paper on Financial Services Policy 2005–2010." Brussels, December.

Gutiérrez de Rozas, Luis. 2007. "Testing for Competition in the Spanish Banking Industry: The Panzar-Rosse Approach Revisited." Banco de España Research Paper WP-0726. (http://ssrn.com/abstract=1005116).

Hauswald, Robert, and Robert Marquez. 2003. "Information Technology and Financial Services Competition." *The Review of Financial Studies* 16(3):921–48.

Hempell, Hannah S. 2002. "Testing for Competition Among German Banks." Economic Research Centre Discussion Paper 04/02. (http://ssrn.com/abstract=304343or%20DOI:%2010.2139/ssrn.304343).

Henry, Peter B. 2006. "Capital Account Liberalization: Theory, Evidence, and Speculation." NBER Working Paper 12698, Cambridge, MA.

IMF (International Monetary Fund). 2007. "Reaping the Benefits of Financial Globalization." Washington, DC: International Monetary Fund. (www.imf.org/external/np/res/docs/2007/0607.htm).

Ivatury, Gautam, and Ignacio Mas. 2008. "The Early Experience with Branchless Banking." April 1, CGAP Focus Note 46, Washington, D.C.

Karceski, Jason, Steven Ongena, and David C. Smith. 2005. "The Impact of Bank Consolidation on Commercial Borrower Welfare." *Journal of Finance* 60: 2043–82.

Kose, M. Ayhan, Eswar Prasad, Kenneth Rogoff, and Shang-Jin Wei. 2008. "Financial Globalization and Macroeconomic Policies." forthcoming. In D. Rodrik, and M. Rosenzweig, eds., *Handbook of Development Economics*

Levine, Ross, 1996. "Foreign banks, financial development, and economic growth." In *International Financial Markets*, ed. by Claude Barfield. Washington, DC: American Enterprise Institute Press.

Laeven, Luc, and Ross Levine. 2007. "Is There a Diversification Discount in Financial Conglomerates?" *Journal of Financial Economics*, forthcoming.

Levine, Ross. 1996. "Foreign Banks, Financial Development, and Economic Growth." In Claude Barfield, ed., *International Financial Markets*. Washington, DC: American Enterprise Institute Press.

Marquez, Robert. 2002. "Competition, Adverse Selection, and Information Dispersion in the Banking Industry." *Review of Financial Studies* 15(3):901–26.

Mas, Ignacio, and Hannah Siedek. 2008. "Banking through Networks of Retail Agents." CGAP Focus Note 47, Washington, DC.

Nakane, Marcio I. 2001. "A Test of Competition in Brazilian Banking." Banco Central do Brasil Working Paper 12. (http://ssrn.com/abstract=273208).

Panzar, John C., and James N. Rosse. 1987. "Testing for 'Monopoly' Equilibrium." *Journal of Industrial Economics* 35:443–56.

Perotti, Enrico C., and Javier Suarez. 2002. "Last Bank Standing: What Do I Gain if You Fail?" *European Economic Review* 469:1599–622.

Petersen, Mitchell A., and Raghuram G. Rajan. 1994. "The Benefits of Lending Relationships: Evidence from Small Business Data." *Journal of Finance* 49:1367–400.

———. 1995. "The Effect of Credit Market Competition on Lending Relationships." *Quarterly Journal of Economics* 110:407–43.

Pindyck, Robert S. 2007. "Governance, Issuance Restrictions, and Competition in Payment Card Networks." MIT Sloan Research Paper 4651-07. (http://ssrn.com/abstract=996722).

Prasad, A., and Saibal Ghosh. 2005. "Competition in Indian Banking." IMF Working Paper 05/141. (http://ssrn.com/abstract=888010).

Rajan, Raghuram. 1992. "Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt." *Journal of Finance* 474:1367–400.

Sapienza, Paola. 2002. "The Effects of Banking Mergers on Loan Contracts." *Journal of Finance* 57: 1891–921.

Schulz, Heiner. 2006. "Foreign Banks in Mexico: New Conquistadors or Agents of Change?" Wharton Financial Institutions Center Working Paper 06-11. (http://ssrn.com/abstract=917149).

Sengupta, Rajdeep. 2007. "Foreign Entry and Bank Competition." *Journal of Financial Economics* 84: 502–28.

Strahan, Philip. 2003. "The Real Effects of US Banking Deregulation." *Federal Reserve Bank of St. Louis Review*, 85(4):111–28.

Tressel, Thierry and, and Thierry Verdier. 2007. "Does Financial Globalization Discipline Politically Connected Firms?" Working Paper, IMF, Washington, DC.

Vives, Xavier. 2001. "Competition in the Changing World of Banking." *Oxford Review of Economic Policy* 17:535–45.

Wong, Jim, T.C. Wong, Tom Fong, and Ka-fai Choi. 2006. "Competition in Hong Kong's Banking Sector: A Panzar-Rosse Assessment." (http://ssrn.com/abstract=1032034).

World Retail Banking Report. 2005. "Produced by Capgemini, ING, and the European Financial Management & Marketing Association EFMA."

Yildirim, H. Semih, and George C. Philippatos. 2007. "Restructuring, Consolidation, and Competition in Latin American Banking Markets." *Journal of Banking and Finance* 31(3):629–39.

Access to Financial Services: Measurement, Impact, and Policies

Thorsten Beck, Asli Demirgüç-Kunt, and Patrick Honohan

In many developing countries less than half the population has access to formal financial services, and in most of Africa less than one in five households has access. Lack of access to finance is often the critical mechanism for generating persistent income inequality, as well as slower economic growth. Hence expanding access remains an important challenge across the world, leaving much for governments to do. However, not all government actions are equally effective and some policies can even be counterproductive. This paper sets out principles for effective government policy on broadening access, drawing on the available evidence and illustrating with examples. The paper concludes with directions for future research. JEL Codes: D31, G20, G21, O12, O16

Financial markets and institutions exist to overcome the effects of information asymmetries and transaction costs that prevent the direct pooling and investment of society's savings. They mobilize savings and provide payments services that facilitate the exchange of goods and services. In addition, they produce and process information about investors and investment projects to guide the allocation of funds, monitor and govern investments, and help diversify, transform, and manage risk.[1] When they work well they provide opportunities for all market participants to take advantage of the best investments by channeling funds to their most productive uses, hence boosting growth, improving income distribution, and reducing poverty. When they do not work well growth opportunities are missed, inequalities persist, and in extreme cases, there can be costly crises.

Until recently econometric research on the performance of formal financial systems around the world has focused mainly on their depth, efficiency, and stability. Cross-country regressions have shown financial depth to be not only pro-growth but also pro-poor: economies with better developed financial systems experience faster drops in income inequality and faster reductions in poverty

doi;10.1093/wbro/lkn008 Advance Access publication February 27, 2009 24:119–145

levels. Much less attention has been devoted to financial outreach and inclusiveness: the extent to which individual firms and households can directly access formal financial services. Even deep financial systems may offer limited outreach. Yet important tasks of a well-functioning financial system are providing savings, payments, and risk-management products to as large a set of participants as possible and seeking out and financing any and all worthwhile growth opportunities. Without inclusive financial systems, poor individuals and small enterprises need to rely on their personal wealth or internal resources to invest in their education, become entrepreneurs, or take advantage of promising growth opportunities. It seems plausible, therefore, that an inclusive financial system might be associated not only with lower social and economic inequality, but also with a more dynamic economy as a whole (Rajan and Zingales 2003).

Modern development theories increasingly emphasize the key role of access to finance: lack of finance is often the critical mechanism for generating persistent income inequality, as well as slower economic growth. That is not to say that more borrowing by poor people or by highly leveraged enterprises is always a good thing. Abuses revealed in the United States sub-prime mortgage crisis of 2007-08 underline the danger of overborrowing, whether by individuals misled through predatory lenders or by over-optimistic entrepreneurs.

Earlier theories postulated that a rise in short-term inequality was an inevitable consequence of the early stages of economic development (Kuznets 1955, 1963). However, modern theory has examined the ways in which inequality can adversely affect growth prospects through limiting human capital accumulation and occupational choices, which implies that wealth redistribution can spur development (Banerjee and Newman 1993; Galor and Zeira 1993). Despite the emphasis that financial market imperfections receive in theory, development economists often take them as given and focus their attention on redistributive public policies to improve wealth distribution and to foster growth.[2] However, financial market imperfections which limit access to finance play an important role in perpetuating inequalities, so that financial sector reforms that promote broader access to financial services should be at the core of the development agenda. Indeed, the task of redistribution may have to be endlessly repeated if financial market frictions are not addressed, damaging incentives to work and save. In contrast, building inclusive financial systems creates positive incentive effects by equalizing and expanding individual opportunities. While theory highlights the risk that selectively increased access could worsen inequality, both cross-country data and evidence from particular policy experiments suggest that a more developed financial system is associated with lower inequality in the medium- to long-term. While still far from conclusive, the bulk of the evidence suggests financial development and improving access to finance is likely to not only accelerate economic growth, but also reduce income inequality and poverty.

