INSIDE

Jacob Yaron,
McDonald Benjamin, and
Stephanie Charitonenko
on Promoting Efficient Rural
Financial Intermediation


Gerard Caprio, Jr., and
Ash Demirgüç-Kunt
on the Role of
Long-Term Finance


John Baffes and
Jacob Meeraman
on Agricultural Subsidization
without Protection?


Harold Alderman and
Kathy Lindert
on the Potential and
Limitations of Self-Targeted
Food Subsidies


Dominique van de Walle
on Targeting Revisited


Bernard Hoekman
on Using International
Institutions to Improve
Public Procurement


Estelle James
on New Models for
Old-Age Security

ISSN 0257-1032
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, at the address in the copyright notice below.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of the data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

The World Bank Research Observer is published twice a year in February and August. Single copies may be purchased for US$12.95. Subscription rates are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Individuals</th>
<th>Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-year subscription</td>
<td>US$25</td>
<td>US$40</td>
</tr>
<tr>
<td>2-year subscription</td>
<td>US$45</td>
<td>US$75</td>
</tr>
<tr>
<td>3-year subscription</td>
<td>US$60</td>
<td>US$100</td>
</tr>
</tbody>
</table>

Orders should be sent to: World Bank Publications, Box 7247-7956, Philadelphia, Pennsylvania 19170-7956, U.S.A. Readers with mailing addresses in developing countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request. Written confirmation is required every two years to renew such subscriptions.

© 1998 The International Bank for Reconstruction and Development / THE WORLD BANK
1818 H Street, N.W., Washington, D.C. 20433, U.S.A.
All rights reserved
Manufactured in the United States of America
ISBN 0-8213-4114-6; ISSN 0257-3032

Material in this journal is copyrighted. Requests for permission to reproduce portions of it should be sent to the Office of the Publisher at the address in the copyright notice above. The World Bank encourages dissemination of its work and will normally give permission promptly and, when the reproduction is for noncommercial purposes, without asking a fee. Permission to copy portions for classroom use is granted through the Copyright Clearance Center, Suite 910, 222 Rosewood Drive, Danvers, Massachusetts 01923, U.S.A.

The World Bank Research Observer is indexed by the Journal of Economic Literature, the Standard Periodical Directory, and other services. It is available in microform through University Microfilms, Inc., 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.
Promoting Efficient Rural Financial Intermediation  
*Jacob Yaron, McDonald Benjamin, and Stephanie Charitonenko*  
147

The Role of Long-Term Finance: Theory and Evidence  
*Gerard Caprio, Jr., and Asli Demirgüç-Kunt*  
171

From Prices to Incomes: Agricultural Subsidization without Protection?  
*John Baffes and Jacob Meerman*  
191

The Potential and Limitations of Self-Targeted Food Subsidies  
*Harold Alderman and Kathy Lindert*  
213

Targeting Revisited  
*Dominique van de Walle*  
231

Using International Institutions to Improve Public Procurement  
*Bernard Hoekman*  
249

New Models for Old-Age Security: Experiments, Evidence, and Unanswered Questions  
*Estelle James*  
271
Promoting Efficient Rural Financial Intermediation

Jacob Yaron • McDonald Benjamin • Stephanie Charitonenko

Although governments have traditionally used subsidized credit programs to promote agricultural growth, this approach has generally failed to improve incomes and alleviate poverty in rural areas. It has also led to the mistaken belief that rural credit programs cannot be profitable. A new approach seeks to raise standards of living in rural areas by casting the government in a very different role—one of setting a favorable legal and policy environment for rural financial markets and addressing specific market failures cost effectively through well-designed and self-sustaining interventions. There is evidence that this approach can be highly successful. The Village Bank system of Bank Rakyat Indonesia has shown that financial services can be extended to millions of low-income rural clients without relying on subsidies. Indeed, the program has generated enormous profits for the bank by using simple, innovative, and largely replicable techniques.

Providing affordable financial services to the rural population has been an important component of development strategy for the last several decades. Direct interventions in rural financial markets to stimulate growth and reduce poverty—through a blend of targeted credit programs, interest subsidies, and other government policies—became widespread in the 1950s, when Keynesian economics inspired many governments to design fiscal interventions at the macroeconomic level. But these direct interventions have generally been disappointing and have tended to retard, rather than promote, the development of financial services in rural areas. One explanation is that these policies were based on serious misconceptions about the real challenges facing rural communities and were directed more toward the symptoms rather than the causes of inadequate rural financial intermediation.

More recent developments in the provision of rural financial services (both savings and credit) have demonstrated that proper institutional design and adherence to appropriate policies pay off handsomely and have the potential to generate substantial achievements in terms of both sustainability and greater institutional outreach.
Although the new approach focuses on the same objectives, that is, income expansion and poverty reduction, the perceived challenges and ways of addressing them are strikingly different. The principal change involves promoting deep and efficient rural financial markets by creating a favorable policy environment, improving the legal and regulatory framework that supports rural financial markets, and addressing specific market failures in cost-effective ways through well-designed, self-sustaining interventions.

The Traditional Approach to Rural Finance

Throughout the world, governments have intervened extensively in financial markets in general and rural financial markets in particular. The extent of intervention has varied from indirect measures aimed at improving the policy environment (for example, by addressing incentive problems and regulating financial intermediaries), to direct steps to increase or supplant credit provided by private lenders. Many countries eager to channel funds to farmers directed private banks to make concessional loans to agriculture, or they established and supported state-owned agricultural credit institutions.

Traditionally, the case for subsidized agricultural credit programs has been based on the following arguments: governments should focus on agriculture to promote rural development; agriculture is undercapitalized; farmers need cheap credit to encourage them to adopt modern technology and to compensate them for policies that are biased in favor of urban dwellers; farmers are too poor to save; and private banks provide little or no credit, forcing small borrowers to use moneylenders who charge usurious interest rates. Donors provided considerable support for subsidized credit; the World Bank, for example, lent $16.5 billion in agricultural credit under largely traditional programs prior to 1992 (World Bank 1993).

These programs have generally had a limited outreach and resulted in huge costs, with little identifiable impact at the farm level. In an extreme example, during the 1980s, one Latin American rural financial institution with more than 500 branches and 27,000 employees received $10.3 billion in fiscal and quasi-fiscal transfers (that is, capital injections and interest subsidies), while recovering only 10–15 percent on its portfolio and serving only 2 percent of the rural population. Elsewhere, government-sponsored rural credit programs and institutions from Peru to Malawi to Indonesia have collapsed under the weight of losses generated by traditional directed credit strategies.

These failures are largely explained by the pursuit of short-term objectives framed in terms of agricultural production gains rather than long-term objectives aimed at the sustained expansion of rural incomes. The excessive focus on disbursing cheap
agricultural credit has typically resulted in programs with a poor credit culture, manifested by a dependency on subsidies, low recovery rates, inadequately diversified portfolios, mistargeting of credit (Khan 1977), and rent-seeking by credit officials and influential farmers (Ladman and Tinnermeier 1984). The tremendous potential for rural savings has also been neglected, and private for-profit financial institutions have been crowded out by state-owned rural financial institutions dependent on government subsidies.

Although directed credit has been heavily criticized (Von Pischke, Adams, and Donald 1983), and more market-friendly approaches have been proposed for some time, many countries have resisted changing the rules under which state-owned financial institutions operate. Nevertheless, major reforms of rural credit systems have been launched in several countries, including India and Mexico, to ensure that public resources are used more effectively, to support the expansion of rural incomes, and to reduce poverty.

The New Approach to Rural Finance

The new approach continues to focus on income expansion and poverty reduction but makes the case for cost-effective alternatives, such as increased investment in rural infrastructure or in human development, to reach these goals. (For the literature on changes in rural finance, see Adams, Graham, and Von Pischke 1984; David and Meyer 1984; González-Vega 1984; and Vogel 1984). Advocates of this approach propose that governments concentrate on establishing a favorable policy environment that facilitates the smooth functioning of rural financial markets while playing a more limited and efficient role in the direct provision of rural financial services. The factors that prevent rural financial markets from operating efficiently are recognized to be broader and include macroeconomic policies, weakly regulated financial sectors, institutional features (legal and regulatory), and specific constraints related to intermediation in rural areas. This approach sees the government’s main task as creating a conducive environment for private intermediaries in rural financial markets (figures 1 and 2). The case for direct interventions depends on whether the objective is general rural income expansion or targeted poverty reduction.

Creating a Conducive Policy Environment

The starting point for formulating policies aimed at increasing rural incomes and reducing poverty is an assessment of the efficiency of markets, particularly rural financial markets, and of the causes of market inefficiencies. Typically, there are weak-
Goal 1: Rural Income Expansion

Enhance the efficiency and completeness of markets

Evaluate policy options

Promote deep and efficient rural financial markets

Promote nonfinancial markets (beyond scope of this paper)

Is the market efficient?

Yes

Maintain existing set of policies

No

Identify probable causes

Is there a poor policy environment?

Yes

Create a favorable policy environment

- Ensure macroeconomic stability
- Remove urban-biased policies
- Promote broad financial sector reforms

No

Is there a weak legal and regulatory framework?

Yes

Improve the legal and regulatory framework

- Improve land titling and registration
- Reform the law of secured transactions
- Deregulate lending by non-deposit-taking institutions

No

Is there an identifiable market failure?

Yes

Can the market failure be removed cost-effectively with a public intervention?

Yes

Based on cost-benefit analysis, select direct interventions to remove market failure

- Support for social intermediation
- Incentives for innovation
- Piloting and dissemination of information

No

No

No

No

Maintain existing set of policies

Figure 2. Decision Tree for Poverty Reduction

Goal II: Rural Poverty Reduction

1. Conduct poverty assessment to characterize the rural poor and the binding constraints they face.

2. Evaluate policy options:
   - Are market-oriented measures sufficient to reduce poverty?
     - Yes
       - Enhance the efficiency and completeness of markets (see figure 1)
     - No
       - Based on cost-effectiveness analysis, select program of direct interventions to reduce poverty of target group

- Pursue comprehensive strategy for developing markets
  - Promote deep and efficient rural financial markets
  - Promote nonfinancial markets (beyond scope of this paper)

- Rural finance interventions
  - Matching grants for community-based funds
  - Support for social intermediation
  - Start-up subsidies for innovative savings and credit programs
  - Lines of credit at market rates

- Alternative interventions
  - Targeted food support
  - Employment-generation via public works programs
  - Investment in rural infrastructure
  - Human resource development in rural areas

nesses in the policy environment that hamper the development of rural markets, including financial markets. For example:

- Unsound macroeconomic policies result in volatility and high real interest rates that can adversely affect all financial intermediaries, while misaligned exchange rates distort price signals and lead financial markets to channel excessive resources to inefficient sectors.
- Development policies biased toward urban areas reduce the profitability of agriculture and nonfarm rural enterprises and devastate rural financial markets. Countries with the highest degree of discrimination against agriculture have had the lowest rates of economic growth (Schiff and Valdés 1992).
- Inadequate regulatory oversight, inappropriate interventions in financial markets, and financial repression increase the risks and constrain the development of financial markets.

Governments can promote financial markets in general by strengthening the supervision and prudential regulation of financial institutions, deregulating interest rates, reducing excessively high reserve requirements, and relaxing credit controls. Governments can also adjust the regulatory framework to facilitate operations in rural areas by community-based, deposit-taking intermediaries. Such policies would combine lower capital requirements with higher capital-asset ratios and more circumscribed permitted activities to minimize regulatory arbitrage (see Berenbach and Churchill 1997 for similar regulatory issues in microfinance).

The Legal and Regulatory Framework

Too often the institutional foundations for financial markets in rural areas are absent. Lenders need a system that provides formal procedures for claims against property and enforcement of financial contracts. The more uncertain and expensive this process, the less willing are lenders to lend (Fleisig and de la Peña 1996). In many countries deficiencies in laws, regulations, and institutions prevent the formal sector from delivering credit to farmers, rural businesses or even nonbank creditors (typically traders), who have many advantages in efficiently reaching poor rural borrowers.

The required changes needed to expand access to credit in rural areas include titling and registering land; reforming the law of secured transactions, such as legally acceptable forms of collateral; establishing legal registries and expanding the scope for private operation; lowering the costs of registration and foreclosure; drafting specific, clear, and limited homestead provisions; and removing interest rate ceilings. Well-designed programs to reform the laws of secured transactions have increased
the supply of credit and lowered interest rates, producing gains over time that have been estimated at several percentage points of gross domestic product GDP (Fleisig and de la Peña 1996). The costs of implementing such legal reform programs are usually remarkably low.

Designing and Justifying Direct Interventions in Rural Financial Markets

What role should government play in rural finance and development? There is growing recognition that governments should first and foremost facilitate the workings of the market so that private participants can allocate resources efficiently in response to price and profit signals.

Risks of Market Failure

At the same time, markets may fail for several reasons, because the assumptions that are required in theory for efficient market-based resource allocation may not hold in practice. For example, individuals may not bear the full benefits or costs of their actions. Or externalities may arise because investors cannot capture the full benefits of their investments if they cannot exclude others from free-riding. Alternatively, individuals may fail to take into account the costs they impose on others when undertaking a given activity. A market may not have sufficient buyers and sellers or permit sufficient ease of entry and exit to ensure an efficient allocation of resources. In many countries market participants may not be able to enter easily into enforceable contracts. Finally, market participants may not be able to ensure against certain contingencies, although efficient markets for pricing and exchanging risks are required for optimal resource allocations when there is uncertainty. Rural financial markets, particularly in developing economies, generally have these shortcomings.

The critical factor that explains the externalities, missing markets, and local (competitive) monopolies in rural financial markets is imperfect information (Stiglitz 1996; Virmani 1982). Financial transactions in a given currency by their very nature involve a contractual exchange of cash for a promise of a future stream of payments, rather than a simultaneous exchange of cash or goods—or both—for goods. The promissory feature of financial transactions makes it essential for participants to be well informed about their counterparts’ ability and willingness to honor contractual obligations. The absence of such information will constrain a lender’s (or depositor’s) ability not only to discern the creditworthiness of potential borrowers (or banks), but also to enforce contracts. These constraints point to an important role for government in regulating financial intermediaries, for in-
stance to limit excessive risk-taking by banks using other people's money, and in providing a sound legal and regulatory framework for enforcing contracts.

**Rethinking Direct Interventions**

Information constraints do not immediately justify direct government interventions in the market, because markets may be constrained-efficient, that is, they maximize incomes subject to the information and other barriers that participants face. For example, in rural areas, poverty, low population density, isolated markets, seasonality, and highly covariant risk such as widespread crop failures in a given region often result in high transaction costs, a lack of traditional collateral, variable incomes, and limited opportunities for diversifying risk. These features differentiate rural financial markets from urban ones and often scare off traditional for-profit financial intermediaries. They do not, however, entail market failures, because these features result in high real costs to society that government interventions would also face. At the same time, there may be clear economic gains that the market has not yet obtained, for example, transactions in which the first entrants bear the full cost of losses but are unable to capture the full benefits of success (box 1).

Careful analyses that identify market failures and specify their causes should precede appropriate interventions to expand rural incomes. Even if a market failure is identified, direct interventions (through subsidies, credit programs, or institutions) are warranted only if the market failure can be addressed cost effectively; thus, the benefits must exceed the costs. A government failure is not a solution to a market failure.

Although policy and regulatory reforms that promote growth are often the most promising way to reduce rural poverty, special interventions may be required if economic growth is not appropriately shared. These interventions are justified based on social norms rather than on market failures (see figure 2). Interventions in rural financial markets, however, are still warranted only if they are the most cost-effective means of reducing poverty.

---

**Box 1. How Imperfect Information Can Generate Market Failures**

It is widely—incorrectly—assumed that if a given activity were profitable, someone in the private sector would have done it already. As Besley (1994) notes, "An inefficiency might develop if individuals hang back waiting for others to try things out. The slow diffusion of certain agricultural technologies has often been attributed to a reluctance to be the first user. An obvious role for government intervention is to subsidize early innovators. Thus experiments in institutional design, such as the Grameen Bank in Bangladesh, might serve as prime candidates for subsidization. Such arguments appear only to justify subsidizing new ventures, however, and subsidies should be phased-out along the way."

---

Two Perspectives on Market Failure and the Argument for Intervention

Stiglitz (1993): "There is a role for the state in financial markets; it is a role motivated by pervasive market failures. In developing countries, market failures are almost undoubtedly greater than in the more developed countries. . . . While limitations on markets are greater in less developed countries than in developed countries, so too, many would argue, are limitations on government. We have argued that government policies can be designed which are attentive to those limitations. . . . What is clear is that a simple ideological commitment to financial market liberalization cannot be derived either from economic theory or be justified by an examination of a broad base of experience. . . ."

Besley (1994): "In summary, there may be good arguments for intervention, and some may be based on market failure. But as one unpacks each argument, the realization grows that, given the current status of empirical evidence on many relevant questions, it is impossible to be categorical that an intervention in the credit markets is justified. Empirical work that can speak to these issues is the next challenge if the theoretical progress on the operation of rural credit markets is to be matched by progress in the policy sphere."

Government interventions in rural financial markets should aim to remove the causes of market failure or poverty, using the most appropriate mix of instruments, such as funding for pilot programs; institutions, such as private financial intermediaries, nongovernmental organizations, or state-owned rural financial institutions; and products, such as credit, savings, guarantees, and insurance. Where appropriate, market failure caused by imperfect information can be addressed by providing seed capital to establish rural financial institutions in remote areas (box 1).

Interventions should always be designed to complement, facilitate, or improve rural financial markets over the long term (box 2). For example, if an initial cost-benefit analysis suggests that a state-owned rural financial institution is a more cost-effective vehicle for promoting rural financial markets than working through private banks or nongovernmental organizations, the government should not later prop up its rural financial institutions with more favorable access to subsidies or concessional funds than are available to other entrepreneurs. On the contrary, competition should be encouraged. Subsidies or grants should generally be restricted to seed capital or be limited by a sunset clause. Finally, the cost of programs to develop rural financial markets should be monitored to assess the cost-effectiveness of the interventions.

There have been widespread failures among state-owned specialized agricultural credit institutions around the world (Adams, Graham, and Von Pischke 1984). These agencies have lacked appropriate governance, capable management, political autonomy, and innovative, efficient operating procedures. They have not addressed information constraints and have been plagued by incentive problems. Taking into account foregone opportunities by private rural financial institutions, the inefficient allocation of resources, and the fiscal costs of propping up loss-making and often
Box 3. Two Good Reasons for Market Interest Rates: Equity and Efficiency

**Equity:** Directed credit programs invariably face the following dilemma: whether to lend to more clients with no subsidy, or to fewer people with a high subsidy per dollar lent. If the issue is perceived as resolving the inadequate access to formal credit of the rural masses, then (on equity grounds), the policy should pursue *increased outreach*—a choice that requires eliminating or minimizing the subsidy per dollar lent.

**Efficiency:** Several studies show that liberalized financial markets generate a more efficient allocation of resources and higher rates of economic growth (King and Levine 1993; Jaramillo Schiantarelli, and Weiss 1993; McKinnon and Shaw 1976). Other studies point to a positive relationship between savings and real interest rates in developing countries (Fry 1988). The importance of financial institutions in offering and charging positive real interest rates is clearly shown in King and Levine (1993), who find that real growth in gross domestic product during 1974–89 for a sample of 76 countries was more than 2 percent higher for those offering the highest deposit interest rates than for those offering the lowest deposit rates. Indeed, growth was negative for the latter group of countries.

corrupt institutions, the economic costs are likely to have far outweighed the benefits of these public-sector institutions. Even worse, the rich have frequently captured the subsidies, compounding problems of poverty and inequality.

State provision of financial services in rural areas is not recommended unless it meets the following strict governance criteria:

- Fully autonomous management that is held accountable for the bank’s financial performance
- Exemption from civil service pay scales to attract and reward quality staff on the basis of the institution’s financial performance
- Insulation from staffing pressures by local authorities, for example through autonomous organizational charts with professional qualifications criteria
- The same freedom to set borrowing and lending rates that apply to commercial banks, so that both deposit and lending rates are at market rates, are usually positive in real terms, and provide an adequate spread to cover costs (box 3)
- Application of international best practice prudential regulatory, accounting, and disclosure practices, and therefore the development of a strong management information system by the rural financial institutions, and both off-site supervision and on-site examinations by the same agency that supervises private banks
- A hard budget constraint
- A clear strategy to develop rural financial markets that are supported only with initial, nonexclusive, time-bound, transparently budgeted subsidies.
Measuring the Performance of Rural Financial Intermediaries

Evaluating the effect of a rural credit program on incomes and poverty is very difficult because it is rarely clear what borrowers would have done in the absence of the program. Therefore, practitioners and academics have developed a new framework for assessing the performance of credit programs. This framework rests on outreach and self-sustainability (Yaron 1992a). It argues that rural financial institutions that provide a broad range of services to the targeted clientele in an efficient manner are likely to have the desired impact of expanding incomes and reducing poverty. Therefore, evaluating their performance based on these criteria provides an easily quantifiable proxy of the impact of rural financial intermediation in lieu of a full benefit-cost analysis (figure 3).

**Figure 3. Assessing the Performance of Rural Financial Institutions**

<table>
<thead>
<tr>
<th>Assessment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-sustainability</td>
</tr>
<tr>
<td>Outreach to target clients</td>
</tr>
</tbody>
</table>

- **Subsidy dependence index**
  - Measures subsidies received against interest earned

**Examples of subsidies:**
- Interest rate subsidy on concessional loans
- Opportunity cost of equity
- Reserve requirement exemptions
- Free equipment provided by government/donors
- Government's assumption of loan losses
- Free training for staff provided by government/donor
- Government assumption of foreign-exchange loans

**Examples of indicators:**
- Market penetration
  - Number and annual growth rate of saving and loan accounts
  - Value and annual growth rate of the loan portfolio and deposits
  - Number of branches and staff
  - Relative income level
  - Value of average loan and range of loan amounts
  - Percentage of rural clients
  - Percentage of female clients
  - Quality of services
  - Transaction costs to clients
  - Flexibility and suitability of services
  - Distribution network

*Source: Yaron, Benjamin, and Pipeck (1997).*
Outreach is measured by a hybrid index comprising several indicators, such as the number of clients, the value of the loan portfolio and its annual growth, the percentage of female clients (where social norms discriminate against women), the average loan size (as a proxy for income level of the clientele), and so on.

Self-sustainability is assessed by calculating the subsidy dependence index, that is, the percentage by which the agency’s average on-lending interest rate would have to increase to make it self-sustainable (Yaron 1992b; Benjamin 1994). Conventional accounting practices fail to reflect most subsidies received by state-owned rural financial institutions or by nongovernmental organizations and therefore do not show the true social costs of maintaining these intermediaries. Yet without this measure it is impossible to determine whether continuing support for those institutions is warranted. Given the prevalence and extent of subsidies, recognizing the subsidy dependence is essential to evaluating the performance of state-owned institutions and nongovernmental organizations. The subsidy-dependence index is instrumental in:

- Relating the total amount of subsidies received by a rural financial institution to its level of activity, represented by the interest earned on its loan portfolio. This exercise is similar to calculations of effective protection, domestic resource cost or job creation cost. It also captures the notion of matching grants by comparing the value of subsidies received against the income earned from clients in the market place.
- Tracking a rural financial institution’s performance in terms of subsidy dependence over time and relative to that of other institutions that provide similar services to a similar clientele.

**Successful Rural Financial Institutions**

Three Asian rural financial institutions are widely considered successful based on the two primary criteria of outreach and self-sustainability: the Bank for Agriculture and Agricultural Cooperatives in Thailand; the Village Banks, or Unit Desas, of Bank Rakyat Indonesia (BRI-UD); and the Grameen Bank in Bangladesh. All have succeeded in providing financial services at unprecedented levels to millions of rural people. The Grameen Bank has reduced its dependence on subsidies, the Bank for Agriculture and Agricultural Cooperatives has benefited from low to modest subsidies, and the Village Bank program of the Bank Rakyat Indonesia has completely eliminated its dependence on subsidies. These successes contrast sharply with the traditional view that heavy subsidies are inevitable in the provision of financial services to rural entrepreneurs. A variety of mechanisms have enhanced the efficiency of these institutions, including the following:

- A high degree of management autonomy in formulating operational policies.
• Policies that provide for staff accountability, investment in human capital, and rewards (monetary incentives and promotions) that are related to sound financial performance and sustainability.
• Innovative, low-cost delivery systems and mobile banking services.
• Innovative and flexible loan terms and conditions adapted to social, economic, and cultural circumstances. For example, all offer weekly or monthly repayment schedules tailored to the clients’ cash flow.
• Close monitoring of loan performance; high, on-time collection rates and low loan losses.
• Development of domestic savings accounts to reduce or eliminate the need for donor funds.
• Positive and often relatively high on-lending rates that ensure an adequate spread.
• Control over administrative expenses and effective use of economies of scale.
• Advanced management information systems that facilitate effective planning, control, and timely monitoring of loan repayments.
• Concentration on rural markets that have relatively high population densities.

A close look at the operations of the BRI-UD shows how it has achieved unparalleled success in rural financial intermediation.

Explaining the Success of the Village Banks Program

This successful public entity has succeeded in reaching financial self-sustainability while providing credit and saving services to rural low-income families that had not previously had access to formal financial services. Moreover, it has achieved an unprecedented level of profitability, earning $177 million in profits during 1996, and a return on average assets of more than 5 percent a year between 1994 and 1996. This level of profitability is rarely found even among financial intermediaries that serve clients who borrow and save vastly larger amounts, and who have much more substantial enforceable collateral.

Background

By the mid-1970s Indonesia’s directed-credit program aimed at channeling funds to rice farmers (under the BIMAS, or Mass Guidance program), had contributed to making the country self-sufficient in rice production. By the early 1980s, however, the program had become increasingly unsustainable as a result of subsidized interest rates, poor loan repayments, and employee incentives directed toward disbursing credit rather than generating profits. In 1984 the government transformed the op-
eration into the BRI-UD system and ordered it either to devise a program to provide rural financial services on a self-sustaining basis or to face closure. With a relatively small initial subsidy in 1984, the new BRI-UD became profitable within eighteen months. By 1996 the bank was a global leader in rural financial intermediation, with 2.5 million loan accounts and 16.2 million deposit accounts, combining substantial market penetration among low-income rural clients with sustained profitability (table 1 and figure 4).

The most fundamental policy change was a shift from disbursing credit to motivating loan recovery and mobilizing savings; that is, to genuine rural financial inter-

Table 1. BRI-UD's Outreach and Financial Self-Sustainability (US$)

<table>
<thead>
<tr>
<th>Outreach</th>
<th>1985</th>
<th>1990</th>
<th>1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average annual loan volume (millions)</td>
<td>162</td>
<td>562</td>
<td>1,178</td>
</tr>
<tr>
<td>Number of outstanding loans (millions)</td>
<td>1.0</td>
<td>1.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Average outstanding loan amount / borrower</td>
<td>162</td>
<td>296</td>
<td>512</td>
</tr>
<tr>
<td>Average annual deposit volume (millions)</td>
<td>49</td>
<td>685</td>
<td>2,382</td>
</tr>
<tr>
<td>Number of deposit accounts (millions)</td>
<td>—</td>
<td>7.3</td>
<td>14.5</td>
</tr>
<tr>
<td>Average deposit amount per saver</td>
<td>—</td>
<td>94</td>
<td>164</td>
</tr>
</tbody>
</table>

Financial self-sustainability (in percentages unless noted)

| Nominal average yield earned on loan portfolio | 27.4 | 31.5 | 31.6 |
| Nominal average interest rate paid on deposits | 10.5 | 11.3 | 9.7  |
| Nominal interest rate spread                  | 16.8 | 20.2 | 21.5 |
| Inflation                                     | 4.7  | 7.4  | 9.4  |
| Real average yield earned on loan portfolio   | 21.7 | 22.4 | 20.2 |
| Real average interest rate paid on deposits   | 5.6  | 3.6  | 0.9  |
| Lowest lending rate needed for financial self-sustainability |
| Nominal                                      | 36.2 | 27.2 | 17.5 |
| Real                                         | 30.1 | 18.4 | 17.2 |

Operating costs as a percentage of:

| Average annual net loan portfolio (LP)         | 20.5 | 12.9 | 12.6 |
| Half of the average annual net LP and deposits | 31.5 | 11.6 | 8.5  |
| Average annual total assets                    | 15.1 | 8.0  | 5.7  |
| Profits ($ millions)                           | -0.8 | 34.3 | 170.2|
| Percentage of profitable units                 | 48.3 | 89.1 | 95.7 |
| Average ann. deposit volume / average ann. LP volume | 0.31 | 1.22 | 2.02 |
| Subsidy dependence index                       | 32.2 | -13.7| -44.5|

--- Not available.

mediation. To broaden and diversify its clientele, the bank targeted the low-income rural population and offered loans for all income-generating activities—a sharp departure from the traditional pattern of lending solely to the agricultural sector, which has accounted for a shrinking share of gross domestic product. BRI-UD underpinned these policy shifts with substantially higher loan and deposit interest rates, while maintaining a sufficient interest-rate spread to cover the high costs of servicing small loans and deposits. The average annual yield on loans has been about 32 percent in recent years; average annual financial costs have been about 10 percent. Innovative incentive systems were introduced for both clients and employees to encourage timely and complete loan repayment (Chaves and González-Vega 1996).

An Autonomous Organizational Structure

The Village Bank system functions as an independent profit center within the Bank Rakyat Indonesia. The Village Banks are free to set their own loan terms, although transfer prices (discussed below) are negotiated with the bank. BRI-UD has developed its own management tools, including an efficient management information system.
to assess performance and a sophisticated employee incentive system to encourage profitability, loan recovery, and savings mobilization.

Since 1984, 10 percent of each unit's annual profit has been distributed to employees as a reward for achieving good collection rates. Bonuses are paid early in the year and are capped at 1.5 month's salary per employee. Because about 96 percent of the units were profitable in 1996, a similar percentage of more than 21,000 employees and trainees are benefiting from this program as well as from additional bonuses for achieving goals that are earned in routine competitions between units. The importance of these incentive schemes cannot be overstated; they clearly set BRI-UD apart from government development banks elsewhere that remunerate their employees on the basis of inadequate civil service pay scales.

Innovations for Rural Customers

The loan application process takes about a week for a new borrower and less time for a repeat customer. Loans are extended on an individual basis, and generally have a maturity of 18 months, with monthly repayments. Collateral is desirable but not mandatory. Loan delivery systems incorporate cost-minimizing features. For example, paperwork is kept to a minimum, and where the volume of business is relatively small, mobile offices provide limited services to clients in outlying areas once or twice a week.

The small average size of loans and relatively high cost of legal procedures make foreclosure prohibitively expensive (although warranted in certain cases to achieve a demonstration effect). Thus the focus is on quality at entry and appropriate incentives for repayment:

- Applicants are prescreened based on available information (gathered from peers and from village leaders), and proposed investments are evaluated.
- Clients are given a substantial incentive to repay through both interest rebates (of about 12 percent a year) and access to additional larger loans contingent on timely repayment.
- Staff incentives are linked to the performance of the loan portfolio, so clients are monitored more closely than is usual.
- Loans are priced to encourage more selective choices of investments and to promote credibility in the institution.

Because BRI-UD is not viewed as just another transient government program, borrowers have a greater incentive to repay their loans and depositors place greater trust and confidence in the institution. Fieldwork throughout the country since 1982 has pointed to extensive demand in rural areas for reliable financial savings facilities, especially for liquid savings accounts (Robinson 1994). Four savings instruments with interest rates that varied substantially with account size and liquidity were of-
ferred in 1986 as part of the new rural savings program. This program was the flagship of the bank’s revamped effort to provide services to rural clients.

Managing Banking Operations

The BRI-UD’s loan-loss treatment is very conservative compared with most state-owned rural financial institutions in other countries. In addition to general loan loss reserves of 3 percent against all outstanding loans that are not yet due and payable, there are reserves of 50 percent against loans less than three months overdue, and reserves of 100 percent against loans that are three months to a year overdue. Loans that are more than one year overdue are fully written off. Asset classification is also conducted conservatively with a view to avoiding hidden rescheduling, or “evergreening,” of the portfolio.

One of the principal advantages of belonging to a nationwide branch network is that the Bank Rakyat Indonesia system serves as a clearing house between cash-surplus and cash-deficit units. Fund transfers carry an interest rate—the transfer price—which is adjusted periodically according to the bank’s overall liquidity position. The transfer price is usually set slightly higher than the top savings rate offered at the units so that those with a surplus of funds can at least cover their interest costs and are not discouraged from mobilizing savings (Charitonenko, Patten, and Yaron 1998).

To improve the quality of bookkeeping and customer service, Bank Rakyat Indonesia has developed a stand-alone, personal computer-based system. By the end of 1995 computerization had been introduced in about 89 percent of the 3,135 Village Banks, facilitating the units’ bookkeeping and management practices.

Measuring Financial Performance

The two most widely used financial ratios for measuring a financial institution’s performance are its return on equity and return on assets. In addition, the subsidy dependence index is increasingly being used to evaluate state-owned financial institutions and nongovernmental organizations that provide financial intermediation services because the profitability of these intermediaries often depends on their access to subsidies.

In the case of BRI-UD, a further refinement is required. Typically, the return on equity compares an institution’s net income in a given fiscal year with its average equity during that fiscal year. In the absence of more detailed data (on the timing and amount of capital injections and dividend payments), average annual equity is generally calculated simply as the sum of start-of-year equity plus year-end equity, divided by two. Complete information is available for the BRI-UD, however: there have been no capital injections or grants; and every year on January 1, all profits earned during the preceding year are transferred to Bank Rakyat Indonesia’s general
account. Because the units do not retain their earnings, it is more meaningful to calculate the return based on start-of-year equity. This refinement has major implications for measuring profitability, as shown in figure 5. The system has been exceptionally profitable by any banking standards. Whereas banks in low-inflation countries might earn 15 to 20 percent (after tax) on their average annual equity, BRI-UD earned more than 60 percent on its average equity in 1990 and 1991. By 1995 this figure had more than doubled to 136 percent. When net income is measured against start-of-year equity, the return on equity increases to about 90 percent in 1990 and 1991, rising to an astounding 407 percent by 1995 (a net income of Rp. 403 billion on a start-of-year capital of Rp. 99 billion).

The relatively low ratio of equity to assets is an important factor in explaining these high returns on equity. If BRI-UD were to become an independent microfinancial intermediary, rather than a profit center within a larger bank, it would have to maintain a significantly higher equity-to-assets ratio (say, 15 percent instead of only 1.4 percent in 1995). The return on assets thus offers a more meaningful indicator of performance.

In competitive financial systems, a 1 percent return on assets is considered an indication of sound financial performance; figures of 2 to 3 percent are often recorded in the better performing commercial banking systems in emerging markets. By contrast, BRI-UD's pre-tax return on assets reached 6.1 percent in 1995, more than double the 2.6 percent return earned in 1991.1 Thus, whatever common financial indicator is used to assess their performance, the units have earned returns on rural financial intermediation that are well above those in the banking industry.

---

Several rural financial institutions in developing countries have reported adequate financial returns using the returns on equity and returns on assets, even though they are in fact dependent on subsidies (these include the Agricultural Development Bank of Jamaica, the Caisse Nationale de Credit Agricole in Morocco, the Grameen Bank in Bangladesh, and the Bank for Agriculture and Agricultural Cooperatives in Thailand). Because the BRI-UD was built on an earlier program's infrastructure of rural branches, it was able to shed its subsidies in only three years (figure 6). In 1995 it achieved a negative subsidy dependence index; the units could have reduced the yield on their loan portfolios by 44.5 percent (from 31.6 to 17.5 percent) and still remained independent of subsidies and earned an adequate market rate of return on equity. This indicates the "real" profitability that has resulted from effective rural finance intermediation and underscores the tremendous potential for efficient and profitable rural finance in other countries.

