Weathering the Storm: The Impact of the East Asian Crisis on Farm Households in Indonesia and Thailand
Fabrizio Bresciani, Gershon Feder, Daniel O. Gilligan, Hanan G. Jacoby, Tongroj Onchan, and Jaime Quizon

Peter R. Fallon and Robert E. B. Lucas

Weak Links in the Chain II: A Prescription for Health Policy in Poor Countries
Deon Filmer, Jeffrey S. Hammer, and Lant H. Pritchett

Public Intervention in Health Insurance Markets: Theory and Four Examples from Latin America
William Jack

Urbanization in Developing Countries
Vernon Henderson

Developing Countries and a New Round of WTO Negotiations
Thomas W. Hertel, Bernard M. Hoekman, and Will Martin
Weathering the Storm: The Impact of the East Asian Crisis on Farm Households in Indonesia and Thailand

Fabrizio Bresciani • Gershon Feder • Daniel O. Gilligan • Hanan G. Jacoby • Tongroj Onchan • Jaime Quizon

This article assesses the impact of the East Asian financial crisis on farm households in two of the region’s most affected countries, Indonesia and Thailand, using detailed household-level survey data collected before and after the crisis began. Although the nature of the shocks in the two countries were similar, the impact on farmers’ income (particularly on distribution) was quite different. In Thailand, poor farmers bore the brunt of the crisis, in part because of their greater reliance on the urban economy, than did poor farmers in Indonesia. Urban-rural links are much weaker in Indonesia. Farmers in both countries, particularly those specializing in export crops, benefited from the currency devaluation. Although there is some evidence that the productivity of the smallest landholders declined over the period in question, it is difficult to attribute this directly to the financial crisis. At least in Thailand, a rural credit crunch does not seem to have materialized.

Now that the East Asian financial crisis has waned, its impact on two of the region’s most affected countries, Indonesia and Thailand, can be more readily assessed. Agriculture is the major employer in these economies, yet little is known about how farm households weathered the crisis. Hyperbolic news reports notwithstanding, many farmers surely benefited from the exchange rate depreciation. Other effects of the crisis, however, may not have been so sanguine. Overall, one would expect considerable variation in the impact of the crisis within the rural sector. Of particular interest to policymakers, given the implications for the design of safety net programs and balanced rural development, is how the rural poor fared relative to better-off households. This article uses detailed household survey data from Indonesia and Thailand collected before and after the onset of the 1997 financial crisis to explore its effect on farm production and income, especially its differential impact on the poor.
Framework for the Study

Both the similarities and the differences between the Indonesian and Thai cases are instructive. In both cases a currency collapse triggered a recession in the urban labor market, although it was more severe in Indonesia. Because farm households' exposure to these shocks varied across the two countries, the impact on incomes and particularly on the distributions of income was decidedly different. To make this argument more precise, consider the main channels through which crises like that of 1997 are transmitted to the countryside.

First there is currency depreciation, leading to higher prices for tradable commodities, such as rice and tree crop products. Against this increase in farm revenue is, all else equal, the higher cost of tradable inputs, most notably fertilizer. The net impact of the depreciation on farm income, given that the most important factors of agricultural production (land and labor) are nontraded, will typically be positive. Of course, farmers growing crops destined for domestic markets may receive a double blow: higher input prices and lower output prices because of the depressing effect of the recession on domestic demand.

After the devaluation, firms have difficulty servicing their debt denominated in foreign currency. A labor market recession, concentrated in urban areas, ensues. The impact on farm households depends on their ties to the urban economy, particularly to the hardest-hit sectors of that economy. If temporary rural-urban migration and remittances to rural families are pervasive, the recession will have a substantial negative effect on farm household income. Strong rural-urban linkages could also induce a fall in rural wages, which would only benefit those (perhaps few) farm households that are net hirers of labor.¹ The main point, however, is that the extent of labor market integration is decisive for transmission to the countryside. Also of possible importance for the distribution of labor market impacts across rural households is whether the recession affected unskilled workers more than skilled workers.

A secondary effect of recession and the consequent fall in cash income from off-farm employment is a reduction in a household's ability to purchase agricultural inputs, which would in turn reduce farm income. A related channel of crisis transmission is credit. The credit crunch that took hold after the onset of the crisis may have stifled the supply of rural lending, just as the demand was increasing due to a drying up of household cash reserves. The result may have been further pressure on the ability to purchase cash inputs.

As this discussion makes clear, a farm household's exposure or vulnerability to a crisis is complex and multifarious, depending on its positions in output, input, labor, and credit markets. Are poorer farm households more vulnerable? Perhaps so, because they tend to depend more heavily on wage income and less on cash-cropping. Yet poor farm households may be less dependent on the urban labor market than their richer counterparts and hence more insulated from the effects of recession. Indeed,
the much greater extent to which rural households in Thailand depend on off-farm income, largely from urban sources, is one of the striking contrasts between Thailand and Indonesia.

The East Asian financial crisis, particularly Indonesia’s experience, has already spawned considerable literature. Fallon and Lucas (also in this volume) review crisis experiences in several countries, but present no new evidence. For Indonesia, Suryahadi, Suharso, and Sumarto (2000); Frankenberg, Thomas, and Beagle (1999); Skoufias, Suryahadi, and Sumarto (1999); and Skoufias (2000) examine changes in household expenditures or poverty indices since the onset of the crisis. Levinsohn, Berry, and Friedman (1999) infer such changes using data on precrisis expenditures and on changes in consumer prices (see Boothe 2000 for an overview). The picture that emerges from these studies is that urban areas, particularly on Java, were hit harder than rural areas, though poverty rose everywhere (see especially Skoufias 2000 on this point).

For Thailand, Kittiprapas and Intaravitak (2000) and the World Bank (2000a) analyze pre- and postcrisis expenditure data, and World Bank (2000b) and Kakwani (1998) investigate changes in employment and earnings based on labor force surveys (see next section). Knowles, Pernia, and Racelis (1999) examine both sources of evidence. The analyses of expenditure data yield somewhat different conclusions, with Kittiprapas and Intaravitak (2000) reporting almost no change in overall poverty (though a slight amelioration in income inequality) and the World Bank (2000a) indicating an increase in poverty concentrated in rural areas.

None of these studies for either country differentiates between farm and other households. Thus, this article fills a gap in the literature by providing detailed information on changes in agricultural production and in sources of income of farm households and by focusing on the distribution of crisis impacts within rural areas. Although Levinsohn, Berry, and Friedman (1999) also ask whether the rural poor in Indonesia were hurt more than the rich (their answer appears to be “no”), they use a very different approach that relies on changes in consumer prices rather than data on preor postcrisis income. Skoufias (2000), meanwhile, looks only at changes in the distribution of income in rural areas using pre- and postcrisis expenditure data.

Having established a framework for thinking about the crisis impacts and having situated the present investigation in the broader crisis literature, this article next examines the data. But before turning to this analysis, it is worth laying out the key stylized facts of the crisis, especially as they relate to agriculture.

Stylized Facts about the Crisis

The most spectacular macroeconomic symptom of the crisis was the currency devaluation (figure 1). In July 1997, the Thai baht was allowed to float, giving way to a
devaluation that ended several years of relative exchange rate stability. In August 1997, Indonesia followed suit, and the rupiah plummeted even more dramatically than the baht, settling at less than a third of its precrisis value against the U.S. dollar. Consumer price inflation surged in both countries, but again much more dramatically in Indonesia, as did nominal interest rates. Private domestic credit contracted sharply as well. Real gross domestic product per capita fell about 10 percent in Thailand and 13 percent in Indonesia from 1997 to 1998, although the agricultural sector fared relatively well in both countries.

Recent analyses of labor force surveys reveal the impact of the crisis on the labor market. For Thailand, Kakwani (1998) shows that recession in the construction sector and, to a lesser extent, in manufacturing severely affected unskilled urban workers. Kittiprapas and Intaravitak (2000) report that between February 1997 and February 1999 unemployment in Thailand rose from 2.2 percent to 5.2 percent, and underemployment (less than 30 hours work per week) increased from 5.5 to 7.6 percent. Bangkok saw the biggest increase in unemployment, but rural areas were also affected. Real wages fell 3.8 percent (see also World Bank 2000b).

For Indonesia, the National Labor Force Surveys showed only a small increase in the unemployment rate, from 4.7 percent to 5.5 percent from 1997 to 1998, al-
though the National Socioeconomic Survey showed a larger increase, from 5.0 percent to 6.8 percent. By contrast, the drop in measured real wages was staggering—about 36 percent in urban areas and 32 percent in rural areas—largely reflecting the surge in inflation and stagnant nominal wages. No doubt the decline was partly a short-run phenomenon; some nominal wage catch-up would be expected in 1998–99. Data from the Farmer Terms of Trade survey (see following discussion) indicate a modest rise in real agricultural wages during 1999, but information on wages for urban sector employment is unavailable.

The currency devaluation translated directly into higher prices to farmers for tradable commodities. Average farmgate prices rose 29 percent in Thailand and 81 percent in Indonesia during 1996–98, roughly mirroring the 68 percent depreciation of the rupiah and 20 percent depreciation of the baht between July 1997 and December 1998 (figure 2). Rice is the most important staple food and a major source of crop income for Indonesian and Thai farmers. Nominal growth in rice prices outstripped inflation in both countries between the onset of the crisis and the end of 1998 (figure 3), though prices settled down by mid-1999, so that the increase over precrisis levels was about the same as for other domestic consumer products. Thus, there was

---

**Figure 2. Farmgate Price Indices, Indonesia and Thailand (1994 = 100)**

![Diagram showing farmgate price indices for Indonesia and Thailand from 1994 to 1999.](image)

*Source: Farmer Terms of Trade survey collected by the Central Statistical Bureau for Indonesia and from Office of Agricultural Economics data for Thailand.*
little change in the relative price of rice for consumers, whereas for the surplus producer, the situation was vastly improved in Indonesia and only temporarily improved in Thailand. Rice prices in Indonesia shot up further in late 1999, but this change came in response to government reforms. It should also be noted that in Thailand international increases in the price of rice were not fully transmitted to the farmgate (Bresciani and others 2000).

Consider what happened to the nominal prices of urea, the fertilizer most commonly used by rice producers in the two countries (figure 4). Government subsidies and falling world prices limited the increase in domestic fertilizer prices during the crisis. In Indonesia, the price of urea rose only 22 percent from the third quarter of 1997 to the same period in 1998. However, fertilizer prices spiked upward after they were liberalized in the food marketing reforms of November 1998. The price of urea more than doubled in the six months from the third quarter of 1998 to the first quarter of 1999. Similarly, in Thailand, the increase in the price of urea in 1997–98 was less than the rate of currency depreciation, due in part to the 19 percent decline in international urea prices in U.S. dollar terms (World Bank 1999). As a result, many farmers enjoyed a significant improvement in terms of trade as commodity price increases outpaced the rise in fertilizer prices.

Source: Farmer Terms of Trade Survey data for Indonesia; Ministry of Agriculture and Cooperatives for Thailand.
Evidence on Crisis Impacts from Household Survey Data

To assess the rural impacts of the crisis, data collection efforts were initiated in Indonesia and Thailand in mid-1999. The samples and the content of the questionnaires had to match a suitable baseline precrisis survey to allow for comparison. Both the Indonesian and Thai surveys ask about land use, crop and livestock production, input use, off-farm income sources, assets, and so forth. Panel data spanning the crisis period are available for about 1,600 rural households in each country. In Thailand, the same 400 villages were sampled in each year, though not necessarily the same households. This does not affect the analysis in this article.

Both surveys have broad regional coverage, but neither is nationally representative—six provinces are surveyed in Indonesia (Central and East Java, Lampung, North and South Sulawesi, and West Nusa Tenggara) and three regions in Thailand (North, Northeast, and Central). (Representativeness within the regions or provinces covered is addressed in the following.) The Indonesian sample covers all rural households, about a fifth of them landless. This group is probably just as representative of the nonlandowning population of Indonesia as the rest of the sample is of the landowning population.

The Thai sample is restricted to households that own land. The consequences of not sampling landless households can be assessed from information in the nation-
ally representative Socioeconomic Survey (ses) for 1998, which collects information on household land ownership. The ses data indicate that about a third of rural households in the North, Northeast, and Central regions do not own land. But real per capita expenditures are almost 60 percent higher for this “landless” group than for landowning households, suggesting that the rural households overlooked by the Thai agricultural survey are, at least on average, not poor. Nevertheless, some of the landless rural poor have clearly been left out, a caveat to be kept in mind throughout the discussion of the analysis and findings.

It is tempting to view any changes that occurred between the baseline surveys—covering 1994–95 in Indonesia and 1995–96 in Thailand—and the follow-up surveys, covering 1998–99, as reflecting the impact of the crisis. However, both countries experienced robust growth right up until mid-1997, which could mask any crisis-induced downturn or exaggerate an upturn. Some of the effects of the crisis, such as the increase in rice prices immediately following the 1997 devaluation, were transitory and would have already dissipated by the time of the follow-up survey. Indeed, much of the 1995–99 increase in rice prices in Thailand (see figure 3) took place before July 1997 and cannot be attributed to the crisis. Also, the main effects of the El Niño drought were felt in 1997–98. In short, caution is required in drawing inferences about the impact of the crisis from these data. It would be more precise to refer to changes during the crisis period.

Choosing the right welfare-ranking criterion is critical for comparing crisis period impacts among poor and nonpoor farmers in each country. A common approach is to use per capita expenditures or income in the baseline period. The risk of misclassification is considerable, however, because of the large transitory component of expenditures and income and the low precision with which they are measured.

The approach taken here is to group households by quintiles according to per capita landholdings adjusted for land fertility. Landholdings are generally measured more precisely than consumption or income, and as the principal household asset in these samples, land is a good indicator of wealth. However, the value of land varies, especially by region, due to differences in fertility. To account for these regional differences, landholdings are adjusted by each region’s average rice yield in the base year normalized by the rice yield in the highest yield region (Central Java in Indonesia and the Central region in Thailand). Multiplying landholdings by this indicator of relative yield creates a measure of fertility-adjusted landholdings. For example, the average yield in North Thailand is almost 90 percent that of the central region, and the Northeast average yield is about 60 percent. Thus, a household in the North with one hectare per capita is placed in the same effective landholdings quintile as a household in the central region with 0.6 hectare per capita. For Indonesia, nonlandowning households are lumped into the lowest adjusted landholding quintile. As mentioned, many of these “landless” households are truly landless laborers, but some also engage in business activities and are not poor.
A final methodological issue is the choice of deflator. For Thailand, where crisis-induced inflation was not so severe and regional markets are well integrated, the national consumer price index (cpi) is used. For Indonesia, the choice of deflator is more complicated because of the high postcrisis inflation and the geographic fragmentation of the country (Suryahadi, Suharso, and Sumarto 2000; Frankenberg, Thomas, and Beagle 1999). The province-specific rural cpi is used, based on the Farmer Terms of Trade survey data collected by the Central Statistical Office. This index more reliably captures changes in rural prices than, for example, the national cpi series, which is based on a sample from urban areas.

Changes in Total Household Income

Real household income is the sum of farm income (see the following discussion), off-farm labor earnings, business income, and government and private transfers. Table 1 reports median real per capita income in the two survey rounds and changes in medians across rounds for each of the adjusted landholdings quintiles. Census-weighted figures are reported to account for the fact that the size of the sample in each region does not reflect the true regional share of the rural population. For example, households on Java are underrepresented in the Indonesian sample. Thus observations are weighted by the actual regional or province share in the rural population divided by the sample share.

Two striking income results emerge. First, rural incomes appear to have risen faster in Indonesia than in Thailand, despite the greater severity of the crisis in Indonesia. However, this finding may be partly explained by the fact that the Indonesian surveys span nearly one extra year of precrisis growth than the Thai surveys. In addition, because the largest portion of agricultural revenues in Indonesia are earned at the end of the rainy season in March, which is near the close of the one-year survey recall period, deflating by average annual prices overstates real incomes. Further,

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poorest)</td>
<td>341</td>
<td>357</td>
</tr>
<tr>
<td>2</td>
<td>304</td>
<td>447</td>
</tr>
<tr>
<td>3</td>
<td>306</td>
<td>455</td>
</tr>
<tr>
<td>4</td>
<td>379</td>
<td>579</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>577</td>
<td>667</td>
</tr>
</tbody>
</table>

Note: See text for explanation of quintile adjustments.
Source: Data for Indonesia are from the 1994–95 and 1998–99 PATANAS surveys; data for Thailand are from the 1995–96 and 1998–99 Ministry of Agriculture and Cooperatives OAC farm household surveys.
because Indonesian inflation was very high in 1999 and much higher than in Thailand, 1999 income (and hence 1995–99 income growth) may be overstated in the Indonesian case.

Second, the relative effect of the crisis on the poorer farmers differs considerably in Thailand and Indonesia. In Thailand, the poorest two quintiles suffered large real income losses, whereas the richest households enjoyed spectacular income gains over the period. Confidence in this finding is bolstered by results showing a similar pattern in per capita consumption expenditures—available in the Thai survey, but not the Indonesian (table 2). Again, the top adjusted landholding quintile in Thailand did extremely well, but the bottom quintiles experienced large declines in real consumption.

The distribution of income gains is far more equitable in Indonesia, with no clear pattern emerging. The bottom landholding quintile appears to gain the least over the 1995–99 period, but recall that some of these households are not poor; indeed 1995 median income in this quintile is higher than that of quintiles two and three.

So, although the crisis period was far from disastrous for farm households overall, it did exacerbate income inequality among farm households in Thailand, but not in Indonesia. The question is why. The answer would seem to lie in a more fine-grained decomposition of income.

**Changes in Farm and Nonfarm Income**

A starting point in explaining the different experiences of rural households in Thailand and Indonesia is looking at how they earn their income. In Thailand, the median share of farm income (defined as the value of crop production minus variable costs and land rent plus net income from livestock and fishponds) in total income increases steadily by quintile, from lowest to highest: 30 percent, 41 percent, 47 percent, 63 percent, and 75 percent. The Indonesian sample includes households that

<table>
<thead>
<tr>
<th>Quintile</th>
<th>1995/96</th>
<th>1998/99</th>
<th>Change (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poorest)</td>
<td>7,222</td>
<td>6,121</td>
<td>-15.2</td>
</tr>
<tr>
<td>2</td>
<td>7,760</td>
<td>5,852</td>
<td>-24.6</td>
</tr>
<tr>
<td>3</td>
<td>8,235</td>
<td>7,358</td>
<td>-10.7</td>
</tr>
<tr>
<td>4</td>
<td>9,520</td>
<td>8,507</td>
<td>-10.6</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>11,360</td>
<td>13,723</td>
<td>20.8</td>
</tr>
</tbody>
</table>

Note: See text for explanation of quintile adjustments.

do not own a farm, so the pattern of farm income shares by increasing landholdings quintile is more stark: 0 percent, 44 percent, 92 percent, 82 percent, and 99 percent.

The fact that the poor, as defined here, are less reliant on farming has profound implications for the distributional impact of the crisis. Richer farm households in these samples are highly exposed to agricultural output and input price shocks; poorer farm households have comparatively greater exposure to labor market shocks. During the crisis period, output and fertilizer prices moved in a direction favorable to farmers, while the labor market performed badly. Incomes of the poor, therefore, would be expected to suffer more than those of the rich. As will be seen, however, the reality is somewhat more complex.