Financial market imperfections—such as information asymmetries and transactions costs—are likely to be especially binding on the talented poor and the micro and small enterprises. Without inclusive financial systems these parties are limited by their lack of collateral, credit histories, and connections, and have only their own savings and earnings.[3] However, this access or outreach dimension of financial development has often been overlooked, mostly because of serious data gaps on the people using financial services, the types and quality of services they receive, and the price they pay, as well as a lack of systematic information on the barriers to broader access. But since the concept of financial access resists a simple quantifiable definition, all of these dimensions need to be examined, along with the causes of all of the barriers—price and non-price—to financial inclusion.

Drawing on a recent comprehensive review of econometric research on the measurement, determinants, and impact of access to finance (World Bank 2007), this paper reflects on what is known about the extent of financial access, its determinants, and the impact of access on growth, equity, and poverty reduction. It also discusses the role of government in advancing financial inclusion both of firms and households. Though much remains to be learned, a significant amount of empirical analysis has been conducted on these issues over the past years. As with any review, taking stock of all this research also allows us to identify the many gaps in our knowledge, which help chart the way for a new generation of research.

Specifically, the remainder of the paper covers the following themes:

- *Measurement.* How well do the financial systems in different countries directly serve the poor households and small enterprises? Who uses which financial services (e.g., deposits, credit, payments, insurance)? What are the chief obstacles and policy barriers to broader access? This section discusses some indicators based on surveys of financial service providers and their regulators, as well as users of these services (firms and households) to illustrate the extent of financial inclusion around the world.
- *Evaluating the impact of access.* How important is access to finance as a constraint to firm growth? What are the channels through which improved access affects firm growth? What is the impact of access to finance for households and micro-enterprises? What aspects of financial sector development matter for broadening access to different types of financial services? What techniques are most effective in ensuring sustainable provision of credit and other financial services on the small scale? This paper synthesizes research on the impact of access on firms and households.
- *Policies to broaden access.* What is the government's role in building inclusive financial systems? Given that financial systems in many developing countries

serve only a small part of the population, expanding access remains an important challenge across the world, leaving much for governments to do. However, not all government actions are equally effective and some policies can even be counterproductive. In this section the paper sets out principles for effective government policy on broadening access, drawing on the available evidence and illustrating with examples.

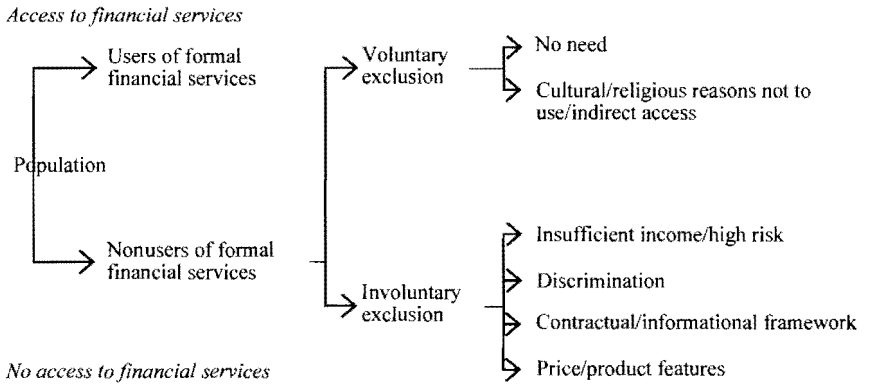
Finally, the paper concludes with directions for further research.

Measurement

While copious amounts of data are available on many aspects of the financial sector, systematic indicators of the inclusiveness of the financial sector are not. Most of the evidence concerning the causal links between financial development, growth, and poverty comes from aggregate data using, for example, financial depth measures (how much finance) rather than outreach or access measures (how many users). Meanwhile, microeconomic studies in the field have tended to use financial or real wealth to proxy for credit constraints.[4] It is only recently that researchers have started to compile cross-country indicators on the outreach and access dimensions of financial development.

It is important to distinguish between access to and use of financial services (figure 1). Critically, non-users of financial services can be differentiated between those that are involuntarily excluded and those that are voluntarily self-excluded. Voluntary self-exclusion can be attributed to a lack of need for financial services, religious or cultural reasons, or indirect access to services through friends and family. In all of these cases, voluntary non-use is driven by lack of demand and

Figure 1. Distinguishing between Access to and use of Financial Services



therefore does not pose problems for policy makers. However, for the involuntarily excluded it is important to distinguish between four different groups in order to formulate proper policy advice. First, there is a group of households and firms that are not considered bankable because their incomes are too low or they pose too high a lending risk. Rather than trying to include them in the financial system, non-lending support mechanisms might be more appropriate. The other three involuntarily excluded groups need to overcome (a) discriminatory policies, (b) deficiencies in the contractual and informational frameworks, or (c) price and product features. Any of these problems can exclude large parts of the population, especially in the developing world, and all call for specific policy actions.

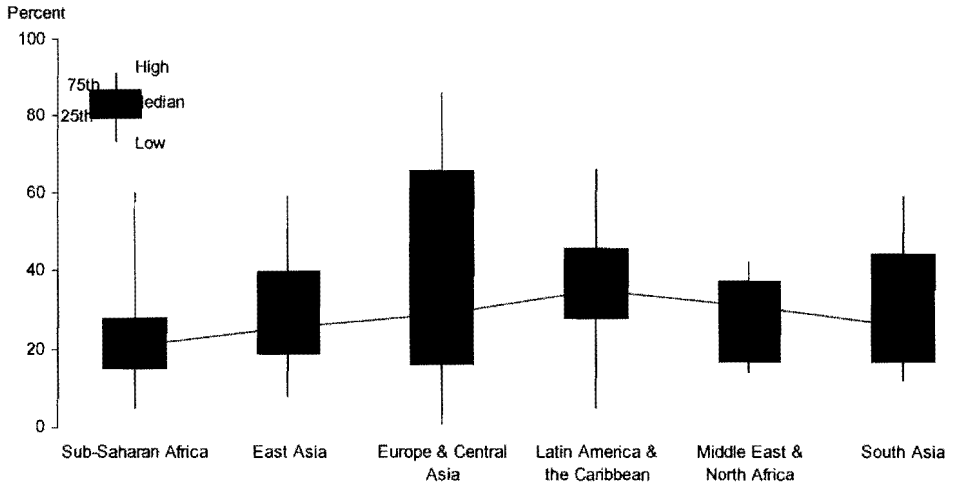
Across these groups, three main approaches to measuring access and usage have produced promising results. The first seeks to count the number of users of basic financial services, the second relies on the subjective assessments of firms as to the quality of the financial services that they obtain, and the third looks at physical and cost barriers to access. Each approach has its shortcomings: in the case of the first approach, the quality and price of the services received by the account holders of different formal or semi-formal financial institutions may vary substantially; in the case of the second, the robustness or interpretability of subjective assessments of service quality may be questionable; and in the case of the third, data on some barriers (such as distance to a bank branch, or documentary requirements to open an account) may be easier to assemble and therefore more complete than data on other barriers. Still, these data help us understand the reasons for financial exclusion and provide hints as to which policies could be helpful in removing barriers and broadening access.

Despite the usefulness of these methods the limitations of available data are striking: even the number of individuals with a bank account is not known from regulatory or industry sources. While we may know how many accounts exist, many individuals and firms have multiple accounts, others have none, and regulatory authorities generally do not collect data on individual account holders. The best data would be generated by a census or survey of users, which would allow researchers to measure financial access across sub-groups. However, few such surveys exist for households and there are problems with cross-country compatibility of the data sets. In the absence of comprehensive micro-data, researchers have sought to create synthetic headline indicators, combining the results of existing surveys with more readily available macro-data on the number of accounts and financial depth indicators (Honohan 2008a). For example, the proportion of households with some access to a bank account can be approximated by a non-linear function of the number of accounts in commercial banks and microfinance institutions (MFIs) and the average size of these accounts: the available survey data on household accounts has shown the validity of this function.

Headline indicators like these indicate that household access to and use of financial services are very limited around the world. Although in several European countries more than 90% of households have a bank account, in many developing countries less than half of households have an account and in many African countries less than one in five households have an account (figure 2).