Although more in-depth impact evaluations are still required, borrowers enjoyed 25 percent growth in real profits, 21 percent growth in household income, and 18 percent growth in employment per enterprise, according to earlier studies (Sutoro and Haryanto 1990, as cited in Boomgard and Angell 1994; Patten and Rosengard 1991). Moreover, women have had greater access to banking service (25 percent) than is common in the Indonesian banking system (Reed and Befus 1993). The profits have also attracted competition from both private and public lenders, particularly at the higher end of the rural financial market, prompting BRI-UD to cut its on-lending rate on its larger loans by 7 percentage points; in some areas it also faces competition for deposits (Ravicz 1998).

![Figure 6. BRI-UD's Subsidy Dependence Index and Return on (Average) Equity](image)

Source: Bank Rakyat Indonesia's financial statements authors' calculations.
<table>
<thead>
<tr>
<th>Attribute</th>
<th>BIMAS Credit Program, 1970-84</th>
<th>BRI-UD, 1984–present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutional objective</td>
<td>Disbursement conduit for subsidized credit</td>
<td>Profitmaking, full-service rural bank</td>
</tr>
<tr>
<td>Financial autonomy</td>
<td>BIMAS windows in BRI branches, with accounts subsumed in BRI branches' financial statements</td>
<td>Distinct profit or loss centers, with separate financial accounting</td>
</tr>
<tr>
<td>Operational autonomy</td>
<td>Limited—borrowers chosen in practice by extension workers of the Ministry of Agriculture, which certified BIMAS participants</td>
<td>Full—borrowers selected on the basis of the financial viability of their farm or off-farm enterprise</td>
</tr>
<tr>
<td>Staff evaluation and accountability</td>
<td>Primarily based on the volume of disbursements or on hectares covered</td>
<td>Primarily based on the profitability of Unit Desas</td>
</tr>
<tr>
<td>Staff incentives</td>
<td>Civil service–like flat salary structure, promotions</td>
<td>Profit-related bonus incentives, promotions</td>
</tr>
<tr>
<td>Target market</td>
<td>Rice farmers</td>
<td>Any income-generating enterprises</td>
</tr>
<tr>
<td>Client incentives</td>
<td>Timely payment incentive: effectively none. Penalty for delinquency: curtailment of further loans, although not well enforced</td>
<td>Timely payment incentive: substantial interest rebate; larger follow-on loans. Penalty for delinquency: curtailment of further loans; incentives well monitored and enforced</td>
</tr>
<tr>
<td>Interest rates</td>
<td>12% (subsidized); below both the inflation rate and the interest rate paid on small savings deposits</td>
<td>Around 30% (not subsidized); well above both the inflation rate and the interest rates paid on small savings deposits</td>
</tr>
<tr>
<td>Main sources of funds</td>
<td>Concessional lines of credit, plus grants</td>
<td>Client deposits at market rates of interest</td>
</tr>
<tr>
<td>Dealing with losses</td>
<td>Soft budget constraint: operating losses covered by government</td>
<td>Hard budget constraint: loss-making operations suspended</td>
</tr>
<tr>
<td>The bottom line</td>
<td>Heavy losses and subsidy dependence</td>
<td>Large profits and self-sustainability</td>
</tr>
</tbody>
</table>

Source: Authors' findings.
Several insights into replicable findings can be gathered by contrasting the Village Bank system with its predecessor, the BIMAS credit program, which incorporated most features of traditional credit programs in other countries (table 2). That is, it offered targeted credit at below-market rates of interest and focused primarily on the volume of disbursements rather than on loan recovery and institutional viability. The lack of attention to the program’s long-term institutional viability encouraged adverse incentives on the part of staff and clients that ultimately led to its demise. Considerable attention must be paid to creating a conducive institutional framework for a public intervention to succeed and therefore for the market to develop. The key elements of such an environment include a hard budget constraint; full operational autonomy (that is, insulation from political interventions); skillful management of information; and a careful alignment of staff and client incentives with long-term institutional objectives.

What Has Happened to the Profits?

The answer to this question is far from reassuring. The vast profits have been used to cross-subsidize Bank Rakyat Indonesia’s wealthier clients. In fact, even as the Village Bank system succeeded, the rest of the bank continued to suffer from low recovery rates. This issue is of the utmost importance, because the enormous size of the cross-subsidy results in regressive income redistribution; year after year, small-scale entrepreneurs subsidize their more affluent countrymen. The rural lending scheme’s very success may have reduced the pressure on the parent bank to achieve an equivalent level of efficiency. In 1984, when BRI-UD was handed an ultimatum to become self-sustaining or face closure, it may have been expedient to leave the bank’s traditional lending to influential borrowers essentially unchanged. But it is clearly time to review these arrangements in light of their substantial economic costs to the country and their perverse effect on poverty reduction objectives.

Considerations for the Future

It is premature to assess the impact of the current financial crisis in Indonesia on the BRI-UD and its clients. Clients have not borrowed in foreign-denominated loans; the portfolios of the individual units were of excellent quality and high liquidity before the crisis—with a 55 percent loan to deposit ratio in 1996—and BRI-UD has enjoyed a flight to quality on the deposit side. Profits had not declined significantly as of 1997, largely due to high transfer prices late in the year for units with surplus liquidity. As an institution, however, Bank Rakyat Indonesia faces rising arrears and foreign exchange losses, reinforcing arguments in favor of broader reforms.
Nonetheless, the BRI-UD experience demonstrates not only that financial services can be extended to low-income rural clients at lower costs than previously thought possible, but that they can in many cases be provided while significantly reducing or even eliminating the need for any subsidies. The challenges that remain are those of strengthening the policy environment, improving the legal and regulatory framework, and adopting appropriate governance arrangements, management principles and operating procedures for interventions that reflect a new and more promising approach to rural finance.

The findings related to the level of subsidy independence are extremely important indications of the potential for improvement in the operations of numerous specialized agricultural credit institutions that still have little outreach to the target clientele and are heavily subsidized. Government policymakers, managers of state-owned rural financial institutions, and directors and managers of for-profit financial intermediaries generally believe that rural financial intermediation is not a profitable proposition and that subsidies are essential to compensate for mandated thin spreads and large loan losses. The experience of the Village Banks program shows that losses are not inevitable and that substantial and consistently increasing outreach can be achieved in rural financial markets in a self-sustaining manner. Indeed, the key lesson is crystal clear: rural finance can be highly profitable, even when it serves low-income clients.

Note

Jacob Yaron is rural finance adviser in the Private Sector Development and Rural Development Departments of the World Bank. McDonald Benjamin is a financial economist in the East Asia and Pacific Region Financial Sector Development Unit of the World Bank. Stephanie Charitonenko is a consultant in the Private Sector Development Department at the World Bank and a graduate student in agricultural economics at Michigan State University. This article draws on material in Yaron, Benjamin, and Piprek (1997).

1. Although BRI-UD, as a profit center within BRI, is not subject to corporate taxation, if a tax rate of 30 percent were applied, its return on investment would decline to 1.8 percent and 4.3 percent for 1991 and 1995, respectively.

2. A negative subsidy dependence index of 44.5 percent has no equal in rural or microfinance.

References

The word “processed” describes informally reproduced works that may not be commonly available through library systems.


*Jacob Yaron, McDonald Benjamin, and Stephanie Charitonenko*


Improving the supply of long-term credit to industrial firms is considered a priority for growth in developing countries. A World Bank multicountry study looks at whether a long-term credit shortage exists and, if so, whether it has had an impact on investment, productivity, and growth. The study finds that even after controlling for the characteristics of individual firms, businesses in developing countries use significantly less long-term debt than their counterparts in industrial countries. Researchers are able to explain the difference in debt composition between industrial and developing countries by firm characteristics; by macroeconomic factors; and, most importantly, by financial development, government subsidies, and legal and institutional factors.

The analysis concludes that long-term finance tends to be associated with higher productivity. An active stock market and an ability to enter into long-term contracts also allow firms to grow at faster rates than they could attain by relying on internal sources of funds and short-term credit alone. Importantly, although government-subsidized credit markets have increased the long-term indebtedness of firms, there is no evidence that these subsidies are associated with the ability of firms to grow faster. Indeed, in some cases subsidies are associated with lower productivity.

The popular view holds that financial markets in developing economies are highly imperfect and, in particular, that the alleged scarcity of long-term finance is a key impediment to greater investment and growth. Indeed, a significant part of the lending by the World Bank and other multilateral development banks is aimed at correcting for the dearth of long-term credit through the creation and encouragement of development finance institutions (DFIs) that could lend funds through loans from financial intermediaries and commercial banks, and recently through guarantees that lengthen the maturity of loans. Yet a recent strand of the finance literature has been studying the forces that determine the maturity structure of a firm’s debt (Berglof and von Thadden 1993; Diamond 1991, 1993; Rajan 1992). In those models
long-term debt is not necessary for acquiring physical capital and may depress a firm’s performance. These analysts would view policy-induced changes in the term structure of finance generally—if not uniformly—with great skepticism.

Notwithstanding the difference of views, attempts to cure the alleged scarcity of long-term credit in developing countries have been plentiful and expensive. By the early 1980s many DFIs were experiencing significant portfolio problems. A 1974 World Bank study of delinquency rates in agricultural lending institutions reported that the average arrears rate was 41 percent. A 1983 report indicated that 39 percent of all DFIs had serious portfolio problems, while another 53 percent faced moderate problems, many of which became more severe in the late 1980s and resulted in a wave of failures by DFIs (Siraj 1983).

Furthermore, in many cases, long-term directed credit failed to reach the intended beneficiaries (Atiyas 1991; World Bank 1989). Once these directed credit programs were established, governments found it politically difficult to reduce support for them, regardless of their cost and inefficiency. Prompted by these problems, the World Bank adopted new guidelines governing lending to DFIs, and World Bank loans to these institutions dropped dramatically, from 11 percent of new credits in fiscal 1989 to only 2.4 percent in fiscal 1993. Controversy continues, however, and both the World Bank and the development community at large are reevaluating mechanisms aimed at increasing the availability of long-term finance or lessening the constraints imposed by its absence.

Although aggregate data and anecdotal evidence suggest that there is less long-term credit in developing countries, even in those countries with low or moderate inflation rates, until recently no attempts had been made to examine the evidence more systematically to see if the scarcity merely reflected the characteristics of firms in poorer economies. This lacuna was understandable even five years ago, because the data were not available. Recently, however, this gap has been filled, first by the availability of data in emerging markets for the top tier of firms listed on stock exchanges, and second by various surveys of listed (and unlisted) firms in selected countries whose governments have sought to understand the impact of a variety of policies. Armed with this data, a number of studies have appeared in the last year, and answers are now available to fill an important gap in our knowledge.

Financing Decisions and Debt Maturity

What does it mean to say that long-term credit is scarce? Typically this question has been answered by asking firms to identify the important constraints on their operations; credit—usually long-term credit—regularly is at or near the top of the list. This approach is unsatisfactory, however, not least because it is unclear what the respondents imagine they will pay for credit. Moreover, it is unclear under what type
of financial system they would be able to obtain short- or long-term credit. Even the most advanced financial system will find some borrowers uncreditworthy or will lend them much less than they might desire or at higher interest rates than they would like. In the case of riskier firms, loans at average market rates are attractive precisely because they convey a subsidy in the form of a lower risk premium than the market would grant them. Whenever the lack of long-term credit constrains many firms from expanding, there are three potential sources of the credit constraints: first, macroeconomic factors that limit supply; second, institutional factors that are specific to the financial sector (often dubbed market imperfections); and third, the characteristics of the firms, or classes of firms, in the country.

One way to interpret scarcity then is by relative access to credit. That is, there is scarcity to the extent that developing country firms find it more difficult to borrow money than do similar firms in industrial countries. In this relative sense, if there is a scarcity, there may be a potential correction. To be sure, any correction may be difficult. For example, it is argued (and confirmed below) that a leading cause for the absence of long-term finance is high inflation and unstable macroeconomic policies. Attempts to increase the supply of long-term credit without addressing the inflation problem could easily prove to be short-lived or costly. Similarly, high real interest rates may reduce the effective demand for credit. Entrepreneurs will say they want more credit, but not at the market price. If the yield curve is upward sloping (meaning that long-term interest rates are higher than short-term rates), the demand for long-term credit will tend to suffer most. Again, addressing the factors that account for high real interest rates may in the long run be more useful than attempts to force banks to make longer-term commitments (Brock 1995). In the 1980s Chile succeeded both in tackling the factors underlying high real rates (an overvalued exchange rate and insolvent banks) and, by moving to a fully funded pension system, in creating a natural source of long-term finance without interfering in credit and investment decisions. The studies reviewed here finesse this issue by limiting the study to countries with relatively stable macroeconomic environments.

Institutional factors generally affect borrowers only until funds are disbursed but are crucial during all phases of a credit relationship for providers of funds, who are concerned about the return on their investment. Diamond (1991) points out that banks use short-term credit as a way to control borrowers and that they tend to use this instrument more frequently in cases in which the financial infrastructure is underdeveloped. Thus if information systems or contract enforcement mechanisms are absent, or accounting and auditing techniques are not adequate, lenders will be unwilling to enter into long-term loans. Ignoring this deficiency and establishing government banks for long-term credit is faster, cheaper, and less difficult than trying to address the information or contract enforcement problems, but government banks will have to cope with the same issues and may have additional incentive problems as well.

Gerard Caprio, Jr., and Asli Demirgüç-Kunt
Finally, the maturity structure differs within an economy depending on the characteristics of domestic firms. Below we review the importance of these firm-specific factors that affect access to long-term finance.

Access within Countries: The Relevance of Firm Characteristics

In the aftermath of the seminal Modigliani-Miller (1958) article, which found that the value of a firm was not affected by its mix of financing, the study of financing choices initially received little attention. As economists and finance experts have renounced the simplifying assumptions of this classic framework, however, they have developed a literature on the maturity structure of firm financing, stressing the different roles played by long- and short-term finance. This literature emphasizes that short-term debt has three effects: it permits loans to be repriced to reflect new information; it increases efficiency by allowing uneconomic projects to be terminated; and it gives managers and owners strong incentives to avoid bad outcomes. In contrast, long-term debt protects the firm from liquidation by imperfectly informed creditors and prevents opportunistic creditors from using the threat of liquidation to expropriate the profits of healthy firms.

Several factors determine the optimal mix of long- and short-term debt. These include the firm’s credit rating, its portfolio of growth opportunities, the profitability of the project, the ability to fund the project through retained earnings, the liquidation value of the assets, the perceived accuracy of financial information, the firm’s size and age, and the level of banking competition. Valuable assets that can serve as collateral ease borrowing constraints considerably. According to Myers (1977) firms can also use their “growth opportunities” as collateral. Firms whose principal asset is the present value of growth opportunities may not be able to borrow that easily, however, because the owner-managers have greater opportunities to divert resources for their own use. As Myers notes, the firm can limit this problem by carrying less debt, by including restrictive covenants in its debt contracts, or by borrowing more short-term debt (which permits the creditors to detect opportunistic behavior relatively quickly). In developing countries, one might expect to find more firms relying on growth opportunities, so this diversion problem could be significant. Moreover, since it is difficult in lower-income countries both to sell shares of stock (one way to lower debt-equity ratios) and to enforce contracts (because regulatory mechanisms are typically less developed), businesses can be expected to use more short-term debt. By sequencing a series of short-term loans, bankers retain control over their clients because the option to discontinue rolling over these loans is easier to exercise and is a more credible near-term threat.

Hart and Moore (1995) find that the faster the returns to investment are realized, the shorter the optimal payment structure will be. Empirically, this suggests a particular relationship among the maturity of debt, purpose of the loan, and the nature
of the firm's assets. Long-term loans are usually used to acquire fixed assets, equipment, and the like. Short-term loans, on the other hand, tend to be used for working capital, such as payroll, inventory, and seasonal imbalances. Collateral usually consists of such things as inventories or accounts receivable. In other words, firms will tend to match the maturity of their assets and liabilities; only firms with long-term assets will tend to have a longer debt maturity structure. If this tendency is born out in developing country experience, it suggests that the most effective way to deal with the market allocation of credit is to take account of the structure of the firms' assets. A program to extend long-term credit to firms with short-term assets may not be welcomed, as it is inconsistent with the desire to balance the maturity of assets and liabilities.

The size of the firm is another key variable. Indeed, the desire to get more credit—particularly long-term credit—to small firms is a justification for a number of credit market interventions. In general there tends to be less information about small firms, not only because they are new, but also because such information is costly to obtain. Thus, even in the most developed financial systems, small- and medium-size enterprises tend to get a larger part of their external financing from banks. Banks overcome some information problems by developing long-term relationships with smaller firms.

The point is that firms in developing countries may have less long-term debt than firms in developed countries simply because they have different characteristics rather than because of deficiencies in credit markets. Moreover, comparisons of debt maturity structures in different countries are more likely to be informative if researchers control for these parameters.

Although numerous empirical papers test the implications of capital structure models, attention has turned only recently to empirical determinants of debt maturity (see Harris and Raviv 1990 for a review of the literature). Titman and Wessels (1988) find that highly leveraged firms tend to issue more long-term and short-term debt but that the mix varies according to the firm's characteristics. Barclay and Smith (1995) report that large firms as well as those that have few growth options have more long-term debt, a finding confirmed by Stohs and Mauer (1996), who note that larger, less risky firms with longer-term asset maturities use longer-term debt.

In these studies, based on U.S. data, the link that stands up most clearly is the matching of firm assets and liabilities. This finding is quite robust in Italy and the United Kingdom (Schiantarelli and Srivastava 1996), where it is also clear that firms with higher profits have access to more long-term credit. World Bank research using developing country data generally confirmed these results. Maturity matching also is evident in Colombia (Calomiris, Halouva, and Ospina 1996), India (Schiantarelli and Sembenedelli 1996), and Ecuador (Jaramillo and Schiantarelli 1996). If maturity matching represents a tendency in both industrial and developing country markets, attempts to stimulate long-term finance may prove to be excessive; firms may take
on long-term debt only if it fits their balance sheet structure, and perhaps only if long-term debt is subsidized, meaning that they can take advantage of a lower risk premium than is available from the market. These country studies confirmed that where financial markets are free from government intervention, they provide more long-term finance to better quality firms, and attempt to monitor lower quality firms more closely by using short-term debt.

The government’s decision to intervene in credit markets should depend on the link between long-term credit and the firm’s performance as well as on equity considerations. Lack of collateral as well as the age of the business may be factors where small firms find it difficult to obtain long-term credit, as in Ecuador, where only 11 percent of very small firms and 17 percent of small firms reported long-term debt every year (1984–88). Larger firms in Ecuador tended to be more profitable, suggesting that the allocation of credit favored firms with better balance sheets. The allocation could also reflect the economic and political power of such firms. A more disturbing aspect was that, regardless of the size of the firm, the amount of long-term credit obtained was unrelated to past profits. Whether this reflects a market failure, the limits of banking (bankers can pick the class or industry, but not individual winners and losers), or excessive intervention (a substantial portion of the debt was subsidized) is not clear.

Access across Countries: The Relevance of Institutional Factors

Financial theory suggests that a major factor in the choice of capital structure is the cost of contracting between firms and their providers of capital. It is the institutions in the economy—legal or financial—that facilitate monitoring and enforcement of financial contracts.

For example, when the legal system is costly or inefficient, short-term debt is more attractive than long-term debt (Hart and Moore 1995; Bolton and Scharfstein 1993). Diamond (1991, 1993) also emphasizes the importance of clear legal rules to govern contract enforcement. This implies that if complicated loan covenants (to anticipate a variety of future outcomes) could be enforced at a lower cost, the risk for the lender would be reduced and the willingness to expand the supply of long-term debt would increase.

Financial institutions. Two types of institutions—financial intermediaries and stock markets—directly influence an enterprise’s choice of financial structure. Financial intermediaries have a comparative advantage in monitoring borrowers because, as Diamond (1984) argues, bankers have economies of scale in obtaining information. They may also have greater incentives to use the information to discipline borrowers than do small investors. By collecting information, monitoring borrowers, and exerting corporate control, a developed banking sector can facilitate access
to external finance—especially long-term finance—by smaller firms that have limited access to alternative means of financing due to information costs.

Large stock markets allow entrepreneurs the opportunity to substitute equity for long-term debt. Moreover, the prices quoted in financial markets also transmit information that is useful to creditors (Grossman 1976; Grossman and Stiglitz 1980). This revelation may make lending to a publicly quoted firm less risky and thereby increase its ability to obtain long-term credit.

**GOVERNMENT PROVISION.** In an effort to promote the availability and use of long-term debt, governments may adopt policies that direct or subsidize long-term capital to favored firms or sectors. Directed credit policies include preferential discount lines from the central bank, portfolio restrictions on private commercial banks, guaranteed credit for public enterprises, and credit lines through development banks. These programs need not always involve financial subsidies, but they frequently do. The degree of these distortions varies from country to country.

For example, Atiyas (1991) and World Bank (1989) provide evidence that directed credit often fails to reach its intended beneficiaries. In many cases, such programs are used not to correct the inadequacies of financial markets, but to channel funds to priority sectors whether or not these are the most productive investments. In many countries, including Brazil, Colombia, India, Kenya, Mexico, and Turkey, government interventions have generated large costs by funding inefficient borrowers and crowding out private credit intermediaries.

Directed credit programs did achieve their legitimate objectives in a few cases. In successful interventions, as in Japan, Korea and Singapore, credit policy priorities are determined as part of a national plan with broad participation. Commercial standards are applied within the priority sectors; once priorities have been established, lending decisions by agencies are shielded from public pressures; interventions that do not work are discontinued (Calomiris and Himmelberg 1993; Stiglitz and Uy 1996). Where political systems do not allow government authorities to develop and implement effective plans for the distribution of industrial credit, however, success may be difficult to achieve.

**THE INSTITUTIONAL ENVIRONMENT.** Several studies explore the effect of the institutional environment on the choice of debt. Hoshi, Kashyap, and Scharfstein (1990) show that membership in industrial groups linked to banks reduces financial constraints on Japanese firms, and Schiantarelli and Sembenelli (1996) find the same benefits flow to Italian firms that are members of large national groups. Calomiris (1993) examined the effect of differences between the banking systems of the United States and Germany and argued that regulatory limitations on the scale and scope of U.S. banks hampered financial coordination and increased the cost of capital for industrialization. Rajan and Zingales (1995) and Demirgüç-Kunt and Maksimovic
(1994) compare the capital structure of firms in five industrial countries and ten developing countries, respectively, and find that institutional differences were crucial in understanding the determinants of capital structure.

Because of data constraints, however, systematic cross-country empirical studies have been few and recent. Demirgüç-Kunt and Maksimovic (1996a, 1996b), who look at debt-equity ratios in 30 industrial and developing countries from 1980 to 1991, find that access to an active stock market increases firms' ability to borrow, especially in countries with developing financial markets. They also report systematic differences in the use of long-term debt between industrial and developing countries, as well as between small and large firms, even after controlling for the characteristics of the enterprises.

The data in figure 1 show that firms in industrial countries clearly have more long-term debt as a proportion of total assets. For example, the long-term debt-to-asset ratio of an average firm in Norway, with a per capita gross domestic product of $20,000, is five times greater than it is in Zimbabwe, with a per capita GDP of less than $1,000. And large firms have more long-term debt as a proportion of total assets and debt than do smaller firms (figure 2).

**Figure 1. Average Long-Term Debt as a Percentage of Total Assets, 1980–91**

![Graph showing average long-term debt as a percentage of total assets for different countries, with large firms having more long-term debt than smaller firms.](source: Demirgüç-Kunt and Maksimovic (1996b). The data set consists of financial statement data for the largest publicly traded corporations in manufacturing.)
Importantly, this lack of long-term finance in developing countries persists even after controlling for firm characteristics and macroeconomic factors. Demirgüç-Kunt and Maksimovic explain this differential by institutional differences, such as the extent of government subsidies, different levels of development of stock markets and banks, and differences in the underlying legal infrastructure. Their results indicate that although policies that help develop legal and financial infrastructure are effective in increasing access to long-term debt, different policies are required to lengthen the debt maturity of large and small firms. Improvements in the legal system, for example, benefit all firms, although the effect is less significant for the smallest firms, which have limited access to the legal system. In the study, legal effectiveness is measured on a six-point scale every year, so, for example, the United States has a value of 6 and India a value of less than 2 for the whole sample period. The results indicate that a 10 percent improvement in the index of legal effectiveness would increase the largest firms’ ratios of long-term debt to total debt by more than 5

Figure 2. Debt Ratios of Small vs. Large Firms for 30 Countries

Note: The figure presents the average debt ratio across 30 countries by firm size. The firms in each country are divided into quartiles by value of total assets, and the average debt ratios of each quartile, calculated across countries, is reported. Countries in the sample are Australia, Austria, Belgium, Brazil, Canada, Germany, Finland, France, Hong Kong, India, Italy, Jordan, Japan, Republic of Korea, Malaysia, Mexico, Netherlands, Norway, New Zealand, Pakistan, Singapore, South Africa, Spain, Switzerland, Sweden, Thailand, Turkey, United Kingdom, United States and Zimbabwe.

Source: Demirgüç-Kunt and Maksimovic (1996b).
percent. Thus, keeping all else constant, if a large Indian firm were to be transplanted to the United States, its ratio of long-term debt to total debt would increase by more than 100 percent. For the smallest firms, a 10 percent improvement in the legal index brings about an increase of 4 percent in their long-term debt ratio. In addition to being smaller in magnitude, this effect is also less significant in statistical terms.

Moreover, the authors find that policies to improve the functioning and liquidity of stock markets would benefit primarily large firms. During the sample period, the United States had one of the most liquid stock markets, with a turnover ratio of 60 percent, while Pakistan had one of the least liquid, with a turnover ratio of 11 percent. (Turnover is the value of shares traded divided by market capitalization.) The results indicate that a 10 percent increase in stock market liquidity leads to less than a 1 percent increase in the ratio of long-term debt to total debt for the largest firms. Holding everything else constant, however, transplanting the largest Pakistani firm to the United States would increase the ratio of the firm’s long-term debt to total debt by 40 percent. This effect is not significant for the smallest firms.

In contrast, policies to improve the banking system would give smaller firms better access to long-term credit. The authors use the ratio of bank assets to gross domestic product as an indicator of banking development. In the sample, Germany has one of the largest banking systems—with a ratio of more than 100 percent—and Zimbabwe one of the smallest, with a ratio of 17 percent. The results indicate that a 10 percent increase in banking size leads to about a 3 percent increase in the ratios of long-term debt to total debt of the smallest firms. And again, keeping everything else constant, a small Zimbabwean firm transplanted to Germany would increase its long-term debt ratio by more than 100 percent. This result is not significant for large firms.

The Effect of Debt Maturity on Performance

In recent years theorists have been studying the forces that determine the maturity structure of a firm’s debt (Brick and Ravid 1985; Diamond 1993; Kale and Noe 1990). This literature provides an interesting perspective on how this choice affects the enterprise’s performance by emphasizing the different control and incentive properties of long- and short-term debt. Indeed, these models highlight several of the undesirable effects of long-term debt.

First, a capital structure that excludes long-term debt may be more efficient. Relying on such debt leads to greater distortions in the risk preferences of owners and managers, providing them with incentives to invest in projects that benefit themselves at the expense of outside investors (Myers 1977). This conflict can be miti-
gated, however, by reducing the overall degree of leverage, or the maturity of debt, since the short maturity limits the period during which an opportunistic firm can exploit its creditors without being in default (Diamond 1991, 1993).

Second, short-term debt may also increase efficiency because of its role in disciplining management (Jensen 1986). Because of the more continuous scrutiny of a firm's operations and the threat of liquidation, short-term debt may in fact constrain wasteful activities.

Third, long-term debt allows management to delay its response to deteriorating market conditions and avoid exiting the market when it is obvious that the firm is failing. This reduction in efficiency is even worse if the loan is subsidized. In addition, when the market for refinancing short-term loans is competitive, long-term capital always increases the firm's financing costs and reduces the incentive to exert the effort to save on interest costs and increase efficiency (Rajan 1992; Ofek 1993).

Fourth, debt maturity is also correlated with credit quality and the profitability of existing projects. In the presence of asymmetric information about borrowers, firms of higher quality should choose short-term debt because they will be able to take advantage of the revelation of future good news (Diamond 1991). This positive information effect outweighs the liquidity risk of not being able to refinance oneself and running the risk of being liquidated by the lender. The opposite is true for firms with lower credit rating. Firms with the lowest credit ratings have access only to short-term debt, leading to a nonlinear relationship between maturity and credit rating.

But shorter debt maturity is not all good. Fear of liquidation may induce firms to avoid investing in profitable projects if the returns accrue in the distant future. Similarly, they may be reluctant to adopt more productive technologies unless they provide an immediate payoff. This shortening of the investment horizon may have negative consequences on overall performance. The faster the return on investment, the shorter will be the optimal payment structure. This provides a rationale for firms with long-term assets to have a longer debt maturity structure (Hart and Moore 1995). If financial markets undersupply long-term credit because banks are unable to internalize the full benefits of monitoring the firm (Mayer 1989; Calomiris and Himmelberg 1993) or because few people participate in financial markets (Diamond 1996), firms with a longer asset maturity may be disadvantaged.

On balance, the theoretical literature is inconclusive on how the maturity of debt affects a firm's performance. Notwithstanding data problems, empirical analysis in this area is difficult because it is not appropriate to draw conclusions about performance by simply treating maturity structure variables as independent, given that theory says that expected growth and profitability also affect the choice of maturity. The recent empirical literature attempts to avoid this simultaneity problem by focusing on performance indicators, such as efficiency measures, that should not play a role in "causing" maturity choice, or by using in-
strumens for maturity choice, such as legal efficiency indicators, that measure the ability to enter into long-term contracts. This literature provides some interesting answers.

Evidence from Country Cases

Most of the empirical work in this area has been on growth and external financing. In a seminal work on the effect of financial constraints on investment decisions, Fazzari, Hubbard, and Petersen (1988) show that investment by U.S. firms is sensitive to cash flow. In later works, Calomiris and Hubbard (1995), Calomiris, Himmelberg, and Wachtel (1995), Carpenter, Fazzari, and Petersen (1994), and Calomiris and Himmelberg (1996) argue that the high shadow cost of external finance will show itself most clearly in the cash flow sensitivity of inventories. But surprisingly little work has been done on the links between debt maturity and performance. Gilson, John, and Lang (1990) find that the more long-term debt a firm has, the more likely it will be to reorganize successfully. Hall (1992) reports that when the ratio of long-term debt to physical capital increases, physical investment and research and development expenditures contract. Atiyas (1991), who investigates directed credit programs in Colombia, reports a negative relationship between long-term indebtedness and efficiency.

More recently, a number of case studies were conducted in industrial and developing countries using information about firms in each country—Schianterelli and Sembenelli (1996) on the United Kingdom and Italy; Jaramillo and Schiantarelli (1996) on Ecuador; Calomiris, Halouva, and Ospina (1996) on Colombia; and Schiantarelli and Srivastava (1996) on India. In general, these studies find that short-term debt has no effect on efficiency and growth. The conventional wisdom that more long-term debt may actually lead to productivity improvements was confirmed in Ecuador, Italy, and the United Kingdom. Echoing the earlier findings for Colombia, however, the positive effect of long-term debt in Italy is substantially reduced—and even reversed—if the debt is subsidized (Schianterelli and Sembenelli 1996). There is no evidence that the total investment is sensitive to the amount of long-term credit.

Evidence from Cross-Country Studies

Rajan and Zingales (1996), who look at industry-level data across 27 countries, find a positive correlation between dependence on external financing and growth. They do not distinguish between external debt and equity, however, or between differences in debt maturity. Demirgüç-Kunt and Maksimovic (1996c) analyze firm-level data for 30 countries and find that an active stock market and the
ability to enter into long-term contracts enable firms to grow at faster rates than they could attain by relying on internal capital and short-term credit alone.

This research examines whether differences in financial and institutional development prevent firms in some countries from investing in potentially profitable growth opportunities. Such an effect, if it exists, would not affect all firms equally. Firms that can finance operations from retained earnings will be minimally affected, whereas firms whose financing needs exceed their internal resources may be severely constrained. To gauge the effect of financial development on the firm's ability to exploit growth options, it is necessary to identify firms that have a need for external financing and examine whether their realized growth depends on the development of financial markets.

To get around this problem, Demirguc-Kunt and Maksimovic calculate growth rates of firms that rely only on internal financing. Their estimate of the constrained growth rate is based on the standard “percentage of sales” approach to financial planning (see Ross, Westerfield, and Jordan 1995). For each firm they define a maximum short-term financed growth rate, which is an estimate of the maximum rate of growth if a firm reinvests all earnings and obtains enough short-term credit to maintain the ratio of its short-term borrowing to assets. They then calculate the proportion of firms that exceed their benchmark growth rates, average these over the 1980–90 period, and relate these excess growth rates to firm characteristics, including asset composition, profitability, and size; macroeconomic indicators, such as inflation and the growth rate of the economy; institutional factors, such as the effectiveness of the legal system and the level of government subsidies to enterprises; and financial indicators, such as banking development and stock market liquidity.

Their results show that the proportion of firms that grow faster than the predicted rate is related to specific features of the legal and financial systems and to the institutional structure of the economy. Even after controlling for firm characteristics, macroeconomic environment, financial development, and the extent of government intervention, the results show that firms with higher long-term debt ratios tend to grow faster than they would if they relied solely on internal resources (figure 3). Because it is difficult to identify the direction of causality in this framework, the authors also test their results by replacing the debt ratio with an index of legal effectiveness that they have shown to be highly correlated with long-term debt (Demirgüç-Kunt and Maksimovic 1996b). The results indicate that in countries whose legal systems score high on an efficiency index—which is expected to be an exogenous indicator of the ability to enter into long-term contracts—a greater proportion of firms grow at faster-than-predicted rates. They also find that an active, although not necessarily large, stock market is associated with faster growth. An additional test of causality using initial values of stock market liquidity and legal effectiveness to predict future growth does not alter the results. The findings
Figure 3. Firm Growth and Long-Term Debt

Excess firm growth

<table>
<thead>
<tr>
<th>Ratio of average long-term debt to total assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12</td>
</tr>
</tbody>
</table>

Note: The y-axis is the proportion of firms growing faster than predicted rates. Predicted rate is the rate at which a firm can grow by only relying on retained earnings and short-term credit. Thirty countries in the sample were divided into three equal groups based on the average long-term debt to total assets (LT/TA) ratios of firms. Countries with the lowest LT/TA are Brazil, Hong Kong, Jordan, Malaysia, Mexico, Pakistan, Thailand, Turkey, the United Kingdom, and Zimbabwe. Australia, Belgium, India, Italy, Japan, Korea, Netherlands, Singapore, South Africa, and Spain are in the second group. The third group comprises Austria, Canada, Finland, France, Germany, New Zealand, Norway, Switzerland, Sweden, and the United States.

Source: Calculations based on Demirguc-Kunt and Maksimovic (1996c).

make it clear that the underdevelopment of financial markets and institutions prevent firms in developing countries from investing in potentially profitable growth opportunities.