In Indonesia, per capita farm income increased during the crisis, but with no clear pattern across adjusted landholdings quintiles (table 3). The growth in farm income for the first quintile, which consists of all households that owned no land in 1995, is due to the initiation of agricultural production. Surprisingly, nonfarm income also grew substantially from 1995 to 1999, driven by growth in business income on Java and labor income on the outer islands. Though some of this income growth may have occurred before the crisis, its magnitude is remarkable in light of the massive real wage declines reported earlier. Evidently, urban-rural linkages are weak in Indonesia. Only the lowest landholdings quintile did not enjoy growth in nonfarm income; the data actually show a considerable decline for this group. Because this quintile derives all its income from nonfarm sources, it naturally performed the worst in total income growth.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real per capita farm income</td>
<td></td>
</tr>
<tr>
<td>1 (poorest)</td>
<td>0.0</td>
<td>6.3</td>
</tr>
<tr>
<td>2</td>
<td>48.2</td>
<td>124.5</td>
</tr>
<tr>
<td>3</td>
<td>175.9</td>
<td>177.7</td>
</tr>
<tr>
<td>4</td>
<td>218.4</td>
<td>324.0</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>410.9</td>
<td>421.1</td>
</tr>
<tr>
<td></td>
<td>Real per capita nonfarm income</td>
<td></td>
</tr>
<tr>
<td>1 (poorest)</td>
<td>311.0</td>
<td>276.7</td>
</tr>
<tr>
<td>2</td>
<td>166.8</td>
<td>188.3</td>
</tr>
<tr>
<td>3</td>
<td>21.4</td>
<td>132.9</td>
</tr>
<tr>
<td>4</td>
<td>41.7</td>
<td>89.4</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>0.0</td>
<td>79.6</td>
</tr>
</tbody>
</table>

Note: See text for explanation of quintile adjustments.
Source: Data for Indonesia are from the 1994–95 and 1998–99 PATANAS surveys; data for Thailand are from the 1995–96 and 1998–99 Ministry of Agriculture and Cooperatives farm household surveys.

Fabrizio Bresciani and others
In Thailand, off-farm income also declined for the lowest quintiles, and so did farm income. It is unclear why farm income declined only for the poorest households. There was no noticeable change in cultivation patterns (the issue of input use is taken up below). However, the question may be of limited practical significance since these households derive a small proportion of their income from farming their own land. For the upper income quintiles, the situation is reversed, with both farm and off-farm income rising. Why off-farm income increased for larger landholders is also a bit of a puzzle. Perhaps these households obtain more of their nonfarm income from skilled labor and small-scale business activities, which were less affected by the recession in Thailand.

The main lesson to be drawn from patterns of income changes is that farm households in Thailand that were most reliant on off-farm income experienced the biggest percentage decline in this income. Though this was also true in Indonesia, the magnitude of the shock to off-farm income was apparently much smaller, with fewer repercussions for rural income inequality. Clearly, it is not simply the extent of income diversification that exposes households to urban-based recessions but also the nature of the diversification. In Thailand, much off-farm income is earned outside the village and in sectors that suffered badly during the crisis, such as construction. In the Indonesian sample, however, the sources of off-farm income were not as closely tied to crisis-affected sectors; nearly a third of households had nonagricultural business income, and more than half the individuals reporting off-farm labor income earned it in agriculture.

Data on remittances speak to the relatively strong links between urban and rural sectors in Thailand and their importance during the crisis (table 4). About a quarter of farm households receive remittances, presumably from out-of-village sources.

<table>
<thead>
<tr>
<th>Quintile</th>
<th>1995–96</th>
<th>1998–99</th>
<th>Change (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poorest)</td>
<td>5,000</td>
<td>2,165</td>
<td>−56.7</td>
</tr>
<tr>
<td>2</td>
<td>4,500</td>
<td>2,952</td>
<td>−34.4</td>
</tr>
<tr>
<td>3</td>
<td>3,333</td>
<td>2,666</td>
<td>−20.0</td>
</tr>
<tr>
<td>4</td>
<td>3,000</td>
<td>2,733</td>
<td>−8.9</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>2,550</td>
<td>3,590</td>
<td>40.8</td>
</tr>
</tbody>
</table>

Notes: Households with positive net remittances only. The 1995–96 survey did not break out data on remittance. A special module attached to the 1999 survey asked about remittances received or sent during the 1995–96 and the 1998–99 crop years.


though not necessarily from urban areas. Among households with positive net remittances, most suffered a downturn in remittances, with the poor especially hard hit. As is the case with off-farm income and possibly for the same reasons, the richest farm households experienced an increase in net remittances. These income gains may also be largely of precrisis origin.

Further support for urban-rural linkage in Thailand comes from migration patterns, which appear to have responded to the crisis-induced recession. A retrospective migration questionnaire included in the Thai survey found that annual migration into rural areas increased considerably after the onset of the crisis period, although the numbers involved are still too small to suggest a major return migration to the countryside (see Bresciani and others 2000 for more details).

**Changes in Crop Production**

For most farming households, the biggest effect of the crisis was a shift—sometimes massive—in the price of tradable commodities relative to nontradables. To understand the impact of this shift on farm income it is necessary to trace its effects through the production process. An interesting contrast between Thailand and Indonesia is the much greater reliance of Thai farmers on rice and consequently the greater homogeneity within the country in agricultural production activities. Well over 80 percent of the Thai households surveyed in each year cultivated rice, and rice accounted for more than half their crop sales. On Java, Indonesia’s rice basket, the production value share of rice is comparable to that in Thailand, but for the Indonesian sample as a whole the (census-weighted) rice share is only about 17 percent.

Despite the (transitory) increase in rice prices, no dramatic changes occurred in cropping patterns in Thailand between the two surveys. Obviously, the scope for substitution into rice (or other export crops) is limited in Thailand. In Indonesia, by contrast, there appears to be some scope for crop substitution in response to changing relative prices and other conditions.

Summary statistics on crop composition for the Indonesian sample shed some light on substitution (table 5). Farm households (that is, excluding business owners with no crop income in at least one survey year) are grouped by primary crops grown: rice, dryland crops, tobacco and sugarcane, and tree crops. Dryland crops include all grains other than rice and vegetables (potatoes, maize, shallots, cabbage, cassava, garlic, soybeans) and are grouped because of the high degree of substitutability in production between many of these crops. Most dryland crops show limited responsiveness to world prices, with the exception of shallots and soybeans. Dryland crop farmers are mostly found in the middle three landholdings quintiles.

Tobacco and sugarcane are combined because they represent the two largest cash crops produced primarily for the domestic market. At that time, all sugar was marketed through official channels with farmgate prices set according to a complicated
Table 5. Crop Composition, Indonesia

<table>
<thead>
<tr>
<th>Commodity group</th>
<th>Share of farm households</th>
<th>Mean area planted (hectares)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td></td>
<td>46.4</td>
</tr>
<tr>
<td>Dryland crops</td>
<td></td>
<td>47.1</td>
</tr>
<tr>
<td>Tobacco and sugarcane</td>
<td></td>
<td>14.2</td>
</tr>
<tr>
<td>Tree crops</td>
<td></td>
<td>50.6</td>
</tr>
</tbody>
</table>


formula that had no direct relationship with the world price. Tobacco production in Indonesia is contracted exclusively by domestic (Kretek) cigarette manufacturers and is not exported. Although there are no restrictions on cigarette imports, higher prices for imports are unlikely to have raised prices for Kretek cigarettes very much.

Most tree crops are tradable, and many experienced a large jump in price after the onset of the crisis. The tree crops with the greatest value of production in 1995 were coffee, cocoa, pepper, coconut, vanilla, and cloves. Tree crop farmers are concentrated in the higher landholdings quintiles.

There was a modest increase in the number of households growing rice, but a large jump in area planted to rice (see table 5). Both the number of households and the area under dryland crop production also grew considerably, coming partly from a 32 percent rise in the number of households growing corn. Some households may have shifted to corn production as a more reliable source of income during the 1997–98 El Niño drought. Many households withdrew from tobacco and sugarcane production, as might be expected, and average area planted declined as well. There was also an increase in households growing tree crops. These findings are generally consistent with the changes in relative output prices over the crisis period.

Farm income per capita grew robustly for tree farmers and rice growers, but that for dryland crops and tobacco and sugar growers stagnated or declined (table 6). Thus, as expected, export orientation of farmers is a powerful determinant of their success in weathering the crisis.

Given the status of rice as the main staple crop in both countries, changes in rice yields during the crisis period warrant consideration. Of particular interest is whether the drying up of off-farm income (especially in Thailand) and any rural credit crunch had deleterious effects on production. In Indonesia, rice yields fell for all but the highest landholdings quintile, and the decline was most severe for the poor (table 7). Evidently, the increase in farm income for rice growers as a group was due to higher rice prices rather than improved yields. The larger decline in yields for the poor may indicate that they faced more production constraints than their better-off counterparts. Thai farmers saw a modest overall increase in rice yields, but as in Indonesia the dis-
Table 6. Median Real per Capita Farm Income by Cropping Group, Indonesia (thousands of 1994–95 rupiah)

<table>
<thead>
<tr>
<th>Primary commodity group</th>
<th>N</th>
<th>1994–95</th>
<th>1998–99</th>
<th>Change (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>362</td>
<td>153</td>
<td>201</td>
<td>31.2</td>
</tr>
<tr>
<td>Dryland</td>
<td>296</td>
<td>208</td>
<td>210</td>
<td>1.1</td>
</tr>
<tr>
<td>Tobacco and sugarcane</td>
<td>126</td>
<td>215</td>
<td>184</td>
<td>−14.4</td>
</tr>
<tr>
<td>Trees</td>
<td>289</td>
<td>189</td>
<td>274</td>
<td>44.5</td>
</tr>
</tbody>
</table>


Distribution of these production gains was inequitable. Households in the upper quintiles reaped higher yields, and the poorest cultivators saw their yields decline, as is consistent with the pattern of farm income changes reported in table 3. Again, this is tantalizing evidence that the crisis impeded the ability of the smallest farmers to produce efficiently, perhaps because credit constraints bind more stringently for these households.

However, other factors might have been at work as well. Drought may have had a differential impact on smaller farmers because of unequal access to reliable irrigation. The greater availability of hired farm labor, and therefore depressed agricultural wages, could have raised productivity on larger farms. The regional distribution of poor farmers does not appear to explain the finding for Thailand because there are at least as many poor farmers in the sample from the North, where yields rose, as there are from the Northeast, where yields fell.

A look at the use of fertilizer, the main cash input for rice production, might provide further insight. A significant fall in fertilizer use for poor farmers, concomitant with a decline in yields, would be compelling evidence of a cash squeeze. For Indonesia, a quantity index for different fertilizer types was constructed. For Thailand, only

Table 7. Median Rice Yields by 1995 Fertility-Adjusted Landholdings Quintile, Indonesia and Thailand (tons per hectare)

<table>
<thead>
<tr>
<th>Quintile</th>
<th>Indonesia</th>
<th>Change (percent)</th>
<th>Thailand</th>
<th>Change (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poorest)</td>
<td>3.448</td>
<td>2.879</td>
<td>−16.5</td>
<td>2.169</td>
</tr>
<tr>
<td>2</td>
<td>3.750</td>
<td>3.200</td>
<td>−14.7</td>
<td>2.306</td>
</tr>
<tr>
<td>3</td>
<td>3.045</td>
<td>2.800</td>
<td>−8.1</td>
<td>1.950</td>
</tr>
<tr>
<td>4</td>
<td>2.993</td>
<td>2.800</td>
<td>−6.5</td>
<td>1.888</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>2.371</td>
<td>2.500</td>
<td>5.4</td>
<td>1.750</td>
</tr>
<tr>
<td>All</td>
<td>2.973</td>
<td>2.857</td>
<td>−3.9</td>
<td>2.000</td>
</tr>
</tbody>
</table>

Source: Data for Indonesia are from the 1994–95 and 1998–99 patanas surveys; data for Thailand are from the 1995–96 and 1998–99 Ministry of Agriculture and Cooperatives farm household surveys.
the household's total expenditure on fertilizer is available and is reported here in nominal terms because nominal fertilizer prices changed only slightly between the two survey periods.

Overall, the evidence is mixed. Fertilizer use declined for all but the bottom and top quintiles in Indonesia (table 8). For the bottom group, the increase was due to the initiation of crop cultivation since the 1995 survey. Perhaps the November 1998 price reform was responsible for the general decline in fertilizer use, but it is difficult to conclude that the poor were differentially affected. Suggestive, though, is the fact that the top landholding quintile was the only one to experience both an increase in fertilizer intensity and an increase in rice yield.

In Thailand, farmers in all quintiles increased their use of fertilizer, as inferred from the fact that their nominal expenditures increased while the price remained nearly constant. The poorest farmers increased their expenditures the least, but the difference in growth of fertilizer intensity is not enough to explain why yields fell on poor farms and rose on richer farms. In any event, the crisis certainly did not precipitate a contraction in fertilizer use in Thailand.

Finally, for Thailand at least, it is possible to assess the importance of the credit channel in propagating the crisis to rural areas (for Indonesia, credit data are not included in the precrisis baseline survey). Data on the total value of outstanding loans per household do not reveal a rural credit crunch, as debt remained fairly steady, or increased, in real terms (table 9). Although supply and demand factors cannot be distinguished with these data, no major contraction in the supply of loans could have occurred without a fall in the real value of debt, for which evidence is lacking. Moreover, changes in credit availability cannot explain differences in performance of poor and nonpoor farmers in Thailand because there is no strong pattern in changes in debt by landholding quintile. About a quarter of farm households with outstanding debt before the crisis

<p>| Table 8. Fertilizer Expenditure by 1995 Fertility-Adjusted Landholdings Quintile, Indonesia and Thailand |
|------------------------------------------------------|------------------------------------------------------|------------------------------------------------------|------------------|</p>
<table>
<thead>
<tr>
<th>Quintile</th>
<th>Indonesia (thousands of real 1994–95 rupiah per hectare)</th>
<th>Thailand (baht per hectare)</th>
<th>Indonesia (thousands of real 1994–95 rupiah per hectare)</th>
<th>Thailand (baht per hectare)</th>
<th>Change (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>875</td>
<td>1,000</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>152.6</td>
<td>124.4</td>
<td>656</td>
<td>800</td>
<td>21.7</td>
</tr>
<tr>
<td>3</td>
<td>146.6</td>
<td>109.3</td>
<td>544</td>
<td>719</td>
<td>32.1</td>
</tr>
<tr>
<td>4</td>
<td>125.0</td>
<td>96.3</td>
<td>488</td>
<td>700</td>
<td>42.8</td>
</tr>
<tr>
<td>5</td>
<td>49.3</td>
<td>54.4</td>
<td>494</td>
<td>625</td>
<td>27.4</td>
</tr>
</tbody>
</table>

*Deflator is a Laspeyeres index based on urea and trisodium phosphate price series. Source: Data for Indonesia are from the 1994–95 and 1998–99 PATANAS surveys; data for Thailand are from the 1995–96 and 1998–99 Ministry of Agriculture and Cooperatives farm household surveys. Fertilizer price series (for Indonesia) are from Farmer Terms of Trade data, Central Statistical Office.

<table>
<thead>
<tr>
<th>Quintile</th>
<th>1995–96</th>
<th>1998–99</th>
<th>Change (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poorest)</td>
<td>24,500</td>
<td>25,088</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>25,000</td>
<td>25,000</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>20,000</td>
<td>26,000</td>
<td>30.0</td>
</tr>
<tr>
<td>4</td>
<td>25,000</td>
<td>25,000</td>
<td>0.0</td>
</tr>
<tr>
<td>5 (richest)</td>
<td>30,000</td>
<td>33,990</td>
<td>13.3</td>
</tr>
</tbody>
</table>


obtained government credit (roughly equal across quintiles). This share changes little in 1998–99, which is consistent with the view that the state Bank of Agriculture and Cooperatives maintained its lending operations throughout the crisis period.

Conclusions and Policy Implications

There is a temptation to generalize about the impact of the East Asian 1997 financial crisis on farm households. Perhaps the main lesson of this analysis is that doing so is fraught with peril. The impacts of the crisis were varied even within countries and all the more so across countries. The net effect of the crisis on household income depends on a broad range of factors; no simple economic classification of households will necessarily capture the differences in exposure to the set of shocks that constituted the crisis.

Nevertheless, classifying households by landholdings, a relatively fixed characteristic, yields useful insights, subject to the aforementioned caveats. The evidence suggests that the smallest farmers and landless households in Indonesia and Thailand bore the brunt of the crisis, but larger farmers fared quite well. The crisis generally depressed the labor market and improved the profitability of export crop cultivation. Small and landless farmers derive most of their income from off-farm sources rather than their own cultivation, and the reverse is true for larger farmers. Evidence from Indonesia supports the view that export orientation determines farmers’ exposure to positive price shocks. Evidence from Thailand corroborates the importance of exposure to negative urban labor market shocks.

The household survey data also provide clues about how agricultural production was affected during the crisis. In Indonesia, where there is more scope for crop substitution than in Thailand, some substitution out of domestically marketed crops seems to have taken place. There is also evidence that profits and yields declined for...
the poorest farmers while rising for the better-off farmers. It is difficult to pinpoint the exact cause. Changes in fertilizer use alone (possibly due to increased cash constraints) are probably not large enough to explain the difference. More research is needed to trace the links between off-farm income and farm production decisions and outcomes.

One can envision two types of rural policy responses to the crisis and to the prospect of future recessions more generally: changes in agricultural policy and changes in social policy targeted to rural areas. During the crisis period, the macroeconomic environment was favorable to agriculture, as this article has made abundantly clear. It could perhaps be argued that farmers would have benefited more, for example, had increases in the international price of rice been fully transmitted to the farmgate in Thailand or had the fertilizer price reform been phased in more gradually in Indonesia. But such initiatives would not have helped the poorest farmers very much, because they are not the main beneficiaries of agricultural policy. Similarly, various support policies, such as targeted and temporarily subsidized credit, would not have greatly benefited the landless or smaller landholders.

Change in social policy, by contrast, is worth careful examination. The evidence presented here, particularly that from Thailand, refutes the view held by some policymakers that the smallholder sector can absorb low-skilled labor dislocated by urban unemployment. Most of these workers are associated with farms too small to make productive use of them. Indeed, real per capita expenditures fell most precipitously for Thai households with the smallest landholdings. Temporary social programs targeted to the rural poor, perhaps using landholdings as a targeting criterion, could be effective in providing support for households adversely affected by recession. Such programs are lacking in Thailand (World Bank 2000; Kittiprapas and Intaravitak 2000) and are poorly targeted in Indonesia (Suryahadi, Suharso, and Sumarto 1999; Gilligan, Jacoby, and Quizon 2000), so there is considerable room for improvement.

It bears emphasis that the situation in rural areas of Thailand and Indonesia is fluid. Indeed, there is evidence from Thailand of a recent downturn in agricultural prices just as the urban economy is picking up (Kittiprapas and Intaravitak 2000). If so, some of the crisis impacts identified here may be at least partly reversed. This is not to argue for inaction, but only to suggest that a focus on immediate measures to deal with the 1997 crisis may be misplaced. A more considered medium- and longer-term policy focus would perhaps be on increasing the access of the rural poor to off-farm employment through improved rural infrastructure and other measure and on upgrading the skills of the rural labor force through better education.

Notes

Fabrizio Bresciani is a graduate student at the University of Maryland, College Park. Gershon Feder is Research Manager in the Development Research Group of the World Bank. Daniel O. Gilligan is a graduate student at the University of Maryland, College Park. Hanan G. Jacoby is Senior Economist
in the Development Research Group of the World Bank. Tongroj Onchan is President of the Mekong Environment and Research Institute. Jaime Quizon is Economist in the Development Research Group of the World Bank. Fabrizio Bresciani can be reached via e-mail at bresciani@arec.umd.edu, Gershon Feder at gfeder@worldbank.org, Daniel O. Gilligan at gilligan@arec.umd.edu, Hanan G. Jacoby at hjacoby@worldbank.org, Tongroj Onchan at onchant@merimekong.org, and Jaime Quizon at jquizon@worldbank.org.

1. A more subtle effect on the rural labor market would go in the opposite direction. Improved farmer terms of trade could raise the demand for agricultural labor and thereby raise rural wages, at least in the short run, where the supply of agricultural labor is not perfectly elastic.

2. Though the price index for Thailand is a national average, in the Indonesian case it is an average of the rural producer price indices for the six provinces included in the Indonesian sample. The primary crops are rice, soybeans, cassava, coffee, pepper, and cloves.

3. More details on the data sets can be found in Gilligan, Jacoby, and Quizon (2000) for Indonesia and in Bresciani and others (2000) for Thailand. In Indonesia, the 1999 survey built on the long running PATANAS panel collected by the Center for Agricultural Socio-Economic Research. In Thailand, the 1999 survey followed a subsample of the 1996 farm household survey collected by the Ministry of Agriculture and Cooperatives. The Mekong Environment and Resource Institute assisted in this effort.