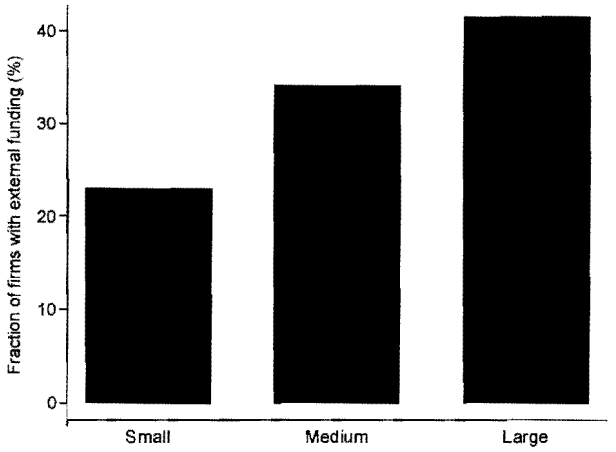
Quite a few systematic surveys of firms, although generally neglecting informal firms, have thrown light on both the financial structure of firms and on their managements' perspectives on service quality. These surveys include the Regional Program on Enterprise Development (RPED) studies for Sub-Saharan Africa in the 1990s, the World Bank- European Bank for Reconstruction and Development Business Environment and Enterprise Performance Survey (BEEPS) for the transition economies, the World Business Environment Survey (WBES) across 80 countries in 1999/2000, and the Investment Climate Assessment (ICA) surveys over the past five years which are available for almost 100 countries. These surveys ask firms to rate the extent to which access to and cost of external finance constitute obstacles to their operation and growth, with higher numbers indicating higher obstacles. In general, small firms in both the WBES and ICA surveys report lack of financing to be one of the most important business constraints they face. One of the most consistent findings of these surveys is that small firms seem to face larger access barriers: for example, fewer than 20 percent of small firms use external finance, about half the rate of large firms

Figure 2. Proportion of Households with an Account in a Financial Institution



Data by country grouped by region.
Source: World Bank (2007).

Figure 3. Percentage of Firms using External Finance, by Firm Size



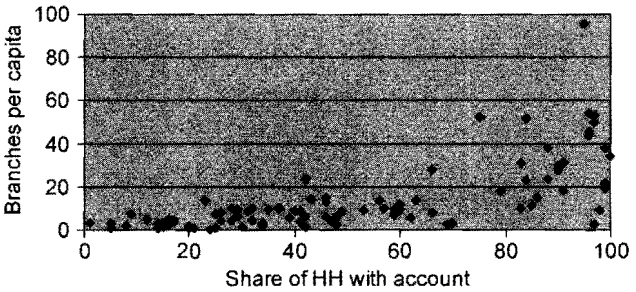
Source: World Bank (2007).

(figure 3). We will discuss in the next section the extent to which the self-reported obstacles and the use of external finance are related to real outcomes.

Geography, or physical access, is among the barriers that prevent small firms and poor households in many developing countries from using financial services. While some services may be accessible over the phone or via the Internet, others require clients to visit a branch or use an ATM. Ideally, we would like to know how far customers are from the location of the nearest branch (or ATM); the density of branches per square kilometer or per capita provide an initial, albeit crude, alternative indicator. For example, while Spain has 96 branches per 100,000 people and 790 branches per 10,000 square kilometers, Ethiopia has less than one branch per 100,000 people and Botswana has one branch per 10,000 square kilometers (Beck, Demirgüç-Kunt, and Martinez Peria 2007). Not surprisingly, the share of households with a financial account tends to be higher in countries with denser branch networks (figure 4).

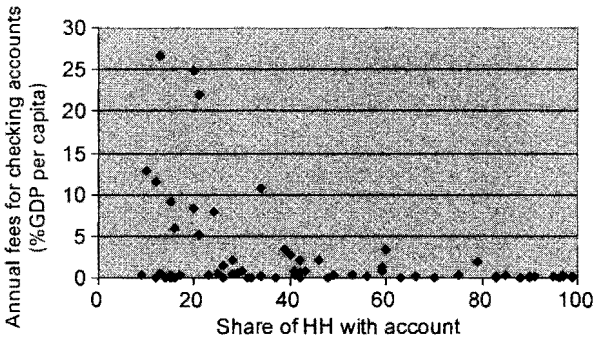
Another barrier is in providing the documents necessary to open an account. Financial institutions usually require one or more documents for identification—such as passports, drivers licenses, pay slips, or proofs of residence—but in many low-income countries a majority of people lack such papers, especially when they are not employed in the formal sector. Furthermore, many institutions have minimum account size requirements or fees: for example, in large parts of Africa it is not unusual for banks to require a minimum deposit equivalent to 50 percent of the population's per capita GDP to open a checking account (Beck, Demirgüç-Kunt, and Martinez Peria 2008). High fees to maintain checking accounts can exclude large parts of the population, as illustrated in figure 5.

Figure 4. Financial Inclusion and Branch Penetration



Source: Beck, Demirgüç-Kunt, and Martinez Peria (forthcoming).

Figure 5. Financial Inclusion and Affordability



Source: Beck, Demirgüç-Kunt, and Martinez Peria (forthcoming).

These barriers to access vary significantly across countries. Lower barriers tend to be associated with more open and competitive banking systems, which are characterized by: private ownership of banks; foreign bank participation; stronger legal, information and physical infrastructures; regulatory and supervisory approaches that rely more heavily on market discipline; and greater transparency and freedom for the media (Beck, Demirgüç-Kunt, and Martinez Peria 2008). While these are simple correlations, they hold even when controlling for the level of economic development, thus providing a sense of what policies are associated with more inclusive financial systems.

The measurements mentioned above all refer to the formal financial sector, reflecting the view that formal finance potentially offers considerable advantages over the informal. The alternative argument—that informal financial systems may substitute for formal—has been canvassed for the case of China by Allen,

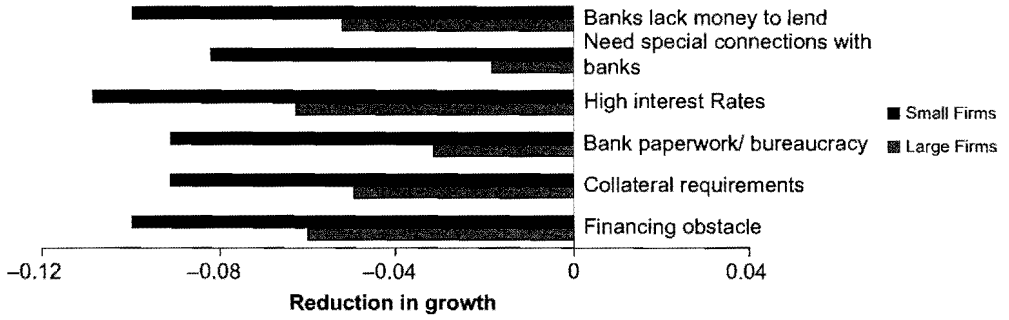
Qian, and Qian (2005, 2008). But their line of reasoning assumes obstacles to formal financial development such as restrictions on entry and pervasive state ownership of banks. Even if informal finance operates in such conditions, it is just a second-best solution. Besides, informal sources of finance vary widely in their effectiveness. Ayyagari, Demirgüç-Kunt, and Maksimovic (2007b) provide evidence from China that, on average for the firms in their sample, access to formal finance was associated with faster firm growth while the use of informal financial sources was not.

However, access indicators are just that—indicators. While they are linked to policy, they are not policy variables. Thus, examining indicators is only the beginning of the effort. To understand the impact of financial access and to design better policy interventions it is necessary to collect and analyze in-depth household and enterprise information on access to and use of financial services. Better data and analysis will help us assess which financial services (savings, credit, payments, or insurance) are the most important for development outcomes and will suggest which cross-country indicators are worth tracking over time.