Do policies intended to increase the availability of long-term debt actually work? Although government subsidies have increased the long-term indebtedness of firms around the world, Demirguc-Kunt and Maksimovic (1996c) find no evidence that government subsidies to firms in the 30 countries in their sample were associated with faster growth (figure 4).

On the contrary, their evidence indicates that although long-term debt is associated with greater numbers of firms growing at rates that are higher than predicted, this result is reversed to the extent the credit is government-subsidized. Country case studies also confirm this effect. And Schiantarelli and Sembenelli (1996) show that the positive effect of debt maturity on firm performance declines as the aggregate proportion of subsidized credit rises—until it is reversed.
Figure 4. The Effect of Subsidized Long-Term Debt on Enterprise Growth

Excess firm growth

<table>
<thead>
<tr>
<th>Subsidized long-term debt</th>
<th>Excess firm growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.23</td>
<td>0.42</td>
</tr>
<tr>
<td>0.55</td>
<td>0.41</td>
</tr>
<tr>
<td>1.29</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Note: The y-axis is the proportion of firms growing faster than predicted rates. Predicted rate is the rate at which a firm can grow by only relying on retained earnings and short-term credit. Twenty-six countries for which data were available were divided into three groups (9, 9, 8) based on a ranking of the extent of subsidized long-term debt obtained by multiplying average LTD/TA ratios of firms by the ratio of government subsidy to gross domestic product. Countries with highest subsidies are Austria, Belgium, Brazil, Finland, India, Italy, Korea, Norway, and Sweden. The middle group includes Canada, France, Germany, Malaysia, Netherlands, New Zealand, Pakistan, Spain, and Switzerland. Finally, Australia, Japan, Mexico, Singapore, Thailand, Turkey, the United Kingdom, and the United States have the lowest amount of subsidized long-term debt.

Source: Calculations based on Demirgüç-Kunt and Maksimovic (1996c).

Lessons for Development Economists

First, even after adjusting for the characteristics of individual firms, long-term credit is scarce in developing countries, particularly for smaller firms. Thus if a firm in a developing country were magically relocated to an industrial economy, other things being equal, it could expect to receive more long-term credit. Second, among manufacturing firms, there is clear evidence that more long-term finance tends to be associated with high productivity. Third, and perhaps most important, this last result is reversed to the extent that such credit is subsidized.

Therefore, although it is worthwhile for governments to attempt to foster the supply of long-term credit, it is crucial that these interventions be crafted with great care—and little subsidy. If the macroeconomic environment is unstable, it
is unlikely that the market will provide long-term finance. When fixed interest rates are offered, savers regularly show that they are averse to putting their assets into long-term instruments; the yield curve is—or may soon become—inverted. When variable interest rates are offered, borrowers, or at least those who intend to repay, will not readily enter into a contract that could leave them bankrupt. Additionally, real sector reforms are important, as they will lead to changes in relative prices and in the performance of firms. If relative prices are clearly out of line, investors will not want to enter into long-term arrangements, so there will not be much demand for long-term credit. Beyond these reforms, however, firm characteristics do not admit to easy change. Institutional differences, such as the adequacy of banking and stock markets and the legal infrastructure, are important in affecting the supply of long-term credit and can be changed, although not overnight. Focusing attention on basic financial infrastructure is a “low distortion” road to achieving more long-term credit. Similarly, improving the legal system and collateral registration would especially redound to the benefit of smaller borrowers, for whom contract enforcement issues (and collateralized finance) tend to be of overriding importance.

Overwhelmingly, long-term finance tends to go to larger companies. That may be good for growth if there is evidence that larger firms tend to be more productive, but societies may be willing to sacrifice some growth for more equity. Unfortunately, many schemes designed to facilitate access to credit for small firms have not achieved their goal. Research shows that banking and stock market development are complementary, probably because each produces and demands better information (Demirgüç-Kunt and Levine 1996). Hence, banking development will not only improve small firms’ access but indirectly help larger firms by leading to greater capital market development as well.

The development of pension funds, insurance, and contractual savings systems are similarly low-distortion paths to stimulating long-term finance (Vittas 1996). Some economists argue that moving to a fully funded old age pension system would induce people to save more, but this has been a subject of considerable controversy on both theoretical and empirical grounds. In the case of Chile, however, a switch from a pay-as-you-go public pension plan to a fully funded private plan increased private savings substantially, although this is likely to be a transitory effect. Even in the absence of a permanent increase in savings, most analysts agree that the development of private contractual savings institutions leads to an increase in the supply of long-term funds because in every country these institutions hold a portfolio dominated by long-term assets. Thus, encouraging mandatory, fully funded pension schemes is an appealing way to encourage such credit indirectly. How rapidly governments will want to move in this direction depends on several variables, not least of which are demographic trends and the likelihood that investment pools can be allocated free of government interference.
Directions for Research

What do we still need to know about long-term credit? First, it is important to confirm the findings reported here. Although deep data sets that include listed and unlisted firms are still relatively rare, testing the findings here with new data as they become available would be appealing and worthwhile. Additionally, it is important to take a closer look at how countries have allocated such credit because it can, but does not always, have a positive effect on performance. Although government interventions have been costly and inefficient in many cases, credit policies in some countries have succeeded by establishing credible mechanisms that ensure proper allocation and repayment of funds (World Bank 1993). In-depth case studies of individual firms would help generate more precise recommendations about how credit policies can achieve optimal growth and equity outcomes; in several countries, long-term credit institutions face insolvency partly as a result of these programs. A careful balancing of the benefits and costs of credit-market interventions is therefore a top research priority.

Notes:

Gerard Caprio, Jr., is a research manager and Asli Demirgüç-Kunt is senior economist in the Development Economics Research Group of the World Bank. They would like to thank Stijn Claessens, Harry Huizinga, Vojislav Maksimovic, Fabio Schiantarelli, and Mary Shirley for helpful comments.

References

The word “processed” describes informally reproduced works that may not be commonly available through library systems.


Gerard Caprio, Jr., and Ati Demirguc-Kunt 189
Drawing from the experience of the direct income support programs recently introduced in the European Union, Mexico, and the United States, this article highlights problems that may arise when the agricultural sector of a developing economy moves from price-based subsidization to less distorted income support. Such programs are a step in the right direction, but as currently implemented, they have many shortcomings. Moreover, developing countries may lack the necessary supporting arrangements needed to make such programs effective. The article argues that the programs should not restrict the use of land, that the programs should last for a stipulated period of time, and that the fiscal costs should be contained by linking income support payments to world prices.

Interventions in pricing policies play a vital role in the performance of agriculture, but they have often had an unfavorable effect on economic development (World Bank 1986; Tyers and Anderson 1992; Krueger, Schiff, and Valdés 1992; Meerman 1997). In industrial countries the allocative inefficiency resulting from the pricing policies of agricultural protectionism puts the deadweight losses to a protectionist government somewhere between 0.5 percent and 3 percent of gross domestic product (GDP) (Burniaux and others 1990). Fiscal costs have also been high and are increasingly viewed as excessive by predominantly urban electorates.

In response, many countries have undertaken structural reforms to stimulate production by liberalizing farm prices and integrating them with those of the world economy. Moreover, responding to commitments under the Uruguay Round of the General Agreement on Tariffs and Trade (GATT) as well as other multilateral and regional arrangements, more countries are expected to embark on similar price and trade reforms. For example, developing countries are committed to reducing average price supports by 13 percent in 10 years (Valdés and McCalla 1997).

The move to eliminate distortions, a step that inevitably implies some redistribution of income, is bound to be resisted by those groups that are penalized. Producers, for instance, will welcome the removal of a commodity tax, but the treasury and...
other groups that have benefited from the tax are likely to oppose it. Farmers will object to proposals to reduce producer subsidies, because the additional costs that result will threaten the profitability of agricultural production. (Winters 1987, 1989–90, and Gardner 1987, 1990, discuss the political economy of agricultural protection.)

To check such adamant political opposition, some governments have attempted to replace price-distorting subsidies with direct income support mechanisms so that current production decisions are independent of—or less dependent on—support prices for specific crops. Because income support programs compensate farmers for the loss of income, they are politically feasible and may make reform easier to accept.

Whether such schemes are in fact effective mechanisms for eliminating price supports is of interest to developing countries for at least two reasons. First, contrary to the widespread belief that industrial countries subsidize agriculture while developing ones tax it, today many developing countries subsidize agriculture—or its specific subsectors. Second, price supports in industrial countries affect developing countries in several ways. Subsidy-induced surpluses in member countries of the Organisation for Economic Co-operation and Development (OECD) depress world prices and reduce the income derived from competing exports from developing nations. The direct losses to foreign producers resulting from U.S. farm programs, for example, are estimated at about $1 billion in 1990. This figure is much higher if one considers the deadweight losses. OECD countries protect agriculture to insulate their markets from external shocks, but such subsidies depress world prices and increase their volatility, thus increasing the pressure for commodity stabilization programs in developing countries. Estimates suggest that the operation of the European Union’s Common Agricultural Policy (CAP) has doubled the variability of world dairy prices while raising prices of wheat and beef by almost 50 percent within the Common Market (Knudsen and others 1990).

Agricultural Protection in Developing Countries

Agricultural protection in developing countries is not uncommon. In a study of eight Latin American countries, Valdés (1996) found that Chile, Colombia, and the Dominican Republic protect agriculture, while Brazil offers minimal protection. Argentina, Ecuador, Paraguay, and Uruguay protect certain subsectors. From 1989 to 1992 Nigeria maintained official prices on wheat and coarse grains at an average of 82 percent and 92 percent, respectively, above their world market prices. Algeria, Morocco, and Tunisia protected the same commodities at a combined average of 34 percent and 13 percent. Mexico's effective tariff equivalents for wheat and coarse grains were 55 percent and 69 percent of world prices in 1989–93 (Ingco 1995).
Extremely high production-supporting subsidies in the agricultural sectors of former centrally planned economies were (and in several cases still are) the rule (Brooks 1993). With the exception of some countries of the former Soviet Union, developing countries that have protected agriculture have usually been importers of the crops protected. In these countries protection has generally taken the form of import restrictions, including tariffs, import prohibitions, or restrictive import licensing. Because the transfer was made directly from consumers to producers, protection in these agricultural importing countries did not require large budget outlays. (The standard analysis of the efficiency, distributive, and fiscal effects of an import tariff is provided in appendix A.)

But in the OECD countries, agricultural protection has usually taken the form of crop prices pegged above world-market levels. This form of protection requires substantial fiscal outlays to finance the subsidies needed to encourage exports of the crop surpluses induced by the high prices. By 1991 the European Union was allocating nearly 1 percent of its GDP (58 billion ECUs) to agriculture, most of it to support the extreme price distortions under the CAP. For example, the wholesale prices of sugar, rice, and butter in 1991 were 308 percent, 171 percent, and 247 percent of respective world market prices (Atkin 1993).

Inputs have been subsidized as well. Governments have held down prices of fertilizer, irrigation, seeds, electricity, credit, and insurance. Knudsen and others (1990) report that in the early and mid-1980s, fertilizer subsidies in Sri Lanka and Turkey cost about 1 percent of total GDP, while those in Côte d’Ivoire, Egypt, Gambia, and Tanzania ranged from 50 percent to 100 percent of the market price of fertilizer.

Who Benefits from Agricultural Protection?

Replacing price supports with direct income support to farmers not only affects the income distribution of the concerned groups (the equity effect), but also generates efficiency gains (the efficiency effect). Whether income should be redistributed (through price supports, tariffs, taxes, lump-sum transfers, and so on) is an equity question in the sense that the marginal welfare of the proposed benefiting group (producers) is in some sense valued more than that of the proposed losing group (consumers or taxpayers). In terms of the demand-supply diagram shown in the appendix (figure A.1), the introduction of a protective tariff on imported agricultural commodities implies that certain producers are valued more than consumers. The efficiency argument, on the other hand, centers on the question of eliminating deadweight losses, that is, increasing total welfare by reallocating capital and labor away from the protected sector (agriculture) to more productive uses. In the real world the question of eliminating agricultural protection must simultaneously address both equity and efficiency.
In theory, such programs can be Pareto-optimal while maximizing GDP. A Pareto optimal policy change is one in which no one is worse off as a result of the change. In the present context achieving such an outcome essentially requires tradeoffs. Those benefiting from the reduction in prices of agricultural commodities due to policy changes give up enough of their increased income to farmers to fully compensate the latter for the loss of income resulting from the policy change. Therefore, replacing price supports with a lump-sum income transfer to farmers can be an attractive policy option. Ideally, producers receive about the same income as before; the treasury is no worse off because it does not spend any more than it did before; and the economy as a whole is better off because resources are allocated more productively. As Gardner (1990:190) puts it: “The existence of deadweight losses from commodity market intervention implies that losers should be able to compensate the gainers a bribe that exceeds their surplus gains, while the losers are better off paying the bribe than enduring the intervention. The maximum size of the net gain is the deadweight-loss triangle. The reasoning, based on the compensation principles mentioned earlier, suggests lump-sum transfers as a policy reform that provides a Pareto improvement.” The appendix provides a geometric version of this statement.

As a practical matter, whatever the aggregate net impact of the change from price to income support, these programs will also redistribute income within the farm community. Table 1 illustrates this point. The table is based on the Mexican proposal of 1994. In that year, Mexico’s price supports on maize averaged about $59 a ton. Average annual national yields per hectare ranged from 1.60 tons (1989) to 2.35 tons (1992). The benefits to farmers were directly related to yields and the amount of land in maize (see the top panel of table 1). The second panel shows the benefits that would have resulted from the proposed plan to provide income support payments of $113 a hectare, up to a ceiling of 100 hectares. The “indifference” yield (that is, the point at which producers would be indifferent between price or income support) was 1.90 tons a hectare, which was also the median yield for 1989–92. The payment at the indifference yield was designed to compensate exactly the loss from eliminating price support. The big losers under the proposal would be high-yield farmers cultivating more than 100 hectares. Subsistence farmers would be the (relative) big winners. To the degree that they consumed the maize they produced, they received no payments under the price support regime. Table 1 shows this effect by assuming that the first hectare of production is not sold but consumed.

Reforming Agricultural Support

The European Union, Mexico, and the United States have all taken recent steps to reform their agricultural price subsidies. How have these policy changes affected equity and efficiency, and what lessons do these reforms have for developing countries?
Table 1. Hypothetical Transfers to Producers under Price and Income Support  
(dollars)

<table>
<thead>
<tr>
<th>Hectares</th>
<th>1.60 tons/hectare</th>
<th>1.90 tons/hectare</th>
<th>2.35 tons/hectare</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assuming price support of $59/ton with a yield of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>4,758</td>
<td>5,650</td>
<td>6,988</td>
</tr>
<tr>
<td>100</td>
<td>9,516</td>
<td>11,300</td>
<td>13,976</td>
</tr>
<tr>
<td>115</td>
<td>10,943</td>
<td>12,995</td>
<td>16,073</td>
</tr>
<tr>
<td>200</td>
<td>19,032</td>
<td>22,600</td>
<td>27,953</td>
</tr>
<tr>
<td></td>
<td>Assuming income support of $113/hectare with a yield of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>113</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>50</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
</tr>
<tr>
<td>100</td>
<td>11,300</td>
<td>11,300</td>
<td>11,300</td>
</tr>
<tr>
<td>115</td>
<td>11,300</td>
<td>11,300</td>
<td>11,300</td>
</tr>
<tr>
<td>200</td>
<td>11,300</td>
<td>11,300</td>
<td>11,300</td>
</tr>
<tr>
<td></td>
<td>Gain or loss as a result of an income support program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>113</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>50</td>
<td>892</td>
<td>0</td>
<td>-1,338</td>
</tr>
<tr>
<td>100</td>
<td>1,784</td>
<td>0</td>
<td>-2,676</td>
</tr>
<tr>
<td>115</td>
<td>357</td>
<td>-1,695</td>
<td>-4,773</td>
</tr>
<tr>
<td>200</td>
<td>-7,732</td>
<td>-11,300</td>
<td>-16,653</td>
</tr>
<tr>
<td></td>
<td>Per hectare gain or loss as a result of an income support program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>113</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>50</td>
<td>18</td>
<td>0</td>
<td>-27</td>
</tr>
<tr>
<td>100</td>
<td>18</td>
<td>0</td>
<td>-27</td>
</tr>
<tr>
<td>115</td>
<td>3</td>
<td>-15</td>
<td>-42</td>
</tr>
<tr>
<td>200</td>
<td>-39</td>
<td>-57</td>
<td>-83</td>
</tr>
</tbody>
</table>

Note: It is assumed that production from the first hectare is consumed by the households. The data resemble the program in Mexico for maize; 100 hectares is the limit of the payment; $59/ton reflects the 1989 price support of maize; $113 was the originally projected per hectare PROCAMPO (Programa Nacional de Modernización del Campo) payment (before the 1994 devaluation); 1.60, 1.90, and 2.35 tons/hectare were the average national maize yields for 1989, the median of 1989–92, and 1991, respectively. At a yield of 1.90 tons a hectare, producers are indifferent between price and direct-income support. This can be seen in the lower panel where producers up to 100 hectares register neither gains nor losses.

Source: Authors’ calculations; see text.

European Union: The 1992 CAP Reform

The costs to the European Union (EU) of the Common Agricultural Policy are of three kinds: Consumers pay far more for food than they need to; the deadweight losses are high—perhaps as much as 3 percent of the EU’s gross domestic product; and budget outlays are high. In recent years the Common Agricultural Policy cost the EU nearly 1 percent of member GDP. These budget costs have become politically
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>To compensate producers for a reduction in support prices</td>
<td>To compensate producers for the elimination of deficiency payments</td>
<td>To compensate producers for the elimination of deficiency payments</td>
</tr>
<tr>
<td>Payment basis</td>
<td>Average acreage in support crops during 1989-91</td>
<td>Average acreage in support crops during 1991-93</td>
<td>Acreage for which deficiency payments were received in any of the past 5 years</td>
</tr>
<tr>
<td>Supported products</td>
<td>Wheat, maize, barley, rye, oats, rapeseed, sunflower, soybeans, dried pulses, beans, tobacco, beef, lamb</td>
<td>Wheat, maize, sorghum, barley, rice, cotton, beans, soybeans, safflower</td>
<td>Wheat, maize, sorghum, barley, rice, cotton, oats</td>
</tr>
<tr>
<td>Time profile</td>
<td>Fixed in nominal terms: no expiration date</td>
<td>Total of 15 years: first 10 years fixed in real terms; declining in final 5 years</td>
<td>Program lapses after 10 years unless extended</td>
</tr>
<tr>
<td>Payment limits</td>
<td>None</td>
<td>$6,700 per farm</td>
<td>$40,000 per farm</td>
</tr>
<tr>
<td>Restrictions on the use of support-crop land</td>
<td>Land should be allocated to support crops; large producers must put into fallow a predetermined level of support-crop land</td>
<td>Land should be allocated to support crops but since 1996 land can be allocated to other agricultural uses</td>
<td>Land should be kept in agricultural uses (excluding fruits and vegetables); use must be in compliance with existing conservation plans</td>
</tr>
<tr>
<td>Other features</td>
<td>Support prices continue for cereals at lower level</td>
<td>“Negotiated” prices in effect for the first 2 transition years of the program; floor prices are retained for maize and beans</td>
<td>Nonrecourse government guaranteed commodity loans are retained in modified form</td>
</tr>
</tbody>
</table>

Note: The upper limit for PROCAMPO payments is 100 hectares and the per hectare payment is currently 484 Mexican pesos, or approximately US$6,700 (at 7.2 pesos/US$). Following the 1994 devaluation, PROCAMPO payments were not fully adjusted to inflation.

Source: Commission of the European Communities (1995); USDA (1996); SARH (no date).
intolerable and have been by far the strongest impetus for reform. In 1993 the EU reduced support prices on grains, oilseeds, and pulses and began to compensate producers by direct payments—based on past acreage in these crops—in conjunction with measures limiting the acreage for current production (table 2). Since that time, producer prices for these crops have declined by one-third.

**Mexico: The 1994 PROCAMPO Program**

Traditionally, CONASUPO (Compania Nacional de Subsistencias Populares), Mexico's agricultural parastatal, has been heavily involved in marketing, transporting, storing, and processing most of the country's agricultural commodities. Maize, beans, and wheat, by far the most important agricultural commodities, have been heavily subsidized through a system of guaranteed producer prices. The government also set panterritorial and pan-seasonal prices, which were usually announced before planting decisions were made. CONASUPO bought unlimited quantities at the guaranteed prices. Because producers knew in advance the price they were going to receive, they could shift production to those crops with the highest degree of relative protection rather than to crops that were the most profitable on the world market. The poorest peasants did not benefit from guaranteed prices because they hardly produced for the market.

In 1994 Mexico introduced a new farm program, PROCAMPO (Programa Nacional de Modernizacion del Campo), to provide income support to grain and oilseed producers—about 90 percent of all Mexican farmers (World Bank 1995; SARH no date). Under this regime prices of the nine crops in the program have become—in law at least—market-driven or autonomous. Thus production and trade should become less distorted. Moreover, PROCAMPO is distributionally more attractive than the earlier guaranteed price support because poor subsistence farmers are eligible for payments and there is a ceiling of 100 hectares on the amount of land that any single farmer can claim to justify payments (see table 2, box 1).

**The United States: The 1996 FAIR Act**

Before 1996 the U.S. government compensated farmers who participated in wheat, feed grains, rice, and cotton programs through a system of so-called deficiency payments. The payment rate was based on the difference between the target price (set by the government) and the higher of the market price or the price at which the government would value crops used as collateral for loans made by the Commodity Credit Corporation, a public entity. The total payment was equal to the payment rate, multiplied by a farm’s eligible payment acreage (usually the amount of land devoted to cultivation of the crop in question) and the program yield (which was established
Box 1. *A Closer Look at Mexico’s Program*

Mexico was not adequately prepared to implement its new income support scheme. Several policy implications can be derived from its experience.

First, the program was announced well in advance of the registration date for eligible producers. The lag allowed many farmers to increase the amount of land used for production of the eligible commodities and thus to increase their future payments. So rather than moving resources to more efficient uses, the scheme was manipulated, initially at least, to move resources into production that was already inefficient. Moreover, because land rights among landowners, tenants, and sharecroppers were unclear, it has been difficult to determine who was entitled to the payment.

Government credibility was also an issue. Initially, some producers did not believe that the government would actually implement the program. Fearing increased taxation, they underreported the amount of land allocated to eligible commodities. Furthermore, the fact that PROCAMPO initially delinked payments from the current use of land but later required that the land continue to be allocated to the eligible crops may have further discredited the government. (In 1996 the government increased the number of eligible crops.)

The macroeconomic environment also played an important role. Before the 1994 devaluation, most grains were highly protected through import restrictions. The devaluation sharply reduced protection relative to world prices. Although prices of tradable inputs also increased, it can be argued that to a large extent the devaluation substituted for PROCAMPO.

Market performance is also relevant. Mexico’s old system of guaranteed prices effectively precluded any kind of autonomous market. Thus markets are still underdeveloped and inefficient. Moving from guaranteed prices to a system of income support is more complicated in Mexico than in countries with a strong tradition of competitive commodity markets.

---

for the particular farm by the Department of Agriculture on the basis of historical yields).

On April 4, 1996, the Federal Agricultural Improvement and Reform Act (FAIR) became law, after the longest debate on a farm bill in U.S. congressional history (USDA 1996). FAIR removed the link between income support payments and farm prices by providing “production flexibility contract payments” for several crops. Participant producers receive these payments as a function of the amount of land registered for government support payments in earlier years. The payments are independent of current production, and farmers therefore have a more flexible incentive structure regarding planting decisions. The payments are fixed annually at a declining rate and under current legislation will end after seven years (see table 2). The practice of leaving land idle, required as part of the previous support programs, has been eliminated and producers are now free to plant any crops on the former contract acres except fruits and vegetables. The result is that producers depend more heavily on the market and also bear greater risk from increased price variability.

---

Assessing the Reform Programs for Efficiency and Equity

On efficiency grounds the three reform programs are improvements. All three are similar in delinking support from current production decisions and in moving toward market-driven prices. All three promote allocative efficiency in that quantity produced and price received are—in varying degrees—Independent of the amount of support. Thus, the three income support schemes are among the less distortionary mechanisms for maintaining agricultural incomes above their market outcomes.

All three programs, however, are less than ideal in that the use of land is not delinked from the program. This requirement probably reflects political considerations, as the payments must be seen to be going to “true” farmers. CAP reform requires either that land remains in production of the crops eligible for support or that producers set aside (hold out of production) a predetermined amount of support-crops land. PROCAMPO also holds land in agricultural production but permits a greater variety of crops to be cultivated. FAIR requires that land be kept in general agricultural use, but cannot be switched to fruits and vegetables, while producers must also comply with existing conservation plans. These restrictions obviously reduce allocative efficiency. In developing countries, where landownership records are poor, they may also pose implementation problems because monitoring compliance with cultivation restrictions would be difficult, especially where the average farm size is small. Monitoring such restrictions may also lead to corruption.

Do these reforms promote equity? Because the payments under the new programs depend on the amount of land used for past agricultural production, large farmers will receive the largest payments (although a limit on the total paid to any single farmer will reduce this effect). In industrial countries high or no payment ceilings have been more or less an acceptable practice, because most farmers have been perceived to have lower standards of living than other citizens. This perception is clearly the case in a number of OECD countries, where one of the most commonly declared objectives of agricultural policy has been a “satisfactory and equitable standard of living for farmers” (Winters 1989–90:241). This perception is probably decreasing. In the EU the belief is gaining ground that the “vast bulk of CAP money goes to farmers, many of whom are well off.” (Financial Times 1997). In many developing countries (especially in South Asia and Latin America), a few people with relatively high incomes own most of the land; many farmers own little or no land. In these cases, where land is distributed very unequally, unlimited payments will normally be a less appropriate criterion.

But such programs may promote equity in those developing countries with a strong correlation between poverty and degree of subsistence production. Under such a program, poor subsistence farmers with land are better off because they can consume
the previously subsidized commodities and receive cash payments at the same time (see first row of table 1). Mexican farmers who own less than two hectares of land receive more than 8 percent of PROCAMPO payments, although they have historically marketed very little and therefore received little support under the old program because the price guarantees applied only to the traded portion of the commodity (Deininger and Heinegg 1995).

**Fiscal Costs**

The fiscal costs of income support programs can be problematic. Thus far, two of the programs have been more costly than the programs they replaced (at prices prevailing at the time of implementation). The cost of the U.S. program in 1996–97 was estimated at about $5.5 billion, as opposed to $4.2 billion in 1994–95 for deficiency payments (USDA 1996). In Mexico the introduction of PROCAMPO almost doubled the transfers to the agricultural sector—from 6.4 billion new pesos in 1993 to 11.7 billion new pesos in 1994 (SARH no date). Moreover, if the programs are expensive initially, their fiscal cost is expected to remain high because payouts are independent of world prices.

The costs of the new program compared with the old price support program depend on actual world prices. Under high world prices the transfers to farmers under price supports would be low—or would even cease. For example, in the United States, because of the 1996 boom in grain prices, support based on deficiency payments would have been $1 billion–$2 billion—a fraction of the $5.5 billion paid under the FAIR act. Under low world prices, price subsidies would be higher, making the new program a bargain.

**Price and Other Risks**

Replacing stable support prices or guaranteed prices with direct income transfers exposes producers to the risk of volatile market prices. The shock from the exposure to risk would be more prevalent in countries where the government assumed all price risks by maintaining panterritorial and panseasonal prices than in countries with other types of protection, such as subsidies per output unit, floor prices, import tariffs, and possibly nontariff barriers.

Short-term price volatility can be alleviated with devices that mitigate market-based risk, either through private initiative or with public assistance. Forward and futures markets are effective tools that can offer both price discovery and hedging not only to producers, but also to merchants or other concerned parties. Participation in existing futures markets by foreign producers and merchants, however, appears to be limited, especially in cases where the costs of producing and marketing the commodity are predominantly in domestic currency (one
would have to hedge both the commodity as well as the exchange rate). Several developing countries are attempting to create their own futures exchange markets, however.

Government-assisted risk-mitigation devices are another option. In 1997 the U.S. Department of Agriculture introduced revenue insurance against both crop failure and falling prices. Initiated on a pilot basis in the states of Nebraska and Iowa, these revenue insurance policies have sold briskly (Economist 1997). Similarly, Mexico offers a guaranteed minimum price to cotton farmers for a predetermined fee through a government organization, ASERCA. The minimum price is based on the New York cotton futures exchange. ASERCA offers the guaranteed price in U.S. dollars and hedges the risk by using the predetermined fees to purchase put options on the exchange for future delivery after the harvest. Should prices fall, ASERCA pays the farmer the difference between the prevailing New York futures and the guarantee price. If prices rise instead, ASERCA makes no payment (Varangis and Larson 1996). Because the entire transaction is denominated in U.S. dollars, cotton producers assume the exchange risk. It should be emphasized, however, that private and government-assisted risk-mitigation devices deal with short-term price volatility, that is, one crop cycle at most.

Generally speaking, there are many ways to reduce risk in addition to formal measures. Farmers can grow a variety of crops with different market and climatic risks. Contract farming at fixed output prices eliminates price risk. An off-farm job for a member of the household is also a possibility.

**Negative Supply Response**

In theory, price and trade reforms can be Pareto improvements. They allow farmers to maintain their incomes at roughly the same levels while inducing growth in the economy through better resource allocation. Nevertheless, the removal of protectionist policies is likely to reduce agricultural output because lower crop prices reduce investment and accelerate the exit from agriculture. This has been the case in the former Soviet economies, frequently dramatically so. In Hungary, for example, market liberalization after 1989 reduced agricultural output by more than one-third and induced the exit from agriculture of one-half the country's farmers.

On balance, the reforms undertaken by the Mexican government in the late 1980s and early 1990s (including the implementation of PROCAMPO) appear to have reduced the profitability of the crop sector (Baffes 1998; World Bank 1996). In the European Union, if recently suggested reforms to the CAP cut preferential assistance to oilseed producers, their income may drop by as much as 20 percent (Oils & Fats International 1997).

Moreover, contraction of the sector inevitably reduces the demand for farm labor, so the income of landless farm workers is likely to decline (at least in the short run), although they may benefit to the degree that the prices of food decline. The removal
of guaranteed prices in Mexico was initially expected to induce a long-term reduction of 6 to 24 percent in the agricultural labor force and a corresponding reduction in grain production of between 11 and 28 percent. (The estimates for the Mexican agricultural sector are from Baffes 1998; Levy and van Wijnbergen 1994; and Burfisher, Robinson, and Thierfelder 1992.) In response to the FAIR Act, U.S. farmers are already shifting land out of maize to soybeans, anticipating changes in the price structures of the two crops.

It is important to be aware of these consequences early on. Otherwise, when the reforms fail to induce a positive supply response or to increase the demand for labor, at least in the short run, the confusion and resistance that is generated could jeopardize the reform process.

Toward a Successful Transition

Although all three programs move in the right direction, the considerations above suggest that they could have been designed and implemented more effectively. For one thing, the programs could have been more comprehensive. In addition reform programs should not restrict land uses, they should not cost more than the subsidization programs they replace, and they should be transitional.

Making Reforms Comprehensive

To realize the full benefits of an income support system, the programs should include all crops and substitute for all existing price support programs. FAIR does not apply to sugar, tobacco, peanuts, and milk, all of which are heavily protected in ways that seriously misallocate resources. And PROCAMPO is restricted to nine commodities; price floors are still in operation for the two basic crops (maize and beans), and several subsidy programs, especially on inputs, are still in place. Although CAP’s reform covers most crops and some livestock, it fails to eliminate price supports and retains many quantitative restrictions.

Lifting Restrictions on Land Use

An income support program should not impose restrictions on the use of land. With the exception of environmental considerations, the main justification for restricting land use is to ensure that program payments go only to bona fide farmers. Yet one important reason for replacing subsidies with income support is to encourage individual producers to use resources as relative prices and comparative advantage dictate. Another is that future payments under the income support scheme could be
used as credit collateral or sold outright as an asset, facilitating investment in profitable activities. However, because it is unlikely that lending institutions would have the capacity (and authority) to verify program participation and program compliance, their value as collateral is limited.

Finally, enforcement of restrictions on land-use entails careful, and in some countries relatively expensive, administrative measures. In Mexico the majority of agricultural extension workers employed by PROCAMPO are engaged in monitoring the enforcement of restrictions on land use.¹

**Keeping the Costs Down**

If they are to be politically feasible, income support payments must be designed so that they do not exceed the costs of the programs they replace. For this reason, governments should link payments inversely to world agricultural prices: when prices are high, producers receive lower income support payments and vice versa. Otherwise if world prices rise, producers not only receive these high prices but in addition are subsidized by the program. This imposes an additional burden on taxpayers who must pay high food prices and are also burdened with the program's costs.

In addition, rather than providing a uniform level of per hectare support, policymakers can set up a declining index with a ceiling. For example, the first 10 hectares receive full support, say $100 a hectare, the next 10 hectares receive less support ($50 a hectare), and so on, effectively increasing the relative support to small producers. If the program is announced before land registration is completed, however, farmers are likely to divide large holdings among family members, in effect neutralizing the effect of a declining index. Similarly, corporate farms can convert to holding companies for a large number of small farms. For the same reasons, such moral hazard also arises with programs that impose ceilings on benefits, such as PROCAMPO and FAIR (see table 2).

Nevertheless limits on the maximum acreage eligible for income support or simple benefit ceilings as in FAIR are equitable devices for containing fiscal costs. Such limits reduce support to larger producers and may promote more equal income distribution. Table 1 shows that such a ceiling can generate substantial savings by reducing the benefits to above-average-yield large producers.

**Limiting the Duration of the Program**

In view of the uncertainties, the programs should be transitional and not permanent. Although PROCAMPO is scheduled to be phased out in 15 years, at which time world prices will prevail, FAIR's language leaves open the question of support when the program expires after 7 years. Thus far, the CAP is not time-bound, although it is
likely to be subject to another round of reforms. Nevertheless, if the final objective is—as it should be—to eliminate support to producers, the programs should wind down to an explicit expiration date.

It is important to distinguish between the short run and the steady state. Consider the Pareto-improvement element of such programs. Under the assumption that price support programs would have continued indefinitely, an explicit time profile for the income support system implies that such programs are not necessarily Pareto improvements. This is the case because producers who previously were receiving support may be worse off after the termination of the income program than they would have been had price subsidization continued.

Establishing Supporting Institutions

The shortcomings discussed above apply to both industrial and developing countries. But weak supporting institutions are likely to be a problem in developing countries. To ensure that payments to producers are fair (that is, based on the amount of land farmed) and are paid in a timely manner, a national land registry must be developed before initiating the program—something that is extremely difficult to achieve in many countries, for example, those in tropical Africa. The government must also have policy credibility if producers are to react as desired. Credibility was a problem in Mexico, where the amount of land in crops was first under-reported in many areas (due to fear of government taxation), and then over-reported. As of 1993, the final determination of eligible areas had not yet occurred. In many instances, eligible areas were “negotiated” at the community level (Salinger, Metzel, and Arndt 1995).

Clearly the macroeconomic environment, and particularly the exchange rate, should be adequate and stable. In some cases eliminating currency overvaluation may make it possible to eliminate protection without fiscal compensation. If substantial devaluation is likely, a feasible approach to eliminating price supports may be to make explicit the benefits to farmers that result from devaluation and include elimination of agricultural price support as part of the macroeconomic reform package. Finally, efficient and integrated local commodity markets are needed to ensure a smooth transition.