4. In the first stage of the Indonesian sampling scheme, villages were chosen to be consistent with the primary crops, topography, and cropping patterns of the region, with an eye toward capturing the diversity of Indonesia’s cropping arrangements. In the second stage, 50 households were randomly selected from each village, so any errors in obtaining a truly representative sample were made at the stage of identifying villages.

5. Though remittances are included in the measure of household income reported above, it is not possible to isolate remittances based on the information in the income module. However, a special module attached to the second survey did ask questions about remittances received or sent during the 1995–96 and the 1998–99 crop years. Given the retrospective nature of the data, though, net remittance income is likely to be imprecisely measured, especially for the 1995–96 crop year. In particular, respondents may have difficulty distinguishing remittances from off-farm earnings of resident family members.

6. Another important crop (and export) in northeast Thailand is tapioca for processors (23 percent of the total cash sales). In the north, corn for animal feed is the next most important commercial crop after rice (6 percent of total cash sales), while in the central region sugarcane for processors is the second most important commercial crop (26 percent of total cash sales).

References

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


International Financial Statistics. In International Monetary Fund [World Bank database online].


Peter R. Fallon • Robert E. B. Lucas

The 1990s have witnessed several financial crises, of which the East Asia and Mexico tequila crises are perhaps the most well known. What impact have these crises had on labor markets, household incomes, and poverty? Total employment fell by much less than production declines and even increased in some cases. However, these aggregates mask considerable churning in employment across sectors, employment status, and location. Economies that experienced the sharpest currency depreciations suffered the deepest cuts in real wages, though deeper cuts in real wages relative to GDP were associated with smaller rises in unemployment. To some extent, families smoothed their incomes through increased labor force participation and private transfers, though the limited evidence available suggests that wealthier families were better able to smooth consumption. The initial impact of the crises was on the urban corporate sector, but rural households were affected as well and in some instances suffered deeper losses than did urban families. School enrollment declined, especially among poorer families, as did use of health facilities, but the impact on children’s nutrition levels appears to vary. Crises have typically proved short-lived, but whether households plunged into poverty during a crisis are able to recover as the economy does remains an open question.

The East Asia crisis erupted with a massive speculative attack on the Thai baht during May 14–15, 1997. During the remainder of 1997 and through 1998, substantial short-term capital outflows accompanied sharp gross domestic product (GDP) declines in Indonesia, the Republic of Korea, Malaysia, and Thailand. Initial estimates for Korea and Malaysia indicate that GDP began to rise again early in 1999. Similar patterns—of short-term capital flight accompanied by declining incomes over a couple of years—were seen in Mexico and Argentina in 1995 and in Turkey in 1994 (table 1) and began to emerge in Brazil in January 1999.

This article looks at how such crises affect labor market conditions, poverty, and income equality, drawing on evidence from these and other economic crises. Al-
Table 1. GDP Growth and Financial Account Balance before and after Crises

<table>
<thead>
<tr>
<th>Country and crisis year</th>
<th>Year relative to crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3</td>
</tr>
<tr>
<td>GDP growth (percent)</td>
<td></td>
</tr>
<tr>
<td>Indonesia, 1998</td>
<td>8.2</td>
</tr>
<tr>
<td>Korea, 1998</td>
<td>8.9</td>
</tr>
<tr>
<td>Malaysia, 1998</td>
<td>9.4</td>
</tr>
<tr>
<td>Thailand, 1997</td>
<td>8.6</td>
</tr>
<tr>
<td>Argentina, 1995</td>
<td>9.6</td>
</tr>
<tr>
<td>Mexico, 1995</td>
<td>3.6</td>
</tr>
<tr>
<td>Turkey, 1994</td>
<td>0.9</td>
</tr>
<tr>
<td>Financial account balance (US$ billions)</td>
<td></td>
</tr>
<tr>
<td>Indonesia, 1998</td>
<td>10.3</td>
</tr>
<tr>
<td>Korea, 1998</td>
<td>17.3</td>
</tr>
<tr>
<td>Malaysia, 1998a</td>
<td>1.0</td>
</tr>
<tr>
<td>Thailand, 1997</td>
<td>12.2</td>
</tr>
<tr>
<td>Argentina, 1995</td>
<td>7.8</td>
</tr>
<tr>
<td>Mexico, 1995</td>
<td>27.0</td>
</tr>
<tr>
<td>Turkey, 1994</td>
<td>-2.4</td>
</tr>
</tbody>
</table>

- Not available.

*Net private short-term capital flows.


Though financial crises may be short-lived, the associated drops in income have been substantial. Moreover, perceptions of these reversals are made worse by the prior experience of rapid growth in most instances examined. 2 Concerns about impacts on poverty during crises and over the longer run are therefore very real.

Placing the discussion in context is a brief review in the next section of stabilization and macroeconomic responses to recent crises. Labor market responses are then reviewed, followed by an examination of inequality, poverty, and basic needs. The article closes with a summary and some thoughts on policy lessons to be derived.

Stabilization Efforts and Macroeconomic Responses

Countries hit by a crisis have not responded uniformly to falling incomes and capital flight. Despite efforts to cut or postpone government development spending, government consumption declined far more sharply than GDP in Indonesia, Malaysia, Mexico, and Turkey as revenue prospects tightened. Smaller cuts in government consumption occurred in Argentina and Thailand, and there was a marked decline in the growth of such spending in Korea. Despite spending cuts, budget surpluses turned

to deficits in Korea, Malaysia, Thailand, and Mexico, although they were only mildly expansionary relative to GDP. In Turkey, spending cuts were large enough that they reduced persistent deficits relative to GDP in both 1994 and 1995.

In Malaysia, Argentina (briefly), and Mexico the real supply of money and quasi-money was cut in response to the crisis; in Indonesia and Thailand the previous rapid growth in the money supply was sharply curtailed. In contrast, Korea and Turkey rapidly expanded the real money supply, but nominal interest rates rose. Nonetheless, local currencies underwent massive depreciations against the U.S. dollar, and these overwhelmed high interest rates, leaving dollar yields negative (except in Argentina, which maintained a currency board). In countries with the largest depreciations, consumer prices rose rapidly, led by tradables. In Mexico, Turkey, and Indonesia (in 1997), the combined effects were a one-year spike of high real interest rates for local consumers. Milder increases occurred elsewhere.

Consumption smoothing may have relieved the impact of declining incomes for some households, but the poor are significantly less able to smooth consumption than are higher income households (Alderman and Paxson 1994). But only in Indonesia did private consumption rise as a fraction of GDP during the crisis. In Malaysia and Mexico, the propensity to consume fell. Some of the decline may reflect rational responses to changes in prices with a currency devaluation. However, consumption smoothing also becomes more difficult when entire communities experience an income shock.

The combination of falling GDP and consumption, sharp devaluation, and rising interest rates was also typical of the economic crises of the 1980s, which were occasioned by macroeconomic imbalances and ensuing structural adjustments. In Latin America, these crises and responses were associated with greater inequality of incomes (World Bank 1999). Labor market adjustments played a key role in this widening inequality (Horton, Kanbur, and Mazumdar 1994).

Labor Market Responses

How did labor markets respond to sharply falling aggregate demand and shifts in relative prices as local currencies depreciated? To answer this question it is useful to think in terms of two groups of employers. For producers of tradables, depreciation of the exchange rate boosts prices, offering an incentive to increase labor in these sectors. For producers of nontradables, the response depends on the extent to which the decline in aggregate demand is offset by a switch in demand toward nontradables as their relative price declines.

Employers in contracting sectors have at least three margins along which they may adjust: they can cut wages, employment, or hours. When rapid inflation accompanies the crisis, nominal wages are quickly undercut unless employers offer compen-
satory pay increments; with less inflation, stickiness in wages can provide some defense for those who manage to remain employed. Employers are usually reluctant to lay off skilled and professional workers during a downturn, especially if the shock is expected to be short-lived. Indeed, implicit contracts between employers and employees (particularly skilled workers and civil servants) may place most of the adjustment burden on employers, as can explicit contracts resulting from collective bargaining. Voluntary or legally mandated severance agreements may also curtail layoffs. However, bankruptcies can render implicit or even explicit understandings moot.

Several factors may constrain employers in tradable sectors from expanding. Hiring is unlikely to take place if the price increase is expected to be temporary. The need to train new employees may favor expanded hours for existing workers. There may be constraints on capacity expansion, including difficulty obtaining trade credit and working capital during a banking crisis. When contagion leads to regional collapse (as in East Asia), competitor countries also devalue, lowering world prices and regional demand, making export expansion difficult. Meanwhile, employment of unpaid family labor is likely to expand to the extent that the opportunity cost of such employment falls. However, this effect may be offset in nontradable sectors by declining demand.

As some sectors expand and others contract, the ability of workers to move into expanding sectors can be critical to the employment situation. Where mobility involves migration, the costs of relocation may be significant. Still, it may be cheaper for members of rural families who moved to town to return to the village if they lose their jobs because of the fixed costs in living separately and lower living costs in rural areas. Decisions on participation and mobility depend to some extent on perceptions of the depth and permanence of the crisis. But although high unemployment rates may discourage labor force participation, family needs to defend declining incomes tend to encourage greater participation. In countries with a significant immigrant workforce, encouraging the return migration of foreign workers offers another mechanism for adjusting the size of the labor force.

With this stylized picture in mind, what does the evidence show about employment, unemployment, and wages and earnings?³

**Employment**

Of the seven crises identified in table 1, only in Korea did employment fall commensurately with GDP (table 2). Indeed, total employment continued to rise through the crises in Indonesia, Mexico, and Turkey, whereas the decline in employment was less than 3 percent in Argentina, Malaysia, and Thailand.

However, these comparatively small changes in total employment hide some important changes in the composition of employment. The direct and induced effects of crises on sectors and regions can vary widely too. Some workers who are laid off find jobs (often at lower pay) in another sector or location, rather than joining the
ranks of the openly unemployed. Pressures on displaced workers to accept work at reduced pay rather than remain unemployed depend on the ability of their family to support them and perceptions about how long the crisis will last. In low-income countries, even short spells of unemployment can impose a harsh penalty on households, with the result that open unemployment is normally low even in severe crises.

Which sectors manage to absorb laid-off workers? Two types of evidence exist: one on which sectors expanded during a crisis and one on employment transitions among individuals during a crisis. The mere fact that employment expands in certain sectors is no guarantee that these sectors are absorbing individuals laid off elsewhere, because labor force participation can expand during a crisis as a strategy for maintaining family earnings as the wages of some members fall. For instance, since the advent of the recent crisis, labor force participation rates have risen in Indonesia among adults ages 15–24 as school enrollment rates have declined (Poppele, Sumarto, and Pritchett 1999). In contrast, in Thailand, labor force participation declined from 1997 to 1998 in both rural and urban areas and for both men and women (Siamwalla 1998). Nonetheless, any sectors with employment expansion are helping reabsorb displaced workers either directly or indirectly by reducing slack in the overall labor market.

Sectoral changes and employment status. In each country listed in table 2 (except Turkey), employment in construction fell as confidence, credit, and investment levels dropped. Manufacturing employment shrank as the corporate sector was hit by the higher costs of imported materials, difficulties obtaining credit, and the rising burden of debt in foreign currency-denominated instruments. Nonetheless, employment in manufacturing shrank less than proportionately to the decline in manufacturing production (except in Korea). Similarly, despite cuts in government consumption spending, employment was cut little (if at all) in government services, public administration and defense, and community, social, and personal services (see table 2).

Employment in agriculture expanded in several cases. In Indonesia, agriculture employment grew more than 13 percent in 1998, from an already high 40 percent of total employment in 1997. Such expansion would be consistent with a sharp depreciation of the real exchange rate, which raises the relative price of tradable agricultural products. How quickly farmers adjust production to these new price incentives depends on the crop cycle, as well as on other external factors.

For nontradable service sectors, the employment consequences of crises appear to be mixed. Services with strong links to the industrial sector, such as transport and the financial sector, appear to be hurt badly. Some nontradable consumer goods and services gain, and some lose. Urban expenditure levels decline precipitously, but some substitution may occur from increasingly high priced imports to domestic goods and services. Perhaps as a result of this ambiguity, only in Mexico does a clear pattern emerge of an increase in employment in other services through the 1995 crisis.
Table 2. Employment Structure and Growth before and after Crises
(percent)

<table>
<thead>
<tr>
<th>Country and crisis year</th>
<th>Sector</th>
<th>Employment share</th>
<th>Employment growth in years relative to crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-3</td>
<td>-2</td>
</tr>
<tr>
<td>Indonesia. 1998</td>
<td>Total</td>
<td>2.4</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>7.0</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>Agriculture</td>
<td>40.7</td>
<td>-0.4</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>4.9</td>
<td>-5.6</td>
</tr>
<tr>
<td></td>
<td>Community services</td>
<td>14.7</td>
<td>-11.3</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>26.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Korea. 1998</td>
<td>Total</td>
<td>2.7</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>21.3</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Agriculture</td>
<td>10.5</td>
<td>-6.3</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>9.5</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>Public administration</td>
<td>3.1</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Othera</td>
<td>55.6</td>
<td>4.9</td>
</tr>
<tr>
<td>Malaysia. 1998</td>
<td>Total</td>
<td>4.1</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>26.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Agriculture</td>
<td>16.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>9.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Government Services</td>
<td>9.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>36.6</td>
<td>-</td>
</tr>
<tr>
<td>Thailand. 1997</td>
<td>Total</td>
<td>-0.2</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>13.4</td>
<td>-2.8</td>
</tr>
<tr>
<td></td>
<td>Agriculture</td>
<td>50.0</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>6.7</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>Community services</td>
<td>12.7</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Othera</td>
<td>17.1</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>0.7</td>
<td>-1.2</td>
</tr>
<tr>
<td>Argentina, 1995</td>
<td>Manufacturing</td>
<td>16.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Agriculture and mining</td>
<td>1.3</td>
<td>-21.0</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>8.6</td>
<td>-7.5</td>
</tr>
<tr>
<td></td>
<td>Public administration</td>
<td>7.8</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>65.6</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>Mexico, 1995</td>
<td>Total</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>15.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Agriculture</td>
<td>26.5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>6.1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Public administration</td>
<td>4.2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>47.4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Turkey, 1994</td>
<td>Total</td>
<td>-2.5</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>15.1</td>
<td>-7.8</td>
</tr>
<tr>
<td></td>
<td>Agriculture</td>
<td>43.4</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>5.8</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>Community services</td>
<td>11.2</td>
<td>-4.4</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>24.5</td>
<td>-2.5</td>
</tr>
</tbody>
</table>

- Not available.

*In Korea in 1998 and 1999 and in Thailand in 1998: Other refers to total minus manufacturing and agriculture only.

Source: ILO various years; World Bank data.
In addition to changes in the sectoral mix of employment, the change in total employment may mask other shifts in employment status. Thailand Labor Force Survey data show almost no change in overall self-employment or unpaid family work in urban or rural areas from 1996 to 1998. In rural areas, a slight rise in both categories of employment among men was offset by a commensurate decline among women (Siamwalla 1998). Similarly, the National Urban Employment Survey in Mexico records only small absolute increases in self-employed workers and unpaid family workers from 1994 to 1995—though these represent large proportional expansions from a small base. In Korea, too, the number of unpaid family workers appears to have risen from 1997 to 1998, though the magnitude is unclear (Atinc and Walton 1999).

In Indonesia, the male wage employment rate fell from 1997 to 1998 in urban and rural areas, while self-employment and unpaid family work increased, though from a low base in 1997. However, total employment rates of men fell during 1998 (see table 3). Among women residing in urban areas in 1998 the overall employment rate was lower in 1998 than in 1997, as both wage employment and self-employment declined, though family work expanded. In rural areas, however, the employment rate (and labor force participation) rose as both self-employment and family work expanded among rural women. Presumably much of this expansion in employment among rural women was in agriculture, and this adaptation played an important role in smoothing rural household incomes.

Employment transitions by individuals. The second, more direct type of evidence concerns the employment transition of individuals during a crisis. Such evidence is comparatively rare, and some of the turnover in employment status in any panel or recall data may simply reflect errors in measurement. However, because there is no reason why errors in measurement should be correlated with the state of the economy, comparisons of employment churning in downturn and upturn phases (as is done with Mexican data below) can be particularly informative.

| Table 3. Employment Status in Indonesia before and after the 1997 Crisis |
|-------------------|---|---|---|---|---|---|---|
|                   | Any | Wage | Self | Family | Any | Wage | Self | Family |
|                   | work| sector | employed | worker | work | sector | employed | worker |
| **Urban**         |     |       |       |       |     |       |       |       |
| Male              | 71.9| 42.6  | 27.2  | 2.1   | -3.1| -8.9  | 4.0   | 19.0   |
| Female            | 37.7| 19.8  | 12.5  | 5.4   | -1.9| -5.1  | -2.4  | 9.3    |
| **Rural**         |     |       |       |       |     |       |       |       |
| Male              | 85.0| 27.9  | 49.8  | 7.3   | -1.2| -12.5 | 3.4   | 11.0   |
| Female            | 51.7| 11.2  | 18.6  | 21.9  | 5.6 | 0.9   | 7.0   | 6.8    |

Source: Smith and others (1999).
Two rounds of the Indonesia Family Life Survey provide insights into the initial impact of the 1998 crisis. An estimated 75 percent of workers who lost their jobs in construction between 1997 and 1998 had found employment in another sector by 1998; half of those who lost jobs in government, finance, or tourism sectors; and 40 percent of those in agriculture. Despite the low rate of reemployment for those losing employment in agriculture, agriculture was the main absorbing sector (as is apparent from table 2). In general, more churning occurred in employment status in rural than urban areas, and shifts between wage employment and self-employment were more common among women than men (Smith and others 1999:table 7). Only 41 percent of men and 50 percent of women who resided in rural areas and were not working in 1997 reported not working in 1998; most of the remaining women in this group became unpaid family workers or self-employed, whereas most men became wage workers.

Using two panel surveys for Mexico from the third quarters of 1994 to 1995 and from 1996 to 1997, Licona (1999) also records significant churning between wage and self-employment in urban areas. As the crisis intensified from 1994 to 1995, job retention was much higher among protected workers (those with social security coverage or in government employment) than unprotected workers (table 4). Still, only 77 percent of those reporting having a job in 1994 reported having one in 1995, some 5 percentage points lower than the retention rate among protected workers in the upswing from 1996 to 1997.

Although transitions from protected to unprotected jobs proved easier than from unprotected to protected, switches in both directions represented a greater fraction

<table>
<thead>
<tr>
<th>Initial employment status</th>
<th>Employment status one year later (percent)</th>
<th>Out of labor force</th>
<th>Unemployed</th>
<th>Protected wage worker</th>
<th>Unprotected wage worker</th>
<th>Self-employed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td></td>
<td>26.1</td>
<td>19.4</td>
<td>22.5</td>
<td>21.5</td>
<td>10.6</td>
</tr>
<tr>
<td>Protected</td>
<td></td>
<td>5.9</td>
<td>4.5</td>
<td>76.6</td>
<td>8.6</td>
<td>4.4</td>
</tr>
<tr>
<td>Unprotected</td>
<td></td>
<td>14.9</td>
<td>7.4</td>
<td>16.6</td>
<td>46.8</td>
<td>14.3</td>
</tr>
<tr>
<td>Self-employed</td>
<td></td>
<td>16.4</td>
<td>3.6</td>
<td>5.6</td>
<td>12.4</td>
<td>62.1</td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td></td>
<td>21.3</td>
<td>12.7</td>
<td>25.4</td>
<td>28.4</td>
<td>11.9</td>
</tr>
<tr>
<td>Protected</td>
<td></td>
<td>5.3</td>
<td>1.9</td>
<td>81.6</td>
<td>7.9</td>
<td>3.1</td>
</tr>
<tr>
<td>Unprotected</td>
<td></td>
<td>15.3</td>
<td>3.2</td>
<td>21.0</td>
<td>46.7</td>
<td>13.6</td>
</tr>
<tr>
<td>Self-employed</td>
<td></td>
<td>15.1</td>
<td>1.4</td>
<td>6.2</td>
<td>13.8</td>
<td>63.2</td>
</tr>
</tbody>
</table>

Note: Protected means having a government job or a job with social security coverage. 
Source: Derived from Licona (1999).
of those changing status in the upswing than in the downswing. Finding wage employment of either variety was difficult in 1994–95. By 1995, less than two-thirds of those reporting an unprotected wage job in 1994 reported having any wage employment by 1995. Approximately the same share of those changing status reported entering some form of self-employment in both periods. However, workers leaving unprotected jobs were more likely to enter self-employment than were those leaving protected jobs. Nonetheless, the net expansion in self-employment does not appear to have been large. In both the downswing and upswing, some 25–30 percent of both protected and unprotected workers who were changing status reported leaving the workforce. However, the share entering unemployment in the downswing were about double the share in the upswing, with protected workers more likely to become unemployed, given job separation (see also Lucas and Verry 1999).