Evaluating the Impact of Access to Finance for Firms

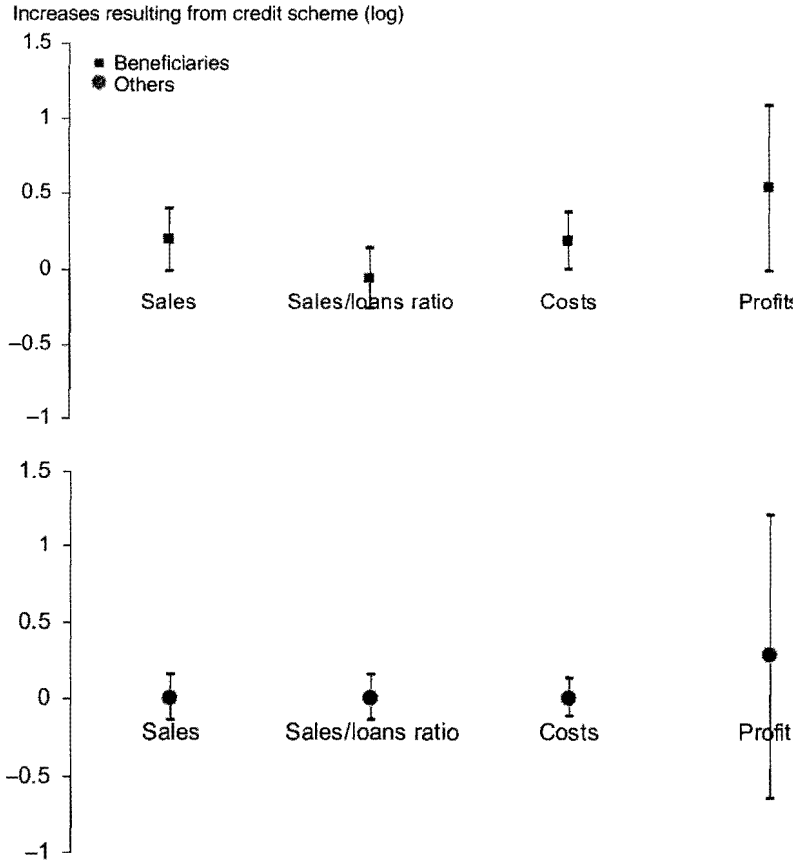
One of the important channels by which finance promotes growth is through the provision of credit to the most promising firms. Recent research utilizing detailed firm-level data and survey information provides direct evidence on how access constraints affect firm growth. Analysis of survey data suggests that firms, particularly small firms, not only often complain about lack of access to finance, but actually have slower growth rates (figure 6; Beck, Demirgüç-Kunt, and Maksimovic 2005; Beck and others 2006). The findings of these broad cross-country regressions are supported by individual case studies utilizing detailed loan and borrower information. Specifically, Banerjee and Duflo (2004) studied detailed loan information on 253 small- and medium-sized borrowers from an Indian bank both before and after they became eligible for a directed credit program.[5] They showed that these firms expanded after becoming eligible, suggesting that they were previously constrained by their lack of credit (figure 7). Experimental evidence from Mexico and Sri Lanka confirms the marginal productivity of micro-entrepreneurs without access to financing (De Mel, McKenzie, and Woodruff 2008a; McKenzie and Woodruff 2008). Micro-entrepreneurs in these two countries were randomly given grants to purchase inputs and saw returns of 5 to 20 percent per month compared to micro-entrepreneurs that did not benefit from these grants. These case studies show that access to external finance has strong positive impacts on firm growth, especially on small and micro-enterprises.

Figure 6. The Effect of Financing Constraints on Growth: Small vs. Large Firms



Source: Beck, Demirgüç-Kunt, and Maksimovic (2005).

Figure 7. Response of Beneficiaries and Nonbeneficiaries under a Credit Scheme



Note: Error bars indicate 95 percent confidence levels.
Source: Based on Banerjee and Duflo (2004).

Access to finance and the associated institutional underpinnings favorably affect firm performance along a number of different channels. Functional improvements in the formal financial sector can reduce financing constraints more for small firms and others who have difficulty in either self-financing or finding private or informal sources of funding. Research indicates that access to finance promotes more start-ups: it is smaller firms that are often the most dynamic and innovative (Klapper, Laeven, and Rajan 2006). Not only do countries with financial barriers lose the growth potential of these enterprises, they also risk missing opportunities to diversify into new areas. Financial inclusion also enables established firms to reach a larger equilibrium size by exploiting growth and investment opportunities (Beck, Demirgüç-Kunt, and Maksimovic 2006). Furthermore, greater financial inclusion allows firms both the choice of more efficient asset portfolios and a greater ability to innovate (Claessens and Laeven 2003; Ayyagari, Demirgüç-Kunt, and Maksimovic 2007a).

If stronger financial systems can promote new-firm entry, enterprise growth, innovation, equilibrium size, and risk reduction, then they will almost inevitably improve aggregate economic performance. It is important to note that finance does not raise aggregate firm performance uniformly, but transforms the structure of the economy by impacting different types of firms in different ways. At any given level of financial development, small firms have more difficulty than large ones in accessing external finance. However, research shows that small firms benefit the most from financial development both in terms of entry and seeing their growth constraints relaxed (Beck, Demirgüç-Kunt, and Maksimovic 2005; Klapper, Laeven, and Rajan 2006). Financial deepening can also increase incentives for firms to incorporate in order to benefit from the resulting opportunities of risk diversification and limited liability (Demirgüç-Kunt, Love, and Maksimovic 2006). Financial deepening can also help foster more independent enterprises, moving economies away from the predominance of family-owned firms or business groups (Rajan and Zingales 2003). Hence, inclusive financial sectors also have critical consequences for the composition and competition in the enterprise sector.

Firms finance their investments and operations in many different ways, reflecting a wide range of both internal and external factors. The availability of external financing depends not only on each firm's individual situation, but on the wider policy and institutional environment supporting the enforceability and liquidity of the contracts that are involved in financing firms. Availability also depends on the existence and effectiveness of a variety of intermediaries and ancillary financial firms that help connect fund providers and users. Bank finance is typically the major source of external finance for all firms, regardless of size (Beck, Demirgüç-Kunt, and Maksimovic 2008). Modern trends toward transactional lending suggest that improvements in information availability (for example, through

development of credit registries) and technological advances in analysis of these improved data (such as use of automated credit appraisal) are likely to improve access of small and medium enterprises (SMEs; Brown, Jappelli, and Pagano, forthcoming). Provided that the relevant laws are in place, asset-based lending—such as factoring, fixed-asset lending, and leasing— are other technologies which can also release sizable financing flows even for small and non-transparent firms.[6]

However, relationship lending—lending based on the loan officer's personal assessment of the borrower and their long-term and repeated contractual arrangements—will remain important in environments with weak infrastructures and informal economic activity. Relationship lending is costly for the lender and requires either high spreads or large volumes to be viable. If the customer's creditworthiness is hard to evaluate, then there may be no alternative to relationship lending. Indeed, limited access to credit in some difficult environments may be attributable to existing intermediaries reluctance to participate in relationship lending on a small scale (Honohan and Beck 2007).

Globalization of finance can also play a part in improving access, by increasing both the flow of investable funds and the efficiency of capital allocation. The most important contribution of international financial service providers, and especially foreign direct investment (FDI), is often their expertise. Considerable South-South technology transfer continues to occur between microfinance providers, reflecting the leadership role that MFIs in developing countries have had in extending access. Only recently many mainstream banks have become interested in profitable provision of financial services to micro, small, and medium enterprises. Their contributions to financial access have always been controversial, however, partly for political reasons. Foreign owners bring capital, technology, know-how, and independence from local business and political elites, but debate continues over whether they have improved access. Most foreign banks are relatively large and do not concentrate on SME lending, choosing instead to stick mainly to the banking needs of larger firms and of individuals with high net worth (Mian 2006). Nonetheless, the increased competition for large customers often drives local banks to focus more on providing profitable services to segments which they had neglected. The balance of a large body of evidence suggests that opening to foreign banks is likely to improve access of SMEs over time, even if the foreign banks often confine their lending to large firms and government. Other evidence, however, has shown that foreign banks use their expertise and technology to go down-market and cater to SMEs' needs (De la Torre, Martinez Peria, and Schmukler 2008). The aggregate evidence is mostly positive: in countries where foreign banks represent a relatively large share of the market, firms are less likely to report access to finance as a problem, regardless of whether they are small, medium, or large (Clarke, Cull, and Martinez Peria 2006). In contrast, the

performance of state-owned banks in this dimension has tended to be poor (La Porta et al., 2002).

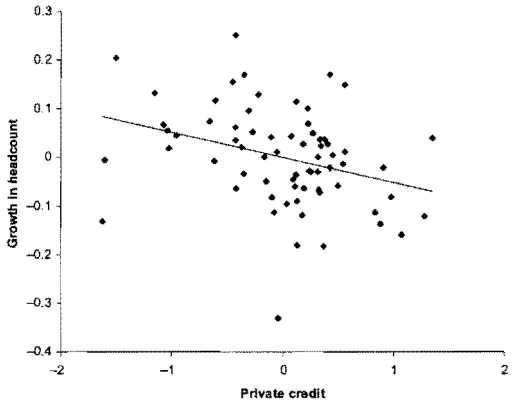
Non-bank finance remains much less important in most developing countries, but it can play an important role in improving the price and availability of long-term credit to small borrowers. Bond finance can provide a useful alternative to bank finance but has limited potential, as shown by the example of the Korean bond market that emerged after a crisis curbed bank lending (Gormley, Johnson, and Rhee 2006). It was mostly larger enterprises that could tap this bond market due to the public's expectations that large enterprises were too big to fail and would be bailed out by government; expectations which were fulfilled after the 1999 collapse of the large company Daewoo.

The emergence of a large market in external equity requires strong investor rights and transparency; these allow for capital inflows that can greatly improve access and lower costs, including for smaller firms which benefit from spill-over effects. This is true both for portfolio equity investments and for FDI and private equity, which are likely to become increasingly important in the future. However, investor rights and transparency might not be enough to foster liquid equity markets: a critical mass of issues, issuers, and investors is also necessary (De la Torre, Gozzi, and Schmukler 2006). While opening up a country's equity market and allowing local firms to list in a foreign stock exchange can both improve access and cost of equity finance for large local firms (Aggarwal, Klapper, and Wysocki 2005) and help import corporate governance (Coffee 2002), it can also result in a loss of liquidity for small local firms (Levine and Schmukler 2007). However, the net benefit is not necessarily negative for small firms: improved access to external finance for large firms may spill over to small firms through trade credit and through forcing internal banks to go down-market as they face competition and lose large clients to equity investors.