Another set of problems stems from uncertain land tenure rights. Who receives the payment may be an issue, particularly in developing countries. Consider a case where a tenant has cultivated the land for the entire period on which the payments are based. If support is based on land previously allocated to supported crops, the landowner will claim the payments. On equally valid grounds, the tenant may claim the payments because it was his or her cultivation that qualifies the land for support. Resolving this issue will often complicate the implementation process. (The United States has developed an elaborate legal definition of “producer”—including land-
owner, share cropper, or rent-paying tenant—that stipulates the conditions of eligibility for receiving benefits under FAIR.)

In the absence of well-defined land tenure rights, appropriate preparation of a transition program may require an extensive survey of the rural economy to identify the distribution of farms operated by owners and those that are not. The survey will indicate the technical feasibility of the program but may also provide information on variables such as the distribution of farm size, yields, and commodities produced. Such information may be needed to design equitable and affordable programs that are politically acceptable; it will also indicate whether direct income support is an appropriate route to reform. These considerations suggest that in many—perhaps most—developing countries, informal land tenure arrangements may preclude subsidies based on land ownership mechanisms. As Binswanger and Deininger (1997:1966) have recently argued: "The costs of maintaining records, negotiating, contracting, and policing property rights can be high and may exceed the value of the land especially in rural areas with low population densities and little market access."

Concluding Remarks

In designing income support programs, it is well to keep in mind, first, that separating the payments from current production decisions implies consequences that are likely to be perceived as negative in many instances, even though economically rational outcomes are fully expected. Producers will frequently face higher risk from increased price variability. Because the ratio of output to input prices will be lower, agricultural output will decline for the crops in question, which in turn may reduce the demand for agricultural labor. As is the case with other types of support, the politically active large producers who will probably receive the lion's share of the support may not be those most in need of support during the transition period. The latter problem has been well recognized and is one of the complicating factors behind attempts to further reform the CAP.

Second, it is unlikely that the conditions and requirements discussed earlier will be fully met, particularly in the developing countries that protect agriculture. Decisionmakers need to be aware of these requirements to ensure that the programs are feasible, quite apart from the question of dealing successfully with the political forces that are involved.

Reforms, therefore, have to be considered within a broad policy perspective. Compensation could, for example, take the form of government-financed investments to benefit producers, such as increased expenditures for rural infrastructure. Most developing countries have an inadequate network of rural roads, inefficient water supply and irrigation networks, and limited electric power. Rural social ser-
VICES ARE ALSO NEGLECTED; EDUCATION AND HEALTH CARE IN RURAL AREAS ARE INFERIOR TO THOSE PROVIDED IN CITIES. SUPPLYING CREDIT THROUGH GROUP-BASED LENDING TO MICRO-ENTERPRISES IN HIGH POPULATION AREAS MAY BE ANOTHER APPROPRIATE INSTRUMENT (BINSWANGER AND LANDELL-MILLS 1995). IN JAPAN, FOR EXAMPLE, THE URUGUAY ROUND AGREEMENT CALLED FOR PERMITTING IMPORTS OF FOREIGN RICE Equal TO 4 PERCENT TO 8 PERCENT OF DOMESTIC CONSUMPTION. ANTICIPATING THAT THIS MARKET OPENING WOULD LEAD TO LOWER DOMESTIC PRICES FOR RICE AND A DECLINE IN FARMERS' INCOME, THE DIET ENACTED A LAW THAT APPROPRIATED MORE THAN 6 TRILLION YEN (ABOUT $60 BILLION) FOR COMPENSATORY MEASURES TO INCREASE FARMERS' INCOMES INDIRECTLY THROUGH IMPROVEMENT IN INFRASTRUCTURE IN RURAL AREAS, ENHANCEMENT OF AGRICULTURAL TECHNOLOGY, ACCESS TO CREDIT, AND SO ON (GOTO 1997).

TO CONCLUDE, IT IS IMPORTANT TO RECALL THAT A DIRECT INCOME SUPPORT PROGRAM IS INTENDED TO PROVIDE A TRANSITION FROM PRICE-DISTORTING SUBSIDIES TO A LIBERALIZED SECTOR THAT ALLOWS RESOURCES TO BE ALLOCATED MORE EFFICIENTLY. IT IS NOT A POVERTY REDUCTION PROGRAM, ALTHOUGH IT CAN INCREASE THE INCOME OF SUBSISTENCE LANDHOLDERS. IT IS NOT AN INVESTMENT PROGRAM BECAUSE IT HAS NO PROVISIONS ON HOW THE MONEY IS TO BE SPENT. IT IS NOT DESIGNED TO INDUCE AGRICULTURAL GROWTH BECAUSE IT LOWERS PRODUCER PRICES. FINALLY, BECAUSE THE PROGRAM IS LINKED TO AN ASSET—LAND—THE LION'S SHARE OF THE PAYMENTS MAY GO TO LARGE PRODUCERS.

APPENDIX A. INCOME REDISTRIBUTION AND EFFICIENCY LOSSES UNDER AN IMPORT TARIFF

Let $D_d$, $S_d$, $P_m$, $S_m$, $Q_0$, $Q_1$, $Q_2$, and $Q_3$ denote domestic demand and supply of the commodity in question, while $P_m$, $S_m$ denotes the world price, which can also be viewed as the perfectly elastic supply of imports (figure A.1). In a closed economy the market clears at quantity and price dictated by the intersection of domestic demand and supply. If the economy is open, price is $P_m'$ and the supply for the commodity becomes $S_d' S_m$. Domestic producers supply $Q_1$, while the remaining $(Q_1 Q_0)$ is imported. This is the competitive outcome.

If the government introduces an import tariff $t$, the effective supply schedule becomes $S_d' S_m$. At $P_1$ (the new price), consumers demand is $Q_2$. Domestic producers supply $Q_3$, while the remaining $(Q_3 Q_2)$ is supplied by imports. An import tariff has several effects. First, consumers pay a higher price (from $P_m$ to $P_1$, and consequently demand less (from $Q_0$ to $Q_2$). Second, domestic producers receive a higher price and hence produce more (from $Q_1$ to $Q_3$). Third, the government receives tariff revenue (the area $BGFC$). To summarize, consumers lose while producers and government gain.

The losses to the three groups, however, exceed the gains. The monetary burden on consumers (reduction in consumer surplus) due to the tariff is the area of $P_m' P_1 FC$.
of which $BGFC$ (the tariff revenue) goes to the government; $P_m P_t G A$ (the change in producer surplus or change in profits) goes to the producers; the triangular area $ABG$ (the excess cost of producing the additional output $Q_1 Q_3$) is the efficiency loss, that is, the additional resources that the country uses to produce domestically the commodity that would have been saved, had the corresponding amount of the commodity been imported. Sometimes this area is termed deadweight loss, or Harberger triangle. Finally, the triangular area $CFD$ denotes an additional welfare loss to consumers, or reduction in consumer surplus that is not offset by government revenue from the tariff. Total deadweight losses equal the sum of the two triangles.

After the introduction of the tariff, the relative size of producer gains and consumer losses depends on demand and supply elasticities and the level of the tariff. If the after-tariff price exceeds the price that prevails if the economy is closed, then imports are not taking place. In this case, the government receives no revenue, while the "transfers" take the form of increased producer profits and increased consumer welfare losses.
For agricultural producers the import-tariff setting is analogous to producer price support with no production controls. A similar outcome arises if, instead of taxing imports, the country limits the imported quantity to, say, $Q_3 - Q_2$. The importers then receive a rent equivalent to the tariff revenue net of the cost of obtaining an import license. Under certain assumptions, importers competing for the import licenses would pay the same in the aggregate for licenses as the tariff forgone.

The fundamental idea behind a Pareto-optimal income support program that compensates for tariff removal goes as follows. After eliminating the tariff, income is redistributed as lump-sum transfers. Producers plus importers (consumers) supply (demand) $OQ_0$ at price $P_m$ as in the no tariff scenario. Producers receive as compensation for their losses the lump-sum transfer $P_mP_tGA$ plus part of the two triangular areas and therefore are better off; the government receives $BGFC$ plus part of the triangular areas and therefore is better off; consumers pay $P_2P_tFC$ and part of the two triangular areas but retain the remaining part of the two triangular areas and therefore are better off. Thus, from an efficiency point of view, a lump-sum transfer can be a Pareto improvement.

Going from price to income support, therefore, can be a win-win move. Even if political forces aim to achieve some such outcome, the “transaction costs” involved in moving to the desired outcome may be high. First, consider the difficulties in measuring areas such as $P,nPPGA$ and $ABG$. Such measurement requires knowledge of the yield as a proxy for land-rent, that is, producer surplus, of each producer. It is difficult for governments to acquire such knowledge. Therefore the program must be designed according to an “average yield,” which, as table 1 indicates, will overcompensate some producers and undercompensate others. An additional difficulty results from the choice of the base period. Because not all producers allocate the same amount of land to a particular crop each period, different periods will affect producers differently.

Second, there are the implications of the vertical linkages in the particular market. Reducing the price from $P_t$ to $P_m$ implies less output and therefore less use of inputs. Because labor is in many instances a major input, the welfare of farm labor is likely to change, independently of what happens to landowners.

Third, there are horizontal linkages. The forces of complementarity and substitutability among products will shift resources from one product market to another (depending on relative prices and the structure of the respective technologies).

Finally, reforms frequently involve altering several policies simultaneously (such as price support, input subsidy, tariffs, and quantitative restrictions). The calculation of the associated benefits and losses by recipient group requires data that practically speaking are virtually impossible to collect.

In summary, the move from price to income support ideally must take into consideration all the difficulties outlined above if the policy change is to be implemented with both increased efficiency and Pareto-optimality. Pareto-optimality as a goal is
helpful in structuring and evaluating proposed reforms. Its achievement is another story.

Notes

John Baffes is with the Development Economics Research Group at the World Bank, and Jacob Meerman is an economist in the Operations Evaluation Department at the World Bank. The authors would like to thank Adolfo Brizzi, Louise Cord, Bruce Gardner, Junichi Goto, Tassos Haniotis, and Alberto Valdés. The article has also greatly benefited from a discussion by participants at a World Bank workshop on February 4, 1997.

1. Land tenure is one of the most difficult aspects of Mexican farm policy. In addition to private farms, there are quasi-communal farms, the ejidos. Created under the Constitution of 1917, ejidos guaranteed all Mexicans land rights through expropriations of large landholdings. But ejido members' rights over land and water use were restricted. Sale or rental of ejido land was prohibited; members could not hire wage workers, and they could not be absent from their farm for more than two years without losing their rights. The ejido system gave members little control in their choice of inputs and outputs (Heath 1990). In 1992 a reform in the Mexican Constitution gave ejidatarios the right to rent and sell land to outsiders with the approval of a majority of ejido members. Land can now be pledged as a collateral. The titling process has been slow, however. By early 1995, only 20 percent of ejidos had been given land titles (De Janvry and others 1995).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.


John Baffes and Jacob Meerman 209


SARH (Secretaria de Agricultura and Recursos Hidraulicos). No date. PROCAMPo: Vamos al Grano para Progresar. Mexico City.


The Potential and Limitations of Self-Targeted Food Subsidies

Harold Alderman • Kathy Lindert

Can self-selection of subsidized commodities be used as a mechanism to transfer income to the poor? Evidence from two self-targeting programs, one in South Africa and one in Tunisia, shows that although self-targeting can clearly improve the distribution of food subsidies to the poorest members of society, its power to alleviate poverty and reduce income disparities is limited by preference patterns, income inequality, and the size of the individual subsidies. Self-targeting through quality and product differentiation can be a useful means to reform existing universal subsidy schemes, but it should be considered a transitional tool while the capacity for implementing more precise mechanisms is developed.

Self-targeting—subsidizing only those commodities the target group has indicated it prefers—has been put in place as a way to reform generalized food subsidies and limit participation in these programs to the intended beneficiaries. By choosing to subsidize commodities consumed primarily by the poor, governments have found that they can improve the targeting—and reduce the costs—of food transfers. Self-targeting is attractive because it sidesteps the difficulty of determining how much income people have. By targeting the behavior of the poor, which they reveal, policymakers can avoid having to screen individuals on the basis of their incomes, which they generally do not divulge.

One of the first examples of self-targeting reported in the literature was an experiment in Bangladesh in which sorghum flour was used as an alternative to wheat in ration shops (Karim, Majid, and Levinson 1980). This successful pilot program proved to be better targeted to the poor population than wheat flour (which was itself self-targeted) because rice is the preferred grain in that country (Ahmed 1988).

Self-targeting can also be achieved by quality differentiation. Tunisia subsidizes foods in several product groups (cereals, cooking oil, sugar, milk) according to the consumption preferences of the rich and the poor for various items in each group (Lindert 1995b; Tuck and Lindert 1996). Both Morocco and Egypt have reduced the...
costs of their extensive subsidy programs by shifting subsidies toward a dark, rough flour that is consumed disproportionately by the poor and shunned by the rich (Lindert 1995a; Ali and Adams 1996). Similarly, yellow maize has been distributed in drought and war relief efforts in Sub-Saharan countries such as Mozambique, where white maize is the preferred staple of the rich (Dorosh, del Ninno, and Sahn 1996).

Although these experiences have been examined in case studies, the literature does not take a broad view of the experience with self-targeting. This article looks at the conceptual, empirical, and practical limitations of self-targeting to determine whether such programs successfully transfer income to the poor and whether they improve the distribution of food subsidies to the intended beneficiaries.

The Self-Targeting Mechanism

Targeted transfer programs that benefit the poor are a core element of the World Bank’s three-pronged strategy to reduce poverty, along with broad-based economic growth and human capital development (World Bank 1990). Such transfers aim to reach a target population among the chronic poor who remain outside the economic growth process and to protect the vulnerable during periods of economic change. The rationale for targeting these transfers is based on the premise that, whereas the costs of safety-net programs are essentially the same for all beneficiaries, the social returns for a given level of benefits are higher for the poor than for the wealthy. Targeting can improve program efficiency and save resources by concentrating expenditures on those who need them the most. It often entails other costs, however, including the administrative costs of screening potential beneficiaries to identify those who qualify (particularly for means-tested transfers), possible economic losses due to disincentive effects, and the potential loss of political support from those who are excluded because they are better-off (Grosh 1994).

Targeting is generally implemented through three types of mechanisms: an individual income (or needs) assessment, geographic and other group indicators, and self-selection. Self-selection, which has been touted as an administratively simple way to direct transfers to those in need, occurs when the benefits of a transfer scheme are ostensibly made available to all consumers, but the program is specifically designed so that the nonpoor elect not to take up these transfers.

Self-targeting is commonly used in food transfer programs, in which product and quality differentiation discourage consumption of the subsidized product by wealthier consumers. Most consumption-based transfers inherently involve a rudimentary degree of self-targeting, because food represents a larger share of the total expenditures of the poor than of the rich. But generalized food subsidies still typically transfer higher absolute benefits to the rich, who purchase larger overall quantities of food. Food transfer programs can be better targeted by selecting certain foodstuffs to carry...
higher subsidies than others. Subsidizing foods that are unattractive to the nontarget group can help prevent consumers who are better-off from trying to capture the benefits (Nichols and Zeckhauser 1982). Those commodities that are consumed less as income rises are excellent candidates for such subsidies.

Public employment programs also rely on self-targeting by establishing time requirements and wage rates that make participation unattractive to higher-income individuals (Ravallion 1991). Such programs, however, are based on a different criterion for self-selection. In the case of subsidized goods, the price of a commodity is generally the same for all consumers, but the demand differs according to income. Self-selection for public works programs does not necessarily require a different demand for leisure among the well-off; wages offer the potential for screening because the price of leisure for this group (their wage opportunity) will generally be higher.

**Advantages and Disadvantages**

Self-targeting has several advantages. First, the information needed to implement self-targeted plans is relatively less cumbersome than that required by other targeting mechanisms. The choice of which commodities to subsidize is generally based on survey data on household behavior (consumption patterns), which is less costly to collect than assessments of individual income or income proxies, both of which are subject to substantial inaccuracies. As a result self-selection avoids the problem of asymmetric information regarding income levels (Besley and Kanbur 1988; Besley and Coate 1991; Lindert 1995b).

Second, self-selection mechanisms have the flexibility to respond to changing economic conditions. Although indicator targeting rarely responds to idiosyncratic (individual) fluctuations in income, individuals can shift their commodity choice rapidly in response to changing circumstances. Similarly, more people may choose a subsidized commodity when prices of higher-quality grades of that food rise. In Bangladesh, for example, purchases of subsidized food from the ration shop were sensitive to the price of rice on the open market (Montgomery 1985). This phenomenon indicates a substantial cross-price response that can enhance the advantage of self-targeting lower grades of a commodity. Because the poor are more sensitive than the rich to changes in food prices, self-targeting is enhanced when prices of nonsubsidized food rise (Timmer and Alderman 1979). Self-targeted programs also may be more easily phased out as incomes climb because beneficiaries voluntarily opt out of the program once they can afford higher-quality foods. This appears to be the case in Pakistan, where demand for rationed flour (which was of an inferior quality) decreased over time, thereby improving targeting (Alderman 1988).

Third, self-targeting may be less vulnerable to bureaucratic corruption and manipulation than other targeting mechanisms (such as means-testing, which requires maintaining a beneficiary roster). Fourth, self-targeting may be less divisive and more
politically acceptable than individual assessment and geographic targeting mechanisms because the decision to participate is made by the individual rather than by the bureaucracy.\(^1\)

Finally, governments also may opt to subsidize commodities that have a higher nutritional content than commodities that are close substitutes. For example, a subsidy on whole wheat flour may shift consumption from more refined flours, with benefits for the long-term health of the population that are in addition to the poverty alleviation objectives of the subsidy.

Self-targeting is not without disadvantages, however. One important drawback to all targeting mechanisms is that some of the very poor may be screened out of the program along with the nonpoor (Subbarao and others 1997). For example, subsidized food products may not be available in poor remote rural areas where subsistence and barter replace cash-based market purchases of food. Experience has also shown that targeting can be imprecise, resulting in large leakages to the nonpoor. This leakage might increase over time if tastes shift toward previously unfamiliar subsidized foods (to subsidized yellow maize, say, in countries where white maize is the norm, or wheat where diets are based on rice). In addition, when lower-quality varieties of commonly consumed food are subsidized, such as high-extraction flour (with some but not all of the bran removed), it can be difficult to disentangle the subsidies on these items from other products made from similar raw materials. This is the case in Morocco, where the subsidy on a high-extraction-rate flour that is well-targeted to the poor is injected at upstream levels (meaning closest to the source) on the marketing chain to simplify the payments process (fewer agents and transactions). This self-targeted subsidy has suffered leakages because a fine, white flour preferred by the rich (farine de luxe) is made from the same raw material as the targeted commodity (Lindert 1995a).

On occasion, self-targeting is achieved by making the process by which a commodity is acquired the means for self-selection. For example, subsidized goods may be available only to those individuals willing to wait in lines or be stigmatized as poor (as when items are available only in state ration shops). This is somewhat different from targeting a commodity that is unlikely to be selected by relatively well-off consumers. In addition, the administrative and other costs of these types of programs result in a deadweight loss to the economy because there is no gain in welfare from this use of resources (Alderman 1987; Ranney and Kushman 1987).

Inherent Limitations to Self-Targeting: Inequality and Demand Parameters

The effectiveness of self-targeting depends on the distribution of income as well as on the commodities being subsidized. A stylized demonstration of these effects on a
country with high inequality and one with low inequality with illustrative elasticities representing different commodities is shown in table 1.²

**Elasticity Effects**

The share of government subsidy expenditures that goes to the poor clearly improves with lower expenditure elasticities, regardless of income distribution (elasticity refers to the change in consumption as a result of a change in income). Table 1 shows that substantial equity gains arise from simply reallocating subsidies within the range of positive expenditure elasticities—such as shifting from a commodity with an elasticity of 1.0 to a good with an elasticity of 0.3 (in both the high- and low-inequality countries). The less demand for a good increases with income, the greater the share of subsidies that will go to the poor. Thus, shifting subsidies from commodities such as dairy products or meat (generally with high income elasticities) to commodities like sugar or grains (generally with lower elasticities) could result in greater targeting of subsidies on the poor. Further gains in targeting transfers to the poor could be achieved by identifying commodities with negative expenditure elasticities, as shown in table 1.

Table 1 also demonstrates the limits on the redistributive power of self-targeted food subsidies. Over a plausible range of income elasticities, it is unlikely that the share of a subsidy going to the poorest two quintiles could be increased much beyond two-thirds of the total transfers (or that accruing to the top three deciles being reduced to less than one-third of the total) by self-selection alone. One illustration of

<table>
<thead>
<tr>
<th>Table 1. Determinants of Self-Targeting of Food Subsidies: Stylized Inequality and Demand Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Share of subsidies received</strong></td>
</tr>
<tr>
<td><strong>High inequality</strong></td>
</tr>
<tr>
<td>Share to poorest 10%</td>
</tr>
<tr>
<td>Share to poorest 20%</td>
</tr>
<tr>
<td>Share to poorest 40%</td>
</tr>
<tr>
<td><strong>Low inequality</strong></td>
</tr>
<tr>
<td>Share to poorest 10%</td>
</tr>
<tr>
<td>Share to poorest 20%</td>
</tr>
<tr>
<td>Share to poorest 40%</td>
</tr>
</tbody>
</table>

*Source: Authors' calculations. Decile rankings are by households. The high-inequality example is based on the 1993 South Africa Living Standards and Development Survey (with a Gini Index of 58.22). The low-inequality example is based on the 1996 Albania Employment and Welfare Survey (with a Gini Index of 27.61).*

---

_Harold Alderman and Kathy Lindert_
plausible ranges of income elasticities for subsidized gradation of commodities comes from Pakistan. In that country a subsidy on an inferior quality of flour with an elasticity of -0.13 achieved real—but modest—fiscal savings over the amount that would have been spent if all wheat flour—with an elasticity of 0.34—was subsidized (Alderman 1988).

Moreover, even when inferior goods exist, the budget shares associated with these items limits the extent to which income is transferred to the poor: these products are often consumed in such a small quantity even by the poor that the implicit income transfer embodied in the subsidy is negligible.

**Inequality Effects**

Although income elasticities indicate how much the purchase of a commodity increases or declines with a change of income, they are not sufficient to indicate the potential of commodity targeting within a country because they do not indicate the degree to which incomes differ. In general, the share of a self-targeted subsidy that goes to the poor tends to increase with higher degrees of inequality. The larger the percentage difference between the incomes of the poor and the nonpoor, the greater the difference in consumption implied by a given negative income elasticity. In the high-inequality country in table 1, shifting subsidies from a luxury good to an inferior good with an elasticity of -0.3 increases the share of the transfer accruing to the poorest by a factor of 10. The difference is only three-fold in the low-inequality country. In both examples, the impact of commodity choice is proportionally lower when the target population is a larger share of the total. The disparity across rows in table 1 is greater for all three target groups when inequality is greater. Hence, the lower the inequality, the lower the scope for self-targeting of commodity subsidies.

**Case Studies: Empirical Impact and Limitations**

Two applications of self-targeting, one in South Africa and one in Tunisia, indicate the potential and limitations of the mechanism to transfer income to the poor. Both programs are fairly well-targeted relative to other examples of self-selection found in the literature (Alderman 1991; Grosh 1994).

**South Africa's Experience with Self-Targeted VAT Exemptions for Food Items**

A prevalent school of thought on fiscal policy suggests that value added taxes (VAT) should be based on efficiency criteria alone. This view holds that equity concerns can be better addressed with targeted income transfers and similar measures. The
VAT, however, often reflects the government's distributorial and fiscal objectives, and for this reason is adjusted to reduce the relative burden on low-income consumers.

The introduction of the VAT in South Africa in 1991 straddled these two viewpoints. Because it was clear that the tax burden would affect the ability of the poorest groups to afford an adequate diet, the government launched a safety net initiative, the National Nutrition and Social Development Programme, aimed at distributing 400 million rand (more than $100 million) annually in community-based food security projects. In addition, to keep their costs down, maize and brown bread were exempted from the VAT shortly after its introduction. By mid-1993, 19 food commodities were exempt, and roughly the same number of additional exemptions had been proposed, including several "luxury" foods, such as meat and dairy products (Alderman and del Ninno 1997).

The fiscal impact of these exemptions is uneven (table 2). The revenue loss from the exemption on maize is similar to the revenues forgone from the exemption on fresh milk and to the combined revenue loss associated with exemptions on brown bread and white bread (for which an exemption has been proposed). The revenue that would be lost from the proposed exemption on meat, however, is roughly equivalent to that for the three other commodities taken together.

REDISTRIBUTIVE POWER OF SELF-TARGETED TAX EXEMPTIONS. The analysis of the distributional impact of VAT exemptions is conceptually similar to the study of the impact of price subsidies. Standard methodologies have been devised to indicate the efficiency of such taxes (in terms of minimizing economic distortions for a given amount of revenue) as well as the equity impacts (Ahmad and Stern 1991; Deaton and Grimard 1992). Applying this methodology to the 1993 South Africa Living Standards and Development Survey (LSDS) shows that the effects of VAT exemptions differ appreciably across commodities (see table 2). 3

Maize and kerosene are clearly the preferred commodities for self-targeted price subsidies (or VAT exemptions). The LSDS defines the poor as the poorest 40 percent of households, or 52.8 percent of the population. Thus, only the existing exemption on maize and the proposed exemption on kerosene deliver a share of tax relief to the poor that is greater than their share in the population. 4 The products with the highest leakage of benefits to the nonpoor are meat, milk, vegetable oil, and other dairy products.

Figure 1 confirms that the highest income groups consume less maize than poorer people. Consumption actually increases up to the poverty line, and then declines. This curvature, however, is not captured in the demand estimation employed nor by many common alternative functional forms. Because the figure only shows consumption by households that use kerosene, it masks the fact that the poor consume most of that fuel. As incomes rise, households tend to shift to other sources for fuel.
Table 2. Distribution of VAT Exemptions in South Africa, 1993

<table>
<thead>
<tr>
<th>Commodity</th>
<th>VAT status</th>
<th>Fiscal revenue loss (millions of rand)</th>
<th>Income elasticity</th>
<th>Budget share (millions of rand)</th>
<th>Value of savings for poor (millions of rand)</th>
<th>Percentage of transfers to poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>Exempt</td>
<td>666.9</td>
<td>0.31</td>
<td>0.00</td>
<td>0.02</td>
<td>6.91</td>
</tr>
<tr>
<td>Fresh milk</td>
<td>Exempt</td>
<td>621.9</td>
<td>1.23</td>
<td>0.78</td>
<td>1.01</td>
<td>2.54</td>
</tr>
<tr>
<td>Beans</td>
<td>Exempt</td>
<td>102.3</td>
<td>0.90</td>
<td>0.47</td>
<td>0.70</td>
<td>0.78</td>
</tr>
<tr>
<td>Vegetable oil</td>
<td>Exempt</td>
<td>179.7</td>
<td>0.63</td>
<td>0.45</td>
<td>0.55</td>
<td>1.20</td>
</tr>
<tr>
<td>Meat</td>
<td>Under consideration</td>
<td>1,807.0</td>
<td>1.29</td>
<td>0.84</td>
<td>1.09</td>
<td>8.27</td>
</tr>
<tr>
<td>Bread</td>
<td>Brown = exempt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(brown, white)</td>
<td>White = under consideration</td>
<td>664.1</td>
<td>0.99</td>
<td>0.50</td>
<td>0.80</td>
<td>4.32</td>
</tr>
<tr>
<td>Other dairy</td>
<td>Under consideration</td>
<td>444.7</td>
<td>1.23</td>
<td>0.78</td>
<td>1.01</td>
<td>2.01</td>
</tr>
<tr>
<td>Sugar</td>
<td>Under consideration</td>
<td>260.1</td>
<td>0.44</td>
<td>0.29</td>
<td>0.28</td>
<td>2.42</td>
</tr>
<tr>
<td>Kerosene</td>
<td>Under consideration</td>
<td>152.4</td>
<td>0.31</td>
<td>-0.40</td>
<td>0.11</td>
<td>1.77</td>
</tr>
</tbody>
</table>

Source: Alderman and del Ninno (1997).
Figure 1. South Africa: Food Expenditures by Income Group

Note: Income is proxied by log per capita expenditures; food expenditures are shown in log form.
Source: Authors' calculations.
for cooking and heating. The bread patterns (which include both white and brown varieties) reflect another common problem with survey data—the inability to distinguish commodities by quality. In this particular case, the approach used by Deaton and Grimard (1992), which uses unit prices to distinguish quality, is not applicable because prices were collected mainly at the community level, and these show that prices of brown and white bread overlap.

South Africa might seem a fairly unique illustration because poverty there has a very strong race-based dimension. As consumption patterns also differ by race, it may appear that the self-targeting parallels race-based targeting—a targeting scheme that, although theoretically possible, is often politically unacceptable. Indeed, the constitution of the Republic of South Africa forbids such targeting. Self-targeting, however, delivers the majority of the VAT exemptions that accrue to the African population to the poor among this group. For example, two-thirds of all maize consumed by Africans is consumed by those Africans in the poorest 40 percent of the total population.

SIZE OF TRANSFER TO THE POOR. Although the tax exemption on maize (and possibly kerosene) appears to be well targeted to the poor, this subsidy has a rather limited ability to transfer income to the poor. Taking the 12 percent of the budget for this commodity for the rural population—most of whom are poor—and multiplying it by the size of the tax (14 percent), one finds that the magnitude of the total transfer in terms of a reduction in the cost of living is a mere 1.7 percent of total expenditures of the poor. Adding the transfer from the subsidy on kerosene, which is an even more minor share of the budgets of the poor, only brings the figure up to 2 percent. Because there are few, if any, other commodities for which a tax exemption would even be neutral in distributional incidence (let alone well-targeted to the poor), it appears that VAT exemptions can have only a modest impact on poverty.

Clearly, alternative price regimes using self-targeting—for example, eliminating the regressive VAT exemptions on fresh milk and vegetable oil and using the revenues for direct subsidies on maize and kerosene—could boost the share of income supported by self-targeted subsidies. Still, such subsidies would be too small to benefit the recipients as much as other targeted transfers in South Africa. For example, although the means- and age-tested income transfer embodied in the old-age pension is targeted as efficiently as the VAT (66.2 percent goes to the poor), the size of these benefits dwarfs the magnitude of income support generated by food subsidies (Case and Deaton, forthcoming). Thus, while the judicious choice of commodities for VAT exemptions can achieve a reasonable degree of targeting, these interventions are still secondary tools in programs designed to provide a safety net for the poor.
Tunisia's Experience with Self-Targeted Food Subsidies

In Tunisia the government moved to a self-targeting framework when severe political and administrative constraints frustrated its efforts to cut existing food subsidies (Lindert 1995b; Tuck and Lindert 1996). In the face of a structural adjustment program in the mid-1980s, fiscal pressures dictated cuts in the extensive generalized food subsidy program that had served for several decades as the primary vehicle for transferring income to the poor. The scheme had indeed become quite costly: by the mid-1980s subsidy outlays had ballooned to more than 4 percent of gross domestic product (GDP) and to 10 percent of total government spending (table 3). It was also inefficient: in 1985 the value of the benefits to rich households was twice that of the benefits transferred to poor households.

Initial attempts to reduce the program in the mid-1980s were met with riots, which forced the government to reinstate the subsidies. Policymakers attempted to develop alternative direct transfer schemes to be targeted using individual assessment

---

### Table 3. Self-Targeted Food Subsidies in Tunisia

<table>
<thead>
<tr>
<th>Year</th>
<th>Product program description</th>
<th>Transfers</th>
<th>Value of transfer to poor as a percentage of total expenditures</th>
<th>Total transfer cost as share of GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>To poorest quintile</td>
<td>To richest quintile</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>Prereform: universal subsidy program</td>
<td>8&lt;sup&gt;a&lt;/sup&gt; 17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15 4</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>Prereform: universal subsidy program</td>
<td>17 20</td>
<td>9 3</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>Self-targeted reform program&lt;sup&gt;c&lt;/sup&gt;</td>
<td>25 14 6 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subsidies on self-targeted products only&lt;sup&gt;d&lt;/sup&gt;</td>
<td>21 18 8 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All food subsidies&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>Simulated results of further self-targeting reforms&lt;sup&gt;f&lt;/sup&gt;</td>
<td>27 13 4 0.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Data from 1985 Institut National de la Statistique (INS) Household Expenditure Survey; poorest expenditure group = 13 percent of population; richest expenditure group = 12 percent of population.

<sup>b</sup> Data from 1990 INS Household Expenditure Survey.

<sup>c</sup> Data from 1993 INS Household Expenditure Survey.

<sup>d</sup> Semolina, large-loaf bread, bulk generic cooking oil, pasteurized reconstituted milk in less-convenient packages.

<sup>e</sup> Self-targeted products listed above plus a number of other items.

<sup>f</sup> Reducing or eliminating subsidies on poorly targeted items that remained subsidized under the 1993 program using elasticities estimated from an almost ideal demand system.

mechanisms in hopes that these could eventually replace the generalized subsidy scheme. High administrative costs, information constraints, and implementation difficulties plagued these programs, however, and leakages to the nonpoor were substantial. Means-tested food stamps were also rejected because of these same administrative constraints and on the grounds that they would be politically unacceptable because they might conjure up images of “wartime” ration cards. The government also explored the possibility of targeting transfers geographically, but this option was abandoned because distinctions between neighborhoods were too obscure in most areas to make geographic targeting effective.

Political, economic, and administrative constraints favored reforms that could be carried out within the existing framework of food price subsidies rather than those that would require an entirely new institutional structure. These concerns led the government to self-target the subsidies to the poor using quality differentiation. One aspect of these reforms involved liberalizing markets to allow private operators to market higher-quality food products to attract wealthier consumers. The government also shifted subsidies to narrowly defined items within a product line that are perceived by consumers to be of a lower quality because they possess certain unattractive features in their packaging or ingredients. Although the intrinsic nutritional value of these products has been preserved, the perceived “inferior” characteristics of these items discourage their consumption by wealthier households.

Self-targeting through quality differentiation was applied to all major subsidized food categories. Subsidies were shifted to semolina, which is disproportionately consumed by the poor, and reduced for pasta and couscous, which do not display self-targeted consumption patterns. Subsidies on baguettes, which were consumed virtually exclusively by the rich, were eliminated, while those on a better-targeted larger loaf of bread were maintained. The bread market was also liberalized to allow private bakers to produce high-quality bread to attract the wealthy customers. Cooking oil subsidies were applied to a generic product purchased from bulk oil drums. Subsidies were 40 percent higher on less-refined brown sugar than on refined white sugar, but all consumers—rich and poor—rejected the brown sugar because they perceived it as “dirty.” Finally, milk subsidies were shifted to reconstituted milk packaged in less convenient half-liter cartons, making it less desirable to the rich, who tend to purchase local fresh milk in bottles and in tetrabriks, a type of carton designed for long storage life.