Migration

Expanding activities may not be located close to the homes of laid-off workers, so willingness to migrate can be critical to employment transitions. Workers displaced from urban jobs must typically relocate to the rural sector to take advantage of any expansion in agricultural employment and associated activities (unless they were previously commuting to town). Displaced urban workers may well elect to rejoin their kin in the villages if only to avoid the additional costs of living in town. There is, however, little systematic evidence on internal migration in response to economic crises.

In Thailand, the crisis is known to have diminished migration to Bangkok by unskilled workers from rural areas, though not by the educated. There are also some indications of substantial reverse migration (Siamwalla 1998; Mahmood 1999). The Indonesia Family Life Survey data indicate that some 6 percent of all prime-age adults moved from urban to rural areas in just one year from 1997 to 1998; just half that number moved in the opposite direction. Thus, net migration to rural areas in response to the crisis was massive.

In Malaysia, the labor market impacts of the crisis were apparently mitigated in part by the nonrenewal of work permits for foreign documented employees, many of whom were previously employed in construction (Pillai 1998). Korea responded to the crisis by repatriating illegal foreign workers (Park 1998). For Indonesia, a major supplier of migrant workers, especially to Malaysia, their mass return undoubtedly exacerbated the downturn in the labor market (Ananta and others 1998).

Unemployment

Despite the fact that the decline in total employment has generally been muted relative to the downturn in manufacturing production, open unemployment rose in the year of crisis in each case in table 5, except in Turkey. It took four years to return to
Table 5. Open Unemployment Rates before and after Crises

<table>
<thead>
<tr>
<th>Country and crisis year</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesia, 1998</td>
<td></td>
<td></td>
<td></td>
<td>4.1</td>
<td>4.7</td>
<td>5.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea, 1998</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
<td>2.6</td>
<td>6.8</td>
<td>6.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia, 1998</td>
<td>3.1</td>
<td></td>
<td>2.5</td>
<td>2.5</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand, 1998</td>
<td>1.1</td>
<td>1.1</td>
<td>0.9</td>
<td>3.7</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina, 1995</td>
<td>7.2</td>
<td>9.1</td>
<td>11.7</td>
<td>15.9</td>
<td>16.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico, 1995</td>
<td>2.8</td>
<td>2.4</td>
<td>3.7</td>
<td>4.7</td>
<td>3.7</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey, 1994</td>
<td>8.4</td>
<td>8.0</td>
<td>8.0</td>
<td>7.6</td>
<td>6.6</td>
<td>5.8</td>
<td>6.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

—, Not available.


prerecession levels following the 1985 recession in Malaysia, but only two years following the 1982 debt crisis in Mexico. In Chile unemployment returned to precrisis levels within three years of the high unemployment episode of 1982.

Open unemployment rates rose with the financial crises of the 1990s, but the absolute rise was substantial only in Argentina and Korea. Helping keep unemployment increases small were reductions in working hours from 1997 to 1998 among both men and women and in rural and urban areas. In Thailand, the largest impact was on short-time work among urban women (Siamwalla 1998).

Wages and Earnings

Despite these adjustments, the main crisis in labor markets was in wages, not employment or unemployment (table 6). In Indonesia and Turkey, inflation dramatically undermined manufacturing wages: real wages fell 44 percent in Indonesia and 31 percent in Turkey in a single year. In Malaysia, real wages declined only slightly in 1998 (as in the recession in 1985), but the effect was intensified because the decline followed a period of very high real wage growth.

Simple ratios of total employment change to GDP change and of real wage change to GDP change in the crisis year show that the impact on total employment was smaller in countries that suffered the deepest wage cuts relative to the decline in GDP (figure 1). Though the number of observations is far too small to generalize, the pattern is suggestive of a potentially important tradeoff that could have significant implications for distributional impacts of crises. Such a tradeoff makes sense intuitively: deeper wage cuts offset the shift in labor demand. It is also clear that countries with the greatest currency depreciations had the largest cuts in real manufacturing wages (figure 2).

The employment cuts reviewed here cover total employment, whereas the wage cuts cover manufacturing alone. How indicative are wage cuts in manufacturing of...
### Table 6. Growth of Real Consumption Wages in Manufacturing before and after the Crises

<table>
<thead>
<tr>
<th>Country and crisis year</th>
<th>Year relative to crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3</td>
</tr>
<tr>
<td>Indonesia, 1998</td>
<td>11.1</td>
</tr>
<tr>
<td>Korea, a 1998</td>
<td>5.2</td>
</tr>
<tr>
<td>Malaysia, 1998</td>
<td>20.9</td>
</tr>
<tr>
<td>Thailand, 1998</td>
<td>10.3</td>
</tr>
<tr>
<td>Argentina, 1995</td>
<td>4.3</td>
</tr>
<tr>
<td>Mexico, 1995</td>
<td>1.6</td>
</tr>
<tr>
<td>Turkey, 1994</td>
<td>21.4</td>
</tr>
</tbody>
</table>

---

*Not available.*

*Industrial sector.*

Source: ILO (various years) and World Bank data; for Indonesia, Islam (1999); for Malaysia, Ministry of Finance, Economic Report.

### Figure 1. Employment and Wage Elasticities Relative to GDP for Selected Crisis Countries

Source: Authors' calculations (tables 1, 2, 6).

those in other sectors of the economy during crises, and what happens to the distribution of wages? Little information seems to be available on this question.

In Indonesia, real wages per hour fell from 1997 to 1998 for men and women, in rural and urban areas. Smith and others (1999), using a nonparametric representation of the wage distribution, find a decline from 1997 to 1998 in the portion of men and women at all wage levels above their respective previous-year medians; the opposite holds for all levels below the medians. Taken together, these indicate widespread wage cuts beyond the manufacturing sector. Among male wage earners, Smith and others (1999:13) found that wage declines were slightly higher in urban areas, largest for the youngest, and uniformly distributed across education groups. Among women, they found that those in the middle of the education distribution suffered the largest declines, along with younger women. The Indonesia Family Life Survey data showed that the relative decline in urban over rural wages was greater for men and women whose wage levels in 1997 were lower. Rural and urban declines were approximately equal for those with higher initial wages.

Given the massive exodus to rural areas already noted and the agricultural base of most of the wage earnings of less-skilled workers in rural areas, the relatively small decline in unskilled wages in rural areas presumably speaks to the expansionary ef-
fects of the exchange rate depreciation on agriculture. Still, the real wages of unskilled workers in rural Indonesia did decline, while the self-employment earnings of men in rural areas (again largely in agriculture) remained largely unaffected in real terms at all levels of earnings (Smith and others 1999). This combination of declining earnings for unskilled agriculture laborers and rising earnings for self-employed farmers (especially net sellers of rice) has strong implications for poverty and income distribution effects (see following discussion). Self-employment earnings fell overall, however, as competition in the urban labor market, presumably combined with declining demand, reduced self-employment earnings in urban areas.

Similarly, in urban Mexico, Licona (1999) records a drop in real monthly earnings from the third quarter of 1994 to the same quarter in 1995. His results suggest some overlap between cuts for wage earners and for the self-employed. Workers in small family businesses fared better than unprotected workers, but less well than protected workers. Single self-employed workers suffered most.

**Household Incomes, Poverty, and Social Spending**

What were the effects of the crisis on household income and distribution and on social spending?

**Poverty and Inequality**

Changes in overall inequality from 1997 to 1998 appear to have been minor in East Asia (table 7). This seems a surprising finding, given the magnitude of changes that occurred and the shifts in inequality other countries experienced during economic crises and subsequent structural adjustment. Inequality in Latin America clearly increased during the crises of the 1980s, though the changes there were measured over longer intervals than the one-year changes reported in table 7.

There may be good reasons to expect different income distribution effects in middle-income then lower-income countries. Bourguignon, de Melo, and Suwa (1991:359) find, from simulations, that “in the standard adjustment package, inequality increased significantly for the Latin American archetype but decreased significantly for the African archetype.” Among the reasons: greater formal sector real wage rigidity and hence higher unemployment in the Latin American archetype exacerbates inequality, whereas the incentives provided to agriculture and the dependence of the poor on agriculture in the African archetype narrows income inequalities.

On this stylized basis, the deep wage cuts in Indonesia, Mexico, and Turkey combined with relatively stable employment should have helped sustain equality through the crisis. In contrast, the relatively high rates of additional unemployment and small wage cuts in Argentina, Korea, and Thailand might have exacerbated inequality. In
Table 7. Poverty Incidence and Gini Coefficients before and after the Crises in Four Southeast Asian Countries

<table>
<thead>
<tr>
<th>Country and crisis year</th>
<th>Poverty headcount</th>
<th>Gini coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Urban</td>
</tr>
<tr>
<td>Indonesia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>11.0</td>
<td>9.2</td>
</tr>
<tr>
<td>1998</td>
<td>13.8</td>
<td>12.0</td>
</tr>
<tr>
<td>1998*</td>
<td>19.9</td>
<td>15.8</td>
</tr>
<tr>
<td>Korea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>2.6</td>
<td>7.5</td>
</tr>
<tr>
<td>1998</td>
<td>7.3</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>9.8</td>
<td>1.2</td>
</tr>
<tr>
<td>1998</td>
<td>12.9</td>
<td>1.5</td>
</tr>
</tbody>
</table>

*Deflation according to price data from Indonesia Family Life Survey 2+.

bFourth quarter; other urban data for Korea in 1997 and 1998 are for first quarter.

Source: For Indonesia, poverty: Frankenberg, Thomas, and Beegle (1999); for Gini: World Bank staff calculations; for Korea, Kakwani and Prescott (1999); for Malaysia, World Bank staff calculations; for Thailand, Kakwani (1999).

four of the seven crisis countries considered here—Indonesia, Mexico, Thailand, and Turkey—agriculture accounts for more than 25 percent of total employment (see table 2); in Malaysia agriculture accounts for nearly 17 percent. In all these countries, agriculture is important economically, and the poor are heavily dependent on agriculture. Bourguignon, de Melo, and Suwa’s (1991) stylization of low-income agricultural economies suggests that currency depreciations would narrow income inequality in these countries.

Indonesia meets both criteria for movement toward greater equality, an outcome supported by the Indonesia Family Life Survey data though not by the Gini coefficients reported in table 7. Thomas and others (1999) derive a Lorenz curve of per capita expenditures for 1998 that is everywhere significantly closer to the diagonal than in 1997. Nonetheless, they estimate that both the top and bottom household quartiles suffered more than households in the middle in declining per capita expenditure from 1997 to 1998. In the bottom quartile, the poorest lost most. In contrast,
in Thailand there is evidence of at least a weak redistribution of incomes from the middle classes to the rich.\(^9\)

Some households seem to have smoothed their incomes during the shock through increased labor force participation or reliance on transfers. In Indonesia, women increased their labor force participation, including unpaid family work. Smith and others (1999) estimate that from 1997 to 1999, the contribution of unpaid family workers to household incomes increased significantly. Quantile regressions on these changes reveal important differences in these effects by income class. In urban areas, upper-income classes benefited from these changes; in rural areas, the poor benefited. Frankenberg, Thomas, and Beegle (1999) note that informal assistance from friends and family members was also important in Indonesia during 1998, with about a quarter of households receiving such assistance. Its median value was considerably higher than that of assistance from formal services. It seems unlikely that most of these transfers came from urban areas because return migration to villages appears to have been the principal strategy of urban migrants from rural areas. Rather, the Family Life Survey data show that per capita expenditure has declined least in communities that are relatively better off (Frankenberg, Thomas, and Beegle 1999). This finding may reflect consumption smoothing within better-off communities, presumably resulting in greater inequality across communities.

Evidence from other contexts indicates that the poor may be less able to smooth consumption. To the extent that interest rates rise during crises, borrowing to smooth consumption becomes more expensive. Many families turn to less formal sources for borrowing funds, though there is little evidence on the impact of crises on such borrowing. A recent survey found that deposits to microfinance institutions continued to rise during the recent East Asian crisis (Atinc and Walton 1999), possibly because they were sounder institutions and rural savers were shifting out of smaller rural banks. This could imply easier access to credit from such sources during a crisis, though recent evidence raises serious questions about whether microcredit institutions actually lend to poorer households (Rai, Topa, and Amin 1999).

It is not surprising to find that the incidence of poverty rose significantly during the crises, though there is no simple association between the severity of declines in GDP and the rise in poverty among the four East Asian countries in table 7. In Indonesia, the increase in overall poverty from 1997 to 1998 and that in urban and rural areas prove very sensitive to the inflation measure adopted. When the increase in the official cost of living is used to deflate expenditures, the overall rise in poverty is not very dramatic and urban poverty rises more than rural poverty. Price data collected in association with the Indonesia Family Life Survey suggest a much higher rate of overall inflation and a sharper rise in poverty.

Thailand experienced wide regional disparities in the poverty impacts of the crisis. The poverty rate dropped from 10.2 percent in 1997 to 9.2 percent in 1998 in
the northern region, while rising dramatically from about 15 percent to 23.2 percent in the northeastern region and from 8.6 percent to 14.8 percent in the southern region (perhaps partially reflecting the fall in world rubber prices). Large regional differences in changes in per capita expenditure emerged in Indonesia as well. Whether the region was a net exporter of rice, and hence benefited from the currency depreciation, seems to have been one factor. Another was the drought of 1997–98, which hurt some regions far more than others.

The impact of currency depreciation on rural areas has generally been mixed. Among self-employed farmers, net sellers of food can be expected to gain as food crop prices rise, while net buyers may well be hurt. The rise in crop prices may translate into greater demand for labor, though labor supply typically increases both through higher participation rates and reduced urban migration. The rising cost of food also erodes agricultural wages.

The Indonesia Family Life Survey data indicate a greater rise in poverty in rural than in urban areas. The same was true in Thailand. In Indonesia, poverty rose more in rural areas despite the sustained real earnings of self-employed men, not only at the mean of self-employment earnings but across the board and despite a greater decline in real urban wages. The explanation may be that a larger portion of rural households was just above the poverty line than before the crisis. In Thailand, Siamwalla (1998) argues that there are not many landless laborers and that the poorest decile of rural households in 1996 derived more of their income from farming than from wage labor. Net buyers of rice and the few households that were more dependent on wage incomes likely bore the brunt of the increase in poverty.

**Social Spending and Safety Nets**

Although the East Asian countries abandoned relative fiscal austerity for counter-recessionary measures and postponed public investment expenditures, real government consumption spending fell or slowed in each of the financial crises considered. Because government-related employment fell little if at all (see table 2), this raises concerns about social spending, public safety nets, and basic needs more generally.

**Health and nutrition.** Public expenditure on health changed little relative to GDP in the four East Asian crisis countries of Indonesia, Korea, Malaysia, and Thailand, which means that it fell in absolute terms (table 8). In Mexico, public spending on health and labor declined 11.6 percent in real terms in 1995 and by a further 5 percent in 1996, though these cuts were less deep than overall cuts in social spending (Lustig 1998).

In Indonesia, use of health facilities by adults and children fell dramatically between 1997 and 1998 as prices on inputs to private facilities rose sharply and stock shortages became common at public facilities (Frankenberg, Thomas, and Beegle 1999).
Table 8. Public Expenditure on Health and Education before and after the 1997 Crises (percent)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesia</td>
<td>0.7</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>1.4</td>
<td>1.2</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Korea</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1</td>
<td>0.6</td>
<td>0.6</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
<td>4.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
<td>1.4</td>
<td>1.3</td>
<td>5.3</td>
<td>4.9</td>
<td>4.0</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Thailand</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.4</td>
<td>1.3</td>
<td>3.4</td>
<td>3.3</td>
<td>3.1</td>
<td>3.4</td>
<td>4.2</td>
</tr>
</tbody>
</table>

As share of GDP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesia</td>
<td>3.2</td>
<td>3.4</td>
<td>3.3</td>
<td>3.1</td>
<td>2.8</td>
<td>6.7</td>
<td>6.7</td>
<td>7.7</td>
<td>3.9</td>
<td>5.2</td>
</tr>
<tr>
<td>Korea</td>
<td>2.0</td>
<td>1.8</td>
<td>2.1</td>
<td>2.2</td>
<td>2.1</td>
<td>19.3</td>
<td>19.2</td>
<td>18.5</td>
<td>15.6</td>
<td>14.4</td>
</tr>
<tr>
<td>Malaysia</td>
<td>5.5</td>
<td>5.5</td>
<td>5.9</td>
<td>6.2</td>
<td>5.6</td>
<td>20.0</td>
<td>21.8</td>
<td>20.9</td>
<td>21.4</td>
<td>21.3</td>
</tr>
<tr>
<td>Thailand</td>
<td>4.4</td>
<td>4.7</td>
<td>4.8</td>
<td>4.9</td>
<td>5.8</td>
<td>14.4</td>
<td>14.4</td>
<td>13.9</td>
<td>13.9</td>
<td>17.5</td>
</tr>
</tbody>
</table>

Source: World Bank staff calculations.
In Thailand, the long-term implications of health spending cuts for the AIDS epidemic has raised concerns (Siamwalla 1998).

There is less evidence on changes in household spending on health care during crises, though the Indonesia Family Life Survey indicates a small cut from 1997 to 1998 in the fraction of household budgets spent on health care in both rural and urban areas. The price of health care rose, implying a real decline in private spending on health care.

But what ultimately matters is whether changes in health spending and nutrition result in greater morbidity or mortality rates. In Indonesia, the changes in health and nutrition status between 1997 and 1998 present a mixed picture. Self-reported health status improved among adults and children, with little difference across income classes or between rural and urban areas. However, self-reported health status can be misleading. More direct measures, such as height for age and weight among children, did not change significantly, and the proportion of children whose weight for height was more than two standard deviations below the median fell. There was, however, a significant decline in mean body mass index among adults and in the proportion of adults in a body mass range considered unhealthy. To Frankenberg, Thomas, and Beegle (1999), this decline coupled with the improvement at the low end of the nutritional spectrum in weight-for-height of children suggests that adults bore a greater share of the nutritional burden imposed by the crisis, through reduced intake or increased energy output (working harder).

The limited evidence on the health and nutritional impacts of crises is mixed. Public and private spending on health care seem to decline, and where food prices increase sharply, nutrition levels may fall sharply as well. The evidence on whether families react by trying to protect the nutritional levels of their children is mixed, despite the potential long-term damage from childhood malnutrition. Where adults absorb most of the nutritional cutbacks, their capacity to work may be affected (see, for example, Thomas and Strauss 1997). However, the evidence from Indonesia indicates no apparent short-term health effects of these changes for adults or children.

Education. In Mexico, government spending on education was cut 9.7 percent in real terms during 1995, though further cuts in education spending were avoided during 1996. In East Asia, public education expenditures also fell relative to GDP in Indonesia, Korea, and Malaysia but rose in Thailand (see table 8). The implications of spending cuts for delivery of education requires more detailed information on which inputs are cut (new school building, number of teachers, materials, university scholarships) and on educational outcomes.

The Indonesia Family Life Survey offers some interesting insights into the impact on outcomes. From 1997 to 1998, enrollment rates declined and dropout rates increased among children ages 7–12 and even more so among youths ages 13–19. Gender differences in these changes were slight. Among children ages 7–12 the
changes in enrollment and dropout rates occurred in rural areas. Among youths ages 13–19, the largest absolute declines in enrollment and increases in dropout rates occurred in urban areas, though after differences in household consumption levels are controlled for, any differences between urban and rural areas are not statistically significant. The decline in enrollment rates was much larger among households in the lowest quartile of per capita consumption as of 1997, and this quartile accounted for almost all of the additional dropout (Frankenberg, Thomas, and Beegle 1999).