Evaluating Impact of Access to Finance for Households

Over the long term, economic growth helps reduce poverty and can be expected to lift the welfare of most households. Finance helps reduce poverty indirectly by fostering economic growth. But does financial deepening help all population segments to the same extent? Evidence suggests that, overall, financial development is not only pro-growth, but also pro-poor. There is econometric evidence that financial development disproportionately boosts the income growth of the lowest income quintile and reduces the share of people living on less than a dollar per day (figure 8; Honohan, 2004; Beck, Demirgüç-Kunt and Levine, 2007). This effect is not only statistically, but also economically significant. Even after controlling for other factors, variation in financial development accounts for 30 percent

Figure 8. Financial Depth and Poverty Alleviation



Note: Data averaged over period 1980–2005, controlling for intial level of headcount.
Source: Beck, Demirguc-Kunt, and Levine (2007).

of the total cross-country variation in changing poverty rates. Consider the example of Chile and Peru. While the share of the population living on less than one dollar per day fell by an average of 14 percent a year in Chile between 1987 and 2000, it rose by a similar rate in neighboring Peru. Cross-country regressions suggest that if Peru had started with as deep a financial system as Chile (private credit of 47 percent rather than 17 percent), its poverty count in 2000 would have been only 5 rather than the actual 10 percent of the population.

In this process, how important is the direct provision of financial services to poor households and individuals? Existing evidence suggests that direct effects of access to finance might be less important than indirect second-round effects created through more efficient product and labor. First, consider the different results from aggregate cross-country regressions and micro-studies. While cross-country comparisons suggest that financial depth (as opposed to financial inclusion) has a statistically and economically strong impact on poverty alleviation, micro-studies studying individuals' credit access without considering spill-over effects can provide only a tenuous picture of profit or welfare outcomes (Morduch 1998; Pitt and Khandker 1998; Coleman 1999; Karlan and Zinman 2006).

Careful country studies provide a different approach to assess the channels through which financial deepening helps reduce poverty. Evidence from the United States' experience suggests that the income distribution decline following branch deregulation was due to the increased participation of unskilled individuals in the labor market, closing the income gap between skilled and unskilled and tightening the income distribution (Beck, Levine, and Levkov 2007).

Financial liberalization had no significant effect on human capital accumulation, nor did the following increase in entrepreneurship contribute to the tightening of the income distribution. Similarly, general equilibrium models (using micro-data for Thailand and taking into account labor market effects) suggest that finance's main impact on income inequality comes not through broadening access to credit, but through higher wages and including a larger share of the population in the formal economy (Gine and Townsend 2004). Hence, the favorable effect of finance on poverty may *not* be coming mainly through direct provision of financial services to the poor. Pro-poor financial policy should therefore certainly not neglect the importance of fostering more efficient capital allocation through competitive and open financial markets.

By no means does this imply that improving access to financial services should not be a policy goal. With as few as 20 to 50 percent of the worldwide population having an account at a formal or semi-formal financial intermediary, there is considerable scope for improvement. Even non-poor households and micro and small enterprises are excluded from all but the most basic financial services (De Mel, McKenzie, and Woodruff 2008b). Therefore, for the most part improving the quality and efficiency of services without broadening access is likely to be insufficient as it will leave large segments of the population, and their talents and innovative capacity, untapped. Providing better financial access to these excluded non-poor micro and small entrepreneurs can have a strongly favorable indirect effect on the poor. Hence, to promote pro-poor growth it is important to broaden the focus of attention from finance for the poor to improving access for all excluded parties (Rajan 2006). However, this evidence also suggests that the discussion should be broadened to financial services other than credit.

There are many reasons why the poor do not have access to financial services. Social as well as physical distance from the formal financial system may matter. The poor may not have anybody in their social network who understands the various services that are available to them. Lack of education may make it difficult for them to fill out loan applications, and the small number of transactions they are likely to undertake may make the loan officers think it is not worthwhile to help them. Mainstream financial institutions are more likely to locate their retail outlets in relatively prosperous neighborhoods, explaining why the poor are often located far from banks. Even if financial service providers are nearby, in some cases poor clients may encounter prejudice, even being refused admission to banking offices. Specifically for access to credit services, there are two important problems. First, the poor have no collateral and cannot borrow against their future income because they tend not to have steady jobs or income streams to keep track of. Second, dealing with small transactions is costly for the financial institutions.

The new wave of specialized microfinance institutions serving the poor has tried to overcome these problems in innovative ways.[7] Loan officers go to the poor, instead of waiting for the poor to come to them. Group lending schemes improve repayment incentives and monitoring through peer pressure, while building support networks and educating borrowers (Ghatak and Guinnane 1999; Karlan and Valdivia 2006; Karlan 2007). Increasing loan sizes as customers continue to borrow and repay reduces default rates. The effectiveness of these innovations in different settings is still being debated. Recently, many MFIs have moved away from group lending products to individual lending, especially in cases where the borrowing needs of customers starts to diverge; initial evidence has shown both techniques to be successful (Gine and Karlan 2006).

Over the past few decades, microfinance institutions have managed to reach millions of clients and have achieved impressive repayment rates, forcing economists to reconsider whether it is really possible to make profits while providing financial services to some of the world's poorest individuals. Indeed, mainstream banks have begun to adopt some of the techniques of the microfinance institutions and to enter some of the same markets. For many, however, the most exciting promise of microfinance is that it could reduce poverty without requiring ongoing subsidies. But has microfinance been able to meet this promise?

While many heartening case studies are cited—from contexts as diverse as slums of Dhaka to villages of Thailand to rural Peru—the overall impact microfinance has had on poverty is still unclear. The uncertainty in evaluating impact is due to methodological difficulties, such as selection bias. Rigorous micro-studies compare groups of borrowers to non-borrowers, controlling for individuals' characteristics and using eligibility criteria or random assignment as identification restriction to overcome problems of unobserved borrower characteristics being correlated with outcomes. While some of these studies have shown a positive impact of access to credit (Karlan and Zinman, forthcoming), some have not (Coleman 1999) and some have depended on the econometric methodology utilized (Morduch 1998; Pitt and Khandker 1998). It is important to note that income is only one measure of welfare in the case of households. Analyses have shown that consumption smoothing, not having to use child labor as buffer in times of negative income or health shocks, and increasing women's participation in family and community decisions are other important welfare indicators. However, these analyses of financial access have mostly used proxy variables, such as durable asset holding and proximity to a bank branch.

Although the attention of microfinance has traditionally focused on the provision of credit for very poor entrepreneurs and enthusiasts often emphasize how the productivity and growth potential of borrowers will be unleashed by microfinance, much of micro-credit is not used for investment. Instead, a sizable fraction of microcredit goes to meet important consumption needs (Johnston and

Morduch 2008). These are not a secondary concern. For poor households, credit is not the only or in many cases the primary financial service they need: good savings and payments services (including international remittances) and insurance may rank higher. For example, one of the reasons why the poor may not save in financial assets may be the lack of appropriate products, such as simple transaction or savings accounts rather than costly checking accounts. Research by Ashraf, Karlan, and Yin (2006a, 2006b, 2006c) has shown that innovative savings products (such as collecting deposits directly from customers) and savings commitment products can increase savings. The demand for microcredit used for consumption purposes could thus signal a demand for more appropriate savings products.

One of the most controversial questions about microfinance is the extent of subsidy required to provide access. Although group lending and other technologies are employed to overcome the obstacles involved in delivering services to the poor, these are nevertheless costly technologies and the high repayment rates have not always translated into profits. Overall, much of the microfinance sector—especially the segment that serves the poorest individuals—still remains heavily dependent on grants and subsidies. Recent research confirms that there is a trade-off between profitability and serving the poorest population segments (Cull, Demirgüç-Kunt, and Morduch 2007).

Then the question remains whether finance for the very poor should be subsidized and whether microfinance is the best way to provide those subsidies. Answering this question requires comparing costs and benefits of subsidies in the financial sector with those in other areas, such as education and infrastructure. There is likely to be a better case for subsidizing savings and payments services, which can be seen as basic services necessary for participation in a modern market economy, compared to credit services. In the case of credit encouraging and taking advantage of technological advances—which are becoming more wide-spread and fast-paced due to globalization—may be more promising than providing subsidies, given the negative incentive effects of subsidies on repayment and the potential disincentives for service providers in adopting market-based innovations.