Tunisia’s self-targeting efforts have proved to be both effective and politically feasible. Reforms have been economically sound: outlays on the subsidy program were cut to 2 percent of GDP in 1993 and 6 percent of total government expenditures (table 3). Moreover, simulations suggest that these reforms would have reduced subsidy outlays even further—to 1.9 percent of GDP and 5.1 percent of public expenditures—if all other factors including GDP, total government expenditures, and population size had remained at 1990 levels.
REDISTRIBUTIVE POWER OF SELF-TARGETED SUBSIDIES. Reforms have also been effective in improving the distribution of the subsidies. Although demand elasticity analysis failed to uncover true “inferior” goods in Tunisian consumption patterns, the rich did not switch to the subsidized products when subsidies were eliminated on higher quality items. Rather, they substituted previously unavailable unsubsidized luxury varieties when they were introduced as part of the reform program. As a result, reforms shifted the subsidy scheme from one that transferred more absolute benefits to the rich than to the poor to one in which the reverse was true. By 1993 the poor benefited 1.2 times more from the program than did the rich (table 3).

Simulations show that additional self-targeting reforms could further reduce costs and improve the coverage of the program. By eliminating all remaining subsidies on products that are not consumed disproportionately by the poor, the cost of the program could be further reduced to 0.8 percent of GDP and to just over 2 percent of total government expenditures (holding all other factors constant). Assuming that the rich would not shift consumption to the remaining subsidized products, the poor would benefit over two times more than the rich. Self-selection through the choice of commodity is thus an effective instrument to improve the target efficiency of universal subsidies.

It is important to note, however, that the efficiency of self-targeted subsidies in Tunisia, while better than the universal program, is not as sharp as that seen in other transfer schemes. A means-tested food stamp program in Jamaica transferred roughly 15 times more benefits to the poorest quintile than to the richest, as did a geographically targeted food supplementation scheme in Peru (Grosh 1994).

SIZE OF TRANSFER TO THE POOR. The amount of income transferred to poor Tunisians through subsidized commodities is limited, as in South Africa, because total purchases of these products are small and the unit subsidies are slight. In Tunisia subsidies on all food products accounted for 8 percent of the budgets of the poor in 1993; of which those specifically targeted to poor consumers accounted for 6 percentage points. In contrast, the VAT exemptions on maize and kerosene in South Africa amount to less than 2 percent of the budgets of poor households. The list of subsidized products, however, still includes large-loaf bread, more of which is consumed by the upper-middle quintiles (third and fourth) than by the poorest 40 percent (although the poorest 20 percent does consume 20 percent more than the richest). If one considers only the transfers from commodities in which the poor consumed more than the average, that is, those that were truly self-targeted to the poor, the program accounted for only 4 percent of the budgets of the poor in 1993. The Tunisia example, therefore, illustrates some of the tradeoffs between finding commodities that are targeted mainly to the poor and finding vehicles to deliver an income transfer.

Harold Alderman and Kathy Lindert
Conclusions

How effective is self-targeting in reducing income disparities and alleviating poverty? Self-targeting is clearly preferable to indiscriminate universal food subsidies. A stylized example of demand parameters and income distributions reveals that self-targeting is more effective when subsidies are focused on products with low or negative expenditure elasticities in countries with higher inequality. In South Africa, for example, a careful selection of the products to be subsidized or exempted from the VAT would clearly improve the efficiency of these transfers. In Tunisia as well, self-targeting reforms had the intended results of reducing the exorbitant costs of universal food subsidies and improving distribution to the poor.

Tunisia’s experience is particularly germane in that self-selection was central to the subsidy reform, but similar savings have been noted in a number of countries. For example, Egypt reduced food subsidies from 19.5 percent of total government expenditures in 1981–82 to 5.3 percent in 1993–94. Although this reduction was fostered by a 10 percent decrease in the eligible population and a partial convergence of domestic and international prices, most of the cost savings were achieved by focusing on commodities that had negative or negligible income elasticities and by eliminating subsidies on items with significantly higher elasticities (Ali and Adams 1996). Thus, subsidies on sugar and coarse flour—and the bread made from it—which together accounted for slightly more than half of the expenditures on subsidies in 1981–82, made up more than 80 percent of the subsidy bill in 1993–94. In Morocco quality differentiation measures, which involved shifting flour subsidies to a coarser grade, reduced the cost of flour subsidies from 1.8 percent of GDP in the mid-1980s to 0.34 percent by 1995 and greatly improved coverage of the targeted population.

Nevertheless, the effectiveness of self-targeted commodity subsidies is limited. Even under “optimal” circumstances (high inequality and low or negative expenditure elasticities), at least one-third of total transfers can be expected to leak to the nonpoor. This result was observed in the case of self-targeted maize subsidies in South Africa and Tunisia. Moreover, self-targeted food subsidies have only a limited impact on poverty alleviation. Our study is consistent with Sah’s (1983) analysis of the limits of redistribution through indirect taxes. Although this investigation does not presume that all revenues are raised by indirect taxes, the amount of income transferred to the poor is constrained by the importance of subsidized products as well as by the size of the subsidies. Self-targeting can play a useful role in reforming existing universal food subsidy programs, until governments develop the administrative capacity to effectively deliver means-tested direct transfer programs to alleviate poverty.
Notes

Harold Alderman is principal economist in the World Bank's Development Economics Research Group. Kathy Lindert is an economist in the Human Development Department of the World Bank's Latin America and Caribbean Regional Office.

1. This is not inherent. There are political costs as well as advantages to subsidizing inferior goods. This may explain why many pilot self-targeted programs have had short lives. The Bangladesh sorghum experiment was discontinued for a variety of reasons after a year of operation. Similarly, Pakistan replaced its self-targeted ration shops with a generalized price subsidy. In Tunisia, the self-targeting of lower-quality school notebooks was abandoned because the stigma associated with visibly separating the rich from the poor, particularly for a durable item consumed by children, was deemed politically unacceptable shortly after the notebooks were introduced.

2. This illustration uses data from South Africa, with a Gini Index of 58.22, to illustrate a high-inequality country, and Albania, with a Gini Index of 27.61, as an example of a low-inequality country. The examples are stylized, with per capita consumption determined by a constant elasticity with respect to per capita expenditure. When this elasticity is zero, per capita consumption is constant. Nevertheless, the poorest deciles of households (ranked in terms of expenditure per adult equivalents) receive more than 10 percent of the subsidy since these deciles have more individuals. Clearly, ranking by decile of individuals would give a neutral pattern.

3. The data set is available to the public on the following site: www.worldbank.org/html/prdph/lsms/country/za94/za94home.html#top.

4. However, it is noteworthy that this distribution pattern would not be expected using income elasticities generated from an almost ideal demand system. A common rule of thumb suggests that if a commodity has an income elasticity of zero, the share of a subsidy or tax exemption to the poor would be the same as its population share, yet the pooled elasticities for both kerosene and maize are slightly greater than zero. Of course, it is not necessary to estimate an income elasticity to indicate the share of benefits accruing to a population group—although price elasticities are needed for a complete analysis of tax efficiency.

5. The general consumer subsidy program covered a wide range of products, including semolinas, couscous, pasta, bread, flour, cooking oil, sugar, and milk. Subsidies were universal and available to anyone who chose to purchase the subsidized products in any quantity desired. The cost of the subsidies was borne entirely by the government (taxpayers), as the transfers covered the gap between producer prices and the artificially low consumer prices.

6. Direct Assistance schemes include the Needy Families Program (cash transfers) and the Union Tunisienne de Solidarité Sociale, which is responsible for low-income food rations and cash transfers for the elderly and handicapped (World Bank 1993).

7. Expenditure elasticities for the well-targeted items estimated using an Almost Ideal Demand System, while not negative, were low, ranging from 0.17 to 0.37 (Lindert 1995b).

8. These benefits are comparable to Tunisia's main other income transfer program at the time, the Needy Families Program. The program suffers from the traditional difficulties associated with indicators targeting: complex administration, failure to update eligibility lists, lack of flexibility, and substantial errors of inclusion and exclusion (Tuck and Lindert 1996).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.


———. 1995b. “Quality Differentiation and the Self-Targeting of Food Subsidies in Developing Countries.” Ph.D. Dissertation, Department of Agricultural Economics, University of California at Davis. Processed.


Dominique van de Walle

Public spending programs aimed at alleviating poverty can either be broadly targeted at categories of spending or narrowly targeted at types of people. Each approach has benefits and costs to the poor. It is often claimed that narrow targeting of the poor will allow governments to reduce poverty more effectively and at lower cost. But narrow targeting often has hidden costs, and once these costs are considered, the most finely targeted policy may not have any more effect on poverty than a broadly targeted one. Both approaches also have hidden benefits, although less is known about their impact. Targeting can help, but it is not a cure-all. Reducing poverty calls for broadly targeted social sector spending combined with narrower targeting of cash and in-kind transfers to specific groups. It is also important for governments to experiment with schemes that offer better incentives, to carefully monitor the costs and outcomes, and to be flexible and pragmatic in their policy responses.

How can public spending assist those who have been left behind by economic growth? It may take a long time for some subgroups in society to catch up. Some, such as the elderly and disabled, may never achieve that goal. Others may well be hurt in the short run by policy reforms that ultimately would benefit the poor. There may be related concerns about regionally unbalanced growth. How can public spending help deal with vulnerability? Incomes can be highly variable over time, particularly in poor rural economies where the population is vulnerable to risks induced by uncertain weather, relative price shifts, or the collapse of community-level support systems during crises. How can public spending meet these objectives adequately with limited resources?

The answer often given to all three questions is “targeting.” Those who have been left behind, or who are vulnerable to risk, can (it is argued) be reached most cost effectively by concentrating limited public resources on narrowly defined “target” groups within society.

In practice, there are two approaches to raising the living standards of the poor with public spending. The first involves targeting types of spending and can be called broad targeting. Under this approach no attempt is made to reach the poor directly as indi-
viduals. Instead, gains are achieved by targeting types of spending that are relatively more important to the poor. Spending on basic social services, such as primary education and primary health care, is one example. Rural development is another. The second approach entails targeting categories of people. Under this approach, which can be called narrow targeting, benefits are intended to be targeted directly to the poor. Such policies include food stamp schemes targeted to poor mothers, innovative microcredit schemes aimed at rural landless women, and development programs that focus on poor geographical areas. Within each of these categories are differences in how much the program relies on administrative targeting and how much it depends on self-selection, based on behavioral responses to the incentives built into the program. Each approach has benefits and costs to the poor and to others.

This article reviews the case for targeting and the relevant recent evidence for developing countries. Much has already been written on this topic. A good starting point is Besley and Kanbur (1993), who cautioned against taking too simplistic a view of targeting and identified the potential costs associated with such efforts. Almost a decade later, it is still common for empirical work and policy discussions to disregard the fact that targeting is almost never costless. But research has progressed on these issues, and the results from careful studies offer some guidance on cost-benefit assessments.

The article revisits the issues surrounding targeted programs in light of the new research and also points to continuing deficiencies in our knowledge. Currently available data and research methods often make it difficult to assess the full costs and benefits of targeting, which are often hidden. Costs and benefits are frequently not properly accounted for in policy assessments and hence are ignored in policy decisions. Researchers are getting better at measuring some factors, such as how utilization of targeted services varies by income group, but other aspects, such as the behavioral responses of participants, administrators, and governments, are still being neglected. People routinely underestimate how difficult it is to target the poor. This point has important implications not only for policies directly aimed at poverty alleviation, but also in areas such as macroeconomic policy. For example, it is often argued that targeted interventions can be used to compensate losers from macroeconomic reforms, yet little consideration is given to how this is to be done. Once the difficulties in practice are acknowledged, it is easy to see that more attention needs to be paid to the timing and sequencing of macroeconomic operations to take into account the government's inability to compensate losers in the short run.

**Broad Targeting**

Although narrow targeting has received the bulk of attention and is the main focus of this paper, the allocation of budgets among categories of public spending can also
be a means of targeting the poor. By looking first at broad targeting, one can also better understand some of the motivation for narrow targeting.

The basic principle underlying broad targeting is that some categories of public spending matter more to the poor than do other categories. For example, spending on basic social services is found to often benefit the poor. Money spent on primary education, for example, is likely to reach more poor children than money spent on secondary or tertiary education, because many poor students will have dropped out of the higher levels, unable or unwilling to afford the opportunity and other costs of continued schooling. Figures 1 and 2 show the distribution across expenditure groups of public spending for health and education in Indonesia, and for education in Côte d'Ivoire. Individuals are ranked by household per capita expenditures and divided into population quintiles or deciles. The utilization of sector services—school attendance at different levels or visits to types of health care facilities—are tabulated for each group and multiplied by the unit costs of providing each specific service to get the total spending implicitly directed to each group. These amounts are expressed on a per capita basis and plotted against the expenditure groups.

In a pattern that is frequently repeated across developing countries, spending on primary education tends to favor the poor, with per capita amounts declining as living standards rise, while spending on higher education, especially tertiary education, benefits primarily the richest group. Similarly, in the health sector, hospital expenditures confer the highest transfers to the top deciles, while spending on health care centers that dispense basic care is much more neutral across expenditure groups. Such evidence has lent credence to the case for broad targeting of public spending to basic social services.

Several additional factors strengthen the case. Better health and basic education, access to safe water, and basic physical infrastructure raise poor people's well-being and may also raise their productivity and incomes. The nonpoor often already have enough of these services, so the added spending reaches the poor. In Malaysia, for example, Hammer, Nabi, and Cercone (1995) find evidence indicating that between 1985 and 1990, marginal expenditures on safe water supply went primarily to households in Malaysia's poorer states. This happened in part because households in the rich states already benefited from this service. It is often argued that basic services tend to be underprovided by the private sector in the absence of government intervention, especially to the poor. For example, poor parents may not be willing to pay the full costs of their children's schooling, particularly when (as is the case for girls in some societies) parents share little in the future benefits of that schooling. Finally, because it makes no socioeconomic distinctions between who can and who cannot participate, broadly targeted spending is perceived to be more politically popular than narrow targeting, and hence more sustainable.

Figures 1 and 2 make the case for broad targeting of public spending to basic social services. How compelling is that evidence? The figures rely on the "benefit
Figure 1. Subsidy Per Capita by Decile, Indonesia 1989

Source: van de Walle (1992, 1994).
incidence” approach to measuring the distributional impact of public spending. Although programs are often analyzed in this way, the method makes a number of strong assumptions that can affect the conclusions drawn and hence the policies advocated (van de Walle 1998). Leaving aside these issues, several other arguments have been made against broad targeting as an instrument for reducing poverty.

The most common criticism is that broad targeting can be expensive. Universal subsidies on basic services also reach many who can afford to pay, as illustrated in figures 1 and 2. Given this “leakage,” a broad targeting program can be a costly way to reduce poverty. Even when the incidence of poverty is high, there are likely to be differences in the severity of poverty, and society is likely to put a priority on helping the worst-off first.

Broad targeting may also be a clumsy instrument for reaching some of the poor. Some groups have specific needs that will not be served by the free provision of safe water or basic social services or other categories of spending that are suited to broad targeting. Similarly, broad targeting may be inefficient in meeting specific objectives. A central role of public spending is to protect the vulnerable and provide insurance for downturns and variability in living standards. In a drought or natural disaster, the provision of basic social services is unlikely to provide the necessary safety
net. Such concerns have prompted policy discussions to turn to narrow targeting as a means of achieving higher impacts on poverty at lower cost.

Narrow Targeting

Narrow targeting is defined as a deliberate attempt to concentrate benefits on poor people—whatever the type of spending. It has been a particularly popular policy proposal in the context of recent worldwide efforts to reduce budget deficits and public spending while still protecting the poor.

Types of Narrow Targeting

Narrowly targeted schemes are based on one of two principles—or a combination of both. The first is indicator targeting (also called categorical targeting). Such a strategy identifies a characteristic of poor people (an indicator) that is highly correlated with low income but that can be observed more easily and more cheaply than can income (Besley and Kanbur 1993). The indicator is then used as a proxy for income to identify and target poor people. A typical example would be the region of residence; geographical targeting has been popular with many governments and in World Bank projects. Alternatively, landholding class, gender, nutritional status, disability, or household size are often used to identify beneficiaries. In the former Soviet Union and East European countries, for example, household size is used to distribute family allowances. Bangladesh’s Grameen Bank combines correlates: microcredit is targeted to rural women from landless or near-landless households.

The second approach is self-targeting. Instead of relying on an administrator to choose participants, these schemes aim to have beneficiaries select themselves through creating incentives that will induce the poor and only the poor to participate. Self-targeting works by incorporating a cost of participation into the design of the scheme. This cost should rise as income rises while the benefits remain the same; thus, depending on where the benefit level is set, the nonpoor are effectively screened out. For example, public employment schemes use work requirements to help screen out the nonpoor; subsidy programs support items that the poor consume but the rich do not; and other controls rely on waiting time, stigma, and lower “packaging” quality of goods and services to dissuade usage by the well-off.

Alderman and Lindert discuss self-targeting through food subsidy schemes in detail in this volume. They describe Tunisia’s approach to converting a fiscally unsustainable universal food subsidy to a smaller, more pro-poor program through self-targeting (also see Tuck and Lindert 1996). Limited by the absence of any real difference in the consumption baskets of different income groups, the Tunisian government’s solution was to differentiate goods not by their intrinsic quality, but
through their outer packaging. "Inferior" goods—perceived by the rich to be of lower quality—were created, and the subsidies were transferred to them. For example, milk was packaged in small, flimsy plastic containers that had to be emptied once opened. Both the size and the inconvenience deterred purchases by the rich. The inconvenience of buying goods in these forms of packaging was the "cost of participation" in this case, and it tended to be higher for the nonpoor. This policy was reinforced by creating unsubsidized superior goods (milk in liter cartons with attractive labeling) and by liberalizing the sale of imports and other high-quality substitutes that appeal to the rich. The policy succeeded in significantly lowering program costs, avoiding political unrest, and concentrating program spending on lower income groups.

The cost of participation promotes self-selection and helps identify who is poor and who is not. It can also be designed to deter the poor from becoming dependent on the scheme (Besley and Coate 1992). The programs are designed to maximize income gains but minimize gains to other aspects of well-being. In this way, they seek to maintain incentives to move out of dependency and poverty. The workhouses for paupers in 19th century England were a classic example of the deterrence argument. The authorities made relief as uncomfortable as possible by requiring beneficiaries to enter the workhouses (Besley, Coate, and Guinnane 1996). One truly had to be desperate. Finally, self-targeted schemes have the added benefit that power is taken away from civil servants and hence incentives toward corruption and favoritism are reduced.

Both types of narrow targeting offer the hope of avoiding two commonly identified errors of targeting: a leakage of benefits to the nonpoor; and imperfect coverage of the poor (Grosh 1995; Cornia and Stewart 1995). Such errors of targeting often influence policy advice. For example, an unfavorable evaluation of the Food for Education program in Bangladesh (World Bank 1997) argued that the benefits reached only 15 percent of pupils and that the leakage was 14 to 26 percent (depending on the target group definition). Others, however, have argued that the low coverage was justified because those who were not covered were unlikely to need any incentive to keep their children in school. Moreover, the leakage to the nonpoor was not high compared to other antipoverty programs (BIDS 1997; Wodon 1998). By these broader criteria, the program looks like a worthy effort. The key point is not whether a scheme avoids errors of targeting, but how well it meets its stated objectives given budget constraints, the information that is available to policymakers, and the behavioral and political responses to targeted interventions.

What Are the Costs of Narrow Targeting?

Narrow targeting, like broad targeting, has costs. Three types of costs are associated with narrow targeting: administrative costs, costs that arise from incentive effects or
behavioral responses, and costs that result from the ramifications of political economy. The first is typically (although not always) easily measured, but the second and third types are often hidden and require closer scrutiny.

**Administrative Costs.** The first cost arises because policymakers have imperfect information, and obtaining that information entails costs. Identifying who is poor with precision and avoiding leakage by reaching only the poor often necessitates high administrative costs. This is a particularly serious problem in developing countries where most of the poor live in rural areas and policymakers do not have accurate information on their incomes. Incomes typically vary considerably over time, so even if reliable data are obtained, the frequent need to reevaluate the information can be expensive. Although targeting specific indicators holds the promise of lowering administrative costs, even the best indicators tend to be imperfectly correlated with low incomes. Combining a few indicators often improves targeting but may affect administrative costs as well (Baker and Grosh 1994).

Leakage and administrative costs can eat into a program’s budget. A means-tested rice subsidy scheme in India’s Andhra Pradesh State is illustrative. Radhakrishna and others (1996) found that for every rupee transferred to a poor person, the program spent 6.4 rupees. Of the additional 5.4 rupees, 3.6 rupees reached the nonpoor and 1.8 rupees were spent on administration. The potential deadweight costs to the poor attributable to incentive effects have not been studied, but if they exist, they would further cut into the amount transferred to the poor. The primary problem appears to be the substantial transfers to the nonpoor. One can surmise, however, that measures to avoid leakage might well result in higher administrative costs rather than greater coverage of the poor.

**Incentive Effects.** Some beneficiaries and nonbeneficiaries may alter their behavior as a result of a policy. Self-targeted schemes often require participants to change their behavior. In the case of indicator targeting, the potential recipients of public spending face incentives to misrepresent their incomes and change their behavior either to become or simply to remain eligible. It can be difficult to find an indicator that can be identified accurately and at little cost, is strongly correlated with income, and cannot easily be manipulated. So, for example, people may change their residence to gain coverage in a plan that is geographically targeted. Programs directed at female-headed households have been blamed for encouraging the breakdown of nuclear family households. Targeting benefits to the unemployed may dampen their efforts to find work.

The incentive problem can be severe for those near an income cutoff point for eligibility. Earn a little less and you qualify for a transfer; earn a little more and you will be disqualified. Indeed, if the loss of transfer is larger than the income gain, then
there will be no incentive to escape poverty. This “poverty trap” arises when target-
ing creates a 100 percent marginal tax rate on the poor.

A few studies have attempted to measure the costs associated with behavioral re-
sponses to policies aimed at raising the incomes of the poor. One study of Sri Lanka’s
targeted food stamp program offers an interesting example in examining whether house-
holds that received food stamps changed their work effort. The targeted scheme was
introduced in 1979 to replace a long-standing general food subsidy. Sahn and Alder-
man (1995) found that workers in recipient households, both men and women, re-
duced the number of hours they worked. On average, rural workers cut back almost
three days a month, effectively reducing the average monthly net household income
gain to 59 rupees, from 91 rupees. This may not be a bad outcome. The recipients’
well-being has assuredly increased. Leisure is valued. Policymakers whose objectives
include welfare more broadly conceived than income alone will be pleased. But if the
policy objective is to maximize income gains, the outcome may disappoint.

Behavioral responses may occur on the part of nonbeneficiaries too. The benefits
of public support can be shared between direct recipients and the donors of private
transfers. For example, a study of South Africa finds that each rand of public pension
support provided to the parents of migrant workers decreased remittances to the
parent from the migrant by at least 0.2 to 0.4 rand (Jensen 1998). This is not
considered a deadweight loss as long as positive social value is attached to the im-
plicit gains to migrant children.

Like indicator targeting, self-selection can entail hidden costs arising out of changes
in behavior induced by participation costs. The latter costs help focus benefits on the
poor, but they also reduce the net transfer benefits, although this fact is often ig-
nored. For example, a review of cost-effectiveness calculations for Bangladesh’s tar-
geted food-grain distribution programs concludes that “programs that have work
requirements or other obligations (such as sending children to school) have proved
to be much more cost-effective than food rations” (Subbarao and others 1997:49).
The claim is based on the fact that “obligations” greatly improve targeting efficiency
and that the underlying cost calculations account only for administrative and food
transfer costs. This conclusion could be quite deceptive because the costs to the poor
of those obligations have been omitted. Forgone incomes could be non-negligible.
The opportunity cost of children’s time and other schooling costs might be large.
The rationale for programs requiring children to attend school in exchange for trans-
fers is precisely that poor parents cannot afford to send their children to school. The
exchange will be acceptable only if the household achieves a net gain. Although the
benefit for participants is positive, that gain is likely to be overestimated by conven-
tional cost-effectiveness calculations.

For example, one study evaluates the value of transfer benefits in a Jamaican school-
feeding program designed to self-select beneficiaries through the nature of the food
provided (Jacoby 1997). Targeting performance is assessed to be high, but so are the costs. Factoring in behavioral responses as well as administrative costs, it costs 2.60 Jamaican dollars to achieve a gain of 1 Jamaican dollar per child.

**Political Economy.** The middle classes are often the primary beneficiaries of public social spending. The poor are largely left out, while the rich have alternatives in the private sector at home or abroad. Spending that is narrowly targeted toward the poor will tend to be associated with a contraction of benefits to the middle class, which is often the government’s most vocal and politically important constituency and which may then cease to support poverty reduction efforts. Especially in situations where the prospective beneficiaries of targeted policies are marginalized, it is possible to end up spending less on the poor (Gelbach and Pritchett 1996). In Sri Lanka, for example, the nominal value of the food stamps was never changed between 1979 and 1989, and their real value was halved through inflation. Some have argued that the value of the food stamps was allowed to erode without resistance because beneficiaries of the program lacked political clout, while the more powerful middle class did not support the targeted scheme (Anand and Kanbur 1990).

The U.S. experience also provides evidence for this view and for the often heard adage “programs for the poor are poor programs.” The electorate supports costly universal programs such as social security, from which it expects to benefit, but not the targeted means-tested welfare programs such as Aid to Families with Dependent Children—a cash transfer program for poor single mothers and their children.

Clearly, many factors influence the political economy of poverty alleviation programs. Under certain sociopolitical environments, taxpayers may be more, not less, inclined to support programs that efficiently reach the deserving poor and to save resources for other publicly provided goods (Besley 1997). Information and public relations campaigns have helped to build societal support. A timely awareness campaign helped secure popular acceptance of Tunisia’s plan to reform a universal food subsidy to a more narrowly targeted scheme (Tuck and Lindert 1996; Alderman and Lindert in this volume). Polls showed that the campaign, which measured the costs of the old program in terms of the roads, hospital beds, and manufacturing jobs that could otherwise have been secured, successfully convinced a majority of the population that the change was necessary and desirable. Strategies to promote the participation and support of diverse groups appear to have helped the sustainability of pro-poor policies within social funds in Bolivia and Zambia (Subbarao and others 1997). Design specifics also matter. For example, the public is often more sympathetic to food and other in-kind transfers and favors work requirements in exchange for public assistance rather than cash payments with no strings attached. Such burdens to beneficiaries are perceived to help weed out the undeserving poor. The macroeconomic environment also mat-
ters. People who are economically content are more likely to back programs that benefit others. The political environment and its determinants are important factors that politicians and policymakers ignore at their peril.

What Are the Benefits of Narrow Targeting?

Common methods of assessment may also obscure some of the potential benefits of narrow targeting. Assessments of the benefits from geographical targeting provide an example. Several studies examine the potential impact on poverty of a predetermined budget optimally allocated across regions. The static gains are often found to be modest, in essence reflecting the fact that the poor are heterogeneous. Not all of them can be identified by the same indicators. For example, even though most countries have regions that are poorer than others, not all the poor live there, nor do all the rich live elsewhere. Hence, geographic targeting will often benefit some of the rich and will bypass—and perhaps even tax—some of the poor who live in the better-off areas (Datt and Ravallion 1993; Ravallion 1993).

Recent work, which allows for the potential dynamic effects of programs, suggests that static assessments may considerably underestimate the benefits. Gains may percolate through and strengthen over time through the positive external effects of development in the poor region on the productivity of private investments by the poor. Measuring such effects is difficult and requires data that are often unavailable. In a study assessing the effects over time of development programs that were geographically targeted to poor areas in China, Jalan and Ravallion (1998) found the expected imperfect coverage of the poor and leakage to the nonpoor. But they also found that the programs had a positive impact on growth rates in the targeted areas. In this case, the short-term gains to the poor were less than the long-term gains.

Similarly, static assessments have tended to underestimate the benefits of land-contingent targeting (meaning that the beneficiaries are selected solely on the basis of land ownership). Many of the poor in rural areas of South Asia are landless. Yet some poor rural households may have significant amounts of land, and more than some of the rural wealthy. In Bangladesh, Ravallion and Sen (1995) concluded that redistributing income from those with ample land to the landless had only a small effect on poverty even though landholding size is positively correlated with incomes. The authors found that the impact was somewhat higher once the higher productivity of the transfers to small—but often credit-constrained—landholders was considered.

Several other potential hidden benefits are still unresearched, including longer-term benefits of policies that succeed in changing the productive abilities of the poor or in increasing their ability to cope with risk and other areas where markets fail; indirect benefits from the assets created by public works schemes as well as benefits from potential impacts on employment and wages outside the schemes; and gains
from receiving public support rather than being dependent on rich relatives or patrons.

A Case Study of a Relief Work Program

In an effort to stave off the potentially disastrous consequences of a bad drought, India’s Maharashtra State introduced a public employment scheme in 1973. The program’s popularity and success in smoothing consumption for the vulnerable led to its permanent deployment as the employment guarantee scheme (EGS). The EGS pays a low wage for unskilled manual work on rural infrastructure projects such as road maintenance, reforestation, and small irrigation works. Anyone who shows up is given work. The idea is that for only the poor will the cost of participation—primarily that one must forgo other employment or leisure—be sufficiently low for them to turn up. But, since few of the able-bodied can afford to be completely idle, this cost will rarely be zero. In assessing impacts on the poor, it is necessary to net out this and any other costs to the poor. The potential longer-term benefits that are not captured by the wages earned must also be factored in. The final impact will be a function of the targeting effectiveness of the indirect gains and of the costs associated with targeting: administrative, behavioral, and political.

Evidence from Two Schemes

There has been much debate in India about two poverty reduction schemes: the Integrated Rural Development Program, a subsidized credit initiative targeted to low-income individuals; and public employment schemes (including the employment guarantee scheme) that rely on self-targeting. Ideally, one would like to compare these programs to see which has a greater impact on poverty. Unfortunately, the data to do so credibly are not available. But the targeting performance of the two programs can be compared for Maharashtra State.

Figure 3 plots the percentages of rural households who participated in either scheme during 1987–88 against household per capita expenditure levels (Ravallion and Datt 1995). If the schemes were well-targeted, participation should eventually fall to zero as expenditures rise. It is clear that participation in the employment guarantee scheme is highest for the poorest households and tapers off with rising expenditure levels. But the integrated rural development program data plots more or less a straight line; participation has no relationship with expenditures. These data suggest that the EGS is much better targeted and that means-testing does not assure better targeting. In this specific case one can only conjecture about why this was so. Some individuals probably understate their incomes to qualify for loans. Another possibility is that
The Costs of Targeting

The EGS is well-targeted; it removes power from administrators, and so is not subject to the same corrupt practices. But this is only one determinant of its impact on poverty. What about the costs of targeting? What benefits accrue to the poor indirectly from public works employment? Let us first consider the political economy of the EGS. Historically, the scheme has enjoyed wide support, in large part because of its indirect benefits. Well-off urban dwellers, whose taxes finance the program, support it because it helps to dampen rural migration into Bombay. Rural elites benefit from the infrastructure that results (better roads, less erosion, reforestation) and from an assured supply of labor because the guarantee of work keeps the rural farm labor force in the area through the lean season. Finally, the guarantee of assistance in times corrupt program administrators may have wide latitude to choose beneficiaries, help their friends or families, and establish patronage relations.

Dominique van de Walle
of need ensures the support of many community members who do not participate in normal times. There would seem to be important lessons here for designing schemes with built-in political sustainability.

The administrative costs in this case equal the scheme’s nonwage costs (material inputs and supervision). The size of these outlays is limited by the EGS charter, which stipulates that wages must account for about two-thirds of the total cost of the project.

Finally, costs may be associated with behavioral responses. Clearly participants would have been engaged in alternative activities had they not been employed by the relief work program. Some of those activities would have been income-earning, and that income clearly must be netted out to calculate the net benefits. But the counterfactual of what the participants would have done without the scheme is difficult to estimate, and with the exception of one study by Ravallion and Datt (1995), no convincing attempts to do so exist for any public works programs. Their study relies on longitudinal data covering a ten-year period for two villages in which some 50 percent of the population was poor by local standards. To estimate forgone incomes, the authors devised a time allocation model to predict the time displaced from income-earning activities due to the EGS. They find relatively low forgone incomes at about 25 percent of the wages earned from the relief work program.

How Did the Poor Fare?

Putting all the costs and the targeting performance together, the study calculates that about 50 percent of the benefits go to the poor and that poverty is reduced. Yet, this outcome is the same as that achievable if the same amount of public funding had been distributed as an untargeted, uniform transfer to all households. The impact on poverty would be just as large. That result is attributable in part to the high incidence of poverty and in part to the simplifying assumptions that the money can be uniformly distributed without administrative expense and without affecting behavior. Still, this hypothetical benchmark shows that concentrating benefits on the poor does not necessarily maximize their gains because it may generate costs.

The key lesson is not that spending should be untargeted, but rather that the case must rest on the net impact on poverty. Policymakers need to pay attention, for example, to the costs associated with project design. In the case of the EGS, the potential indirect benefits must be assessed as well. Benefits from the assets created, reduced risks, and potential effects on other wages could be substantial. There is anecdotal evidence for all three, but little research to draw on. The case for the EGS may ultimately rest on the size of its indirect benefits.
Conclusions

Targeting will often be desirable, but it should not be prejudged. Each case should be examined as carefully as feasible. The motivation for narrow targeting is often that broad targeting is costly, especially because of leakage. In trying to target narrowly, however, costs can be created that can wipe out the benefits. Both types of targeting have costs for the poor. Those incurred by narrow targeting are just more hidden and thus more often ignored. One can easily overestimate the gains to the poor from narrow targeting.

Both broad and narrow targeting may also have hidden benefits, notably when there are dynamic gains from relaxing the constraints that poor people face. Even less is known about benefits than about the costs. Knowledge about both costs and benefits suffers from inadequacies in data and methods of analysis and from the specific context of results. It is clear that the conventional static incidence picture is not going to capture the hidden costs and benefits from behavioral responses and impacts that occur over time. Conventional methods of analysis should thus not be the sole basis for decisionmaking. Reasonably reliable short-cut methods need to be pursued side-by-side with more rigorous and costly evaluations to test those methods.

The choice between broad or narrow targeting is not clear-cut. Ideally, targeted schemes should be designed so that incentives for escaping poverty are not destroyed and with better incentives both for participants and for their administrators. One should look for design features that encourage the poor to select themselves and do not discourage them from escaping poverty by their own means. The conditions conducive to the success of self-targeting will not always be present, however. Furthermore, there may be no indicators that are fixed, easily observed, and good proxies for low incomes. The choice between broad or narrow targeting will depend on country circumstances including constraints on policy instruments, administrative and institutional. It will also depend on policy objectives. Indeed, although the appropriate combination will differ across countries, the two approaches will often be complementary. The best approach may often be a combination of broad targeting of social sector and basic infrastructure spending with narrow targeting of transfers for neglected groups and objectives.

Achieving better outcomes for poor people from public spending may not be easy, cheap, or rapid, and this fact may have implications for other areas of policy. For example, macroeconomic operations may have to be timed and sequenced to allow for the frequent inability to compensate losers in the short term, particularly when administrative capabilities for compensation are weak. Governments may also have to engage in institutional strengthening and capacity building. Building safety nets is a long-term investment; one cannot simply wait for a crisis—the institutional capacity must already be in place if there is any hope of responding in time.
Designing effective programs for the poor also requires collecting appropriate data for understanding poverty and monitoring economy-wide, as well as targeted, policy impacts. Policymakers need to experiment widely, collect information, monitor costs and outcomes of specific schemes, conduct impact evaluations, form careful assessments of whether schemes are working, and be ready to revise and change them if necessary. Careful, rigorous evaluations will not be logistically or economically practical for each new intervention, but periodic, strategically chosen efforts can provide valuable feedback and lessons for design and implementation. There are almost certainly no easy solutions.