Such impacts of crises as decreased enrollment and increased dropout rates are not always observed. In Mexico during the 1982 crisis, high school dropout rates increased slightly while primary school dropout rates fell. Both changes were part of longer-term trends throughout the 1980s (Lustig 1992).

More important than the increase in dropout rates is whether students later return to continue their education. Similarly, a decline in enrollment may reflect delayed entry rather than a failure to ever attend school. Both dropping out temporarily and postponing entry can impose significant costs on lifetime earnings, though this cost is much lower than the impact of permanent withdrawal.

**Social safety nets.** In Mexico, social spending fell 12 percent in real terms in 1995 and another 15 percent in 1996. The government shifted resources out of other anti-poverty programs to a short-term employment program in 1995, creating an estimated half a million jobs. Some 70 percent of these jobs were in rural areas and paid 80 percent of the minimum wage (Lustig 1998).

The East Asian governments significantly increased the budgetary share of safety net spending in response to the crisis. But overall budgetary spending fell, and absolute spending on safety nets was low relative to GDP. Korea had the largest proportional increase in safety net spending, from nearly 0 to 5 percent of the budget. Since 1998 the Korean government has expanded the coverage and budget allocation of its livelihood program. Though the real value of benefits has been maintained for original beneficiaries, only 7 percent of the new poor appear to be covered. Thus the portion of the poor covered fell from 32 percent in 1997 to 17.3 percent in 1998. A public workfare program was also introduced, offering a wage rate below the going market rate. Applications have risen along with the unemployment rate, but eligibility conditions ruled out many who were willing to accept the lower wage (Subbarao 1999). In Indonesia, where public spending on safety nets rose from nearly 0 to 3.6 percent of the budget, ongoing programs frequently pay more than the local going wage (Atinc and Walton 1999). In Malaysia, the safety net has held steady at 0.16 percent of the budget.

To some extent cuts in public social spending may be offset by increased private transfers. However, the limited available evidence suggests that increases in private
transfers are far from sufficient to compensate for public cuts, even in normal times (Cox and Jimenez 1992). Moreover, during a crisis, urban migrants are hardly well placed to increase their remittances to offset declining government transfer programs.

Summary and Conclusions

The dominant labor market effect of the financial crises of the 1990s was a cut in real consumption wages, rather than in employment or hours of work, though unemployment did emerge. Urban self-employment pay seems to have fallen along with wages. Cross-country experience reveals strong positive associations between depreciation of the exchange rate and the cut in real wages and between the cut in real wages relative to the decline in GDP and loss in employment. Allowing high levels of unemployment during a crisis may prove regressive, widening the income distribution. The price of avoiding this may be the acceptance of currency devaluation and the associated loss in real wages.

In a few cases, total employment increased during the crisis, and in others the overall decline in employment was small relative to the decline in GDP. But this does not mean that the changes in employment were negligible. A great deal of turnover in employment accompanied the crises, with movement across sectors and across formal wage jobs, more casual wage employment, and self-employment. This churning was critical to the ability to sustain or even expand overall employment, revealing a considerable degree of flexibility in labor markets. This flexibility may have been particularly high among less-skilled workers, who presumably lowered reservation wages in response to pressures to maintain family incomes.

The initial impact of the crises was on the corporate sectors, especially manufacturing and construction employment. Unpaid family employment expanded, as the need to sustain family incomes rose. Expansion in agricultural employment, with return migration to the villages, also played a significant role in sustaining total employment.

Families smoothed incomes by increasing labor force participation (by women in particular) and private transfers. Only in Indonesia was consumption smoothing observed in the aggregate; indeed, the propensity to consume actually declined during the crises in some cases. Even in Indonesia, consumption smoothing seems to have been concentrated in communities that were better off before the crisis. Poor families are far less able to smooth consumption during idiosyncratic shocks, and this pattern may well be reinforced when entire communities are in shock. Tightening monetary policy to raise interest rates and defend the exchange rate raises the cost of borrowing to smooth consumption, although by increasing expectations of financial defaults and of declining future output, tight monetary policies may
instead weaken the exchange rate (World Bank 1999). To the extent that the poor borrow, it is from informal sources. There is some evidence that loss of confidence in the formal banking sector led to significant transfers of funds to less formal institutions, but the links between formal and informal credit markets during crises remain largely undocumented.

Although the initial impact of financial crises is on the urban corporate sector, some evidence indicates that the incidence of poverty increases more in rural than in urban areas, though there are wide regional disparities within countries. Where agricultural employment is significant, the poor are concentrated in rural areas and depend on agriculture for their livelihoods. Currency depreciation raises the prices of tradables, including food crops, and there is evidence that small farmers can benefit. However, two groups among the rural poor can be hurt by these price increments: net buyers of food crops and landless agricultural laborers.

Despite mildly rising budget deficits and attempts to sustain government consumption spending, real social spending fell, reflecting in part the political infeasibility of cutting public sector employment. In Indonesia, use of health facilities declined markedly with the cuts in health spending, but no increase in self-reported illness occurred, and there was no worsening of nutrition-related problems among children. In Indonesia, adults appear to have borne the brunt of any nutrition-related problems, whereas in Mexico during the crises of the 1980s, girls may have been particularly deprived. Cuts in education spending in Indonesia combined with increased pressures to earn resulted in lower school enrollments, especially among the poorest families, and increased dropout rates, despite new public programs to avoid this.

Most countries introduced a wide range of programs to alleviate poverty. Potentially high startup costs and the brevity of most crises suggest a review of the cost-effectiveness of some of these programs would be warranted. Targeting the new poor is made difficult by an apparent lack of regional correlation between precrisis poverty and additional poverty. Public works programs may be a more cost-effective mechanism, though some new programs offered wages that were above crisis levels or imposed eligibility criteria that precluded participation by newly laid-off workers (Datt and Ravallion 1994). If poverty relief programs are shown to be cost-effective, they should be established during normal periods to avoid the startup costs incurred during the initial phases of a crisis.

Most crises are short. Nonetheless, there are at least three reasons why the short-term poverty impacts of economic crises may have long-term implications even after the economy recovers. First, some workers who lose their jobs during a crisis may not be reemployed in the same field during the recovery. Second, families forced to liquidate assets to smooth consumption may be unable to regain their former livelihood. Third, any declines in nutrition, health, and continuity of schooling may have long-term consequences for labor productivity. Such threats of long-run poverty traps from even a short-lived crisis are in urgent need of further study.
Notes

Peter Fallon is Lead Economist in the Economic Policy Group, Poverty Reduction and Economic Management Network, World Bank. Robert Lucas is Professor of Economics, Boston University. The authors are grateful to Mustapha Kamel Nabli, Duncan Thomas, and three anonymous referees for helpful comments on an earlier draft. Peter Fallon can be reached via e-mail at p.r.fallon@verizon.net, and Robert Lucas at rlucas@bu.edu.

1. The precise year of crisis ought not to be interpreted too strictly in table 1 or subsequent tables, because much can depend on timing of events within a year.

2. Note also that the short-run nature of crises is not irrelevant to production within a year or two. If a comparison is made between an economy that grows steadily at 5 percent a year versus one that suffers a 5 percent decline for two years then returns to 5 percent growth, the latter has a level of production that is permanently 22 percent lower.

3. Most of the evidence on the impact of crises compares measures before and after the crisis. However, some care needs to be taken in interpreting these changes when other elements, besides the direct consequences of the crisis, may blur the picture. For instance, Indonesia suffered from the impact of El Niño simultaneously with the financial crisis (see Datt and Ravallion 1997).

4. In contrast, lack of credit and access to imported materials often seem to constrain the ability of manufacturers to respond quickly to devaluation of the currency. In East Asia in 1998, this effect has been exacerbated by the simultaneous collapse in regional markets.

5. Results in this paragraph are based on a private communication from Duncan Thomas, for which the authors are most grateful.

6. Private communication from Duncan Thomas, for which we are most grateful.

7. Note, however, that mean wages in manufacturing may present a biased picture of the change in wage of employees. For example if, during a recession, low-paid workers are laid off first, then the average wage of those in work may decline by less than the real wage of a typical worker. See Levy and Newman (1989).

8. In a simple hyperbolic regression of the wage elasticity, from table 7, on a constant and one over the employment elasticity, the t-statistic for the coefficient on employment elasticity is -2.59, which is statistically significant at the 95 percent level.

9. There appears to be almost no direct evidence on the distributional consequences of windfall losses and gains to owners of property in declining and expanding sectors during an economic crisis. Given that organized industry typically suffers the major impact we can expect the principal losses in property incomes to fall on the urban elite. However, middle-income groups with property in small-scale urban retail and other service activities may also lose. To the extent that agriculture expands, rural land owners may however prove to be major gainers.

10. The Indonesia Family Life Survey also provides an opportunity to examine poverty transitions during the 1998 crisis because a panel of households was interviewed before and after the crisis. Even during the crisis, many households transition into and out of poverty; for example two-thirds of the poor exited poverty in 1997–98.

11. In contrast, Teruel (1998) relates household expenditures to household composition in four separate years from 1984 to 1994 in Mexico and concludes that during the crisis years women and children (especially girls) were deprived.

References

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


Weak Links in the Chain II: A Prescription for Health Policy in Poor Countries

Deon Filmer • Jeffrey S. Hammer • Lant H. Pritchett

In an earlier article, the authors outline some reasons for the disappointingly small effects of primary health care programs and identified two weak links standing between spending and increased health care. The first was the inability to translate public expenditure on health care into real services due to inherent difficulties of monitoring and controlling the behavior of public employees. The second was the "crowding out" of private markets for health care, markets that exist predominantly at the primary health care level.

This article presents an approach to public policy in health that comes directly from the literature on public economics. It identifies two characteristic market failures in health. The first is the existence of large externalities in the control of many infectious diseases that are mostly addressed by standard public health interventions. The second is the widespread breakdown of insurance markets that leave people exposed to catastrophic financial losses. Other essential considerations in setting priorities in health are the degree to which policies address poverty and inequality and the practicality of implementing policies given limited administrative capacities. Priorities based on these criteria tend to differ substantially from those commonly prescribed by the international community.

In a previous article in this journal, we raise questions about the prevailing orthodoxy concerning appropriate health policy in developing countries (Filmer, Hammer, and Pritchett 2000). That article questions the strategy of promoting of primary health care (PHC) for virtually all countries. It discusses the disappointing experience with this approach in some countries and concludes that it should not be universally promoted because the success of PHC activities is likely to be highly context specific. That article raises doubts as to the universal applicability of the approach. This article addresses the question "What is to be done?" or, given the wide variety of circumstances in different developing countries. "How do we go about determining what is to be done?" We propose a return to first principles suggested by the standard literature of public economics.
What is PHC? Although there have been many interpretations of what constitutes PHC, the approach is best summarized by the World Health Organization's (WHO) definition:¹

Primary Health Care is essential health care based on practical, scientific and socially acceptable methods and technology. It is made universally accessible to individuals and families in the community through their full participation and at an affordable cost to the community and country. Primary health care is the central function and main focus of the country’s health system and of the social and economic development of the community. It is the first contact of the individual, the family and the community with the national health system, bringing health care as close as possible to where people live and work, and constituting the first element of a continuing health care process.

Primary health care rests on the following eight elements:

- education on prevailing health problems and methods of preventing and controlling them;
- promotion of food supply and proper nutrition;
- adequate supply of safe water and basic sanitation;
- maternal and child health care, including family planning;
- immunization against the major infectious diseases;
- prevention and control of locally endemic diseases;
- appropriate treatment of common diseases and injuries; and,
- provision of essential drugs.

The PHC paradigm using this definition has been quite influential in the dialogue on health spending between the international health community and the governments of poor countries, if not in the actual policies implemented. Indeed, some of the words used in WHO’s definition have appeared in Ministry of Health plans (for an example, see Zambia 1994).

The definition does not remove all elements of ambiguity. For example, the statement does not specify who is supposed to provide these services, that is, whether the “country’s health system” is based on public finance and/or provision or on the private sector. It does not define the term “appropriate treatment of common diseases” so as to rule out anything in particular. However, the most common interpretation of this statement gives priority to public financing of these elements, usually implies public provision, and excludes public financing of secondary or tertiary hospital care.

The included activities are a mixed bag of services: those that might be called traditional public health services, such as health education, sanitation and infectious disease control; and others that involve first-level curative care. Our first article noted

---

¹ The World Bank Research Observer, vol. 17, no. 1 (Spring 2002)
the widely variable but often disappointing experience with the curative care component of PHC. It identifies two weak links in the chain of reasoning that lead from public spending on primary health care activities to people's actual health status. The first problem is the difficulty that ministries of health have found in translating public money and program objectives into real health services of adequate quality. Although it is underestimated in virtually all economic analyses of projects, the consequences of not providing proper incentives to public employees to fulfill their duties are particularly serious in the health sector due to the nature of the service. It is a very individual-oriented service requiring specialized labor and both inputs and outputs—particularly in the quality dimension—are difficult to observe. In addition, for doctors on salary, the conflicting motives of performing duties at public clinics versus developing a private business (paid by patients seen or hours worked) give inherent incentives to shirk on the former obligation.

The second problem is that the type of health care that is typically provided at PHC centers, relatively inexpensive (and therefore cost-effective) curative care, is precisely the sort of care for which there is an active private market in the vast majority of poor countries. With a direct substitute for the service, it is possible that the expansion of publicly provided care will crowd out private actors. The result may be that the net increase in services is highly attenuated. Filmer, Hammer, and Pritchett (2000) document the degree of this crowding out in a review of demand studies that analyzes such policy options as raising clinic fees or reducing distance (bringing health care as close as possible to where people live and work). We find that the degree to which visits to private providers were reduced varied widely—from near 0 to fully 100 percent with most clustering between 40 and 60 percent—when prices at public facilities were reduced or distances to them increased. The range of results is not surprising because it depends entirely on local configurations of demand and (private) supply elasticities that could certainly vary widely.

Therefore, money spent on PHC may not translate into real services and, even if it does, the net addition to services provided is attenuated by the reduction of private substitutes. These effects explain the observed, limited impact of the curative component of PHC services.

In this article, we propose that public intervention in the health sector be guided by the same principles that apply to any other sector. Namely, public interventions, including spending, should be judged to the extent that they improve efficiency and the distribution of income, particularly by improving the living conditions of the poorest. This is supplemented with commonsense notions of the relative difficulty of implementing alternatives. Therefore, we discuss the characteristic market failures of the sector and their implied policy solutions, the efficacy of PHC as a poverty-reduction strategy relative to alternatives, the relative difficulty of implementing policies of varying complexity, and how these three concerns stack up against conventional wisdom in health policy.
Efficiency and Market Failure

The standard normative approach to guiding public sector intervention would be to determine those areas of expenditures that yield the greatest improvements in welfare per public dollar spent. Our earlier article tells half the story: assess the increase in services provided publicly net of private displacement effects. But how much is this extra amount worth? If markets were working well, the answer would be "not much" because the marginal value of the service would equal its marginal cost. It is only when markets are not working that the government can actually improve matters by intervening. The existence of such market failures induces a gap between the private value and the social value of services. The size of this gap measures the value of any additional services induced by government policy; in fact, it is only these gaps between social and private values that matter in calculations of benefits from projects (Hammer 1997).

Although health care markets can go wrong in a number of ways, not all market failures are created equal. Two broad categories of market failures with substantial welfare losses characterize the sector. First, there are some activities that combat infectious disease and entail large externalities. Second, there is the virtual absence of private health insurance markets.

Some health-related services, usually those associated with traditional public health activities, are almost pure public goods—those that the private sector will not provide at all because they are nonexcludable and nonrival. The best example is certain forms of vector (pest) control, but we might also include the generation of basic knowledge (where there is no way to restrict the use of information created by research). The difference between social and private value for public goods is equal to their entire value because they would not be provided at all without government intervention.

But public goods are simply an extreme case of the more common form of market failure: externalities, the effects on people other than the ones making the decision to purchase or produce a given good or service. The social value of these services exceeds the private value, but the private value is not zero. Public involvement in the control of infectious diseases, vaccinations, or educational campaigns addresses these problems. Urban sanitation and guarantees of safe water supply are particularly important. The intervention need not involve much public expenditure, but rather regulation and supervision of a monopoly supplier or a public authority with substantial cost recovery policies. The social benefit from increased provision is only part of the total benefit, and this varies from nearly all of the benefit (public goods) to almost none (private goods). The importance of government involvement varies sensitively to the specific nature of the goods being produced.

The second characteristic market failure of the sector is the uncertainty of the demand for the service combined with the limited scope (if not total absence) of health
insurance markets. Speaking of the industrial countries (and thus ignoring the importance of traditional public health), Arrow (1963:946) contends that "all the special features of this industry in fact stem from the prevalence of uncertainty." A plausible theoretical explanation is that the breakdown of markets is due primarily to the asymmetric information problems of adverse selection and moral hazard. Without insurance, the observed demand for relatively expensive procedures may be much lower than is socially optimal. But what kind of services would be most affected by the absence of insurance coverage? Only expensive care would likely be affected because the cost of the procedure must be large enough such that the insurance value exceeds the administrative costs of the insurance scheme. A potential solution is to try to correct insurance markets or even circumvent private insurance by providing public health insurance. Administering such programs can be extremely difficult, and partial or faulty implementation may have many unintended consequences (Gertler and Solon 2000). Issues of the provision of health insurance go beyond the scope of this article. However, in the absence of health insurance, the policy response may need to be the provision of hospitals.

Other possible market failures in the health sector are associated with imperfect information, besides those that generate the breakdown of insurance markets. However, the argument that "imperfect information" is a cause of market failure must be treated with caution. Consumers are almost never completely knowledgeable of every potentially relevant feature of any good. No market is perfect, but the degree of imperfection and the welfare loss associated with it may be large or small. Only the sources of large welfare losses will be high priorities for intervention.

Potentially large failures associated with information occur when problems of asymmetric information are present. One area frequently mentioned in the literature is based on asymmetric information in the medical service market. Medical practitioners, acting as agents for the patients' well-being and having influence over patients' decisions, have financial interests that do not necessarily coincide with those of their clients. Supplier-induced demand is therefore always a potential problem in the medical marketplace. Its effect on welfare is subject to debate and its magnitude is an empirical issue. In poor countries where there is a strong presumption (at least among health professionals) that demand for services is generally too low, the supplier-induced-demand argument does not sound as convincing as in the industrial countries, particularly the United States, where provider decisions interact with the moral hazard induced by third-party payment schemes. The poor-country variant of this might be in the overuse of unqualified practitioners or some types of traditional healers.

Finally, the problem of the general lack of information may induce inappropriate demand. That is, there might be too little demand for broccoli and immunizations, and too much demand for tobacco and, perhaps, traditional healers. Again, the argument must be made with caution. It is strongest where there is no commercial
(private) product associated with the action: hand washing is less likely to be promoted than cold remedies. It is also most persuasive where there is little or no media access: societies with a large illiterate population and with little radio or television penetration may not hear information that the media would spread.

Note that the emphasis here is not just on health status but on welfare (in an economist's sense of the word) for defining the value of services over their market-supplied levels. People value good health, and to a large extent, higher welfare and better health coincide. However, sometimes they diverge. In some circumstances, health improvement might be possible, but it may come at too high a cost in terms of other things people value. For example, perfectly rational individuals, fully informed as to the impact of diet on health, might still choose to eat more fatty foods or sweets than are consistent with optimal health. The desire for goods other than health status opens up the possibility of governments improving health with current resources. However, in some circumstances, there is doubt about the desirability of doing so because it may require excessive resources or involve such a sacrifice in other dimensions of well-being that it would actually reduce welfare.

The reverse is possible as well: there can be justifiable health sector interventions that do not affect health status. The peace of mind brought by improving insurance against catastrophic financial loss may not translate into improved health status, but it may well be an important outcome of health policy. Similarly, improving proximity or amenities in health services to the extent that people desire and are willing to pay for them improves welfare but not necessarily health.