Perhaps more importantly, as we already discussed, the greatest benefits for poor households require a strategy that goes well beyond credit. It is not only the poor that lack access to formal financial services. The limited access to financial services by non-poor entrepreneurs is likely to be even more important for growth and overall poverty reduction. There are also good political economy reasons to focus on ways to make financial services available for all: defining the problem more broadly would help mobilize the efforts of a much more powerful political constituency, increasing the likelihood of success.

Policies to Broaden Access

Since expanding access remains an important challenge even in developed economies, it is not enough to say that the market will provide. Market failures related to information gaps, the need for coordination on collective action, and concentrations of power mean that governments everywhere have an important role to play in building inclusive financial systems (Beck and de la Torre 2007). However, not all government action is equally effective and some policies can even be counterproductive. Direct government interventions to support access require a careful evaluation which is often missing.

Even the most efficient financial system, supported by a strong contractual and information infrastructure, faces limitations. Not all would-be borrowers are creditworthy and there are numerous examples of national welfares that have been damaged by overly relaxed credit policies. Access to formal payment and savings services can approach universality as economies develop, although not everyone will or should qualify for credit. For example the sub-prime crisis in the United States graphically illustrates the consequences of encouraging low-income households to borrow beyond their ability to repay.

An underlying, albeit often long-term, goal is deep institutional reform ensuring security of property rights against expropriation by the state. Prioritizing some institutional reforms over others, however, would help focus reform efforts and produce impact in the short- to medium-term. Recent evidence suggests that, in low-income countries, it is the information infrastructures that matter most, while in high-income countries enforcement of creditor rights is more important. Cross-country variation in financial depth can be explained in low-income countries by the existence of credit information systems but not by the efficiency in contract enforcement, and in the case of high-income countries results are reversed (Djankov, McLiesh, and Shleifer 2007). The recent case of a Guatemalan microfinance institution that joined a credit bureau demonstrates the positive effects that introducing credit registries has on reducing adverse selection and moral hazard. Given that borrowers were only informed that their information was shared after the fact, this entry allowed researchers to identify and quantify the dampening effect of credit information sharing on loan default rates (de Janvry, Sadoulet, and McIntosh 2006).

But even within the contractual framework there are certain shortcuts to long-term institution building. In relatively underdeveloped institutional environments procedures that enable individual lenders to recover on debt contracts (for example, those related to collateral) are more important in boosting bank lending than those procedures mainly concerned with resolving conflicts between multiple claimants (for example, bankruptcy codes; Haselmann, Pistor, and Vig 2006). Given that it is potentially easier to build credit registries and reform

procedures related to collateral than to make lasting improvements in the enforcement of creditor rights and bankruptcy codes, these are important findings for prioritizing reform efforts. Introducing expedited mechanisms for loan recovery can be helpful, as shown in the example of India where a new mechanism bypassing dysfunctional court procedures increased loan recoveries and reduced interest rates for borrowers (Visaria, forthcoming).

Results can be produced relatively fast by encouraging both improvements in specific infrastructures (particularly in information and debt recovery) and the launch of financial market activities that can allow technology to bring down transaction costs. Some examples of these market activities are as follows: establishing credit registries or issuing individual identification numbers to establish credit histories; reducing costs of registering or repossessing collateral; and introducing specific legislation to underpin modern financial technology, from leasing and factoring to electronic-finance and mobile-finance. These can produce results relatively fast, as the success of m-finance in many Sub-Saharan African countries has shown, most recently MPesa in Kenya (Porteous 2006).

Encouraging openness and competition is also an essential part of broadening access, as it both encourages incumbent institutions to seek out profitable ways of providing services to the previously excluded segments of the population and increases the speed with which access-improving new technologies are adopted. Foreign banks can play an important role in fostering competition and expanding access (Claessens, Demirgüç-Kunt, and Huizinga 2001; Claessens and Laeven 2004).

In this process, providing the private sector with the right incentives is key, hence the importance of good prudential regulations. Competition that helps foster access can also result in reckless or improper expansion if not accompanied by a proper regulatory and supervisory framework. As increasingly complex international regulations such as Basel II are imposed on banks to help minimize the risk of costly bank failures, it is important to ensure that these arrangements do not inadvertently penalize small borrowers by failing to make full allowance for the risk-pooling potential of a portfolio of SME loans. Research suggests that while banks making small loans have to set aside larger provisions against the higher expected loan losses from small loans—and therefore need to charge higher rates of interest to cover these provisions—they should need relatively less capital to cover the upper tail of the distribution and support the risk that losses will exceed their expected value (i.e., to cover what are sometimes known as "unexpected" loan losses; Adasme, Majnoni, and Uribe 2006).

A variety of other regulatory measures is needed to support wider access. But some policies that are still widely used do not work. For example interest ceilings fail to adequately provide consumer protection against abusive lending, as banks replace interest with fees and other charges. Increased transparency,

formalization, and enforcement of lender responsibility offer a more coherent approach, along with support for the over-indebted (Honohan 2004). However, delivering all of this can be administratively demanding.

The scope for direct government interventions in improving access is more limited than often believed. There is a large body of evidence that suggests interventions to provide credit through government-owned financial institutions have generally not been successful (La Porta, Lopez-de-Silanes, and Zamarripa 2003; Levy-Yeyati and Micco, 2007). One of the reasons is that lending decisions are based on the political cycle rather than socio-economic fundamentals, as both cross-country evidence and a carefully executed case study for India show (Cole 2004; Dinç 2005).

In non-lending services, the experience of government-owned banks has been more mixed. A handful of governmental financial institutions have moved away from credit and evolved into providers of more complex financial services, entering into public/private partnerships to help overcome coordination failures, first-mover disincentives, and obstacles to risk sharing and distribution (de la Torre, Gozzi, and Schmukler 2007). A good example is the setup of an electronic factoring platform by a Mexican development bank (NAFIN) that brings together small suppliers, large purchasers, and banks. Ultimately, these successful initiatives could have been undertaken by private capital, but the state had a useful role in jump-starting these services. Direct intervention through taxes and subsidies can be effective in certain circumstances, but experience suggests that this intervention is more likely to have significant unintended consequences in finance compared to other sectors.

With direct and directed lending programs discredited in recent years, partial credit guarantees have become the direct intervention mechanism of choice for SME credit activists. Some seem to be functioning well, breaking even financially thanks to the incentive structure built into the contract between the guarantor and the intermediary banks. For example, the Chilean scheme has the intermediary banks bidding for the percentage rate of guarantee and they can adjust the premium charged on the basis of each intermediary's claims record. This has resulted not only in higher lending by beneficiaries, but in a reduction of loan losses (Cowan, Drexler, and Yañez 2008). However, other partial credit guarantees have been poorly structured, embodying sizable hidden subsidies and benefiting mainly those who do not need the subsidy. A careful study of the French guarantee schemes shows that, on the one hand, lending to beneficiaries has increased while no new borrower has benefited. On the other hand, loan losses rose, suggesting that increased risk-taking resulted in high costs for taxpayers (Lelarge, Sraer, and Thesmar 2008). The temptation for an activist government to underprice guarantees (especially for long-term loans when this will not be detected for years) does present fiscal hazards similar to those which have undermined so

many development banks in the past. In the absence of thorough economic evaluations of most schemes, their net effect in cost-benefit terms remains unclear (Honohan 2008b).

If the interest of powerful incumbents is threatened by the emergence of new entrants financed by a system that has improved access and outreach, lobbying by those incumbents can block the needed reforms (Perotti and Volpin 2004). A comprehensive financial sector reform approach aiming at better access must take these political realities into account. Given that both financial inclusion and benefits from broader access go well beyond ensuring financial services for the poor, defining the access agenda more broadly to include the middle class will help mobilize greater political support for advancing the agenda around the world (Rajan 2006).

Looking Forward: Directions for Future Research

While this paper reviews and highlights a large body of research, it also identifies many gaps in our knowledge. Much more research is needed to measure and track access to financial services, to evaluate its impact on development outcomes, and to design and evaluate policy interventions.

New development theory links the dynamics of income distribution and aggregate growth in unified models. However, while there are good conceptual reasons for believing financial market frictions exert a first order impact on the persistence of relative income dynamics, there is too little theory examining how reducing these frictions impacts the opportunities faced by individuals and the evolution of relative income levels. Future theoretical work could usefully study and provide new insights on the impact of financial sector policies on growth and income distribution within the context of these models.

Lack of systematic information on access is one of the reasons why there has been limited empirical research on access. The efforts described above in developing cross-country indicators of access are only first steps in this direction. This work should be continued and expanded, increasing coverage of countries, institutions, and types of available services. Building data sets that benchmark countries annually would help focus policymaker attention and allow us to track and evaluate reform efforts to broaden access.