Note

Dominique van de Walle is senior economist in the Development Economics Research Group. The author would like to thank Harold Alderman, Jennie Litvack, and Martin Ravallion for their useful comments.

1. There is a sense in which policies that are typically categorized in the literature as broad targeting are conceptually similar to indicator targeting when the indicator is demand for a certain type of public service.

2. Geographical targeting had been examined for India (Datt and Ravallion 1993), Indonesia (Ravallion 1993), and several Latin American countries (Baker and Grosh 1994). Targeting the landless in Bangladesh is explored by Ravallion and Sen (1995). Family allowances in Hungary are explored by van de Valle, Ravallion, and Gautam (1994).

References

The word “processed” describes informally reproduced works that may not be commonly available through library systems.

Alderman, Harold, and Lindert, Kathy. 1998. This volume.


Using International Institutions to Improve Public Procurement

Bernard Hoekman

The World Trade Organization’s voluntary rules on government procurement are a useful mechanism for ensuring that public procurement procedures are efficient. They also provide an opportunity to reduce the uncertainty of participants by increasing transparency and accountability. Yet most developing countries have chosen not to subject their procurement policies to international disciplines and multilateral surveillance. Their reasons may include an unfamiliarity with the government procurement agreement (GPA); a perception that the potential payoffs are small; a desire to discriminate in favor of domestic firms; or the successful opposition of groups that benefit from the current regimes. Although the economic rationales for abstaining from the GPA are not compelling, a quid pro quo for accession may be needed to overcome opposition by special interests. Developing country procurement markets are large enough that governments may be able to make accession to the GPA conditional on temporary exceptions to multilateral disciplines or on better access to export markets.

All over the world government agencies purchase the goods and services necessary to provide the public with education, defense, utilities, infrastructure, public health, and so forth. The public procurement associated with these expenditures often represents a significant share of a country’s gross domestic product. To maximize the use of scarce financial resources, governments have developed procedures and mechanisms to ensure that public entities procure these collective goods and services at least cost and in a fair and transparent manner. Some countries have addressed the issue of procurement efficiency head-on; South Africa, for example, wrote a section on procurement principles into its 1994 constitution requiring the government to pursue a fair, public, and competitive procurement process under the direction of independent and impartial tender boards that are obligated to record decisions and make them available to interested parties (Transparency International 1997).
Many procurement systems attempt to mimic the workings of the market by requiring that public entities seek competitive bids from potential suppliers of goods and services. Over time, an increasing number of governments have also pursued more far-reaching efforts to subject production units directly to competitive forces by privatizing state-owned enterprises, encouraging competitive entry into sectors traditionally reserved for the state (for instance, utilities), and by contracting out activities to the private sector. In a survey of the empirical literature on the impact of competitive tendering and outsourcing, Domberger, Hall, and Lee (1995) conclude that savings on the order of 20 percent are common and do not come at the expense of quality.

In the late 1970s several high-income countries negotiated an agreement on government procurement under the auspices of the General Agreement on Tariffs and Trade (GATT). The Government Procurement Agreement (GPA) extends the basic principles of the GATT—nondiscrimination, national treatment, and transparency—to the tendering procedures of specified government entities. Members of the GPA are Canada, the European Union, Hong Kong, Israel, Japan, Republic of Korea, Norway, Singapore, Switzerland, and the United States. Although many developing country governments have made efforts to reduce the cost of providing public services, they have refrained from signing the GPA. This is surprising, because the GPA appears to be a useful mechanism for ensuring that government procurement procedures maximize value for money. The benefits are likely to include not only a reduction in procurement costs but, perhaps more important, positive spillover effects that result from transparency and accountability.

This article examines multilateral government procurement regulations from a developing country perspective. It assumes that governments are interested in achieving and maintaining efficient, transparent, and accountable procurement procedures and asks why developing countries have not signed on to the GPA. This is an important topic because the multilateral rules of the game must be consistent with "best practices." If there are good reasons to question the economics of the GPA from a development and growth perspective, the rules should be revised. If not, the issue becomes one of political economy. Efforts must then be directed at identifying and overcoming the resistance to membership by those groups in society that benefit from the status quo at the expense of the community at large.

Developing countries can expect to face increasing pressure on the subject of government procurement in bilateral trade relations with industrial nations. The United States has played a leading role in this connection, making public procurement practices a priority and linking them to the broader issue of combating corruption. In April 1996, largely at the initiative of the United States, members of the Organisation for Economic Co-operation and Development (OECD) agreed that firms would not be permitted to write off bribes against tax obligations; in May 1997 they recommended classifying bribery of foreign officials as a criminal offense under national
legislation (World Bank 1997). At the December 1996 meeting of the World Trade Organization (WTO) in Singapore, a working group was created to conduct a study on transparency in government procurement practices and develop "elements for inclusion in an appropriate agreement" (WTO 1996, p.7). Although the GPA is not mentioned, the most straightforward way for developing countries to deal with procurement-related concerns is to join the GPA. As far as the United States is concerned, this is clearly the ultimate objective; in the meantime, an interim agreement on transparency, openness, and due process is seen as "an important step toward a more comprehensive multilateral agreement in the WTO" (USTR 1997, p. 4).

The WTO Agreement on Government Procurement

The GPA applies to laws, regulations, procedures, and practices pertaining to any government procurement, including purchases, rentals, leases, and lease-purchase agreements, with or without the option to buy. The procurement agreement covers only entities listed by each signatory in three categories: central government; subnational entities; and others (these are primarily utilities that may be partially or wholly privately owned). Procurement is subject to the GPA as long as its value exceeds certain specified thresholds and the goods or services in question are not exempt from the Agreement. The threshold for central governments is generally Special Drawing Rights (SDR) 130,000; for local governments SDR 200,000; and for other entities usually SDR 400,000. For most members, construction contracts are covered only if they exceed SDR 5 million. (In 1997 one SDR was equivalent to $1.35.)

All procurement of goods is covered unless otherwise specified. (In most instances military procurement is the exception.) In the case of services, only those products that are explicitly listed by each member country are subject to the GPA's rules. In practice these commitments closely parallel those of the World Trade Organization members under the General Agreement on Trade in Services (GATS), which liberalizes access to service markets generally (see Hoekman and Kostecki 1995 for an introduction to the GATS and references to the literature).

General Provisions of the GPA

The primary obligations imposed by the GPA are transparency and nondiscrimination. This policy extends not only to imports but also to goods and services provided by local subsidiaries of foreign firms. Preferential prices, offsets (measures, such as domestic content requirements and technology licensing, that encourage local development), and similar policies that discriminate in favor of domestic firms are in principle prohibited; competitive tendering procedures are encouraged. These include open tendering, where any supplier may respond to a published call for ten-
ders, or selective tendering, where bids are restricted to prequalified suppliers who have demonstrated that they meet technical competence norms. Limited tendering, under which potential suppliers are directly solicited to bid by the procuring entity, may be used in only three circumstances: situations in which no responses have been received to an open or selective call for tenders, cases of urgency, and orders for additional deliveries by an original supplier.

Calls for open tenders must be published in all cases, stating the mode of procurement, its nature and quantity, dates of delivery, economic and technical requirements, amounts and terms of payment, and so on. Individual suppliers may not be given information that could have the effect of precluding competition. Entities are obliged to award contracts to the tenderer who “has been determined to be fully capable of undertaking the contract” and who is either the lowest cost or most advantageous supplier, according to the evaluation criteria set forth in the notices. Because an evaluation of which tender comes closest to satisfying the criteria is open to a considerable degree of discretion, much depends on how the criteria in the notices or tender documentation are worded. It is a violation of the procurement agreement to determine that a tender is the most advantageous on the basis of criteria that are not specified.

The nature of procurement is such that unless rapid remedial action can be taken to intervene in the procurement process, firms are unlikely to contest perceived violations of the rules. A unique feature of the GPA is that it requires members to establish bid protest or challenge procedures, under which bidders can correct breaches of the GPA in order to preserve commercial opportunities. Such measures may involve suspension of the procurement process, reopening of the tender procedure, or the award of compensation for loss or damages. This is a key dimension of the GPA because it gives firms an incentive to defend their interests. The domestic challenge mechanism is complemented by the WTO’s multilateral dispute settlement process. To ensure transparency and facilitate the application of these procedures, procuring entities must provide information explaining why a supplier’s application to qualify was rejected; why an existing qualification was terminated; and why a tender was not selected. They must also identify the winning bidder and clarify the characteristics and relative advantages of the tender selected.

Developing Country Provisions

Although in principle the GPA prohibits signatories from discriminating in favor of domestic firms, developing countries may negotiate mutually acceptable exclusions from the rules on national treatment for certain entities, products, or services (Article V). Such negotiations may also be initiated after signing the procurement agreement. That option, however, is limited to certain entities, products or services, and the scope to pursue such policies is therefore inherently limited by the relative nego-
tiating power of the country seeking to apply them. Developing countries may also, at the time of accession, negotiate conditions for the use of offsets. Although offset requirements explicitly allow for de facto discrimination against foreign suppliers, the requirements may be used only to qualify for the procurement process and not as criteria for awarding contracts. Thus, if a firm offers local content that greatly exceeds the minimum requirements, that aspect of the offer may not be a factor in awarding a contract.

Economic Issues

The major substantive disciplines imposed by the GPA are nondiscrimination, competitive tendering, and transparency, complemented by the domestic and multilateral enforcement mechanisms. What is the economic rationale for these rules? Are there reasons to conclude that they may not be in the interest of developing countries? There are two potential sources of concern: first, discrimination may be needed to ensure least-cost procurement; and second, the net economic payoffs associated with the rules may be too small.

Is Nondiscrimination Always Optimal?

Intuitively the nondiscrimination rule appears to be unambiguously beneficial because it should intensify competition and thereby minimize the costs of procurement. As is often the case in economics, this proposition is not necessarily true. Discrimination may be necessary to minimize average procurement costs. Discriminating against foreign bidders may be welfare improving if domestic firms are at a competitive disadvantage in certain areas (that is, if they are higher-cost producers) and only a limited number of firms (foreign and domestic) bid for a contract. In such situations, foreign firms may exploit their cost advantage by bidding just below the amount they expect domestic firms to bid (McAfee and McMillan 1989). Although the foreign firm will be the lowest bidder, the bid may be substantially above the firm's actual cost. A policy that gives preferences to domestic firms may then induce foreign firms to lower their bids by the extent of the preference margin. If so, procurement favoritism increases national welfare. Even if the cost structure of domestic and foreign firms are identical and account is taken of the social cost of distortionary taxation, discrimination may be rational simply because foreign profits do not contribute to domestic welfare (Branco 1994). Thus favoritism can be used as a rational profit-shifting strategy whenever there is imperfect competition. Shifting demand to domestic firms may also reduce price-cost margins as domestic output expands (Chen 1995).

Even when there are many potential suppliers for a contract, which should guarantee the lowest costs, discrimination may be beneficial in other ways. If, for in-
stance, the products to be procured are intangible or if enforcing contract compliance is a problem, public agencies may need to grant excess profits to contractors to get them to deliver (Laffont and Tirole 1991; Rothenberg 1993). Moreover, if there is an incentive to pay a premium over the suppliers' cost to ensure contract performance, the required premium may increase as the number of potential bidders rises (because each supplier will take into account the higher probability of not getting repeat business) (Breton and Salmon 1995). Minimizing the costs of procurement in such settings may require limiting the number of potential suppliers. If so, governments can be expected to favor domestic over foreign suppliers. The sourcing costs will not be affected, while political benefits may arise because domestic firms are part of their constituencies. Such situations are more likely in the case of service procurement, because of their intangible nature. Services are often the largest category of government purchases—increasingly so in countries that have been pursuing outsourcing and contracting strategies. In the United States, for example, most federal nondefense procurement is for services (Francois, Nelson, and Palmeter 1997).

Problems of asymmetric information may also reinforce the decision to choose local suppliers in order to reduce monitoring costs. Such proximity incentives make it more difficult for foreign firms to bid successfully, even when they are not faced with formal discrimination, and may encourage them to contest procurement markets by investing directly in developing countries. This incentive is, of course, not procurement-specific but applies in all instances where buyers prefer to deal with "local" suppliers. The policy issues then are how to decide whether suppliers are local "enough" and what—if any—barriers exist against foreign direct investment.

Although discriminatory procurement may lower procurement costs in some situations, studies suggest that the net welfare benefits are likely to be modest at best; increased prices will tend to offset any cost savings (Deltas and Evenett 1997). The net welfare impact depends on the government's objective function, particularly the relative importance given to domestic industry profits as opposed to expected procurement costs. The latter will generally be a multiple of the former. In many situations the information required to determine whether discrimination is beneficial is not likely to be available. Even if it is, the wording of general regulations calling for discrimination—for instance, a universal price preference of 15 percent—is too rigid. In many cases markets will be competitive and products will be relatively homogeneous, so these considerations do not arise.

Nonetheless, in principle there is a potential tradeoff between the GPA's nondiscrimination rules and economic efficiency. One way to address this tension would be to give governments the discretion to apply a "rule of reason" and require them to rationalize their decisions. This approach, however, introduces the potential for arbitrary decisions and makes it difficult—if not impossible—to apply the GPA's enforcement and dispute settlement provisions.
Experience with competitive procurement regimes indicates that in most situations competition is the best rule of thumb; competitive and transparent procurement regimes generate substantial cost savings. For example, Transparency International (1997) notes that noncompetitive procedures may increase procurement costs as much as 30 percent. Estimates of the cost savings associated with international competitive bids for World Bank loans are in the same range. In the case of a recent balance of payments loan to the Russian Federation financed by the World Bank, procurement costs under competitive bidding procedures were on average 30 to 40 percent below the costs of identical items acquired in the past without such bidding and as much as 75 percent less for certain pharmaceutical products (World Bank 1994).

In practice discrimination may be motivated not by considerations of cost or contract compliance but by a desire to promote domestic industry. Much has been written about the pros and cons of protecting infant industries. What matters in the current context is not whether there are rationales for intervention on infant industry grounds, but that procurement favoritism is unlikely to be the optimal policy to use in pursuit of this objective. In effect, it is equivalent to a subsidy financed by domestic taxpayers, because government purchasing costs increase and must be financed. But it is less transparent than a direct subsidy from the budget and as a result provides greater potential for rent-seeking and corruption. Debroy and Pursell (1997) provide an interesting review of India's use of procurement regimes to support infant industry policies—a policy the government recently has begun to reverse.

Costs and Benefits

Even if it is accepted that nondiscrimination is the best rule of thumb and that it will produce significant cost savings, the net benefits may be minor because of the costs incurred. These costs can be substantial, as entities must satisfy many procedural requirements to ensure due process and transparency. These procedures are undoubtedly burdensome, but they have advantages as well. Nondiscrimination, transparency, and accountability provisions may constrain rent-seeking activities. Allowing procuring entities to discriminate may facilitate bribery of procurement officials. Although the issue of corruption extends beyond procurement, rent-seeking in the public purchasing context is particularly prominent because the amounts involved are significant and foreign interests are frequently affected. Case studies have demonstrated that corruption can increase the costs of a project by as much as 25 to 50 percent (Wade 1982; Rose-Ackerman 1995a). Corruption and rent-seeking reduce economic growth, distort resource allocation, and result in higher taxes or, more commonly, deficit financing (Mauro 1995; Murphy, Schleifer, and Vishny 1993; Schleifer and Vishny 1993; Bardhan 1997).
Abstracting from differences in cultural norms across countries, effective anticorruption strategies must reduce the magnitude of the benefits that can be granted by officials, increase the costs of bribery for the private sector, and limit the market power of officials (Rose-Ackerman 1995a, 1995b; Bardhan 1997). Of the various strategies and suggestions offered in the literature, the following are particularly relevant for procurement: effective deterrents through ex post punishments that exceed the gains realized (including banning firms caught in attempts to engage in bribery from bidding for contracts for a number of years); the creation of external monitoring devices and institutions (including encouragement and protection of “whistleblowers”); public transparency mechanisms (published audits by independent auditors, a free press); privatization and hard budget constraints; requiring the use of standardized products and goods that have well-established market positions; the use of general retail-wholesale market prices for goods similar to those to be procured as comparators; and the use of incentives to encourage bidders to complain if they suspect corruption.

Although the GPA lacks an explicit corruption standard or norm, it is consistent with—or embodies many of—these principles. Of particular importance are the challenge procedures that allow firms to protest before the decision process is completed, as well as thereafter. Multilateral monitoring and the threat of WTO dispute settlement procedures will also help to ensure that entities abide by the GPA’s substantive and procedural disciplines. Alam (1995) argues that the opportunities for losers to take countervailing actions can constrain rent-seeking activities. In the context of procurement, the set of losers is usually small. In discretionary, nontransparent procurement systems, firms that lose out have little incentive to protest irregularities because they fear being blacklisted. The GPA’s rules are designed to maximize the incentives to obtain and use information concerning possible violations, but given the sunk costs of participating in the bidding process, such a protest is only viable if the expected returns outweigh the costs of protesting. Ensuring that this is the case may be difficult.

Operation of the GPA

The GPA requires signatories to report annual procurement statistics to the WTO Committee on Government Procurement, which provides oversight for the GPA. This committee, which comprises representatives from all member countries, meets periodically in Geneva. Signatories began reporting statistics for the year 1983. Because the latest year available for some countries is 1992, the analysis here is restricted to that period—a time during which the GPA applied only to the procurement of goods by central government entities. As a result, most procurement activities were excluded. In the case of the United States, for example, Francois, Nelson,
and Palmeter (1997) note that in 1993 goods accounted for less than 5 percent of
total federal nondefense-related purchases.

Coverage

In 1992 the total procurement of goods by the entities covered by the GPA was about
$62 billion. Average annual purchases that year by the United States, the largest
procurement market, were some $29 billion (table 1). This compares to total pur-
chases of $16 billion by the European Union (EU)-12 entities, $9.2 billion by Japan,
and $1.6 billion by Canada. For the purposes of cross-country comparisons, it
is helpful to relate these numbers to total central government expenditures on
nondefense-related goods and services and to capital expenditures by the central gov-
ernment. Relative to total central government expenditures on goods and services—
which includes items such as wages—purchases under the GPA by large countries
such as France, Germany, Italy, Japan, and the United Kingdom tend to be below
average; the United States and the Nordic countries are above average. On both
measures, Germany and Italy have the lowest ratios, suggesting they may have sought
to limit the coverage of their GPA obligations. Israel and Singapore have scheduled
substantially fewer procuring entities than the OECD countries.

For a number of reasons it is difficult to estimate how much additional pro-
curement was brought under the GPA’s umbrella in 1996, when it was extended
to cover services and subnational entities. First, it is not always clear whether
specific services are covered or whether the nondiscrimination rule applies to all
services. (Many countries made their commitments conditional upon reciproc-
ity.) Second, no information on the size of the average contract and the types
of goods and services purchased is available. The best estimates of central govern-
ment nondefense expenditures by GPA signatories puts the total at about $2.1
trillion (table 1); subnational government bodies add at least another $1 trillion
(IMF 1996). Not all of this procurement is available for international competi-
tion. In particular, the average subnational government contract is likely to be
smaller than those of central government entities. If it is assumed that one-third
of total outlays by central and subnational governments could be subjected to
GPA rules, the total potential market would be close to $1 trillion a year. Applying
the historical GPA rate of 50 percent for the share of procurement of goods
that falls below the threshold value, some $500 billion could be open to interna-
tional competitive bidding (table 2). This is likely to be an overestimate, how-
ever, as threshold values for subnational government entities and services con-
tracts are significantly higher than those that apply to central government entities.
Because the thresholds for any construction contracts are at least SDR 5 million,
and as much as SDR 15 million in Korea and Japan, $300 billion is a more real-
istic figure.

Bernard Hoekman
Table 1. *GDP and Central Government Expenditures, 1992*  
(millions of dollars)

<table>
<thead>
<tr>
<th>Country</th>
<th>Gross domestic product (A)</th>
<th>Total government expenditure (B)</th>
<th>Total nondefense expenditures (C)</th>
<th>Capital expenditures (D)</th>
<th>Value of total contracts under GDP(^a) relative to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>185,235</td>
<td>40,282</td>
<td>29,842</td>
<td>4,750</td>
<td>433</td>
</tr>
<tr>
<td>Belgium(^b)</td>
<td>218,836</td>
<td>64,197</td>
<td>44,540</td>
<td>4,973</td>
<td>407</td>
</tr>
<tr>
<td>Canada</td>
<td>493,602</td>
<td>128,155</td>
<td>96,959</td>
<td>1,965</td>
<td>2,399</td>
</tr>
<tr>
<td>Denmark</td>
<td>123,546</td>
<td>47,219</td>
<td>35,782</td>
<td>1,539</td>
<td>1,646</td>
</tr>
<tr>
<td>Finland</td>
<td>93,869</td>
<td>30,020</td>
<td>23,045</td>
<td>1,912</td>
<td>834</td>
</tr>
<tr>
<td>France(^b)</td>
<td>1,319,883</td>
<td>256,711</td>
<td>178,104</td>
<td>19,273</td>
<td>3,279</td>
</tr>
<tr>
<td>Germany(^b)</td>
<td>1,789,261</td>
<td>215,669</td>
<td>149,629</td>
<td>35,832</td>
<td>2,055</td>
</tr>
<tr>
<td>Ireland</td>
<td>43,294</td>
<td>16,986</td>
<td>13,236</td>
<td>1,269</td>
<td>208</td>
</tr>
<tr>
<td>Israel</td>
<td>69,762</td>
<td>25,270</td>
<td>13,001</td>
<td>3,500</td>
<td>68</td>
</tr>
<tr>
<td>Italy(^b)</td>
<td>1,222,962</td>
<td>335,439</td>
<td>232,725</td>
<td>17,983</td>
<td>1,994</td>
</tr>
<tr>
<td>Japan(^b)</td>
<td>3,670,979</td>
<td>479,158</td>
<td>332,436</td>
<td>58,360</td>
<td>9,507</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>12,638</td>
<td>2,394</td>
<td>1,784</td>
<td>364</td>
<td>34</td>
</tr>
<tr>
<td>Netherlands</td>
<td>320,290</td>
<td>92,698</td>
<td>75,644</td>
<td>6,505</td>
<td>1,281</td>
</tr>
<tr>
<td>Norway</td>
<td>112,906</td>
<td>25,399</td>
<td>19,276</td>
<td>1,255</td>
<td>775</td>
</tr>
<tr>
<td>Singapore</td>
<td>46,025</td>
<td>8,889</td>
<td>3,424</td>
<td>2,272</td>
<td>30</td>
</tr>
<tr>
<td>Sweden</td>
<td>220,834</td>
<td>75,172</td>
<td>61,158</td>
<td>3,386</td>
<td>1,162</td>
</tr>
<tr>
<td>Switzerland (1984)</td>
<td>90,649</td>
<td>8,973</td>
<td>5,784</td>
<td>1,096</td>
<td>252</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>903,126</td>
<td>338,259</td>
<td>211,108</td>
<td>33,755</td>
<td>5,740</td>
</tr>
<tr>
<td>United States</td>
<td>5,920,199</td>
<td>1,037,354</td>
<td>631,924</td>
<td>56,354</td>
<td>29,120</td>
</tr>
<tr>
<td>Total</td>
<td>3,228,246</td>
<td>2,159,399</td>
<td>256,343</td>
<td>61,227</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Unweighted average  

\(a\) Average 1991 and 1992. 
\(b\) Estimated: Switzerland, 1984; Hong Kong has been excluded because of data inconsistencies. 

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>179.4</td>
<td>403.9</td>
<td>43.2</td>
<td>43.2</td>
<td>51.3</td>
<td>1.5</td>
<td>46.9</td>
<td>43.5</td>
</tr>
<tr>
<td>Belgiuma</td>
<td>129.9</td>
<td>407.2</td>
<td>10.3</td>
<td>8.6</td>
<td>100.0</td>
<td>100.0</td>
<td>40.0</td>
<td>57.5</td>
</tr>
<tr>
<td>Canada</td>
<td>968.5</td>
<td>2,163.9</td>
<td>9.7</td>
<td>7.4</td>
<td>92.3</td>
<td>77.4</td>
<td>42.9</td>
<td>46.0</td>
</tr>
<tr>
<td>Denmarka</td>
<td>447.9</td>
<td>1,651.7</td>
<td>4.0</td>
<td>1.4</td>
<td>95.6</td>
<td>79.8</td>
<td>5.8</td>
<td>12.6</td>
</tr>
<tr>
<td>Finland</td>
<td>256.3</td>
<td>800.8</td>
<td>0.3</td>
<td>0.1</td>
<td>95.7</td>
<td>69.1</td>
<td>39.4</td>
<td>18.1</td>
</tr>
<tr>
<td>Francea</td>
<td>937.9</td>
<td>3,089.5</td>
<td>33.3</td>
<td>29.6</td>
<td>97.6</td>
<td>97.2</td>
<td>35.1</td>
<td>85.5</td>
</tr>
<tr>
<td>Germanya</td>
<td>845.6</td>
<td>2,029.3</td>
<td>13.3</td>
<td>24.0</td>
<td>97.2</td>
<td>99.1</td>
<td>24.9</td>
<td>61.7</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>154.5</td>
<td>348.6</td>
<td>22.1</td>
<td>28.2</td>
<td>5.8</td>
<td>3.2</td>
<td>67.7</td>
<td>71.2</td>
</tr>
<tr>
<td>Irelanda</td>
<td>47.8</td>
<td>185.8</td>
<td>1.8</td>
<td>5.6</td>
<td>100.0</td>
<td>84.8</td>
<td>20.6</td>
<td>29.1</td>
</tr>
<tr>
<td>Israel</td>
<td>30.1</td>
<td>67.9</td>
<td>5.0</td>
<td>3.3</td>
<td>13.7</td>
<td>25.6</td>
<td>84.0</td>
<td>91.1</td>
</tr>
<tr>
<td>Italya</td>
<td>382.4</td>
<td>1,937.5</td>
<td>1.0</td>
<td>5.9</td>
<td>99.2</td>
<td>97.8</td>
<td>5.0</td>
<td>60.5</td>
</tr>
<tr>
<td>Japan</td>
<td>3,379.4</td>
<td>9,274.1</td>
<td>12.3</td>
<td>21.0</td>
<td>86.1</td>
<td>85.5</td>
<td>37.4</td>
<td>43.7</td>
</tr>
<tr>
<td>Luxembourga</td>
<td>14.9</td>
<td>35.8</td>
<td>39.8</td>
<td>23.9</td>
<td>100.0</td>
<td>97.8</td>
<td>30.9</td>
<td>28.4</td>
</tr>
<tr>
<td>Netherlandsa</td>
<td>547.1</td>
<td>1,407.7</td>
<td>16.8</td>
<td>13.5</td>
<td>98.7</td>
<td>86.4</td>
<td>12.1</td>
<td>37.7</td>
</tr>
<tr>
<td>Norway</td>
<td>384.7</td>
<td>721.4</td>
<td>7.1</td>
<td>8.3</td>
<td>48.6</td>
<td>40.4</td>
<td>44.5</td>
<td>58.9</td>
</tr>
<tr>
<td>Singapore</td>
<td>48.5</td>
<td>31.0</td>
<td>0.0</td>
<td>0.0</td>
<td>45.1</td>
<td>67.3</td>
<td>57.9</td>
<td>51.1</td>
</tr>
<tr>
<td>Sweden</td>
<td>624.4</td>
<td>1,200.5</td>
<td>4.0</td>
<td>4.8</td>
<td>43.7</td>
<td>43.6</td>
<td>27.9</td>
<td>36.9</td>
</tr>
<tr>
<td>Switzerland</td>
<td>265.0</td>
<td>806.2</td>
<td>25.1</td>
<td>27.7</td>
<td>35.7</td>
<td>38.3</td>
<td>43.8</td>
<td>39.8</td>
</tr>
<tr>
<td>United Kingdoma</td>
<td>1,329.7</td>
<td>5,375.0</td>
<td>13.7</td>
<td>6.6</td>
<td>99.4</td>
<td>98.2</td>
<td>29.3</td>
<td>49.0</td>
</tr>
<tr>
<td>United States</td>
<td>24,080.8</td>
<td>28,891.2</td>
<td>11.5</td>
<td>9.6</td>
<td>86.3</td>
<td>90.4</td>
<td>80.1</td>
<td>66.8</td>
</tr>
<tr>
<td>Memo: EU countries</td>
<td>7,967.2</td>
<td>16,119.4</td>
<td>19.0</td>
<td>15.0</td>
<td>98.3</td>
<td>94.9</td>
<td>21.2</td>
<td>54.1</td>
</tr>
<tr>
<td>Average unweighted</td>
<td>13.1</td>
<td>13.0</td>
<td>74.6</td>
<td>71.2</td>
<td>38.8</td>
<td>49.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average weighted by value of procurement</td>
<td>13.3</td>
<td>14.1</td>
<td>86.1</td>
<td>87.3</td>
<td>59.7</td>
<td>57.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Foreign Sourcing

Smaller countries tend to procure more goods and services on international markets than do large countries. If Canada, the European Union, Japan, and the United States are excluded, about 60 percent of the purchases of goods by central government entities under the GPA are from national suppliers, compared with more than 90 percent for the large players. Unfortunately, EU statistics define "domestic" as within the European Union, thus skewing self-sufficiency ratios (see table 2). In interpreting these statistics, note that no distinction is made between domestic firms and foreign firms that have established a local presence. As long as the share of foreign direct investment is the same in large countries as in smaller ones, cross-country comparisons should not be affected.

In Japan, the United States, and the EU countries, the share of total procurement from domestic firms was virtually unchanged during 1983–92. The average weighted share across all GPA members is relatively constant; the unweighted average fell by 3 percentage points, from 74.6 percent to 71.2 percent (see table 2). Among smaller countries, however (with the exception of Singapore and Switzerland), the share of procurement from domestic sources has declined over time. Although such changes cannot be attributed solely to the GPA because other developments—such as the North American Free Trade Agreement, the EU procurement liberalization, and unilateral deregulation and privatization policies—also played a role, the finding appears to be robust. But for most of the large players there has been no change.

This conclusion is supported by the few empirical studies on government procurement practices in OECD countries. These use a methodology suggested by Baldwin (1970) and Baldwin and Richardson (1972), which assumes that in the absence of discriminatory policy, government entities would behave in the same way as do private firms. Thus, government imports of a good as a share of total consumption would equal that of the private sector as a whole. The difference between the private sector import propensity and the actual import share of total government consumption can be called a preference margin. (A positive preference margin implies that the government share is lower than the private share.) Baldwin (1970) estimated that the preference margin in the United States was some 20 percent in 1958. After adjusting for the fact that certain large import items such as oil were not subject to discriminatory policies, the margin for the residual set of covered goods increased to some 40 percent. More recent estimates (1992) show the U.S. preference margin was 16.3 percent (Francois, Nelson, and Palmetter 1997). On a sectoral level, positive margins in OECD countries may be as high as 50 percent (table 3). Margins are invariably the highest for procurement of services.

The Baldwin-Richardson methodology is obviously sensitive to the assumption that, other things being equal, the government would import the same share of a
Table 3. *Estimated Preference Margins for Core Government Purchasing, 1992*
(Baldwin-Richardson 1972 Approach)

<table>
<thead>
<tr>
<th>Country or region</th>
<th>Machinery</th>
<th>Other goods</th>
<th>Trade, transport, communication</th>
<th>Utilities</th>
<th>Other services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>39.6</td>
</tr>
<tr>
<td>United States</td>
<td>18.4</td>
<td>17.9</td>
<td>—</td>
<td>18.8</td>
<td>42.6</td>
</tr>
<tr>
<td>Western Europe</td>
<td>—</td>
<td>9.2</td>
<td>13.7</td>
<td>14.9</td>
<td>48.3</td>
</tr>
<tr>
<td>Japan</td>
<td>—</td>
<td>32.0</td>
<td>26.2</td>
<td>34.0</td>
<td>46.6</td>
</tr>
<tr>
<td>Australia</td>
<td>49.8</td>
<td>49.7</td>
<td>—</td>
<td>—</td>
<td>41.5</td>
</tr>
<tr>
<td>New Zealand</td>
<td>13.9</td>
<td>19.7</td>
<td>49.8</td>
<td>—</td>
<td>50.0</td>
</tr>
<tr>
<td>Korea, Republic of</td>
<td>30.6</td>
<td>20.8</td>
<td>—</td>
<td>—</td>
<td>48.2</td>
</tr>
</tbody>
</table>

— Less than or equal to zero.


...good as the private sector, and that all differences can be attributed to formal or informal preference policies. One source of bias that arises in this connection is that private sector imports may be distorted because of tariffs and other trade policies. Alternatively, private sector demand for certain products may be very low or even zero because government is the dominant supplier of output that embodies particular products (defense, utilities, certain types of transport services, and so on).

**Limited Tendering**

Limited tendering procedures (which involve an entity contacting and negotiating with potential suppliers individually) are permissible only under certain conditions. The use of limited tendering varies across signatories, from a reported low of zero (Singapore) to a high of more than 30 percent on average for France, Hong Kong, Italy, and Switzerland (see table 2). Across all signatories the average share of limited tendering was about 13 percent. This practice has been falling over time in the EU-12 (by about 7 percent a year on average during 1983–92), but rising in the United States. By 1992 limited tendering in the EU and in the United States stood at 10 percent, largely as a result of a significant decline in this practice by France and Germany during the 1980s. Japan’s limited tendering rose from around 12 percent during 1983–85 to 21 percent during 1990–92. Hong Kong and Switzerland make even more intensive use of such procurement mechanisms.

This type of data is not available for developing countries. Indeed, little is known about the product composition of procurement across developing countries, the average contract size, or the “import propensity” of government entities. Indeed, one benefit of membership in the GPA would be that data on procurement flows would...
have to be collected and reported to the WTO. This information would be valuable for researchers as well as for the governments concerned.

**Transparency, Disputes, and the Challenge Mechanism**

If procurement procedures are to function efficiently, potential suppliers need to be aware of demand and have sufficient time to respond to calls for tenders. The GPA spells out these requirements, but it does not monitor compliance; that is left to the private sector. As a result, violations may be overlooked. Even in the EU, where procurement opportunities must be published in the *Official Journal of the European Communities* and can be obtained through electronic networks, a recent study found that procuring entities often failed to publish the information or provided insufficient time for responses. Moreover, many potential suppliers did not routinely monitor the *Official Journal* and were thus unaware of the potential market (Gordon, Rimmer, and Arrowsmith 1997). In part these problems result from ambiguities in the wording of the regulations on EU procurement.

The GPA's challenge mechanism is untested as yet, because it only came into force in 1996. The key requirement for these procedures to be effective is that participants can either obtain very rapid intervention by a judicial or administrative body or have the opportunity to obtain significant financial compensation after the fact for violations of procurement disciplines. (Once a tender has been closed and a contract awarded, reopening the proceedings may be difficult.) A concern with this process is its potential to give rise to problems of moral hazard and excessive litigation. Ensuring that firms have access to an effective and speedy mechanism through which they can challenge a procurement process or award is therefore very important. A recent evaluation of EU procedures and disciplines—which were an important model for much of the GPA—suggests that EU suppliers have not used the challenge mechanism to great effect because it is perceived as being too slow (Gordon, Rimmer, and Arrowsmith 1997).