The schematic in figure 1 summarizes some of the implications of the above arguments. The figure classifies various health sector services along two dimensions. The first is the degree to which their markets are subject to serious failures that yield a large degree of distortion between public and private valuation. The second is the degree to which public intervention, provision of service or subsidies, can be expected to lead to increased total or net use of services. In general, larger increases will be associated with services having more elastic demand and less elastic alternative supply since a subsidy will induce larger demand and entail less crowding out of the private sector in these circumstances. Improvement in welfare is closely approximated by the product of the degree of distortion (the difference in social and equilibrium valuation of services) and the change in the use of the services induced by policy.

The highest-priority items from a welfare point of view are those in the upper left quadrant (I). The lower left quadrant (II) shows services in which there are large market failures but little responsiveness to public policy. The upper right quadrant (III) shows areas in which use of services may be greatly influenced by public action but that have little effect on welfare or health. As we argue in our previous article, demand for treatment of minor ailments is likely to be more highly elastic than for more serious illnesses. The lower right quadrant (IV) reflects relatively inexpensive curative care in clinics. Not much is to be expected from public provision because
Figure 1. Classification of Health Interventions by Responsiveness to Policy and Degree of Distortion

<table>
<thead>
<tr>
<th>High responsiveness to policy</th>
<th>Low responsiveness to policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>High degree of distortion</td>
<td>Low degree of distortion</td>
</tr>
</tbody>
</table>

I
Vector control, vaccinations, urban sanitation, hospital care when there is no insurance for catastrophic loss, research, information on treatment prices and options, information on basic hygiene

III
Low-value or ineffective public goods, routine clinical care when the perceived risk from not treating is low

II
Emergency care, serious but affordable care, information from noncredible government

IV
Routine curative care when the perceived risk from not treating is high

Note: Classification of health interventions by responsiveness to policy is measured by the change in equilibrium consumption as a result of a change in policy. The degree of distortion is the difference between private and social valuation of services.

people are likely to seek care anyway and the scope for market failure is limited as well.

It is important to note again that the health implications of these interventions are not completely coincident with welfare effects. For example, the absence of an insurance market leads to the possibility of large gains in welfare from public provision of hospital services at subsidized rates. Whether this gain shows up in terms of improved health status or in terms of increased peace of mind depends on the circumstances. If people tend to sell assets to get hospital care or to go into debt when serious illness strikes, there will be no health effect but a significant welfare impact. If the lack of insurance reduces access to life-saving care, the effect will show up in health status.

How does PHC fit into this picture? Different parts of the PHC package have different impacts. High subsidies for primary-level curative care will tend to be covered by
the low elasticity and low distortion quadrant of figure 1 (quadrant IV). For most serious conditions with cheap curative care treatments, demand will be relatively inelastic and there will be more substitution with the private sector. Public subsidy for this type of service will result in a transfer of income from taxpayers to patients but will have few efficiency benefits and little impact on health because these are the services for which private markets exist.

Clinical services for less serious ailments will tend to fall in quadrant III. The subsidization of first-contact clinical services without fees to screen for severity may be the most serious public policy issue raised by this quadrant. Substantial resources may be used in administrative costs and the provider's time in subsidizing relatively minor ailments. In Indonesia, many of the visits to district hospitals were for muscle aches and skin rashes (World Bank 1994). These also tended to be the ailments that fell most significantly when fees were raised. Private practitioners are well situated to handle these kinds of ailments. In the United States, the most rapid increases in Medicare payments have been for home-care services, including housekeeping, and are quite elastic. In poor countries, the satisfaction of consumer demand for clinical services very likely crowds out the delivery of more population-based public health by eating up budgets both of money (Gertler and Hammer 1997) and of time (Hammer and Jack 2001). The fact that such demand is elastic implies that the quantitative effect and loss of welfare are large.

In sum, there is reason to doubt the likely impact on health and welfare of the curative care components of PHC strategies that often account for the bulk of their expense.

Equity: A Tradeoff with Efficiency?

Some would argue that the preceding discussion is irrelevant because public intervention in the health sector, particularly at the primary level, is justified on the basis of its impact on poverty or at least on the health status of the poor. Therefore, reversing a PHC commitment because it provides low-quality services or because it provides the “wrong services” from a public goods/welfare analysis misses the point, which is that such expenditures are simply a transfer to the poor. In that case, the policy must be compared with other possible transfer mechanisms. As it turns out, some forms of health expenditure can be justified on both equity and efficiency grounds, such as those related to communicable disease control. Others, such as hospital care, involve conflicting equity and efficiency effects, creating a tradeoff to be resolved. PHC again falls into a murkier middle ground.

As far as curative care is concerned, table 1 compiles results on how the benefits of public spending on health are distributed across individuals in each per capita income (or consumption) quintile in 12 countries. Overall public spending on health
Table 1. The Distributional Incidence of Public Spending on Health

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Poorest</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Richest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>1991</td>
<td>1</td>
<td>0.62a</td>
<td>0.62a</td>
<td>0.62a</td>
<td>0.18</td>
</tr>
<tr>
<td>Brazil</td>
<td>1990</td>
<td>1</td>
<td>2.25</td>
<td>3.75</td>
<td>3.13</td>
<td>2.50</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1995</td>
<td>1</td>
<td>1.23</td>
<td>1.62</td>
<td>2.00</td>
<td>1.92</td>
</tr>
<tr>
<td>Chile</td>
<td>1982</td>
<td>1</td>
<td>1.02a</td>
<td>1.02a</td>
<td>1.02a</td>
<td>0.50</td>
</tr>
<tr>
<td>Ghana</td>
<td>1994</td>
<td>1</td>
<td>1.25</td>
<td>1.58</td>
<td>1.75</td>
<td>2.75</td>
</tr>
<tr>
<td>Indonesia</td>
<td>1987</td>
<td>1</td>
<td>1.17</td>
<td>1.58</td>
<td>2.25</td>
<td>2.42</td>
</tr>
<tr>
<td>Kenya</td>
<td>1992</td>
<td>1</td>
<td>1.21</td>
<td>1.57</td>
<td>1.57</td>
<td>1.71</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1989</td>
<td>1</td>
<td>0.69a</td>
<td>0.69a</td>
<td>0.69a</td>
<td>0.38</td>
</tr>
<tr>
<td>Mongolia</td>
<td>1995</td>
<td>1</td>
<td>1.11</td>
<td>1.06</td>
<td>1.09</td>
<td>1.34</td>
</tr>
<tr>
<td>South Africa</td>
<td>1993</td>
<td>1</td>
<td>1.40a</td>
<td>1.40a</td>
<td>1.40a</td>
<td>1.06</td>
</tr>
<tr>
<td>Uruguaya</td>
<td>1989</td>
<td>1</td>
<td>0.57</td>
<td>0.46</td>
<td>0.38</td>
<td>0.30</td>
</tr>
<tr>
<td>Vietnam</td>
<td>1993</td>
<td>1</td>
<td>1.33</td>
<td>1.75</td>
<td>1.83</td>
<td>2.42</td>
</tr>
</tbody>
</table>

Note: Values show the amount of public spending received by each income quintile relative to the poorest quintile.

*Distribution across these quintiles is distinguished in the original source.

Quintiles are defined on household basis, not per capita basis. See Filmer, Hammer and Pritchett (1998) for additional details on the results of the studies underlying these data.

Source: For Argentina, Brazil, Chile, Colombia, Honduras, Kenya, Madagascar, Mongolia, Nicaragua, South Africa, Tanzania, Uruguay, and Vietnam, World Bank poverty assessments and country studies; for Guyana, Jamaica, Trinidad and Tobago, and St Lucia, Baker (1997); for Bulgaria and Ghana, Demery (1997); for Malaysia, Hammer, Nabi, and Cercone (1995); and for Indonesia, van de Walle (1994).

is propoor in only 3 of 12 cases (Argentina, Uruguay, and Malaysia). Only in these three did the fourth quintile receive lower per person benefits than the bottom quintile. Often the richest quintile will have low usage of public facilities because it makes up a large part of the clientele of the private sector. But the main beneficiaries of public subsidies are not among the poorest. This should not be surprising because the income elasticity of demand for health is widely estimated to be very high both within and between countries. Most estimates are in the neighborhood of 1.5. It is unclear why a high-income-elasticity good would be singled out for subsidization for the sake of the poor.

Typically, the distribution of benefits is less equal than a uniform transfer would be, but more equal than the distribution of income (so that public spending on health financed by a proportional tax would be progressive). Moreover, from the analysis above, it should be obvious that even if the distribution of public spending on health were uniform across income groups, the impact of public spending on health status would be larger for the poorest. The net impact on consumption of health services of
a given amount of public benefits is likely to be larger for the poorest because the
displacement effects are smaller. This is consistent with the findings of Bidani and
Ravallion (1997) that public spending has no demonstrable impact on the nonpoor
but is important for the poor.

Table 1 shows how the distribution of public health spending varies across coun-
tries, a pattern that is true not only for health. Recent compilations show that for
government subsidies to education, the ratio of the benefit received by the richest
quintile to the benefit received by the poorest quintile ranges across the same order
of magnitude as that for health, approximately 0.8 to 5 (Castro-Leal and others 1999;
Li, Steele, and Glewwe 1999). In general, however, public subsidies to primary edu-

cation are better targeted at poor households. In a review of nine Sub-Saharan Afri-
can countries, Castro-Leal and others (1999) find that in all but two the benefits of
public subsidies to primary education accrue to the poorest quintile more than they
do to the richest. For public subsidies of food, the experience is that in the best cases
the benefits are close to being uniformly distributed. When the item is an inferior good
(that is, a commodity on which households will spend less as income increases), the
poor will capture relatively more of the subsidy. However, because these are rare cases
that do not usually consume much of the household’s budget, the value of the trans-
fer is quite low (Alderman and Lindert 1998; Grosh 1994). These two articles also
suggest that the distribution of benefits range substantially, although the order of
magnitude of this range is similar to that for health spending.

Evaluating the full distributional impact of public spending on health is even more
difficult than calculating the incidence of expenditures because the distributio-

nal impact of raising revenue needs to be evaluated as well. The net welfare effect of using
a regressive tax to fund progressive expenditures is very hard to evaluate. Combin-
ing the incidence of raising and spending revenue is rarely done, mainly because of
the lack of data. Table 2 summarizes the findings of two recent studies that have at-
ttempted to combine available data to evaluate the incidence. Both studies find that
the incidence of taxes was roughly proportional but public expenditures were pro-
gressive, leaving the overall incidence progressive. That is, as a share of household
income (or expenditures), the poor benefit more from the combined effect of taxes and
public expenditures. In Africa, where tax incidence can sometimes be regressive due
to reliance on agricultural export and other trade and consumption taxes, the net
effect can be very bad (World Bank 1991).

A frequent argument is that expenditures on PHC are more pro-poor than are ag-
gregate health expenditures, which include hospitals and the like. Table 3 shows the
ratio of the share of benefits received by each quintile from two different types of spend-
ing. For instance, in Indonesia, 18 percent of the benefits of spending on public health
centers accrues to the poorest quintile, and only 8 percent of the spending on hospi-
tals does. Thus, although PHC is (slightly) less progressive than a uniform transfer,
the ratio of benefits of the two types of spending is 2.25 for the poorest quintile. In
Table 2. The Net Effect of Taxes and Spending as a Share of Per Capita Household Expenditure, Mongolia and the Philippines, 1990s

| Quintile | Mongolia | | | Philippines | | | | Combined | | | | Combined |
|-----------|----------|---|---|-----------|---|---|---|-----------|---|---|---|-----------|---|---|---|
|           | Taxes    | Expenditures* | Combined incidence | Decile | Taxes | Expenditures* | Combined incidence |
| Poorest   | 0.080    | -0.238       | -0.158             | Poorest | 0.208 | -0.469       | -0.261             |
| 2nd       | 0.090    | -0.168       | -0.078             | 2nd     | 0.205 | -0.222       | -0.017             |
| 3rd       | 0.070    | -0.100       | -0.030             | 3rd     | 0.201 | -0.175       | 0.026              |
| 4th       | 0.100    | -0.092       | 0.008              | 4th     | 0.200 | -0.144       | 0.056              |
| Richest   | 0.100    | -0.064       | 0.036              | 5th     | 0.198 | -0.122       | 0.076              |
|           |          |              |                    | 6th     | 0.199 | -0.102       | 0.097              |
|           |          |              |                    | 7th     | 0.201 | -0.087       | 0.114              |
|           |          |              |                    | 8th     | 0.197 | -0.069       | 0.128              |
|           |          |              |                    | 9th     | 0.197 | -0.051       | 0.146              |
| Richiest  |          |              |                    | Richiest| 0.196 | -0.001       | 0.195              |

*Expenditures include health and education in Mongolia and health, education, and infrastructure in the Philippines.


Table 3. How the Richest and Poorest Quintiles Benefit from Different Types of Public Spending on Health, Selected Countries, 1990s

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Type of public spending</th>
<th>Primary or health center</th>
<th>Primary or health center</th>
<th>Ratio of primary or health center to hospital</th>
<th>Poorest</th>
<th>Richest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subsidies that go to the poorest quintile (percent)</td>
<td>Subsidies that go to the richest quintile (percent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1995</td>
<td>Primary/hospital</td>
<td>11</td>
<td>27</td>
<td>1.45</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>1994</td>
<td>Primary/hospital outpatient</td>
<td>13</td>
<td>35</td>
<td>0.77</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Guyana</td>
<td>1994</td>
<td>Health center/hospital</td>
<td>19</td>
<td>12</td>
<td>1.47</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>1987</td>
<td>Health center/hospital</td>
<td>8</td>
<td>17</td>
<td>2.25</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Jamaica</td>
<td>1989/92</td>
<td>Health center/hospital</td>
<td>19</td>
<td>11</td>
<td>1.32</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>1992</td>
<td>Health center/hospital</td>
<td>13</td>
<td>13</td>
<td>1.85</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>1993/94</td>
<td>Health center/hospital</td>
<td>11</td>
<td>21</td>
<td>1.64</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>1992</td>
<td>Health center/hospital</td>
<td>17</td>
<td>36</td>
<td>0.47</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>St. Lucia</td>
<td>1995</td>
<td>Health center/hospital</td>
<td>26</td>
<td>17</td>
<td>0.96</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>1993</td>
<td>Community health center/hospital outpatient</td>
<td>9</td>
<td>10</td>
<td>2.11</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>

7 of 10 cases, the poorest quintile benefits proportionately more from the lower-level facilities than from hospitals, and the richest quintile benefits proportionately more from spending on hospitals.

However, the number of cases for which the poorest quintile receives more than its population share from PHC is equal to that where it does not (see Filmer, Hammer, and Pritchett 1998 for more on the data underlying table 3). For instance, in Bulgaria, the poorest quintile receives only 16 percent of the benefits of public spending on primary facilities, whereas in Guyana it receives 28 percent of the benefits of public spending on public health centers. In both cases, however, the ratio of this share to the share received from spending on hospitals is about the same (1.45 and 1.47, respectively). Thus, recent studies tend to confirm previous findings about the favorable distributional impact of lower-level spending versus hospital care. This is not because PHC is always strongly propoor but because hospital spending is nearly always strongly prorich.

The same pattern probably extends to the comparison between the clinical components of PHC and traditional public health interventions aimed primarily at infectious disease (vector control, immunization, and sanitation). We must say "probably" because almost by definition it is difficult to assess the distributional consequences of spending on true public goods (Cornes 1995). However, existing evidence indicates that the poor suffer disproportionately from infectious diseases and would benefit most from their control. For example, in India, estimates of the prevalence of tuberculosis vary by a factor of seven between the poorest and richest deciles (by one measure of wealth); estimates for malaria vary by a factor of three. In contrast, the prevalence of blindness, overwhelmingly caused by cataracts, a chronic condition of old age, is only 40 percent higher among the poor relative to the richest of these groups (World Bank 1998; Bonilla-Chacin and Hammer 1999). Although the poor suffer from virtually everything more than do others, the differential burden is the greatest in the communicable diseases. A reallocation of health resources from public hospital services to PHC would likely improve the distribution of benefits but at the expense of corrections to the insurance problem that public hospitals might solve. A reallocation from population-based services to PHC-type clinical services, however, would result in losses in terms of both equity and efficiency.

The problems of poverty and the lack of insurance are interrelated in two additional ways. First, due to their lower incomes, poor people will find a much lower level of expenditures catastrophic. Therefore, in principle, the poor could benefit at least as much as others from the financial protection of subsidized hospital-based services provided that the management or political economy problem of ensuring access to such services can be solved (admittedly an enormous proviso). For example, a recent study calculates the value of the "risk premium" for coverage of costs of inpatient versus outpatient services by income group in India (World Bank 1998). In the absence of an insurance market, the risk premium measures the welfare loss of facing
risk in monetary terms. The study finds that the welfare loss of risk exposure—measured at the average cost of inpatient services—was as much as 60 percent of the expected cost of these services for each poor person, three times higher than for people with twice the average poor person’s income. In this sense, insurance is more important for the poor. Devarajan and Hammer (1998) make similar calculations for emergency maternity services in Indonesia. Alternatively, treatments for the kinds of problems that disproportionately hurt the poor might justify lower caps on payments in public facilities.

The second way poverty and lack of insurance are interrelated is that catastrophic health events may be responsible for people falling into poverty. Two aspects of catastrophes need to be separated: the financial burden of paying for medical treatment and the loss of earning capacity from disability. On the financial burden, Narayan (2000) finds that poor people have a common fear of the possibility of having to resort to distress sales of major assets, such as livestock, as a result of an episode of bad health. Such actions could lead to outright destitution for the family. Once again, this implies that relatively expensive procedures need to be covered publicly when insurance is absent. Few people are pushed into poverty or forced to sell cattle due to payments for PHC-type activities. It is the unexpected hospital bill that would do so.

On the loss of earning capacity, the essential problem is the absence of disability insurance. Subsidized health care may have little or nothing to do with correcting this market failure, except in those cases where the postponement of medical care due to anticipated costs leads to disability. This is an interesting avenue for future research, but it requires attention to the details of the timing of the search for health care, a degree of subtlety captured by few, if any, demand studies. Gertler and Molyneaux (1997) indicate that the disability-inducing effect of poor health rather than the financial burden of out-of-pocket medical expenses leads to increased poverty in Indonesia. Smith (1999) finds similar evidence in the United States. This implies that policies to mitigate increases in poverty due to medical problems may have nothing to do with the provision of medical care.

Whether PHC is a good means of redistribution needs to be evaluated country by country and against other such programs. As shown above, the success of targeting, even for primary or primary-style services, varies widely across countries. Any particular country arguing for PHC as a redistributive mechanism must be careful that it is indeed achieving that aim. Moreover, there are other means outside the health sector for redistribution within a country. If the argument for PHC is based on its redistributive properties, it needs to be compared with other antipoverty schemes, some of which may be more (or less) successful at targeting and some of which may be more (or less) feasible.

Effectively targeted programs may be appealing because they appear to maximize the poverty impact of a fixed budget; however, they may be politically unsustainable. If a large part of the motivation behind public spending on health is redistribution,
then why is the current incidence of public spending not well targeted toward the poor? The reason may lie in the fact that when the number of recipients of a publicly provided benefit falls, political support for the program may disappear. Gelbach and Pritchett (1997) construct a simple economic model of transfers with voting and show how, under reasonable assumptions, the welfare-maximizing outcome for the poor is a universal transfer because when benefits are targeted to the poor, the nonpoor will vote to reduce the overall budget devoted to transfers.

**Implementation: What Can Governments Really Promise?**

Efficiency and equity are the bread and butter of economic analyses of policy. A less commonly analyzed area—but one that is certainly of concern to real-life policy decisions—is the assessment of the relative difficulty of implementation of different programs. We do not have the same standard tools of analysis for addressing implementation as for equity and efficiency. Our earlier article discusses certain incentive problems that plague the health sector. Here we consider how the curative care aspects of PHC differ from other health policies in terms of monitoring, enforcement and implementation.