While cross-country indicators of access are useful for benchmarking, micro data at the household and enterprise level is required to be able to assess the impact of access on outcomes such as growth and poverty reduction. There are few household surveys focusing on financial services. Efforts to collect this data systematically around the world are important in improving our understanding of access. Indeed, household surveys are often the only way to get detailed

information on who uses which financial services from which types of institutions, including informal ones.

Emerging evidence suggests that financial development reduces income inequality and poverty, yet we are still far from a complete understanding of the channels through which this effect operates. We are more advanced in understanding the finance-growth channel: a clear and important role for firms' access to finance has been established from promoting entrepreneurship and innovation to improve asset allocation and firm growth. But how does finance influence income distribution? How important is direct provision of finance for the poor? Is it more important to improve the functioning of the financial system to foster access to its existing firm and household clients or is it more important to broaden access to the underserved (including the non-poor who are often excluded in many developing countries)? Results of general equilibrium models and evidence at the aggregate level hint that direct access of the poor may be less important, and the knowledge that a large proportion of the non-poor are also excluded in many developing countries suggests that just improving efficiency may not be enough. Of course, efficiency and access dimensions of finance are also linked; in many countries improving efficiency would necessarily entail broader access beyond concentrated incumbents. More research is needed to sort out these effects.

In evaluating impact, randomized field experiments are promising. By introducing a random component to assignment of financial products, such as financial literacy training or random variation in the terms or availability of credit to micro-entrepreneurs and households, research can illustrate how removing barriers and improving access affects growth and household welfare. More experiments need to be conducted in different country contexts, focusing on different dimensions of access. Ultimately, it is this welfare impact that should inform which access indicators should be tracked and how policy should be designed.

Policies to broaden access can take many forms, from improvements in the functioning of mainstream finance to innovations in microfinance. Lack of careful evaluation of different interventions makes it difficult to assess their impact and draw broader lessons. More research in this area would also help improve design of policy interventions to build more inclusive financial systems.

Notes

Thorsten Beck is a Professor at the University of Tilburg, Netherlands (T.Beck@uvt.nl). Asli Demirgüç-Kunt (corresponding author, ademirguckunt@worldbank.org) is the Senior Research Manager of Finance and Private Sector in World Bank's Research Department. Patrick Honohan (phonohan@tcd.ie) is a Professor at Trinity College in Dublin, Ireland.We thank Emmanuel Y. Jimenez and three anonymous referees for helpful comments. This paper's findings,

interpretations, and conclusions are entirely those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

1. See Levine (2005) for an overview of the theoretical and empirical literature.
2. See Demirgűç-Kunt and Levine (2007) for an overview.
3. For example, Sachs writes: "When people are . . . utterly destitute, they need their entire income, or more, just to survive. There is no margin of income above survival that can be invested for the future. This is the main reason why the poorest of the poor are most prone to becoming trapped with low or negative economic growth rates. They are too poor to save for the future and thereby accumulate the capital that could pull them out of their current misery." (2005: 56–57)
4. For cross-country analysis, see Beck, Demirgüçc-Kunt, and Levine (2007) and Honohan (2004). For micro-level analysis, see among others Jacoby (1994), Guarcello, Mealli, and Rosati (2003), Jacoby and Skoufias (1997), and Beegle, Dehejia, and Gatti (2007) on the relation between durable asset holding, education, and child labor.
5. The size definition of Bannerjee and Duflo's program was changed in 1998, which enabled a new group of medium-sized firms to obtain loans at subsidized interest rates.
6. See Berger and Udell (2006) for a discussion.
7. See Armendariz de Aghion and Morduch (2005) for an overview.

References

Adasme, Osvaldo, Giovanni Majnoni, and Myriam Uribe. 2006. "Access and Risk: Friends or Foes? Lessons from Chile." Policy Research Working Paper 4003. World Bank, Washington, D.C.

Aggarwal, Reena, Leora Klapper, and Peter Wysocki. 2005. "Portfolio Preferences of Foreign Institutional Investors." *Journal of Banking and Finance* 29(12):2919–46.

Allen, Franklin, Jun Qian, and Meijun Qian. 2005. "Law, Finance, and Economic Growth in China." *Journal of Financial Economics* 77(1):57–116.

_____. 2008. "China's Financial System: Past, Present and Future." InL. Brandt, and T. Rawski eds.,. *China's Economic Transition: Origins, Mechanisms, and Consequences.* Cambridge, U.K.: Cambridge University Press.

Armendariz de Aghion, Beatriz, and Jonathan Morduch. 2005. *The Economics of Mirofinance.* Cambridge, Mass.: MIT Press.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006a. "Deposit Collectors." *Advances in Economic Analysis and Policy* 6 (2):1483–83.

_____. 2006b. "Female Empowerment: Further Evidence from a Commitment Savings Product in the Philippines." Working Paper. Yale University, Economics Department, New Haven, Conn.

_____. 2006c. "Tying Odysseus to the Mast: Evidence from a Savings Commitment Product in the Philippines." *Quarterly Journal of Economics* 121 (2):635–72.

Ayyagari, Meghana, Aslı Demirgüç-Kunt, and Vojislav Maksimovic. 2007a. "Firm Innovation in Emerging Markets: Role of Governance and Finance." Policy Research Working Paper 4157. World Bank, Washington, D.C.

_____. 2007b. "Formal versus Informal Finance: Evidence from China." Mimeo. World Bank, Washington, D.C.

Banerjee, Abhijit V., and Esther Duflo. 2004. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program." Discussion Paper 4681.Centre for Economic Policy Research.

Banerjee, Abhijit V., and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101 (2):274–98.

Beck, Thorsten, and Augusto de la Torre. 2007. "The Basic Analytics of Access to Financial Services." *Financial Markets, Institutions and Instruments* 16 (2):79–117.

Beck, Thorsten, Demirgüç-Kunt Aslı, Laeven Luc, and Vojislav Maksimovic. 2006. "The Determinants of Financing Obstacles." *Journal of International Money and Finance* 25 (6): 932–52.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Ross Levine. 2007. "Finance, Inequality, and the Poor." *Journal of Economic Growth* 12 (1):27–49.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Vojislav Maksimovic. 2005. "Financial and Legal Constraints to Firm Growth: Does Firm Size Matter?" *Journal of Finance* 60 (1):137–77.

———. 2006. "The Influence of Financial and Legal Institutions on Firm Size." *Journal of Banking and Finance* 30 (11):2995–3015.

———. 2008. "Financing Patterns Around the World: Are Small Firms Different?" *Journal of Financial Economics* 89 (3):467–87.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Maria Soledad Martinez Peria. 2008. "Banking Services for Everyone? Barriers to Bank Access and Use around the World."*World Bank Economic Review* 22 (3):397–430.

———. 2007. "Reaching Out: Access to and Use of Banking Services across Countries." *Journal of Financial Economics* 85 (1):234–66.

Beck, Thorsten, Ross Levine, and Alex Levkov. 2007. "Big Bad Banks? The Impact of U.S. Branch Deregulation in Income Distribution." Policy Research Working Paper 4330. World Bank, Washington, D.C.

Beegle, Kathleen, Rajeev Dehejia, and Roberta Gatti. 2007. "Child Labor and Agricultural Shocks." *Journal of Development Economics* 81 (1): 80–96.

Berger, Allen N., and Gregory F. Udell. 2006. "A More Complete Conceptual Framework for SME Financing." *Journal of Banking and Finance* 30 (11):2945–66.

Brown, Michael, Tulio Jappelli, and Marco Pagano. Forthcoming. "Information Sharing and Credit: Firm-Level Evidence from Transition Countries." *Journal of Financial Intermediation.*

Claessens, Stijn, Aslı Demirgüç-Kunt, and Harry Huizinga. 2001. "How Does Foreign Entry Affect the Domestic Banking Market?" *Journal of Banking and Finance* 25 (5):891–911.

Claessens, Stijn, and Luc Laeven. 2003. "Financial Development, Property Rights, and Growth." *Journal of Finance* 58 (6):2401–36.

———. 2004. "What Drives Bank Competition? Some International Evidence." *Journal of Money, Credit, and Banking* 36 (3):563–83.

Clarke, George R. G., Robert Cull, and Maria Soledad Martinez Peria. 2006. "Foreign Bank Participation and Access to Credit Across Firms in Developing Countries." *Journal of Comparative Economics* 34 (4):774–95.

Coffee, John C.. 2002. "Racing towards the Top?: The Impact of Cross-Listings and Stock Market Competition on International Corporate Governance." Columbia Law and Economics Working Paper No. 205 (May 30). Available at SSRN: http://ssrn.com/abstract=315840

Cole, Shawn. 2004. "Fixing Market Failures or Fixing Elections? Agricultural Credit in India." Mimeo. Harvard Business School, Cambridge, Mass.

Coleman, B.. 1999. "The Impact of Group Lending in Northeast Thailand." *Journal of Development Economics* 60 (1):105–42.