**Negotiating Accession: Considerations of Political Economy**

As noted earlier, membership in the GPA did not alter the sourcing patterns of large countries, but small countries purchased more from foreign suppliers over time. These findings are consistent with economic theory. The larger the country, the greater the number of potential domestic suppliers and the higher the probability that domestic firms attain minimum efficient scale. Thus, large countries can be expected to continue to buy predominantly from national suppliers even if they abide fully by the GPA. Small countries, in contrast, will generally have fewer national firms that can provide the goods needed by the government at least cost, especially specialized,
capital-intensive items where scale economies are important (such as telecommunications, transport, or power-generating equipment).

The economic rationale for refusing to join the GPA is weak and may in large part be driven by political economy factors. The issue then is to seek to offset the opposition of groups that would like to maintain the status quo. The problem is similar to that facing trade reformers: incumbent industries and procurement officials have more of an individual incentive to oppose reform than losers from the status quo (taxpayers, consumers) have to push for it. In the context of trade liberalization, the problem arising from the asymmetric distribution of costs and benefits can be addressed by providing those who stand to gain from reform a greater incentive to engage in the political process in pursuit of their interests. One way to mobilize support for liberalization is to promise exporters better access to foreign markets, something that the government can demand as a quid pro quo from its trading partners for liberalizing the domestic market. But such access will not have much of an impact in increasing developing countries’ share of OECD procurement markets. Although some countries may have a comparative advantage in providing certain services, service industries are much more subject to restrictive regulations than goods markets. There is no formal constraint on demanding other concessions, however. Developing countries could bargain for access to markets that are of interest to them. Although trade barriers are relatively low in industrial countries, a significant amount of tariff escalation remains. Tariffs and other trade barriers tend to be substantially higher on labor-intensive products such as footwear or clothing than on other types of manufactures. Achieving reductions in these barriers is difficult. Linking membership in the GPA to reductions in trade barriers by OECD countries may help mobilize support for domestic liberalization of procurement markets.

Another way to attenuate opposition is to change the rules of the game in a manner that meets some of the concerns associated with the implementation of the GPA. For example, developing countries could be allowed to apply price preferences that favor procurement from national suppliers. Although a strong presumption exists that the GPA’s nondiscrimination, transparency, and enforcement provisions will benefit developing countries, many nations have procurement regimes that provide preferences for domestic firms. Such provisions may be motivated by infant industry arguments; they may reflect an attempt to offset the higher input costs of domestic bidders that result from other policies (for example, high import tariffs); or, as noted earlier, they may be a response to imperfect competition or asymmetric information. In principle, the optimal policy to deal with distortions such as tariffs is to eliminate or significantly reduce them, thereby eliminating domestic preferences as well. Some countries have embarked on such a process, but local content requirements and price preferences continue to prevail in many jurisdictions.

The GPA could be amended to allow developing countries to establish price preferences as long as these are transparent and accountable. Although the case for policies
Table 4. Government Expenditures and the Role of Official Development Assistance
(millions of dollars)

<table>
<thead>
<tr>
<th>Country group</th>
<th>GDP 1992</th>
<th>Total nondefense expenditure on goods and services</th>
<th>Capital expenditure</th>
<th>Multilateral development bank loans, 1992a</th>
<th>Development Assistance Committee loans, 1992b</th>
<th>Ratio multilateral development bank loans/total nondefense expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of low-income countries (N=46)</td>
<td>23,704</td>
<td>2,053</td>
<td>1,637</td>
<td></td>
<td></td>
<td>18.1</td>
</tr>
<tr>
<td>Total low-income countries</td>
<td>1,090,384</td>
<td>94,447</td>
<td>75,296</td>
<td>17,076</td>
<td>32,931</td>
<td></td>
</tr>
<tr>
<td>Average of lower middle-income countries (N=42)</td>
<td>31,151</td>
<td>2,237</td>
<td>921</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total lower middle-income countries</td>
<td>1,308,362</td>
<td>93,960</td>
<td>38,682</td>
<td>13,972</td>
<td>15,337</td>
<td>14.9</td>
</tr>
<tr>
<td>Average of upper middle-income countries (N=16)</td>
<td>80,384</td>
<td>6,602</td>
<td>1,496</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of upper middle-income countries</td>
<td>1,286,158</td>
<td>105,631</td>
<td>23,943</td>
<td>6,135</td>
<td>1,239</td>
<td>5.8</td>
</tr>
</tbody>
</table>

a. Refers to loan approvals.
b. Development Assistance Committee data pertains to net disbursements and includes bilateral assistance.
that discriminate in favor of domestic suppliers is not compelling, such an approach may be a useful transition to encourage membership in the GPA. Multilateral financial institutions, including the World Bank allow such preferences subject to certain conditions and limits. The World Bank permits a maximum price preference of 15 percent for procurement of goods, and 7.5 percent for public works projects. Provisions for the use of price preferences are also included in the Model Law on Procurement of the UN Committee on International Trade Law (Beveglia-Zampetti 1997). An advantage of price preferences over other policies that favor domestic procurement is that they permit the price mechanism to continue to work. As such, they are superior to local content and offset requirements, policies that the GPA currently allows developing countries to apply. One option would be an agreement to convert offset and similar policies into price preferences. This would be akin to the conversion of quotas and related measures to tariffs in the Uruguay Round for agriculture (see Hoekman and Kostecki 1995). Even if this meant that price preferences in some countries were relatively high, they would become the focal point for future multilateral negotiations to reduce discrimination. Of course, the use of such preferences should be optional and should not be used to create a preference margin that exceeds the level currently implied by existing policies. Nor should tariff conversion be considered in instances where preference policies do not already exist.

**Potential Negotiating Leverage**

What negotiating leverage could developing countries exercise to pursue better access to foreign markets or temporary exceptions to GPA disciplines? Given the mercantilist nature of bargaining in the WTO context, the size of nonmember (developing country) procurement markets is perhaps the best indicator of leverage. Unfortunately, the data on the size of procurement markets are very poor. Total central government expenditures on goods and services by non-OECD countries in the early 1990s was some $300 billion (table 4). This figure includes conservative estimates for several large countries that do not report data, including India, Indonesia, Pakistan, and South Africa. Data on expenditures by subnational government entities in developing countries are even patchier. If it is assumed that the purchases by such entities are equal to central government expenditures, the total will be at least $600 billion. How much of this is available to foreign suppliers depends on the share of contracts that exceeds the GPA’s threshold values. These thresholds are likely to have a greater effect in developing countries as the average size of each contract can be expected to be less than in a high-income country.

It is also important to recognize that many expenditures by developing country governments are financed through official development assistance funds, both bilateral and multilateral. Official bilateral development aid is often tied to buying goods and services from the donor country; and the recipient government cannot subject
aid-financed purchases to international competition. Absent an agreement from donors to eliminate tied aid, such projects would be exempt from GPA rules. Conversely, procurement financed through multilateral development assistance usually already is subject to international competitive bidding.

Available data on the relative importance of aid flows as a share of government expenditure suggest that aid finances a significant share of the total purchases of goods and services by developing country governments (equivalent to 35 percent of the total in low-income countries, 16 percent in lower middle-income nations, and 6 percent in upper middle-income economies; see table 4). This suggests that much of the procurement by poor countries either cannot be subjected to the GPA or is already subject to international competitive bidding. In practice, however, the GPA is expected to interest the higher-income developing countries in which aid plays only a minor role. It can also be noted that what matters in the WTO context is not only the actual policy stance of a government, but also the extent to which countries bind (lock in) their policies by making formal commitments not to become more restrictive in the future. Thus, even if countries are already relatively open, this does not mean they have nothing to offer. An agreement to lock in a policy regime that is already nondiscriminatory has value in the WTO setting. Negotiating leverage for developing countries is therefore substantial.

Conclusion

Government procurement is a key interface between the public and private sectors. Not only is public purchasing of fundamental importance in ensuring that government gets the best value for its money, but procurement practices figure prominently in the way potential investors and civil society at large view a country. Many developing countries have adopted procurement legislation and regulations that aim to ensure that public entities purchase goods and services through an open and competitive process. To what extent actual practice is consistent with the formal rules and principles is often difficult to determine, in part because the incentives to contest violations of the formal rules of the game are often small. The GPA provides a unique international mechanism through which governments can credibly commit themselves to a transparent and competitive procurement regime and provide participants with more effective enforcement mechanisms than may be available under the status quo. Enforcement is a necessary condition for any set of rules to be effective. This applies as much to industrialized as to developing countries.2

Governments desiring to minimize procurement costs have options that go beyond the adoption of efficient purchasing practices and related procurement regimes. These include the privatization of government operations and private sector participation in markets presently serviced by public entities. The potential benefits of such
initiatives are increasingly recognized. Procurement disciplines may be a second-best option—if not redundant—if market forces can be applied to the activities of public entities. The greater the extent to which public entities operate in a competitive environment and are subject to hard budget constraints, the less likely that inefficient purchasing decisions will be made. Where public entities have a dominant position, however, competition laws and policies are required to ensure that they do not abuse their market power. The same is true of collusion and other restrictive business practices, matters on which the GPA has nothing to say (Wood 1997). This does not imply that procurement regimes will become irrelevant; quite the contrary. Governments will remain important purchasers of goods and services. Indeed, the more activities are contracted out by government agencies, the greater the need for transparent and efficient procedures to allocate contracts.

More research is required to determine the extent to which current procurement regimes in developing countries are compatible with the GPA. Surprisingly little is known about actual practices and their economic impact. Examples include data on the composition of procurement, the import propensity of government entities compared with those of the private sector, the extent to which procurement costs diverge from market price-based comparisons, whether differences in ownership and market structure across countries affect procurement efficiency of similar entities or contracts, and how frequently negotiations are reopened after contracts have been awarded.3

For developing countries, the main quid pro quo for membership in the GPA is reciprocal access to the procurement markets of existing members. This is not a very compelling incentive. Seeking better access to export markets more generally appears to be the best approach, complemented by efforts to seek transitional periods in which to phase in those aspects of the GPA, such as challenge mechanisms, where institutional strengthening will be required.

Notes

Bernard Hoekman is senior economist with the Development Research Group of the World Bank and research fellow at the Centre for Economic Policy Research in London. He would like to thank Marie-Helene Le Manchec and Faten Hatab for excellent research assistance; Annet Blank and Dick Westin for help in obtaining data; and Simon Evenett, Garry Pursell, and Alan Winters for helpful comments.

1. Data reported in IMF (1996) on the average ratio of expenditure on goods and services to total central government expenditures of all the developing countries reporting both variables was used to estimate data for missing countries.

2. In the EU, for example, many members states have been slow to implement European Community directives relating to procurement, and the directives are often violated. In 1994 the Commission of the European Communities initiated some 250 legal actions against members states or entities for inconsistency with procurement rules (Financial Times, November 23, 1996).

3. This is potentially a significant loophole in any procurement regime because it may reflect or induce corruption and collusion between the entity and a favored supplier, who will bid low to win

Bernard Hoekman
the tender in the knowledge that the contract will be renegotiated ex post. This is alleged to be a prevalent practice in some European countries that belong to the GPA (Financial Times, November 23, 1994, p. 2).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.


New Models for Old-Age Security: Experiments, Evidence, and Unanswered Questions

Estelle James

The escalating costs of traditional social security systems are forcing countries to reevaluate the formal programs that provide income maintenance support to the aging. This article suggests a reform strategy built around three systems, or “pillars,” to provide old-age security—a public pillar with mandatory participation, a private, mandatory savings plan, and a voluntary savings system. Three variations of this model are being implemented in different countries: the Latin American model, in which individual workers choose an investment manager for their retirement funds; the OECD model, in which employers, union trustees, or both choose the investment manager for an entire company or occupation; and the Swedish notional account model, a reformed pay-as-you-go first pillar that may be supplemented by a second, funded pillar. Preliminary empirical evidence on the efficiency and growth effects of pension reform, mostly from Chile, indicates that the impact on national saving and financial market development and, through these, economic growth, has been positive and possibly large. Problems concerning high administrative costs and regulations that distort investment decisions remain to be resolved, however.

In the next 35 years the proportion of the world’s population that is more than 60 years old will nearly double, from 9.5 to 16 percent (figure 1). With rapid increases in life expectancy and declines in fertility rates, the population in developing countries is aging much more rapidly than are populations in industrial countries. By 2030, as today’s young working-age people near retirement, 80 percent of the elderly will live in what today are developing countries, and it is essential that policymakers start to plan for the care of their aging populations.

Many of these countries now rely heavily on family assistance and private, voluntary old-age support. But cross-sectional analysis shows that public spending on for-
mal pension plans increases exponentially as populations age (figure 2). In some industrial countries, for example, it now exceeds 15 percent of gross domestic product (GDP) and will soon reach that level in many more countries as the demographic transition proceeds.

With such large sums involved, the way this money is generated and spent can affect the entire economy by influencing productivity, factor supplies, and therefore the size of the GDP. For example, high payroll taxes can lead to unemployment, deficit financing can fuel inflation, and prefunding pension expenditures can be part of a plan to increase national saving. Countries with larger private pension funds have lower public spending, and these two types of spending may have different effects on the broader economy (figure 3). Therefore, two criteria should be used to shape and evaluate these programs: they should protect the old in an equitable way; and they should promote, or at least not hinder, economic growth—which is important for both the old and the young.

In the past most government old-age security systems were pay-as-you-go plans; workers were taxed today to pay pensions to old people today. This article argues
Figure 2. Relationship between Percentage of the Population over 60 Years Old and Public Pension Spending

Pension spending as percentage of GDP


Figure 3. Relationship between Public Pension Spending and Private Pension Assets

Actual-predicted pension/share of GDP

Source: World Bank data.
that because aging is a predictable life experience, saving during younger years can self-insure a large part of old-age security, shifting consumption from younger productive years to older years when consumption exceeds income. Myopia among workers may require that retirement saving be mandatory, but relying to some extent on self-insurance and saving for old age may reduce many of the incentive problems associated with tax and transfer pay-as-you-go systems.

Another aspect of old-age security systems requires pooling risks and insuring or redistributing across individuals because some people will retire early with disabilities, die young and leave dependents, live longer than average and run out of resources, or earn very low lifetime incomes that are insufficient to support them during their working and nonworking lives. This is the rationale for providing a combination of mandatory self-insurance and insurance across individuals in a multipillar system that puts greater emphasis on saving, has separate financing and managerial mechanisms for redistribution and saving, and shares responsibility between the private and public sectors.

During the past few years, many countries have indeed been adopting multipillar old-age security systems. Although structural change is always difficult, the experience of these countries shows that it is possible, that it takes somewhat different forms in different places, and that it usually involves transition costs that are spread over several generations. Preliminary empirical evidence suggests a positive impact on efficiency and growth. But it also brings to the fore new problems—high administrative costs and regulatory regimes that distort investment—that remain to be solved.

Problems with Old Systems

Most formal systems of old-age security today are publicly managed, pay "defined benefits" (meaning a payout formula based on the worker's earnings and years of service), and are financed by payroll taxes on a pay-as-you-go basis. It is now widely recognized that these systems generate many problems, including

- High and rising payroll taxes that may increase unemployment
- Evasion and escape to the informal sector, where productivity is lower
- Early retirement, which reduces the supply of experienced labor
- Misallocation of public resources as scarce tax revenues are used for pensions rather than for education, health, or infrastructure
- Lost opportunities to increase long-term saving
- Failure to redistribute to low-income groups
- Unintended intergenerational transfers (often to high-income groups)
- The growth of a large implicit public pension debt and financing gap that makes the current system unsustainable in many countries.
As a result, existing systems have not always protected the old and are particularly unlikely to protect those who grow old in the future. Moreover, they often have failed to distribute benefits in an equitable way and have hindered economic growth. In addition, they simply are not sustainable in their present form. Not each problem exists in every country, but they are found in most countries, both industrial and developing. This prevalence suggests that these problems are not accidental, but inherent in the economics and politics of pay-as-you-go defined benefit plans—the model preferred by politicians who find it simpler to promise short-term benefits at the expense of large long-term costs.

The Multipillar System

To avoid these dangers, the World Bank has been recommending and many countries have been moving toward a system in which some of an individual’s pension is financed by preretirement savings, which are privately managed. Specifically, these new arrangements contain three pillars:

- A mandatory, publicly managed, tax-financed pillar for redistribution
- A mandatory, privately managed, fully funded pillar for savings
- A voluntary pillar for people who want more protection in their old age.

The first pillar resembles existing public pension plans, but it is smaller and focuses on redistribution—providing a social safety net for the old, particularly those whose lifetime income was low. The benefit formula can be flat (uniform for everyone or related to years of covered employment, as in Argentina and the United Kingdom), it can be means- and asset-tested (as in Australia), or it can provide a minimum pension guarantee (as in Chile). The first alternative provides additional co-insurance and redistribution to lower-middle-class workers, while the last option is obviously cheaper. In some cases (Australia, Chile), the first pillar is financed out of general revenues rather than through a payroll tax. Because this pillar is of limited scope and has a broad tax base, the tax rates needed to support it are much lower than the public system requires in most countries today.

The second pillar differs dramatically from traditional systems. It links benefits actuarially to contributions as in a defined contribution plan, is fully funded, and is privately and competitively managed. (In such a plan the contribution is defined, and the future pension depends on accumulated contributions plus investment returns. In a fully funded system assets are always sufficient to cover future liabilities.) Essentially, people are required to save for their old age.

A third pillar, voluntary saving and annuities, offers supplemental retirement income for people who want more generous old-age pensions.
The most innovative and controversial of these arrangements is the second pillar, so it is worth examining the rationale for its characteristics:

- Why mandatory? The rationale here is myopia—a significant number of people may be shortsighted, may not save enough for their old age on a voluntary basis, and may become a burden on society at large when they grow old.
- Why defined contribution? The close link between contributions and benefits in this plan should discourage evasion, escape to the informal sector, and other labor market distortions because people are less likely to regard their contribution as a tax. And those who do evade bear the cost in the form of lower benefits rather than passing the costs on to others and undermining the financial viability of the scheme. Because the pension is acquired on actuarially fair terms, given the age and accumulation of the worker, these plans are likely to deter early retirement and to raise the normal retirement age automatically as longevity increases—without involving the government in a difficult political decision.
- Why fully funded? First, prefunding makes the costs clear up front so that countries are not tempted to make promises today that they will be unable to keep tomorrow. Second, it avoids large payroll tax increases that are needed in a pay-as-you-go system as populations age. Third, it prevents large, inadvertent intergenerational transfers from young people to older workers. Once an unfunded system is set in motion, intergenerational transfers occur automatically as a result of the aging and maturation process, sometimes in ways that people did not expect and would not have chosen. For example, the early generations to be covered (including its rich members) gain, while later generations (including its poor members) lose, even though they did not have a chance to participate in the political decision that produced this contract. Full funding eliminates such undesirable transfers. And finally, funding may be used to help build long-term national savings. These savings can enhance the productivity of future workers, they can be embedded in consumer durables that provide a stream of future services, and they can be invested abroad and redeemed to finance purchases of consumer goods. Thus, saving can be an important ingredient of a long-term strategy for providing additional domestic consumption when the dependency rate increases.
- Why privately managed? This maximizes the likelihood that economic rather than political objectives will determine the investment strategy, thereby producing the best allocation of capital and the highest return on savings; and it helps countries, especially middle-income countries, develop their financial markets. Empirical data show that publicly managed pension reserves typically earn low, even negative, returns, largely because public managers are required to invest in government securities or loans to failing state enterprises at low nominal interest rates that become negative real rates during inflationary periods. The
hidden and exclusive access to these funds makes it easier for governments to run large deficits or to spend more wastefully than they could if they had to rely on a source of funds for which they were more accountable.

Competitively managed, funded pension plans, in contrast, are more likely to be invested in a mixture of public and corporate bonds, equities, and real estate, thereby earning a higher rate of return. Private pension funds can enjoy the benefits of investment diversification, including international diversification, which protects them from inflation and other country-specific risks and thus enables them to increase their yield and reduce their risk. Private pension funds build constituencies that help them resist political manipulation. They spur financial market development by creating a demand for new financial instruments and institutions. Three caveats obtain, however: countries must have at least rudimentary capital markets before they can put the funded pillar in place; considerable government regulation and regulatory capacity are needed to prevent fraud and excessive risk; and if this regulation is excessive or misdirected, financial markets and investment policies will not be optimal.

All three pillars co-insure against the many risks that old people face, particularly the risk stemming from uncertainty about the future economy or polity—such as breakdowns of the market or the government, changes in relative prices of labor and capital, or a deterioration in the position of a particular country—by diversifying across types of management (public and private), sources of finance (from labor and capital), and investment strategies (equities and bonds, domestic and international). Risk diversification is especially important given the long time periods and great uncertainty involved. Whatever unpredictable disasters occur in the future, as they surely will, this diversified system is most likely to continue providing protection based on the old adage “don’t put all your eggs in one basket.” (See World Bank 1994 for more details. For a quantification of the welfare gains from diversification, see Pujol 1996).

How Have Countries Reformed?

During the past decade, and particularly during the past five years, several countries have adopted variations on this multipillar system. The three major variations are the Latin American (individual account) model, the Organisation for Economic Co-operation and Development (OECD) (employer-sponsored) model and the Swedish (notional defined contribution) model. Experience with these variations shows that pension reform is possible, even in democracies, but that it takes somewhat different forms as a result of differing initial conditions and political economies.
For example, different conditions led Argentina to choose a relatively large public pillar, whereas Peru decided against a public pillar. Chile and Australia chose a much larger private pillar than Mexico and Argentina. The United Kingdom and Switzerland built on a history of employer-sponsored plans; Australia and Denmark built upon widespread, union-negotiated plans, and Sweden and Italy adopted a defined contribution plan that remains largely pay-as-you-go.

One of the most important initial conditions that influences the shape of the reform is the implicit pension debt—the present value of the pensions that are owed to current pensioners and to workers according to their years of participation in the old system. This debt is inherent in pay-as-you-go systems, where workers expect to get a specified pension in return for their contributions. But assets are not accumulated to cover this debt; instead, the obligation is covered by implicit government IOUs. In many countries, the implicit debt exceeds the country's conventional explicit debt (backed by government bonds) and in some cases exceeds 200 percent of GDP (table 1). It is especially large in countries with high coverage, generous benefits, and older populations. Although this debt is not always legally binding, it tends to be socially and politically binding; governments cannot easily renege on these obligations. Countries that do not want to make their debt transparent frequently shy away from a shift to a funded pillar because it makes at least part of the implicit debt explicit. Most developing countries have little pension debt because of their low

<table>
<thead>
<tr>
<th>Country</th>
<th>Implicit pension debt as a percentage of gross annual product</th>
<th>Percentage of population over 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senegal</td>
<td>27</td>
<td>4.3</td>
</tr>
<tr>
<td>Mali</td>
<td>26</td>
<td>4.9</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>15</td>
<td>5.0</td>
</tr>
<tr>
<td>Venezuela</td>
<td>30</td>
<td>5.6</td>
</tr>
<tr>
<td>Peru</td>
<td>37</td>
<td>5.8</td>
</tr>
<tr>
<td>Cameroon</td>
<td>44</td>
<td>5.8</td>
</tr>
<tr>
<td>Congo</td>
<td>30</td>
<td>6.1</td>
</tr>
<tr>
<td>Brazil</td>
<td>187</td>
<td>6.7</td>
</tr>
<tr>
<td>Turkey</td>
<td>72</td>
<td>7.1</td>
</tr>
<tr>
<td>Albania</td>
<td>67</td>
<td>8.1</td>
</tr>
<tr>
<td>China</td>
<td>63</td>
<td>8.9</td>
</tr>
<tr>
<td>Uruguay</td>
<td>214</td>
<td>16.4</td>
</tr>
<tr>
<td>Croatia</td>
<td>350</td>
<td>17.8</td>
</tr>
<tr>
<td>Ukraine</td>
<td>141</td>
<td>18.7</td>
</tr>
<tr>
<td>Hungary</td>
<td>213</td>
<td>19.3</td>
</tr>
</tbody>
</table>

Note: Assuming a 4 percent discount rate.
Source: Kane and Palacios (1996); World Bank data for Albania, Burkina Faso, Congo, and Mali.
coverage rates and are therefore in the enviable position of being able to change to a partially funded system before the debt becomes unmanageable.

The Latin American versus the OECD Models

Chile pioneered the Latin American model in 1980, and its initial success there led Argentina, Bolivia, Colombia, Mexico, Peru, and Uruguay to adopt similar plans in the 1990s. Hungary and Kazakhstan were the first countries outside the region to adopt the model, and it is one of three options proposed by the Social Security Advisory Committee in the United States. In this model, each worker chooses the investment managers of his or her own individual defined contribution retirement account.

By comparison, the OECD model builds on existing employer-sponsored pension plans. These plans simply became mandatory instead of voluntary in Australia, Denmark, Switzerland, (and de facto by collective bargaining in the Netherlands); in the United Kingdom, employer-sponsored plans became an attractive optional alternative to the state plan. Under this model the employer or a combination of employer and union trustees chooses the investment manager for each company or occupational group. These plans thus benefit from economies of scale and financial expertise and possibly from lower marketing costs (although this has yet to be proven). The OECD model, however, introduces a principal-agent problem; that is, the employer or union representative selects the investment manager, but the workers bear the risk. The choice may not be in the worker’s best interest and may not maximize net returns. For this reason, workers in OECD model plans may ultimately demand more individual choice.

For example, Australia is now permitting workers to put their retirement savings into special bank accounts. In the United Kingdom employers were initially permitted to opt out of the state earnings-related plan, but subsequently workers were given the right to opt out of their employer’s plan in favor of their own personal retirement plan. Unscrupulous insurance company salesmen then persuaded workers to purchase individual annuity plans, when, in fact, they would have been better off staying in the employer’s defined benefit plan. The incident, which led to lawsuits and a government inquiry, illustrates the point that worker choice makes consumer information imperative and opens the door to the probability that some mistakes will be made (Johnson, forthcoming; Whitehouse 1998).

In most of the OECD model countries cited above—unlike those in Latin America—a modest public pillar with a small pension debt and little or no payroll tax financing was already in place when the new system was adopted. Thus they could simply retain it and build the second pillar alongside the first. They had no trouble financing the transition because accrued rights were small and the contributions to the second pillar were added on, rather than being diverted from the first pillar.
For example, Australia had a means- and asset-tested first pillar, financed out of general revenues, to which it simply added a mandatory, employer-based, funded pillar financed by payroll contributions. General revenues also financed Denmark's flat benefit in the public pillar, now being reduced, while an occupational funded pillar is added. In the United Kingdom the state earnings-related pension had been initiated just a few years before Margaret Thatcher decided to end it by encouraging employers and workers to opt out; the accumulated rights were still very small. In Switzerland employer-run plans that already existed in many firms became mandatory alongside the public pillar.

In contrast, bloated public pillars in the Latin American countries meant that the first pillar had to be downsized and redesigned to create space for the second pillar. When a worker switched to the new system, he was given credit for his past service under the old system while part of his future contributions were diverted to the new second pillar. These countries had to find the money to continue paying the promised benefits to current pensioners and older workers (the implicit pension debt) under the old system, while part of the payroll tax flowing in was diverted to funded individual accounts—a problem that has become known as “financing the transition.” Most countries that reform in the future will have to solve this problem.

The Swedish Model: Notional Defined Contribution Plans

Many countries with large public pillars and implicit pension debts have found it exceedingly difficult to make the transition to a partially funded system with a mandatory private pillar, in part because of the financing problem, but also because of the political interests associated with existing institutions. This explains the third group of reforming countries—those that feature notional defined contribution plans. In this plan the worker has an individual account that is credited with his contributions plus interest. The accumulation is notional, however, rather than actual, since the money paid in by workers is immediately paid out to pensioners rather than being invested; in fact, the system remains pay-as-you-go. Upon retirement, the notional accumulation is converted into a real annuity, supposedly on actuarially fair terms. Thus the notional defined contribution plan is essentially a reformed pay-as-you-go pillar (sometimes accompanied by a second, funded pillar). The problem of transition costs is substantially avoided.

Sweden developed this system, although it has not yet been implemented there. Shortly after Sweden acted, Italy adopted the system, but with a long transition period. In both cases, the first pillar is to be converted into a notional defined contribution plan, buttressed by a redistributive guaranteed pension. The system is also being implemented by Latvia, which hopes to save enough money from reducing
evasion and early retirement eventually to start a funded pillar. Poland plans a new system with a notional defined contribution first pillar and a funded second pillar. Outside of Europe, China has a notional defined contribution system, de facto. In principle China wants to start a second pillar made up of funded individual accounts, but many cities have been unable thus far to finance the transition, so the individual accounts remain largely notional.

The notional defined contribution system was designed to capture some of the advantages of linking benefits closely to contributions within each cohort. Most important, it reduces idiosyncratic intracohort inequities and labor market distortions, including incentives to evade—providing the notional interest rate is close to the market interest rate. For example, early and late years of contributions receive the same rate of return, and workers with flat age-earnings profiles receive the same rate of return as those workers with steep profiles; this equality is not true of most defined benefit plans. In addition linking benefits to contributions makes the system more sustainable and avoids the selection problem that occurs when low return people evade but high return people stay in the system.

Furthermore, the notional defined contribution system discourages early retirement because workers automatically receive lower benefits if they retire early; and the costs of early retirement are internalized rather than being passed on to others. For the same reason, it automatically induces workers to retire later as longevity increases, thereby avoiding the difficult political decision to raise the retirement age, a step that is periodically necessary in defined benefit plans.

The notional defined contribution plan is not inherently redistributive, however, so it does not accomplish the first pillar task of protecting low-wage earners. For this purpose a redistributive “O pillar” must be added that will guarantee a minimum pension to workers whose own defined contribution pension falls below a specified minimum. Such a guarantee assures a minimum pension even if the employee contributed very little. If the redistributive O pillar is large, it may override the link between contributions and benefits. In Sweden, for example, the high level of the guaranteed minimum pension in the O pillar makes the defined contribution component irrelevant for the majority of workers—especially women, who are more likely to have spent only a part of their adult lives in the labor force.

A bigger failing is that the notional system does not capture the benefits of funding, because there are no funds. That is, the system serves as the first (pay-as-you-go) pillar and crowds out the opportunity for a large funded pillar. Intergenerational transfers remain, saving is not augmented, and financial markets do not develop. Most important, as the dependency rate increases, the contribution rate would have to increase to keep the system solvent in the absence of prefunding. These younger cohorts may have to “save” a much larger amount for their old age than is optimal.

Estelle James
for them in order to cover benefits promised to older cohorts. In that case, the incentives for evasion and escape to the informal sector would be strong.

Sweden plans to build a buffer fund to reduce the need for large tax increases for the first pillar as its population ages. But this buffer fund will be a publicly managed overlay, because the individual accounts remain notional. This situation raises all the problems, summarized earlier, concerning political manipulation and poor allocation of publicly managed funds. The pillar will be supplemented by a small second pillar funded at 2.5 percent of payroll.

How are the notional interest rate and the conversion rate of notional capital into annuities determined? If the notional interest rate is higher than the market rate, it will be a costly guarantee for the government to fulfill. If it is less than the market rate, the contribution is more likely to be regarded as a tax, so labor market distortions are likely to reappear, and pressures may arise for an increase. Typically the interest rate is set equal to some exogenous rate to insulate it from such political manipulation; nevertheless, the possibility remains that a future government will discard this connection and arbitrarily change the rate. Most commonly thus far, the notional rate has been tied to the growth in the per capita wage, or the covered wage bill—supposedly an equilibrating device. If the wage bill increases, so too do contributions, and therefore the ability to impute interest will be high. This means, however, that when the working-age cohort is large and growing (for example, the baby boomers), the imputed interest rate is high, and the pension debt increases rapidly. But when the working age cohort declines (generation x), so too does the notional interest rate. This generation must then pay a high contribution rate to cover the pension debt and will receive a low notional interest rate—fertile grounds for evasion and questionable from the viewpoint of intergenerational equity. Thus using wage bill growth as the notional interest rate does not appear to be an equilibrating device (see Schwarz and Valdes, forthcoming).

The conversion factor into annuities supposedly depends on expected longevity upon retirement. Because the process is notional, however, it too is highly subject to political manipulation. For example, the government can decide to grant notional credit for noncontributing years (a common problem in old pay-as-you-go defined benefit systems), it can impute a low or a high future interest rate into the calculation, and it can fail to adjust the conversion factor when life expectancy increases. In the absence of market discipline, implicit taxes or subsidies can creep in that interfere with the labor market efficiency effects of the new system. Because the government sets both the conversion factor and the interest rate, the notional defined contribution may be thought of as a pay-as-you-go defined benefit in which the benefit is defined in a new way.

In sum, the notional defined contribution system is attractive to countries that have very large implicit pension debts, especially those that are unwilling to incur an explicit fiscal deficit to pay off these obligations. It may be a politically convenient way to
reduce benefits in inflated programs and to equalize the retirement age for men and women (as in Latvia). In such cases it may lay the groundwork for savings that eventually enable the growth of a funded second pillar; but until that happens, it should be recognized as a reform of the first pillar rather than as an introduction of a multipillar system. (For details on the Swedish reform, see Sunden 1998; on Latvia, see Fox 1998; on Poland, see Rutkowsky and others 1997; and on Italy, see Hamann 1997).

How Have Countries Financed the Transition?

When countries with a large pay-as-you-go pension debt shift to a multipillar system that includes a funded component, some of the contribution usually is shifted to individual accounts. Some other revenue source must then be found to cover the resulting financing gap between the remaining revenues and the expenditures needed to pay retirees. The only countries to have experienced this problem are those that have followed the Latin American model. How did these countries finance the transition? Three basic methods were used: reducing the value of the pension debt and the financing gap, finding alternative revenue sources to pay it off, and, finally, resorting to the general borrowing and taxation powers of the treasury.

Reducing the Implicit Debt and the Financing Gap

The implicit social security debt and the gap between payroll taxes and expenditures can be reduced in several ways. First a country can take certain steps before the transition, such as downsizing benefits provided under the old system, raising the retirement age and the penalties for early retirement, tightening eligibility for disability benefits, and changing the indexation method to price indexation, so the outstanding debt, whether implicit or explicit, will be smaller. Argentina, Chile, and Uruguay followed this strategy, which may be indispensable to a good pension reform. It cuts the benefits that must be paid to those who stay in the old system, as well as the compensation owed to those who switch to the new system, and it increases the probability that workers will switch. Otherwise, there is a risk that the government will pay excessive amounts for benefits that never should have been promised in the first place, and it will be more difficult than before to escape from these promises.

Second, the government can acknowledge the value of the pension earned thus far by issuing a recognition bond (as in Chile) or a promise of a compensatory pension (as in Argentina) to each worker who switches to the new system. This step postpones the day when cash will be needed, because the recognition bond cannot be cashed until the worker retires, and the compensatory pension is gradually paid off over the entire retirement period of the worker. Besides extending the pay-off period, the issuance of the recognition bond provides another opportunity to reduce
the debt. A legally binding piece of paper, the bond gives the worker greater certainty that the pension debt will eventually be repaid, and in return for reducing uncertainty, the government can downsize the face value of or interest rate on the bond (as in Peru). The face value can be further reduced if workers have more faith in the new system than in the old one; they (especially young workers) will then be willing to switch even with little compensation for their past service. By choosing the minimum terms that are needed to convince the desired number of workers to switch, a government can substantially downsize the recognized debt and save on its transition costs (as in Hungary).