A common problem in PHC systems is the difficulty of staffing facilities in rural areas. Statistics on this are hard to find, although anecdotal evidence is enormous. Positions in rural areas are often vacant for long periods of time in many countries. Data from Indonesia (World Bank 1994) show that in the remote province of Irian Jaya, the rate of unfilled posts is as high as 60 percent. Worse than vacant posts (because vacancies do not necessarily cost money) posted medical personnel are often not present at all. In an intensive study of a PHC post in Bihar, India, Khan, Prasad, and Quaiser (1987) find that three of the four medical officers assigned to the post were not seen in the month of the researchers’ visit. Two did not live near the PHC location and were busy with their own private practices elsewhere. They did, however, draw their salary. The officer in charge did not complain, according to nurses, because the presence of the other doctors would have cut into his own private practice.

This is not an idiosyncratic problem. Medical personnel are highly educated relative to the rest of the population in all countries, and they prefer urban life for many reasons, from income-earning opportunities to urban amenities to better educational opportunities for their own children. It is always difficult to induce medical personnel to live in rural areas. In a “willingness to accept” study of Indonesian medical school graduates, Chomitz and others (1998) find that the amount of pay required to induce relocation to the outer islands is multiples of current wage rates (for students who had not come from those places originally). Poorer countries have difficulties retaining even less well-educated providers because their training is still enough to earn a premium in urban markets.
Once they have staff in PHC centers, countries may have problems providing conscientious care and monitoring the behavior of relatively elite members of society in providing a service with many unobservable characteristics. In addition, in many PHC settings, the medical personnel must do a variety of tasks—providing primary care as well as public health activities. Hammer and Jack (2001) note that when both activities are the responsibility of the facility, local pressure tends to bias time spent toward curative care and that it is hard to create incentives for maintaining the population-based activities. The fact of the matter is that it is not easy to monitor and regulate the behavior of decentralized, complex activities.

How does this compare with other kinds of health sector policies? There is little systematic evidence on this. However, single-purpose, campaign-style activities, such as immunization drives or infrastructure investments for water and sanitation, do not require continual staffing of rural clinics with educated personnel. For example, a campaign to increase polio vaccinations has been very successful in India, but it only asks personnel to go to rural areas for two or three days a month rather than to move their families (Deshpande 1999). Most traditional public health activities—health education campaigns, maintaining water and sanitation infrastructure, health inspections and immunizations—can be done with episodic visits rather than permanent residence. Permanent residence, required for PHC staffing, may well be better for the rural population but is much harder to achieve.

Similarly, it may be more feasible to maintain attendance by medical personnel by organizing them in hospitals rather than small clinics. Doctors tend to like working in hospitals, so attendance is easier to ensure, as is peer monitoring of professional work. When asked about the determinants of job satisfaction, Indian doctors ranked interaction with colleagues and access to equipment and materials that gave them the ability to make use of their training as most important (World Bank 2001). These are much more likely to be found in hospital settings.

We confess ignorance of the institutional reforms needed to improve incentives for public providers. The answer to the question of which institutional design— corporatization, regulation, or subcontracting to private providers or nongovernmental organizations—would work better in any given setting requires more in-depth work. How to get better service from public officials—whether health care workers, teachers, or those providing core central government services—should be a high priority for research in public economics. However, it is likely that managing a network of PHC clinics is particularly difficult relative to other types of government intervention.

Summary

Table 4 summarizes our argument. It contrasts six different approaches to health policy as far as the relative importance each attaches to three gross categories of health

Deon Filmer, Jeffrey S. Hammer, and Lant H. Pritchett
Table 4. Relative Priorities and Tradeoffs in Health Policy under Different Approaches

<table>
<thead>
<tr>
<th>Priority</th>
<th>Traditional public health</th>
<th>Primary healthcare, clinic-based</th>
<th>Hospital-based care</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Status quo (varies widely)</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>2. Alma Ata (ideal)</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>3. Alma Ata (real)</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>4. Economic efficiency</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>5. Economic rationale</td>
<td>Higher</td>
<td>Varies</td>
<td>Not so high</td>
</tr>
<tr>
<td>(efficiency and equity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Full public sector rationale</td>
<td>Highest</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>(efficiency, equity and implementability)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

describe expenditure: traditional public health interventions that are not clinic-based and/or have large externalities associated with them, the clinic-based health care component of PHC, and publicly provided hospital care.

The first row in table 4 describes the status quo in many developing countries with heavy emphasis on hospital care. Although countries vary widely in this regard, the generalization is not far wrong and serves as a foil for the second row. The second row characterizes the emphases in the Alma Ata convention, which reversed the priorities. The third row suggests an acknowledgment that, for a variety of reasons, the primary care component has tended to dominate discussions of health policy, particularly in the selective PHC approach.

The fourth through sixth rows summarize the economic arguments for the three different sets of policy options incorporating successively more comprehensive considerations. The fourth row highlights the areas in which the largest market failures in terms of welfare losses are likely to be found. These are the delivery of services that are most like public goods and, unless the endemic problems of insurance markets are corrected directly (a very difficult task), the delivery of expensive services that generally need to be done in hospitals. These would be the areas of emphasis on strict economic efficiency grounds. Note the contrast with the PHC approach. Although the fourth row emphasizes traditional public health, it questions the priority of inexpensive curative care, an area in which the large private sector competes. Although those markets are not perfect, they are better than those characterizing the other two areas of policy.

In the standard economic rationale that includes equity considerations (the fifth row), the story is less clear. Although the rationale for traditional public health interventions to combat infectious disease is reinforced, the relative weight on primary and higher-level care shifts away from hospitals. How much it changes depends on the importance of the insurance problem (and the extent to which it cannot be handled any other way) relative to the bad distributional consequences of subsidized hospitals (and the extent to which referrals to hospitals cannot be restricted to those really requiring hospital treatment).
The sixth row reflects the argument that running a PHC network is simply hard to do. This is the reason why quality is low and clinics go understaffed and underused. The entries for the other two columns are frankly speculative and represent our judgment that both traditional public health programs and hospitals are more manageable (there have been failures in these areas as well). At this point, we merely note that staffing and maintaining a wide network of clinics appear much more difficult than running fewer, more easily monitored operations.

Taken together, the efficiency, equity and implementability of policies stand as a challenge to conventional wisdom concerning PHC. The relative emphasis of the conventional wisdom, especially as often presented (the second row), is directly at odds with the economic rationale for public sector involvement.

Conclusions

If the answer to the question “What should be done?” is not “It depends,” either the question was trivial or the answer was wrong. By the same token, to answer “It depends” without saying on what it depends and in what measure is equally uninteresting. Neither theory nor empirical outcomes support the obvious policy of reallocating resources from “ineffective” tertiary to “effective” primary care. There are three “depends” that must factor into public policy.

First, health policy depends on the anticipated efficacy of the public sector under existing institutional arrangements. If this is low—and it has been extremely low in many developing (and more than a few industrial) country settings—then adopting strategies that are intensive in public sector capacity are of dubious validity. Providing a population-wide network of primary-level facilities that provide quality clinical care and integrate into a comprehensive chain of referral is an extraordinarily capacity-intensive task. Kerala might be able to do it, Bihar certainly cannot.

Second, health policy depends on the underlying justification for public intervention. If it is because government is providing a pure public good and individual cost recovery is impossible, then there is no alternative to the public sector (for example, vector control or disease surveillance). However, if a supply of services would be forthcoming if there were effective demand (for example, clinical services), then even in the presence of externalities (for example, immunizations), public provision may not be the best way to increase consumption. Alternatives that leave power and choices in the hands of consumers might be preferred.

In low-income countries with low capacity in the public sector (which may well be most low-income countries), the highest priority should be basic public health, control of infectious diseases where possible, and those programs that have a track record of effective administration (for example, vaccination campaigns). The stance toward inexpensive curative services must be very carefully crafted. We have ar...
gued that public provision in these countries is extremely hard to manage. But usually in the same countries the private sector is also of very low quality. Capacity rises together in both public and private sectors. In some circumstances, leaving such services to the market (or at the very least charging for them so they do not drain public resources) may be the lesser of two evils and would at least allow countries to cover the crucial public goods. In countries with slightly higher capacity, the focus should be to regulate the market and perhaps provide demand-based instruments.

Third, the impact of health policy depends on how responsive individuals' decisions are to public actions. Health care services that are cheap and critical are extremely unlikely to be sensitive to price, except for the very poorest of the poorest countries. In higher-income countries that are further along in the epidemiological transition, the development of mechanisms to pool risk is the key element and expansion of primary curative services is unlikely to be important.

In sum, we emphatically do not defend the common developing country status quo in which the public spends large amounts on ineffective secondary and tertiary facilities that serve primarily a rich, urban clientele. That said, there are very few instances where the conventional prescription for the government to supply PHC, as currently applied, as its main strategy will be the right one.

Notes
Deon Filmer is Senior Economist and Jeffrey Hammer is Lead Economist with the Development Research Group at the World Bank. Lant Pritchett, on leave from the World Bank, is Lecturer at the Kennedy School of Government. This article is a revision of "Health Policy in Poor Countries: Weak Links in the Chain," Policy Research Working Paper No. 1874, DECRG, World Bank, January 1998. This article grew out of an earlier collaboration between the authors and Maureen Lewis and Samuel Lieberman. The authors would like to thank Phil Musgrove, Martin Ravallion, and Susan Stout for helpful discussions and comments. Deon Filmer can be reached via e-mail at dfilmer@worldbank.org, Jeffrey Hammer at jhammer@worldbank.org, and Lant Pritchett at lant_pritchett@harvard.edu.


References


Bonilla-Chacin, Maria Eugenia, and Jeffrey Hammer. 1999. “Life and Death among the Poor.” World 

Castro-Leal, Florencia, Julia Dayton, Lionel Demery, and Kalpana Mehra. 1999. “Public Social Spend- 


Johns Hopkins University Press.

World Bank, Washington, D.C.

Deshpande, Vidya. 1999. “Polio Cases at All Time Low.” *Financial Express* (India), December 26, 
1999.

ington, D.C.


Filmer, Deon, Jeffrey Hammer, and Lant Pritchett. 1998. “Health Policy in Poor Countries: Weak Links 


Be Better.” Mimeo, DECRG, World Bank, Washington, D.C.

Gertler, Paul, and Jeffrey Hammer. 1997. “Strategies for Pricing Publicly Delivered Health Care Ser-

for Publicly Delivered Health Care Services: Utilization, Health Outcomes and Private Provider Re-
sponse.” RAND, Santa Monica, Calif. Processed.

Countries?” Mimeo, University of California at Berkeley and University of the Philippines. Processed.


Developing Countries: Contracts, Competition and Cost Control.” Policy Research Paper No. 2547, 
World Bank, Washington, D.C.

Expenditures in Malaysia.” In D. van de Walle and K. Nead, eds., *Public Spending and the Poor: Theory 

Khan, M. E., C. V. S. Prasad, and Neshat Quaiser. 1987. “Reasons for Under-Utilization of Health Ser-
vices—A Case Study of a PHC in a Tribal Area of Bihar.” *Demography India* 2:177–95.


Public Intervention in Health Insurance Markets: Theory and Four Examples from Latin America

William Jack

This article examines rationales for public intervention in health insurance markets from the perspective of public economics. It draws on the literature of organizational design to examine alternative public intervention strategies, including issues of contracting, purchaser provider splits, and regulation of competition. Health insurance reforms in four Latin American countries are then considered in light of the insights provided by the theoretical literature.

Health care expenses and lost labor earnings due to illness—not to mention the direct effects of feeling lousy and dying young—represent a major source of risk for individuals and families. Exposure to such risks is costly in itself (if individuals are risk averse), but can also have long-term effects, especially on the poor. Selling assets, withdrawing children from school to care for ill parents, and exiting the labor market can leave low-income families trapped in poverty. This article addresses the role of government in spreading and reducing health risks with particular emphasis on the design and organization of the relevant institutions in Latin America.

Faced with wide disparities in both health needs and access to medical care across regions and income groups, and with continuing pressures on public finances arising from the macroeconomic crises of the 1980s and 1990s, a number of countries in the region have adopted wide-ranging health sector reforms that continue today (Greene, Zevallos, and Suarez 1999). Generally, among the higher-income countries, there has been a move toward extending explicit insurance coverage to those outside the formal labor market. At the same time, these countries have examined the ways in which insurance and health care have been delivered and have instituted reforms that are meant to improve allocative and production efficiency in the sector. Lower-income countries in the region have not proceeded as far in terms of explicit health insurance reform, which requires a certain administrative capacity, and have tended to concentrate on running public hospitals and clinics better.
Large-scale changes in health insurance and health care markets inevitably involve significant public intervention. This article examines the arguments in favor of such intervention from a public economics perspective. Having identified market failure and redistributional rationales for public intervention, it addresses the important issue of how the government should intervene. This is effectively a question of organizational design, incorporating ideas from industrial organization, contract theory, and theory of the firm. The article undertakes a detailed examination of the reforms pursued in Colombia, Argentina, Brazil, and Chile. These countries followed strategies that reflect a variety of routes toward the goals of expanding formal insurance coverage and improving the efficiency of health service delivery.

Reasons for Public Intervention in the Health Insurance Sector

The theoretical literature on the performance of insurance markets is well developed. However, not all of the market failures that may arise in such markets necessarily justify public intervention. This section examines the efficiency and equity reasons for intervention in health insurance markets, paying specific attention to the informational constraints facing governments.

Market Failure in the Health Insurance Sector

Inefficiencies in health insurance markets derive primarily from information asymmetries and imperfect competition and less from standard public goods and externality characteristics.

Moral hazard and adverse selection. The role of information in the performance of insurance markets has been widely appreciated. In the health insurance literature, Feldstein (1973), Pauly (1968), and Zeckhauser (1970) show how asymmetric information at the ex post stage—that is, after an insured event has occurred—can lead to overconsumption of care and the costs of this ex post moral hazard are offset by reducing the level of insurance. A similar inefficiency results from ex interim moral hazard, when individuals fail to take precautionary actions after an insurance contract is signed. Unfortunately, there is little the government can do to correct these inefficiencies. Only by taxing or subsidizing goods in related markets (for example, cigarettes and immunizations, respectively) can it indirectly alter incentives in a welfare-improving fashion (Greenwald and Stiglitz 1986).

Although moral hazard derives from asymmetric information that is generated after individuals enter into insurance contracts, adverse selection occurs in markets where information is held asymmetrically at the date of contracting. A competitive insurance market in a population with heterogeneous ex ante risk character-
istics may perform inefficiently if insurance contracts cannot be differentiated on the basis of these risks. Either relatively low-risk individuals will be driven from the market (Akerlof 1970) or they will be constrained to purchase incomplete coverage (Rothschild and Stiglitz 1976). A number of studies have provided evidence of the existence of adverse selection in insurance markets (Cutler and Zeckhauser 1997; Cutler and Reber 1998). However, government intervention cannot easily correct these market failures. Universal and uniform coverage can be mandated, but the resulting resource and risk allocations are not Pareto-comparable with the initial equilibrium.

**Imperfect competition.** The models of adverse selection reviewed above identify failures of competitive insurance markets. But even in the absence of adverse selection problems, insurance markets may yield socially suboptimal resource and risk allocations if firms have market power. Such market power may derive from information imperfections on the demand side, contributing to switching costs (which make it difficult for new firms to attract customers). Increasing returns in administrative costs suggest that a somewhat concentrated industry is likely to be observed in equilibrium.

In standard industrial organization models, although market power typically leads to allocative inefficiency, competition is generally welfare improving. However, in insurance markets with information asymmetries, competition may sometimes have negative effects on allocative efficiency. For example, insurers may reduce the quality of coverage to attract low-risk individuals (Jack 2001). This kind of active (as opposed to adverse) selection might suggest public intervention to control the extent (or at least type) of competition in the insurance market.

**Consumer protection through regulation.** Consumers are subject to potential exploitation by health insurers in two respects. First, insurers may provide or finance low-quality care; and second, they may behave imprudently in their roles as financial managers. Both of these issues suggest a role for direct regulation, but many countries may lack the capacity for such regulation. The first problem might require devoting significant medical expertise to checking up on the actions of providers of care. The second would call for employing individuals with significant financial sector experience to check the insurers' books. When countries do not have enough trained individuals to perform the underlying tasks of medical care delivery and financial risk management, diverting resources to regulation entails a high opportunity cost.

Moral hazard (that is, overconsumption of health care) can be mitigated by basing insurance coverage not on incurred medical expenses but on the basis of a physician's judgment of a patient's need. Such a contract is efficient as long as information about needs is held symmetrically by all parties. In practice, the physician is the primary source of this information, so that when acting as the patient's agent, the physician
confers an information advantage on the patient vis-à-vis the insurer. However, when acting for the insurer, the physician may put the patient at a disadvantage and warranted treatments could be withheld. Public intervention—including monitoring of physicians’ decisions and actions—might then be required to maintain quality.

In some respects, insurance companies perform similar functions to banks. Banks facilitate intertemporal trades (saving and dissaving) implemented through contemporary interpersonal trades (lending and borrowing). Insurance companies facilitate trades between uncertain states of nature implemented through interpersonal pooling of current risks. Similarly, just like banks, insurance companies hold financial assets that must be invested by managers.

Dewatripont and Tirole (1994) use an incomplete contracts model to show that bank managers can be given appropriate incentives to perform by transferring control from equity holders (who have relatively weak incentives to interfere with management) to debt holders (whose incentives to interfere are stronger) when bank performance as measured by the value of assets is poor. The role of government is then to act as a representative of small, uncoordinated debtors, and the theory rationalizes public takeovers of distressed banks as a means of providing incentives to managers.

Similarly, if the value of an insurance company’s assets falls enough, the government may wish to intervene on behalf of small policyholders and take over the administrative functions (perhaps contracting out such administration to another healthy insurer).

**Equity**

At a conceptual level, being at high risk of needing medical attention reduces an individual’s available (expected) consumption opportunities. Consequently, the government may wish to redistribute resources between individuals with identical money incomes but different health risks. One way of implementing such redistribution is through uniform pricing (known as community rating) of insurance policies across individuals. However, requiring private firms to community rate may only exacerbate the selection problems that already exist.

It is important to note that, even if selection issues were unimportant, it would still be only second best to require uniform insurance pricing. The first-best policy, of course, is to redistribute income (lump sum) from low risks to high risks and require each to buy insurance at the actuarially fair price (that is, to allow price discrimination by firms). Such redistribution is notoriously difficult, even more so when income inequality itself is high, as it is in many Latin American countries.

Henriet and Rochet (1999) analyze the optimality of a uniform public insurance system when individuals differ in both their health risks and incomes. They find that in the absence of moral hazard, a comprehensive policy providing full insurance to all individuals is part of an optimal tax and insurance system. This result relies to some
extent on the assumption that individuals with different incomes nonetheless face the same potential losses, albeit with different probabilities. In fact, the poor may prefer additional income transfers and less extensive public insurance to being offered the same level of coverage as the nonpoor.

In the presence of health risk and income differentials across the population, and in the absence of first-best redistributive taxation, governments will likely wish to couple a progressive general tax (for example, an income tax) with a system of health insurance (privately or publicly supplied) that delivers subsidized insurance to the poor but allows coverage to increase with income. One means of effecting such a graduated insurance profile is to have a mixed public/private system of insurance in which the government provides (or mandates) a given base level of insurance and individuals are permitted to top off their coverage through private purchases (Besley and Coate 1991) or to opt out of the public system and purchase private insurance (Gouveia 1997).

The Nature of Government Intervention

The arguments above suggest that, due to market inefficiencies and redistributive concerns, governments may wish to control individuals’ choices about insurance in certain ways. But the discussion does not explain how such control over choices should be effected. This section examines the effects of alternative public interventions—including explicit contractual arrangements between insurer and provider, organizational choices (for example, vertical integration) and competition—on the costs and quality of insurance.

Contractual Arrangements: Motivating Physicians

Physicians, like other workers, need to be motivated to make decisions that appropriately trade off patient benefits and costs. These decisions require effort; when physician effort is not directly purchasable (that is, contractible), incentives may be difficult to generate without exposing the physician to undesirable risk. Two extreme cases are the salaried physician and the decentralized fund holder. Under the first arrangement, an insurer (possibly the government) pays the physician a fixed amount, independent of the physician’s supply of effort, and reimburses nonphysician expenses (for example, laboratory tests). At the margin, doctors will tend to substitute out of personal effort and into complementary inputs and will face little risk. A decentralized fund holder, who is given a fixed budget to finance all incurred costs (including the cost of the physician’s own effort), will have strong incentives to choose the right input mix but might be exposed to considerable risk. The tradeoff, as in any moral hazard problem, is between incentives and risk.