Cowan, Kevin, Alejandro Drexler, and Alvaro Yañez. 2008. "The Effect of Partial Credit Guarantees on the Credit Market for Small Businesses." Mimeo. Central Bank of Chile.

Cull, Robert, Aslı Demirgüç-Kunt, and Jonathan Morduch. 2007. "Financial Performance and Outreach: A Global Analysis of Leading Microbanks." *Economic Journal* 117 (517):F107–F133.

de Janvry, A., E. Sadoulet, and C. McIntosh. 2006. "From Private to Public Reputation inMicrofinance Lending: An Experiment in Borrower Response." Mimeo. University of California at Berkeley and San Diego.

De la Torre, Augusto, Juan Carlos Gozzi, and Sergio Schmukler. 2006. "Capital Market Development. Whither Latin America?" World Bank, Washington, D.C.

———. 2007. "Innovative Experiences in Access to Finance: Market Friendly Roles for the Visible Hand?" Policy Research Working Paper 4326. World Bank, Washington, D.C.

De la Torre, Augusto, Maria Soledad Martinez Peria, and Sergio Schmukler. 2008. "Bank Involvement with SMEs: Beyond Relationship Lending." Mimeo. World Bank, Washington, D.C.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008a. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *Quarterly Journal of Economics* 123 (4):1329–72.

———. 2008b. "Getting Credit to High Return Microenterprises: The Role of an Information Intervention." Mimeo. World Bank, Washington, D.C.

Demirgüç-Kunt, Aslı, and Ross Levine. 2007. "Finance and Economic Opportunity." Working Paper. World Bank, Development Research Group, Washington, D.C.

Demirgüç-Kunt, Aslı, Inessa Love, and Vojislav Maksimovic. 2006. "Business Environment and the Incorporation Decision." *Journal of Banking and Finance* 30 (11):2967–93.

Dinç, Serdar. 2005. "Politicians and Banks: Political Influences on Government-Owned Banks in Emerging Countries." *Journal of Financial Economics* 77(2): 453–479.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. "Private Credit in 129 Countries." *Journal of Financial Economics* 84 (2): 299–329.

Galor, Oded, and Joseph Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Economic Studies* 60 (1):35–52.

Ghatak, Maithreesh, and Timothy Guinnane. 1999. "The Economics of Lending with Joint Liability: Theory and Practice." *Journal of Development Economics* 60 (1):195–228.

Gine, Xavier, and Dean Karlan. 2006. "Group vs. Individual Liability: A Field Experiment in the Philippines." Policy Research Working Paper 4008. World Bank, Washington, D.C.

Gine, Xavier, and Robert Townsend. 2004. "Evaluation of Financial Liberalization: A General Equilibrium Model with Constrained Occupation Choice." *Journal of Development Economics* 74 (2):269–307.

Gormley, Todd A., Simon H. Johnson, and Changyong Rhee. 2006. "Corporate Bonds: a Spare Tire in Emerging Markets?" http://ssrn.com/abstract=891258

Guarcello, Lorenzo, Fabrizia Mealli, and Furio Rosati. 2003. "Household Vulnerability and Child Labour: The Effects of Shocks, Credit Rationing, and Insurance." Working Paper. Understanding Children's Work Research Project, Washington D.C.

Haselmann, Rainer F. H., Katharina Pistor, and Vikrant Vig. 2006. "How Law Affects Lending." Columbia Law and Economics Working Paper 285. Columbia University, N.Y.

Honohan, Patrick. 2004. "Financial Sector Policy and the Poor." Working Paper 43. World Bank, Washington, D.C.

———. 2008a. "Household Financial Assets in the Process of Development."*Journal of Banking and Finance* 32 (11):2493–500.

———. 2008b. "Partial Credit Guarantees – Principles and Practice." *Journal of Financial Stability*

Honohan, Patrick, and Thorsten Beck. 2007. *Making Finance Work for Africa.* Washington, D.C.: World Bank.

Jacoby, Hanan G., 1994. "Borrowing Constraints and Progress through School: Evidence from Peru." *Review of Economic Studies* 76 (1):151–60.

Jacoby, Hanan G., and Emmanuel Skoufias. 1997. "Risk, Financial Markets, and Human Capital in a Developing Country." *Review of Economic Studies* 64 (3):311–35.

Johnston, Don, and Jonathan Morduch. 2008. "Microcredit vs. Microsaving: Evidence from Indonesia." *World Bank Economic Review* 22 (3):517–37.

Karlan, Dean. 2007. "Social Connections and Group Banking." *Economic Journal* 117 (517): F52–F84.

Karlan, Dean, and Martin Valdivia. 2006. "Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions." Yale University, Department of Economics, New Haven, Conn.

Karlan, Dean, and Jonathan Zinman. Forthcoming. "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment." *Econometrica.*

Klapper, Leora, Luc Laeven, and Raghuram Rajan. 2006. "Entry Regulation as a Barrier to Entrepreneurship." *Journal of Financial Economics* 82 (3):591–629

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

———. 1963. "Quantitative Aspects of the Economic Growth of Nations." *Economic Development and Cultural Change* 11 (2):1–80.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer, 2002. "Government Ownership of Commercial Banks," *The Journal of Finance* 57 (1):265–301.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Guillermo Zamarripa. 2003. "Related Lending." *Quarterly Journal of Economics* 118 (1):231–68.

Lelarge, Claire, David Sraer, and David Thesmar. 2008. "Entrepreneurship and Credit Constraints. Evidence from a French Loan Guarantee Program." Mimeo, University of California Berkeley.

Levine, Ross. 2005. "Finance and Growth: Theory and Evidence." InPhilippe Aghion, and Steven Durlauf eds., *Handbook of Economic Growth.* Amsterdam: North-Holland Elsevier Publishers.

Levine, Ross E., and Sergio L. Schmukler. 2007. "Internationalization and Stock Market Liquidity." *Review of Finance* 10 (1):153–87.

Levy-Yeyati, Eduardo, and Alejandro Micco. 2007. "Concentration and Foreign Penetration in Latin American Banking Sectors: Impact on Competition and Risk." *Journal of Banking and Finance* 31 (6):1633–47.

McKenzie, David J., and Christopher Woodruff. 2008. "Experimental Evidence on Returns to Capital and Access to Finance in Mexico." *World Bank Economic Review* 22 (3):457–82.

Mian, Atif. 2006. "Distance Constraints: The Limits of Foreign Lending in Poor Economies." *Journal of Finance* 61 (3):1465–1505.

Morduch, Jonathan. 1998. "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh." Princeton University, Department of Economics, Princeton, N.J.

Perotti, Enrico C., and Paolo F. Volpin. 2004. "Lobbying on Entry." Centre for Economic Policy Research 4519. World Bank, Washington, D.C.

Pitt, M. M., and S. R. 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.

Porteous, David. 2006. "The Enabling Environment for Mobile Banking in Africa." Bankable Frontier Associates. http://www.bankablefrontier.com/assets/ee.mobil.banking.report.v3.1.pdf.

Rajan, Raghuram. 2006. "Separate and Unequal." *Finance and Development* 43 (1):56–57.

Rajan, Raghuram, and Luigi Zingales. 2003. *Saving Capitalism from the Capitalists.* New York: Crown Business.

Sachs, Jeffrey. 2005. *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin Press.

Visaria, Sujata. Forthcoming. "Legal Reform and Loan Repayment: Microeconomic Impact of Debt Recovery Tribunals in India." *American Economic Journal: Applied Economics.*

World Bank. 2007. *Finance for All? Policies and Pitfalls in Expanding Access.* Washington, D.C.

THE WORLD BANK

1818 H Street NW
Washington, DC 20433, USA
World Wide Web: http://www.worldbank.org/
E-mail: researchobserver@worldbank.org

ISBN-13: 978-0-19-957449-0
9 780199 574490

THE WORLD BANK

Research Observer

The World Bank Research Observer is intended for anyone who has a professional interest in development. *Observer* articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the *Observer* considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the Editor. Please direct all editorial correspondence to the Editor, *The World Bank Research Observer*, 1818 H Street, NW, Washington, DC 20433, USA.

For more information, please visit the Web sites of the *Research Observer* at www.wbro.oxfordjournals.org, the World Bank at www.worldbank.org, and Oxford University Press at www.oxfordjournals.org.

THE WORLD BANK

Research Observer

Volume 24 • Number 1 • February 2009

Rural Poverty: Old Challenges in New Contexts
Stefan Dercon 1

Symposium on Evaluation

Evaluation in the Practice of Development
Martin Ravallion 29

Timing and Duration of Exposure in Evaluations of Social Programs
Elizabeth M. King and Jere R. Behrman 55

Symposium on Financial Sector

Competition in the Financial Sector: Overview of Competition Policies
Stijn Claessens 83

Access to Financial Services: Measurement, Impact, and Policies
Thorsten Beck, Asli Demirgüç-Kunt, and Patrick Honohan 119