Third, a government can keep some workers—and their contributions—in the old system. This may be accomplished by excluding some workers, such as the military or the police, from the new system (as in Chile). Argentina gave all workers a choice but made the new system attractive mainly to young workers. Colombia operates the old system side by side with the new one, and workers are permitted to switch back and forth. In Uruguay the new funded pillar is compulsory only for rich and young workers and is voluntary for others. The financing gap is reduced because those employees who remain in the old system continue to contribute to it. The serious danger with this option is that, in an effort to solve a short-run cash-flow problem, these countries have increased their long-term implicit debt by keeping participants in a financially unsound pay-as-you-go system; this solution may turn out to be unsustainable.

Fourth, a government can retain a large pay-as-you-go component in the new system, so that some revenues continue to flow into the public pillar. Argentina offers a moderate-sized flat benefit in its new public pillar rather than the narrower minimum pension guarantee used in Chile. In Argentina about 60 percent of the total contribution flows into the public pillar. In addition, workers can choose between a funded and a pay-as-you-go option for the second pillar. The inflow of funds to the first pillar and the pay-as-you-go second pillar help pay current pensioners and, eventually, the compensatory pension. But if either pillar offers benefits that are too generous (actuarially unsound), the reform will not be sustainable in the long run—a danger that Argentina faces.

Finding Alternative Revenue Sources

Governments can also pursue policies to offset the revenue gap. One way is to use an existing treasury surplus to pay off part of the pension debt. Chile took this path, but most countries are burdened with fiscal deficits rather than surpluses. Alternatively, countries that have a surplus in the social security system can use it to pay off part of the debt. The Latin systems generally did not have a surplus, but the U.S. social security trust fund could be used in this way, if the United States were to make a transition. In cases where public enterprises are being privatized, some of the pro-
ceeds can be used to pay off the pension debt—a cancellation of long-term assets against long-term liabilities. Peru followed this strategy, Poland is considering it, and Bolivia is also using privatization assets for pension reform.

Measures to reduce evasion and increase coverage will increase system revenues. Although Argentina’s plan incorporated such measures, no reduction in evasion has yet materialized. China is considering financing the transition by bringing all workers in township and village enterprises (a rapidly growing group) into the new partially funded system. This expansion of coverage would help to pay off the accumulated pension debt but simultaneously would create new debt to cover the newly enrolled workers who will eventually demand their pensions, and may lead to evasion in the interim. Thus, this strategy produces short-term revenues but runs the risk of undermining the long-run sustainability of the plan and its credibility unless the payroll tax and promised benefits are low.

**Using General Borrowing and Taxation**

General treasury debt can be used to cover the remaining cash gap in the short run. Because money is fungible, it is not clear to what extent resources for pension reform have come from debt as opposed to other general revenue sources, but government borrowing has usually increased in the early years of reform. In countries with a large implicit pension debt, the use of temporary debt finance is almost inevitable to mitigate the heavy double burden of taxation on the transition generation of workers. Some of this debt may be sold to the pension funds in the new second pillar; government debt and bank deposits have been the largest initial investments of the new pension funds. An important proviso is that government bond sales should be open, transparent, and carry the market interest rate. Pension funds should not be compelled to purchase government bonds. All Latin American countries limit pension funds’ overseas investments, however, which virtually ensures that they will have large investments in domestic government bonds.

Is this temporary debt finance problematic? Financial markets might react negatively if they were not previously aware of the size of the implicit pension debt, or if they believed the obligation to repay it was “soft” and has now become “hard,” and if either of these beliefs increases the expected default risk on regular bonds. Two pieces of evidence suggest that, so far, the financial market response has been positive. First, the International Monetary Fund recently adopted the position that debt finance earmarked for a pension transition should be allowed beyond the permissible ceiling for other debt, because it is a swap of implicit for explicit debt in the short run and is intended to reduce the overall debt and will thus improve fiscal solvency in the long run. Second, for much the same reason, Hungary’s credit rating from Moody’s improved after it adopted its pension reform, even though the reform entailed an increase in the explicit debt.
Eventually, the debt should be paid off through taxation. Otherwise there will be no increase in national savings (additional private saving will be offset by additional public dissaving if the implicit debt is simply changed to an on-going explicit debt). The redemption of the debt through tax revenues can be spread over a long period of time—but the longer the payoff the slower the benefits of increased national saving for productive investment. It has been estimated that if half the current pay-as-you-go system in the United States were converted to a funded system, it would take 70 years to pay off the financing gap with a payroll tax rate of about 1.5 percent (Gramlich 1996); it would take roughly the same amount of time in China (Friedman and others 1996).

How Large Are the Efficiency and Growth Effects of Alternative Systems?

The chief theoretical argument for the recommended multipillar system is that it will have a positive effect on efficiency and growth because the old system introduced—or failed to remove—distortions that the reforms will eliminate. A second argument is that the multipillar approach will enhance the financial sustainability of the old-age system and thereby provide better protection for the elderly in the long run. A third argument is that it will improve intergenerational equity.

Efficiency and growth effects are notoriously difficult to quantify and prove, in part because relatively little experience and data are available and in part because, even with the data, it would be difficult to build models that capture all the complex dynamic interactions; that is, it is difficult to specify the counterfactual. Pension reform has several different potential efficiency effects; usually studies focus on one of these while ignoring or holding the others constant. For example, general equilibrium models that analyze labor supply effects often assume perfect capital markets and thereby limit the predicted increases in savings, and vice versa. In this section I summarize the limited empirical research that has been done on these topics, concentrating on the simulated effects in countries that have been considering structural reforms and econometric estimation of the actual effects in Chile, the country that has the longest track record with a reformed system. In general, the beneficial labor market effects come from shifting to a defined contribution system from a defined benefit plan; the beneficial impact on savings comes from shifting to a funded old-age security plan from a pay-as-you-go system; and the financial market impact comes from managing these funds privately.

First, a brief comment on the distinction between efficiency and growth. Greater efficiency, for example, due to a reduction in labor market distortions, increases the
level of output. If some of the increased output is plowed back into investment, as would often be the case, growth also increases. Growth can also be increased without an increase in efficiency. For example, an increase in savings (and consequently growth) may simply indicate an intergenerational or life cycle redistribution that does not increase efficiency because it does not make (or have the potential to make) everyone better off. But such an increase enhances efficiency if the initial rate of saving was suboptimal because of public or private myopia or because of a tax wedge between private and social returns to investment. Both of these conditions are usually alleged as a justification for mandatory retirement saving plans, in which case they would expand both efficiency and growth.

Avoiding Labor Market Distortions

One problem in pay-as-you-go defined benefit systems is the possibility that the high payroll tax will lead to labor market inefficiencies (stemming from distorted decisions about labor force participation, age of retirement, hours worked, choice of job and location, degree of effort, form of compensation, and so on), whereas the contribution in a defined contribution system may be regarded as saving rather than as a tax. Only fragmentary evidence is available about the effect of pension reform on most of these actions. For example, Wise (1997) shows that the labor force participation rate of older men is highly sensitive to the implicit social security tax on labor, stemming from the absence of actuarial penalties on early retirement—the loss of generous defined benefits during years when they continue working induces most workers to stop working before they reach age 60. Countries that have a larger actuarial adjustment in their systems, hence a lower implicit tax on labor, have higher labor force participation rates of older men. Funded defined contribution plans automatically build in this actuarial adjustment, so by extension they should deter early retirement and its negative impact on GDP and the financial solvency of the scheme.

The distortionary labor market effects of traditional systems may be larger in developing countries because escape to the informal sector is easier there, both for workers and their employers. Productivity in the informal sector may be lower because firms have less access to product and credit markets or because technological change is embodied in capital in the formal sector and has an external effect on labor productivity throughout the economy (as discussed in the endogenous growth literature). In addition, regulations that set a minimum wage and other benefits in the covered sector may lead to a wedge between wages and productivity in the formal versus the informal sectors. In simulations for a representative economy, Corsetti and Schmidt-Hebbel (1997) show that a payroll tax rate of 20 percent could cause a massive (47 percent) shift to the informal sector, thereby reducing economywide
growth by more than 1 percent annually. In many Latin American countries the informal sector and small firms in the quasi-informal sector do indeed absorb more than half of the labor force (ILO 1996). Although many other forces are at work, a shift to a defined contribution system, where benefits are closely linked to contributions, might reduce these incentives for informality.

What light does Chile's experience throw on this issue? Between 1980 and 1990, when the average share of informal employment in Latin America increased from 26 percent to 31 percent, it dropped from 36 percent to 31 percent in Chile. Unemployment in Chile fell and wages rose. Edwards (1997) shows that, given reasonable assumptions about the elasticity of labor demand in the two sectors, the pension reform was responsible for a decline of between 2.2 and 3.6 percent in unemployment and an increase of 5 to 8 percent in wages.

In evaluating these numbers and their applicability to other countries, it is important to realize that a shift to defined contribution may not always have this salutary effect. For example, myopic workers may continue to evade contributions because they will not have access to their mandatory savings for many years. In periods when investment returns are low, workers may be especially tempted to evade, preferring to consume or to invest in education, housing, or consumer durables. In Chile returns have been high (more than 12 percent real during the first 15 years), encouraging compliance of most workers. If the payroll tax for pensions is only a part of the total payroll tax, the incentive to escape to the informal sector may remain strong; again, this phenomenon did not apply in Chile, where the total payroll tax was relatively small. Indeed, preliminary evidence from Argentina suggests that evasion has not declined since the new system was established (Valdes-Prieto, forthcoming). In contrast, Chamorro (1992) and Schmidt-Hebbel (1996) found that only 5 percent of potential contributors in Chile had dropped out of the system. (Chile does not even attempt to cover the self-employed, who make up the largest share of evaders in other countries.) It is difficult to be conclusive about this, because it is hard to separate evasion from normal labor force withdrawals and exogenous shifts into self-employment.

Escape to the informal sector under a funded defined contribution plan does not have the same negative effects on the system's sustainability as it does under a pay-as-you-go defined benefit plan, because the costs are borne by the worker in the form of lower benefits rather than being passed on to others in the form of a higher contribution rate. This is a big plus. Nevertheless, it still creates the same problem for labor allocation and productivity and an even greater problem for the workers who may not have an adequate pension and may become a charge on the public treasury when they grow old. So although the initial evidence from Chile is encouraging, it is important to analyze the data on evasion, wages, and employment carefully to determine whether these results are robust and generalizable.
**Increased National Saving**

A major rationale for fully funded pension plans is that they increase long-term national saving, with positive effects on growth and efficiency. Such saving is important because most savings stay in the country of origin and most of a country's productive investment comes from its own saving, despite the development of global capital markets.

When a country without a prior pay-as-you-go system institutes a multipillar system, consumption will decrease and saving will increase if the mandatory saving rate exceeds the voluntary rate. When a country with an existing pay-as-you-go system replaces it with a multipillar system, national saving increases if benefits are cut or taxes are increased, usually to cover transition costs. In both cases, putting part of the contribution into the worker's own mandatory saving account may be more politically acceptable and less economically distortionary than increasing saving through high taxes that go into the general treasury.

But this increase might not materialize. Mandatory saving may not increase total private saving if individuals find ways to offset required saving by reducing voluntary saving or accumulated assets. In that case, capital may increase in the mandatory pillar but decline in the voluntary pillar. With perfect capital markets, private saving will not increase at all, because people will simply borrow against their mandatory pension saving. A positive saving effect ultimately depends on the assumptions that voluntary long-term saving and assets are small and that borrowing opportunities are limited for substantial groups within the population. The low asset condition probably holds for most slow-growing economies, the limited borrowing condition for most developing countries, and both conditions for low-income households in most countries.

Public saving matters as well as private saving. On one hand, pension reform may reduce public dissaving as governments no longer need to borrow to cover escalating pension costs, but on the other hand, it may increase public dissaving if the build-up of pension reserves relaxes fiscal discipline and makes it easier for governments to run large deficits. If the transition is fully financed by borrowing, government dissaving will offset private saving, and the expected increase in national saving will simply not occur. But if it is financed through taxes or cutbacks in other government expenditures, public saving increases national saving. Estimating the impact on public saving therefore requires modeling government behavior—how governments will behave after pension reform and how they might have behaved in the absence of reform.

Several simulations have projected the impact on saving of a shift to a fully funded scheme. Not surprisingly, the results turn out to be highly dependent on the assumptions, especially the assumptions about the crowding out of voluntary saving and the method of financing the transition. Underlining the importance of the former, simulations of a representative economy indicate that a tax-financed transition to a fully funded system in the presence of credit constraints on consumers (implying
low crowd-out) will increase output by 22 percent and welfare by 16 percent in the long run, while the gain is only 2 percent without credit constraints (Cifuentes and Valdes-Prieto 1997).

In planning its new funded pillar, Australia assumed that half of the mandatory saving would be offset by reduced voluntary saving (the crowding-out effect) in the case of newly covered workers and even more for workers who already were covered by voluntary occupational plans. This assumption implied that when the contribution rate reached 12 percent, national saving would rise by 1.5 percent of GDP, almost doubling the current net national saving rate, which is 2.2 percent of GDP. (The gross national saving rate is about 15 percent of GDP.) Australia did not have to borrow to pay off a pension debt because the second pillar was an add-on rather than a diversion of previous contributions. Although the tax deductibility of contributions was initially projected to cause some government dissaving, in the long run the decreased burden on the means-tested public pension is expected to reduce government dissaving. One of the main effects of the reform may be to shift the allocation of private saving away from home ownership, which is now the predominant investment, because of tax inducements, and toward other, more productive, investments (Bateman and Piggott, forthcoming).

In his simulations for Mexico, Ayala (1996) assumes a 30–40 percent rate of crowding out. If the transition is tax-financed, or if it is debt-financed and Ricardian equivalence holds (so that private saving goes up to offset public dissaving), total saving rises by 0.4 percent to 2.1 percent of GDP, a magnitude similar to that expected in Australia. If the transition is debt-financed and Ricardian equivalence does not hold, the impact on total saving is much smaller, even negative in some years, although positive overall during the next 30 years.

Only Chile has had a mandatory saving plan long enough to permit an estimate of its effects. Data from Chile are problematic, and the savings ratio is erratic, complicating the analysis and making the results highly sensitive to the starting date for comparisons. According to Corsetti and Schmidt-Hebbel (1997), private sector saving as a percentage of GDP increased from almost zero in 1979–81 to 17 percent in 1990–92, while private consumption decreased commensurately. Their reduced-form two-stage-least-squares regressions attribute half of the decline in the private consumption ratio to the growth of Chile’s funded pension plans and correlated developments, such as capital market deepening. Time series regression analyses by Haindl (1996) indicate that pension reform accounts for 6.6 of the 9.9 percentage-point increase in the national saving rate in Chile (from 16.7 percent of GDP in 1976–80 to 26.6 percent in 1990–94). Of the 6.6 point increase, 3.1 points were attributable to the direct impact of pension saving; the remaining 3.5 points were attributable to the financial market deepening caused by the pension fund (4.2 percentage points), offset slightly by a crowding-out effect caused by borrowing constraints (0.7 points). Using an error correction model, Morande (1996) also finds a significant positive
effect of a pension fund dummy on private saving from 1960 to 1995. He speculates that the financial market deepening caused by pension reform may have made voluntary saving less likely to be crowded out by, and therefore less sensitive to fluctuations in, foreign saving, thus making the country’s supply of investible resources less dependent on foreign capital.

Agosin, Crespi, and Letelier (1996) are more skeptical; they find that increased private saving resulted from an increase in corporate saving, from 6 percent of GDP in 1978–85 to 23 percent in 1994—a response, they believe, to the lack of foreign credit and the privatization of public enterprises. (Of course, privatization was itself facilitated by the pension reform, illustrating the complex interactions among these variables.) Voluntary household saving was negative (about 4 percent of GDP throughout this entire period), indicating consumer dissaving or borrowing. Forced saving through the new pension system gradually grew to almost 4 percent of GDP, however, and was not offset by greater voluntary dissaving (presumably because credit constraints had already been exhausted). This 4 percent magnitude is roughly consistent with the findings of Bosworth and Marfan (1994) that pension reform increased saving by 3 percent of GDP. The risk remains that the growth of consumer credit, possibly fueled by the pension reform, could increase consumer dissaving and offset some of these gains in the future (Holzmann 1996).

Instead of focusing on enhanced private saving, other studies emphasize the impact of pension reform on public saving and dissaving. Chile financed a pension transition in part through deficit finance, which decreased national saving. The fiscal costs of the transition may have canceled out the positive effect on private saving initially (Agosin, Crespi, and Letelier 1996). Observing that the pension-related deficits of the government (payments to pensioners from the old system plus redemptions of recognition bonds for new pensioners who had switched) were larger than the inflows to the new pension funds until 1989, Holzmann concludes that during the 1980s the new pension system had a negative effect on national saving. He appears, however, to overlook the fact that redeemed recognition bonds became part of private pension saving and were not immediately consumed. Correcting for this point alone generates a positive savings effect as early as 1985.

More important, a simple accounting exercise neglects the disciplining effect that pension reform might have had on other government taxes and expenditures. Chile ran an increasing surplus during this period, possibly to help cover the transition costs. Since 1987 the consolidated government budget has been in surplus, which quickly exceeded 5 percent of GDP. In addition Chile accumulated a large budgetary surplus in preparation for the reform, thereby reducing its need for deficit finance. How large the current or past surplus would have been otherwise is unknown, but to the degree that the pension reform was financed by increasing general taxes, cutting other public spending or accumulating a prior surplus, transition costs did not decrease public saving. Moreover, the transition costs are short run, while the increased
private saving may persist in the long run. As a result of all these factors, total national saving in Chile is currently much higher than it was before reform.

Given the high correlation between pension reform and other policies that are often simultaneous, the controversy surrounding the determinants of private saving (for example, which variables are endogenous?), and the even greater uncertainty about the determinants of public saving (what is the counterfactual?), all these econometric and simulation results are highly sensitive to the model's specifications. Nevertheless, preliminary evidence indicates that pension reform can have beneficial effects on long-term national saving, especially if it is accompanied by broader policies designed to constrain consumer and government borrowing.

Financial Market Development

One reason for favoring private management of pension funds is that it will develop a set of financial institutions—investment managers, insurance companies, and banks—that are essential for economic development. A funded pillar, if competitively managed and well regulated, can enable the financial market to grow in safety, size, depth, and complexity. In developing countries, where private saving is already high, one of the main effects of a funded pillar may be to shift savings out of land and jewelry and into long-term financial market investments that are better for the broader economy.

Even in Australia the financial market is expected to grow as a result of the mandatory second pillar. As noted earlier, some private saving may be redirected out of owner-occupied housing into the financial markets. Insurance companies are expanding, developing new products, including annuities, to meet the anticipated demand stemming from pension funds (Bateman and Piggott, forthcoming). In Switzerland too, the growth of the life insurance industry, investment companies, and mutual funds have been spurred by mandatory funded pension plans. And corporate governance has been gradually changing, as institutional investors have demanded disclosure and better performance (Hepp, forthcoming). All these changes enhance efficiency.

But the strongest evidence for this expected growth effect comes from Chile. During the five years preceding the adoption of its new system, Chile prepared the groundwork by organizing a primary market for treasury bonds, reforming the laws governing mutual funds, corporations and securities, privatizing banks, authorizing a price-indexed mortgage bond market, and liberalizing the provision of insurance and reinsurance (Valdes-Prieto, forthcoming). Once the system was introduced, this process continued; financial markets became more liquid as stock market trades increased; demand was created for the equities of newly privatized state enterprises; information disclosure and credit-rating institutions developed; the variety of financial instruments grew, and asset pricing improved. In several of the studies summa-
rized above, financial market deepening associated with pension reform was given credit for the observed increase in private saving. Econometric analysis suggests that financial market efficiency induced by the pension reform (and related factors) increased total factor productivity 1 percent a year, or half of the increase in total factor productivity (Holzmann 1996).

In sum, a small but growing body of empirical evidence indicates that pension reform has produced positive efficiency and growth effects. That is, the impact on saving, productivity, output, and welfare may be high relative to exogenous sources of growth and other policies available to increase growth.

Several caveats are essential in interpreting this evidence. First, because specifying the counterfactual is difficult, these results are highly sensitive to the assumptions that are made. In particular, the econometric analyses for Chile are subject to omitted variable bias, and the simulation results depend heavily on assumptions about crowding out, transition costs, and rates of return. Second, the growth impact also depends on key policy decisions in setting up the new system, such as the question of how high the required contribution rate will be, the proportion of the multipillar system to be funded and defined contribution, and how the transition will be financed. Debt finance may be necessary for political purposes, but some degree of tax finance is necessary to meet the economic objectives, and of course some taxes have better efficiency properties than others. Third, it is important to remember that, even if it claims to use a general equilibrium model, each study typically deals with only one possible source of growth, so many of these results are partially additive—that is, the total growth effect is the sum of the separate effects on labor market distortions, early retirement, escape to the informal sector, capital accumulation, financial market development, and other sources of growth. So if each separate effect increases GDP in amounts ranging from 1 to 10 percent, their sum may increase GDP much more.

New Problems and Issues for Further Research

Although many efficiency gains seem to have been achieved, the new systems have also created problems that must be solved and related research that needs to be done. The problems involve high administrative costs, financial market distortions, and distributional effects.

Administrative Costs

The big advantage of private investment over public is the likelihood that it will produce a better allocation of capital and therefore higher returns for the fund and growth for the economy. Decentralized systems also may charge high administrative
fees, however, partly because of high marketing costs in competitive industries. Costs sometimes produce important side benefits, such as consumer information and increased compliance, but this does not appear to be the case in most countries that have recently reformed.

Preliminary evidence indicates that workers are ill informed and do not make decisions based on fees or investment returns, and that pension funds incur high sales commissions and other marketing costs to attract them. In Chile and other Latin American countries, fees are front-loaded, meaning that workers pay a one-time fee on new contributions rather than an annual fee based on assets. (This system was probably adopted because the new system initially had no assets.) Specifically, this one-time fee is about 2 percent of wages or 15–20 percent of new contributions, and about one-third of this fee is for marketing.

These numbers appear very high. To understand their impact on net returns, these one-time charges on contributions must be converted into their equivalents in annual charges on assets, a conversion that depends on how large the assets are relative to the contributions. Obviously, for accounts that have small accumulated assets (young workers with few years of contributions), the one-time fee will be high relative to assets. For accounts that have built up substantial assets over the years, however, the fee will be small relative to assets.

Simulations show that if the current fee schedule is maintained, the average Chilean worker who contributes for 40 years will pay the equivalent of less than 1 percent of assets a year. This is approximately the same amount mutual funds charge for voluntary retirement savings accounts in the United States; it is not excessive, from the lifetime point of view, in comparison to a competitive market retail price for individuals. Moreover, it is not excessive in comparison to a less expensive system that produces much lower gross and net returns (such as publicly managed reserves in Singapore and the United States social security trust fund). Competition may lower costs further in the long run.

Nonetheless, this fee structure is an apparent problem in the early years of a new system, when all accounts are small. It is a real problem for workers who will be in the system for only 20 to 30 years, such as workers who were relatively old upon the date of reform; simulations show that these workers pay a much higher lifetime fee as a percentage of assets. It is a problem for transient workers who move in and out of the labor force, such as women, because they may never have a chance to accumulate 40 years of contributions. The higher lifetime fee as a percentage of assets and hence the lower net return received by these groups is a matter of concern on equity grounds in a mandatory system. On both equity and efficiency grounds, it is questionable whether the government should compel all workers, including those who are very risk-averse, to incur these costs with certainty while the benefits are uncertain. Besides the equity consideration is the practical consideration that high costs may lead those groups most affected to evade. Moreover, it would be desirable to find ways to
increase administrative efficiency for all workers to increase their rates of return and replacement rates.

Some analysts believe that administrative costs would be lower under a group plan, and they thus favor choice by the employer or union. Such group plans may be better positioned to benefit from economies of scale in decisionmaking, greater financial expertise, and lower marketing costs (for a discussion of scale economies see James and Palacios 1995; Mitchell 1996a). This is one reason why employers or unions choose the investment manager in OECD model countries. Because employers or union representatives make the investment decision while workers bear the risk, such plans can also open the door to financial abuse and principal-agent problems: employers might choose investment managers or strategies that benefit them even if their choice implies lower returns for their workers.

For example, lower "wholesale" charges appear to be available for large group [401(k)] plans in the United States, but not all employers have gone to the effort of obtaining these rates. In Switzerland employers tend to place retirement funds at banks with which they have longstanding financial relationships, without exploring other options carefully (Hepp, forthcoming). One of the worst cases of employer abuse of worker retirement funds was the Maxwell scandal in the United Kingdom; but individual choice also led to a scandal in that same country, as uninformed workers were induced to abandon their employers' plans and purchase financially disastrous policies by unscrupulous insurance company salesmen (Johnson, forthcoming). Basing the second pillar on occupational plans is especially a problem for mobile workers, who may end up with many small costly accounts unless these can be consolidated in one personal account.

Anecdotal evidence about costs and returns to group choice versus individual choice is available, but a careful empirical study has yet to be done. Meanwhile, the principle-agent problem makes it likely that political pressures will develop to give workers the right to opt out of employer pension plans into their own personal retirement savings plan in most mandatory systems; this has already happened in the United Kingdom and Australia.

A third alternative may be desirable in small countries where economies of scale do not allow markets to support many pension companies efficiently, in countries with undeveloped financial markets that want to attract investment expertise and minimize start-up costs, and in countries with low contribution rates to the second pillar. Instead of open entry, the government might auction off operating rights to a limited number of investment companies, among whom workers then choose. The contract could specify the maximum risk, offer a reward for high returns, and choose the winners based on who charges the lowest administrative fees.

The voluntary Thrift Saving Plan for United States federal employees uses a competitive bidding process to choose its money managers, at a total cost of less than 10 basis points (0.1 percent). An auction process was recently used in Bolivia, which as
a result expects to have much lower administrative costs than Chile does. Another alternative is to set a low fee ceiling and open entry to all qualified pension fund managers willing to abide by that limit. Sweden plans to use a variation on this theme for its new second pillar—centralized collection and record-keeping, while workers choose among mutual funds that have reached an agreement on fees with the central agency.

The dangers here are the difficulties in insulating the auction and investment process from political manipulation, corruption, and collusion, and in incorporating incentives for good performance when entry and price are limited. Moreover, these mechanisms may feature less consumer education and service along with lower marketing expenditures. The advantages are that much lower costs, allowing an increase in net rates of return and replacement rates of as much as 20–40 percent, can be achieved if the process is well handled.

To sum up, one could construct a continuum with considerable choice, competition, political insulation, and relatively high administrative costs on one end and limited choice and competition with lower costs on the other end, with each arrangement having different implications for political insulation, rates of return, and other kinds of service. Countries could then choose which mix of costs and benefits they prefer. Thus many additional measures can be and are being considered to economize on costs, and their effects should appear over the next decade. The impact of alternative institutional arrangements on administrative costs in the second (decentralized funded) pillar has heretofore received little attention. Society could certainly benefit from careful analytic and empirical studies in this area.

Financial Market Distortions

Multipillar systems have justifiably been given credit for stimulating the growth of financial markets in middle-income countries and thus promoting economic growth. As these systems have been implemented, however, they have distorted the operations of financial markets in various ways.

This problem results because policymakers want workers to make investment decisions and bear the corresponding risk, but they also want to limit this risk to avoid a disaster. Relatedly, the government must set certain investment constraints and offer guarantees to overcome political opposition to reform. The contradiction here can potentially lead to malfunctioning markets, particularly if the pension funds are relatively large players in the market. Although the risk-reducing benefits of international diversification and diversification into private sector securities is one of the rationales for pension reform, in fact most countries require or strongly encourage domestic investments, with a heavy concentration in government bonds.

This ambivalence can be seen in Chile and several other Latin American countries where pension funds are heavily penalized if they deviate more than 2 percentage
points from the group mean. As a result, funds have been accused of herding behavior, as each tries to look very much like the others. Rather than having a choice of different points on the risk-return frontier, stemming from differing asset allocations—as would be the case in a well-functioning financial market—workers have the much less meaningful choice of companies that provide the same asset allocation and risk-return mix. Because workers are required to invest in one fund instead of diversifying among several and thereby reducing their risk, the lack of meaningful portfolio differences among them means that gains from diversification would be small in any event.

In Mexico all workers are required to enter the new system, but those currently in the labor force may return to the old pay-as-you-go system upon retirement if they will fare better under it. This option was included to acknowledge the “acquired rights” of workers and therefore avoid a legal challenge to the reform efforts. But it creates an obvious moral hazard problem: workers have an incentive to gamble with their pension funds, accepting too much risk because they are substantially protected from loss. The Mexican authorities have avoided the problems by greatly limiting the choice of investment strategies: at least 65 percent of all assets must be invested in government bonds (currently the funds have 99 percent of their assets in government bonds) and international investments are proscribed. Because workers have no real choice of portfolios, moral hazard is avoided; but the flow of pension funds to the financial market and the private sector is also avoided.

Bolivia initially intended to invest most of its revenues from privatization (targeted for pension reform) abroad, to protect it from excessive government borrowing and other country-specific risk. To overcome union opposition to the reforms, however, the government had to pay off the implicit debt of the complementary pensions that unions had negotiated in the past. To cover these and other expenditures, the final arrangements decreed that initially almost all of the privatization assets would be invested domestically, in government bonds. In Uruguay, to help cover transition costs, pension funds are required to put at least 80 percent of new assets in special issue government bonds.

Regulations in Switzerland require a 4 percent nominal guaranteed rate of return in their second pillars, thereby leading to a very conservative investment strategy, consisting largely of bonds. Until recently, providers of second pillar pensions for civil servants in the Netherlands faced little competition, again leading to low rates of return that might have been off or at a corner of the risk-return frontier.

These distortions should not be exaggerated, because the guarantees and limits on competition and portfolio diversification are likely to fall through time, as the schemes mature. Chile started with rigid restrictions but has gradually opened up the system to greater diversification, including international investment. Mexico is now considering allowing each pension fund to offer more than one portfolio, together with worker diversification among different portfolios. Another possibility under consid-
eration is to allow pension funds to differentiate their asset allocation strategies and corresponding benchmarks (if available), applying different risk limits depending on the type of portfolio chosen. For example, the Thrift Saving Plan in the United States offers portfolios that concentrate on bonds, stocks, and international investments, with different degrees of risk implied by each. This choice allows workers to pick their preferred point on the risk-return frontier and should help the financial markets to operate better, but it also requires substantial worker education as well as greater diversity of financial instruments than currently exists in many developing countries.

The Distributional Impact of Pension Reform

Although this paper has focused on the efficiency and growth effects of pension reform, an equally important topic is the impact of reform on equity. Because traditional pension systems are typically both inefficient and inequitable, they offer an opportunity to improve both. Which multipillar systems have actually succeeded in achieving a better distributional outcome is not known, however. Closer examination suggests that the devil is in the details, and some of the results may be surprising.

For example, in Chile's public pillar, workers are eligible for a minimum pension guarantee of about 27 percent of the average wage after 20 years of contributions, meaning that the government tops up the benefits of these workers to the guaranteed point if their own accumulation does not suffice. The main beneficiaries here will be lower earners who worked only 20 years, disproportionately females, who have limited labor market attachment, while workers who remain in the formal sector for a full career are unlikely to receive this subsidy. In contrast, in Argentina a flat benefit of about 28 percent of the average wage is paid to all workers who have at least 30 years of contributions (plus an additional 1 percent for every year above 30 up to 45). The main recipients will be workers who spent most of their adult lives in the formal labor sector, and (in sharp contrast to Chile) women are unlikely to qualify. In the United Kingdom, which pays a flat benefit about half the size of Argentina's (as a proportion of the average wage) but does not set a required number of contributory years, the big gainers are people who work few years and live long lives, such as women.

The setup of the second pillar also has distributional consequences. If flat fees per account are permitted, net returns are reduced for low earners more than for high earners. Flat fees were charged by Chilean funds initially, but the unfavorable publicity they encountered was one factor leading them to drop this practice. Some funds in Mexico now use flat fees. If low-income workers tend to choose more risk-averse investment strategies than high-income workers, they will have lower replacement rates in the future. The distributional issue is explored further in a separate paper (James 1997) and certainly merits additional empirical research.
Conclusion

Averting the Old Age Crisis (World Bank 1994) argued that old-age security systems with a large funded defined contribution component, decentralized competitive fund management, and a social safety net, are most likely to promote economic growth, provide acceptable income to the old, and reduce risk by diversification. During the past five years, the move toward multipillar systems has accelerated. With the aging of the global population, it has become increasingly important to choose a reliable and cost-effective method of old-age support. As economic growth slows and financial markets open up, it has become increasingly important to raise productivity through improved incentives in the labor market and through the accumulation of capital that can be allocated to its most efficient uses. To reduce income disparities, it has become increasingly important to provide additional protection to low-income wage-earners who have grown old. A multipillar system that includes a mandatory, publicly managed, tax-financed defined benefit pillar for redistribution, a mandatory, privately managed, funded defined contribution pillar to manage peoples' retirement savings, and a voluntary pillar for people who are willing to pay for more security, has seemed to many countries the most likely way to accomplish these objectives.

Thus several Latin American, OECD, and transition countries have already adopted multipillar systems, and they are under serious consideration in many more. Preliminary evidence from Chile, the only country that has had this system in place long enough for empirical studies to be conducted, supports the existence of a positive growth effect, stemming from increased labor market efficiency, mobilization of long-term saving, and financial market development.

Countries with a large implicit pension debt and an accompanying set of social security institutions are having trouble overcoming political opposition and financing the transition, however. Developing countries are fortunate in that they are at a relatively early stage and can choose a preferred multipillar system almost from the start, before these obstacles arise.

Notes

Estelle James is lead economist in the Development Economics Research Group at the World Bank. She would like to acknowledge the assistance of Robert Palacios, who gathered the data for figure 3, and Cheikh Kane, who collaborated with Mr. Palacios in the preparation of table 1.

1. For further details on the Latin American and OECD reforms see Bateman and Piggott (forthcoming); Cerda and Grandolini (1997); Hepp (forthcoming); Johnson (forthcoming); Mitchell (1996b); Palacios and Rocha (1997); Quieser (1998); Rofman and Bertin (1997); Valdes-Prieto (forthcoming); Von Gersdorff (1997); and Whitehouse (1998). For another summary of structural and piecemeal reforms see Demirgüç-Kunt and Schwarz (1997).
References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.


Estelle James
Knowledge for Development

The twenty-first edition of the World Bank's World Development Report examines the importance of knowledge to development. An appendix to the Report presents a set of Selected World Development Indicators—social and economic statistics for more than 200 countries.

Forthcoming October 1, 1998.


Global Economic Prospects 1998 and the Developing Countries

The ninth edition of this annual report examines the East Asian financial crisis and its effects on developing country prospects worldwide. GEP 1998 also explores the policy options available to domestic and international policymakers in preventing future financial crises.

Forthcoming in November 1998.

World Bank Publications
US customers, contact The World Bank, P.O. Box 960, Herndon, VA 20172-0960. Phone: (703) 661-1580, Fax: (703) 661-1501. Shipping and handling: US$5.00. For airmail delivery outside the US, charges are US$13.00 for one item plus US$6.00 for each additional item. Payment by check drawn on a US bank payable to the World Bank or by VISA, MasterCard, or American Express. Customers outside the US may also contact their local World Bank distributor. Visit our website at www.worldbank.org and click on Publications.
Would you like to have more information about

THE WORLD BANK

Research Observer?

You can view the tables of contents and the abstracts of articles from current and past issues at:


We’re interested in your comments. Please write us at:

Editor, The World Bank Research Observer
The World Bank
1818 H Street, N.W. (MC-4-392)
Washington, DC 20433