William Jack
This tradeoff at the provider level is not specific to the public sector. Private insurance companies also must induce physicians to implement insurance contracts at minimum cost, and so face a similar contracting problem. However, some endogenous differences across the public and private sectors may emerge that imply different contractual relationships between payers (the government or insurance companies) and providers of medical care. For example, if public insurance is aimed at the poor, then in the absence of accurate eligibility tests, self-selection constraints may require that the public system provide a relatively low quality of service. Inducing low effort from physicians who provide services to the public system might be easier than inducing higher effort. Higher service quality is a characteristic of private insurance companies that, in equilibrium, serve the nonpoor. Thus, it is likely that incentive schemes are relatively flat for physicians serving the public insurance system and that compensation schedules employed by private insurers are steeper. Even inducing low effort might be difficult when services are provided to the poor because providers might have to live in rural areas or poor urban ones. Hammer and Jack (forthcoming) describe some models addressing incentive issues in these cases.

Another reason that public and private insurance systems may provide different incentive schemes to physicians is that physicians may represent a heterogeneous group. If they differ in their aversion to risk, ethical priorities, or job satisfaction, it may be optimal to offer one kind of compensation contract to one group of physicians and another kind to a second group. However, these arguments suggest reasons for different ways of paying physicians in the delivery of health insurance, but they are not necessarily the outcome of a public/private mix of insurance provision.

**Purchaser-Provider Split versus Vertical Integration**

Instead of writing a detailed contractual agreement between insurer and physician, the two parties instead might decide to integrate into a single organization and rely on bargaining protocols to determine the allocation of rents. Traditionally, in many countries in Latin America and elsewhere, public insurance systems have been highly vertically integrated. However, recent reforms have focused on separating the functions of insurance and provision, through the so-called purchaser-provider split wherein explicit contractual arrangements govern relationships between insurers and providers.

By contrast, traditional private insurance was of the fee-for-service type (that is, indemnity plans), whereby a physician would send a bill to the insurer for covered services. This is one kind of explicit contract. Over time, however, private insurance companies have moved toward a more integrated organizational structure, bringing physicians in house or at least adopting long-term contractual relationships with them. This apparent anomaly between the evolution of the organization of public and private systems can be understood in a number of ways, including soft budget constraints and common agency.
Soft budget constraints in the public sector. Both institutional developments (purchaser-provider split and organizational integration) might represent attempts to provide physicians with stronger incentives, within constraints that differ between the public and private sectors. For example, it might be difficult for a government bureaucracy to commit to funding a public sector employee prospectively. Future renegotiation in the event of high costs (when it is difficult to withhold extra funding) or low costs (when it is difficult to resist the temptation to expropriate profits) induces a soft budget constraint that limits incentives. A purchaser-provider split might insulate a public payer from the incentive to renegotiate a contract and would harden an otherwise endogenously soft budget constraint. Conversely, the easiest way for a private payer to provide a physician with incentives might be to make the physician a shareholder in the insurance firm. (Of course, there are obvious free-rider problems when physicians are paid on the basis of group profits and not individual contributions thereto.)

Common agency in public institutions. Another way to understand the opposing directions of reforms in the public and private sectors is to look more closely at the nature of the purchaser-provider split arrangements. In particular, instead of contracting directly with individual physicians, a public insurance system might sign contracts with groups of physicians, indeed, often with managed care organizations. This suggests that the function that is being contracted out from the public system is the management of physician services. Having a formal arm’s-length contract between the public sector and the manager of physician services might be an effective way of improving the incentives of such a manager. For example, Dixit (1997) has shown that when a manager reports to multiple noncooperative principals with heterogeneous objectives, such as elected officials in public office, equilibrium incentives are low powered. Requiring an explicit contract could facilitate the cooperation of the principals, leading to higher-powered incentives being given to the manager.

Explicit arm’s length contractual arrangements might also be a way of limiting the scope of a manager’s activities. Tirole (1994) suggests the usefulness of limiting the objectives of public sector decisionmakers. Dewatripont, Jewitt, and Tirole (1999) formally elaborate a model of career concerns. Providing a manager with a well-defined mission makes it easier to induce effort. In Dixit’s (1997) analysis, the narrowing of a manager’s objectives is effected by reducing the number of competing principals to whom the manager answers.

Contractual incompleteness. An alternative literature examines issues of contracting out versus internal provision—that is, vertical integration—starting from the presumption that contracts are necessarily incomplete. Even if choices are observable by both parties to a contract, if they are not verifiable by a third party and if the contract is consequently unenforceable, then institutional arrangements can have sub-
stantive effects on incentives. In particular, ownership of productive assets can matter when explicit contracts are unavailable. Private contractors are those who own the assets they use to produce services (for example, hospitals); public servants do not have the implied control rights over asset use.

Hart, Shleifer, and Vishny (1997) present a model of service provision when quality and cost are noncontractible. If cost is noncontractible, then procurement contracts like those studied by Laffont and Tirole (1993) are not feasible and a fixed price contract must be used. By definition, public sector employees cannot retain ownership of any quality innovations they generate, but private sector providers can. Incentives for quality innovations are thus greater in the private sector. Incentives for cost reduction are also greater in the private sector, but cost control is associated with lower quality. Thus, the private sector will always (in the model) produce at lower cost but could produce higher or lower quality. When one of the ways of reducing costs is to actively select easy-to-treat patients ahead of expensive cases, the social aspects of quality can be severely affected by incentives to control costs.

Some might argue that innovations in medical care are very important and those in insurance administration are less so. This would argue in favor of private provision of physician services under the condition that active selection could be controlled adequately. However, in countries with large sections of the population uninsured against health needs that are susceptible to standard treatments, innovation in insurance delivery may have high social payoffs, in which case (regulated) private provision may then be favored.

**Competition**

By allowing consumers to exert discipline on providers, competition can strengthen incentives for quality and cost-reducing effort. Competition among suppliers should not necessarily be identified with private supply. Indeed, the government of the United Kingdom has attempted to induce competition among public providers by developing the so-called quasi-market (Le Grand 1991). Even when consumers do not face financial incentives to choose wisely between suppliers, consumers might still induce effort and quality provision if their choices affect the payoffs to providers. Thus, Halonen and Propper (1999) model the impact of competition on quality when a public sector payer pays providers on behalf of consumers who are free to choose their supplier. The essential feature of their model is that when providers’ objectives are not coincident with consumers’ (on average), allowing consumer choice can help realign providers’ incentives. The benefits of competition are of course limited by the elasticity of demand.

Competition between public and private providers might also be beneficial. A common argument in favor of large purchasing groups is the monopsony power they can wield in negotiating supply contracts. However, as Propper and Green (1999) point out, there is no particular merit in such actions—market power is inefficient, whether
it is on the supply side or the demand side. They suggest that under such arrange-
ments, staff will either be of poor quality (good staff will be driven from the market by
low wages) or employment contracts will permit outside earnings with little or no
monitoring, weakening incentives for performance of primary job tasks. Introduc-
ing competition from the private sector may thus lead to higher public sector wages
and costs, but with a net welfare gain (Danzon 1992).

Introducing public provision into a private market may be beneficial if it serves the
purpose of making a minimum quality standard credible. For example, Ronnen
(1991) uses a model of vertical product differentiation to show that a minimum qual-
ity standard can raise the quality provided by all market participants and lower equi-
librium hedonic prices (that is, prices adjusted for quality). By restricting product
differentiation and intensifying ex post price competition, an appropriately chosen
standard makes all consumers better off. In principle, such a standard could be im-
posed by the government without recourse to public provision. However, if quality
is difficult to monitor and if lapses are costly to penalize, public provision of the stan-
dard quality at a minimal price could act as a substitute for direct monitoring. In
general, the idea that public and private sector quality can act as strategic comple-
ments should be viewed with some caution. Standard equilibrium analysis suggests
the need to anticipate potential crowding out of private sector supply by public sec-
tor provision (Hammer 1997). In the case of nondifferentiated goods, this is indeed
appropriate and it would be expected that if the publicly provided quality level was
too high, then private supply would dry up.

Institutional Evolution in Latin America

Over the past 20 years, several Latin American countries have embarked on wide-
ranging reforms of their health insurance and delivery systems. Countries have
moved away from integrated public provision of insurance and health care and to-
ward more decentralized provision, sometimes incorporating private sector involve-
ment, in pursuit of expanded coverage and more efficient delivery. Major health
insurance reform, like health care itself, appears to be a luxury good; the poorer
countries in the region have focused on more basic challenges in terms of primary
care delivery. This section reviews the experiences of four countries that have adopted
significant reforms—Colombia, Argentina, Brazil, and Chile. For a more complete
survey of health systems in Latin America, see Bengoa and others (1998).

Colombia

The health sector reforms Colombia initiated in the early 1990s represent possibly
the most ambitious policy interventions undertaken since the 1950s and 1960s,
when countries in the region implemented large-scale social security mechanisms. Before the reforms, Colombia had a centralized, budget-financed, poorly organized public health delivery system that consisted of two uncoordinated bodies: the social security institutions that provided subsidized health services to formal sector workers, and the Ministry of Health, which provided public health inputs and subsidized hospital care for those who did not use the social security system.

The general goal of the Colombian reforms was to ensure a basic level of coverage for all individuals that could be supplemented by those willing and able to pay more. This conforms with the role of subsidized health care as a redistributive instrument. At the same time, the reforms attempt to reduce supply inefficiencies by encouraging alternative provider payment systems and allowing consumer choice. Thus, to implement the equity objective of universal coverage, the country adopted such techniques as competition and contracting.

There has indeed been a marked increase in formal coverage of the population, particularly among lower-income groups. Figure 1 shows coverage rates by income quintile in 1993 and 1997 (Uribe, Londono, and Jaramillo 1999). Overall, the proportion of individuals with insurance doubled in this period, from 23.7 to 57.2 percent, with the largest proportionate gains among the poor.

Insurance coverage is allocated to consumers through two regimes that together approximate a simple two-level voucher system. Formal sector workers and their families receive an implicit voucher for insurance that covers a wide range of services.

---

Figure 1. Health Insurance in Colombia: Proportion of Population with Coverage

![Figure 1](image-url)
Others (essentially the poor) receive an implicit voucher for a less generous package of insurance. The first regime is referred to as the contributory regimen and the second as the subsidized regimen. On the financing side, participants in the contributory regimen are required to pay a 12 percent payroll tax to help finance health care. Participants in the subsidized regimen make a means-tested contribution (that may be zero in many cases) to support health insurance costs.

Participants in the contributory regimen can use their vouchers to buy insurance from Empresas Promotores de Salud (EPS), which are essentially private sector insurance companies. The EPS can cash in the voucher with the government (via the Fondo de Solidaridad y Garantía, FOSYGA) for a fixed amount that is adjusted for some of the risk attributes of the consumer. Participants in the subsidized regimen can use their implicit vouchers to purchase (less generous) insurance either from EPS or Empresas Solidarias de Salud (ESS). ESS are medical care purchasing organizations that subnational governments must set up to facilitate the coverage of the self-employed and nonsalaried workers who may not be able to access EPS.

The implicit vouchers have fixed monetary values, so there is little effective price competition among EPS and ESS. Instead, the EPS compete on the basis of the level of insurance they provide, as determined by the copayment rates and the quality and range of services offered. Thus, the standard packages of services defined for participants in each regimen act as basic plans that can be supplemented by insurers to attract clients. This kind of arrangement encourages efficient provision of insurance because the insurer retains any efficiency gains.

However, if demand is not elastic in response to quality changes, incentives to control costs might outweigh incentives to improve or maintain quality. Such incentives could be particularly strong in the subsidized regimen whose participants have less access to alternative providers. Similarly, in an attempt to attract inexpensive clients, insurance providers might bundle a low-quality standard package with high-quality additional services. For example, La Forgia (1998:257) reports concern over “the practice of EPS’s to integrate the [standard package] with complementary plans, thereby undermining competition for a homogeneous service plan.”

On the supply side, EPS and ESS contract with hospital and physician groups, including private sector Instituciones Prestadoras de Servicios (IPS), and formerly public sector but now autonomous Empresas Sociales del Estado (ESE). The financial aspects of such contracts are less precisely regulated than the demand-side transactions between consumers and EPS/ESS, although the law attempts to encourage innovative payment methods to encourage provider effort and efficiency (for example, capitation and diagnosis-related group (DRG)–based payments).

One aspect of the reforms that has received considerable attention is the fact that medical care providers are paid on the basis of demand. That is, insurers contract with providers to supply services for the covered clients, based on expected use (in the case of capitated payments) or realized use (in the case of fee-for-service and DRG-based con-
tracts). In the past, when the government paid for services, payments were nearly always purely prospective, deriving from budgetary allocations to hospitals and provincial health ministries. Demand-side financing is meant to impose discipline on providers by making them suffer financial losses as patients switch in response to low quality.

The formal separation of purchaser and provider is mixed in the Colombian system. On the one hand, under the proposed reforms (when fully implemented) the government will have little direct role in the provision of either health insurance or health care. The FOSYGA acts as a clearinghouse for taxes paid by individuals and transfers made to EPS and ESS. Apart from monitoring the quality of the insurance and medical services provided, the public sector will not actively perform any insurance purchasing role, this function being delegated to consumers. On the other hand, the degree of integration of the delivery of insurance and medical care varies widely. Some purchasers (EPS and ESS) contract at arm’s length with provider networks (IPS and ESE), whereas others effectively own such networks along the lines of health maintenance organizations. Even if an EPS owns or has close contacts with a particular provider network, it is required by law to offer the services of at least one other IPS to consumers to increase ex post quality competition among providers.

Argentina

Argentina’s formal health insurance system was well developed in the late 1980s. All employees were obliged to be insured by the so-called obra social that covered the sector in which they worked, effectively prohibiting formal sector workers from choosing their health insurer. In essence, the obras sociales were and remain nonprofit insurance companies owned by the relevant labor union. There are about 360 such obras, covering about 10 million individuals and their families. In addition, each of the 24 provinces of the country operates an obra provincial, covering about 5 million public sector employees and their dependents. Obras are funded on the basis of compulsory payroll taxes.

Retired workers and pensioners—about 4 million individuals—received health insurance coverage through the Integrated Program of Medical Care (Programa de Asistencia Médica Integral; PAMI), operated by the National Social Service Institute for Retirees and Pensioners (Instituto Nacional de Servicios Sociales para Jubilados y Pensionados). These services were funded by payroll taxes and taxes on pension benefits. The obras and PAMI combined covered about 61 percent of the population. Another 2 million people were covered by private, for-profit, prepaid insurance plans (pre-pagas) and another 1 million people received insurance through small insurers known as mutuales, of which there were around 1,000. Figure 2 shows the distribution of coverage across types of insurance. The 25 percent of the population who
lacked formal insurance would typically seek care through public hospitals operated by provincial and municipal governments.

The main focus of recent reforms has been on the insurance market and less on the organization of and payment for medical care. Within the insurance market, the focus has been on improving the efficiency of coverage, as opposed to extending formal coverage to the uninsured. However, over the long term, insurance reform aims to attain universal coverage through competitive provision of at least a minimum level of insurance. This focus on production efficiency derives from the belief that lack of consumer choice of obra, coupled with a weak regulatory framework, has led to poor quality coverage, financial instability, and political cronyism.

The existing fragmented structure—consisting of a public system for pensioners and retired individuals, a noncompetitive but decentralized system of obligatory insurance for formal sector workers, an unregulated competitive system for others able to pay, and a public hospital system providing insurance through the provision of low-quality services—is to be unified across consumers. That is, the characteristics of insurance providers and the environment in which they operate are to be independent of the identities of the individuals they cover. This represents a rational separation of the organization of production (determined by internal efficiency considerations) and the allocation of consumption (determined by allocative efficiency and equity concerns). The major incentive instruments are to be consumer choice and, where that is ineffective or imperfect, regulation of quality and financial soundness.
Competition among obras has been introduced; by mid-1999, most people were aware that they had a choice of insurer. Consolidation in the industry is clearly required and has been partially achieved through mergers of obras. (The number of registered obras fell from 360 in 1997 to 294 in 1999.) Insurers are required to offer a standard health benefits package (Programa Medico Obligatorio, PMO). In addition to representing a minimum level of insurance that may facilitate redistribution in general, the PMO has enabled consumers to make relatively easy comparisons of obras. However, formal coverage for the poor is not automatic, and ensuring the quality of services offered to those with low incomes may be difficult under the PMO. This is especially true because the effective price paid for coverage is 90 percent of the 8 percent payroll tax (the other 10 percent goes into a reinsurance/redistribution fund), so vertical segmentation of the market seems likely to occur.

So far, competition between obras and pre-pagas is limited. Individuals who are eligible for coverage by an obra are not permitted to obtain formal insurance from a pre-paga. However, some schemes have developed whereby an obra will subcontract with a pre-paga, thus allowing an individual to remain formally in the obra sector while effectively receiving insurance from outside.

In summary, Argentina's health insurance reforms have focused primarily on improving the efficiency of insurance delivery, mainly by increasing the role of competition among insurance providers. There has been relatively little in the way of either provider payment reform or the extension of coverage to marginalized groups.

Brazil

Health insurance in Brazil is something of a mix between a nominally comprehensive public system and a large and active private system. Unlike Argentina, Chile, and Colombia, which have tried various ways to implement privately provided mandatory insurance coverage, Brazil has opted to provide universal public insurance while encouraging use of the private sector as an alternative. Consumers have the freedom to choose among private plans and whether to use the public system. The big difference is that in opting out of the public system, individuals do not take their financing with them, so consumer choice provides little incentive for the public system to maintain quality. In reality, the public system acts as a floor, available to all but used primarily by individuals with low income.

Government intervention in the health insurance market through direct provision (of insurance) is motivated on distributional grounds, and the reforms have addressed the efficiency with which such public insurance is provided. In particular, the health sector reforms have concentrated on the organization of public service delivery, particularly through decentralization, and the public system's payment of providers. Until recently, relatively less attention had been paid to the regulation of private insurance providers.
Until 1993, public insurance was implemented through Instituto Nacional de Assisência Médica e Previdência Social (INAMPS), the medical arm of social security. INAMPS originally covered only formal sector workers, but under the new constitution of 1988 and subsequent legislation, coverage was nominally extended to the whole population. This insurance has been provided (at times inadequately) through a mix of subsidized public hospital care and reimbursement of privately provided care.

A comprehensive reform of the health system was instigated in the early 1980s. The World Bank (1994) provides a full review of the history of the reforms. These reforms were organizational in nature, relating to the coordination of INAMPS with the Ministry of Health, the decentralization of INAMPS functions to states and their partial recentralization, and finally the abolition and integration of INAMPS into the Ministry of Health under the umbrella of the Secretaria de Ações de Saúde (SAS). The main function of the SAS is to transfer funds to state health secretariats.

INAMPS contracts with private sector providers were first on a fee-for-service basis but later used a U.S. Medicare-type prospective payment system. There are currently two prospective payment systems in operation, for outpatient and inpatient services, respectively, although the cost control attributes of prospective payment have been ineffective largely because of a lack of monitoring and evaluation by the public payer (World Bank 1994).

The payment of medical providers and hospitals on the basis of utilization represents a limited version of contracting out. This reimbursement mechanism does not represent the contracting out of the management of individuals' health care needs in any meaningful sense. Nor does it represent the contracting out of insurance. Thus, although the private sector is heavily involved in the delivery of services financed by public insurance, providers have relatively weak incentives to focus on health outcomes.

Private insurance covered about 25 percent of the population in the mid-1990s. A number of factors have brought about the expansion in private coverage over the past 30 years, including growing incomes (in the 1970s), a tax deduction for out-of-pocket expenditures and premiums, and the deteriorating quality of the public system.

Private insurance is provided through four alternative types of organizations. The largest and historically most important is the prepaid group practice, which is similar to the health maintenance organization model in the United States, and had about 47 percent of the private insurance market in 1991. Medical cooperatives, which contract with preferred providers, captured 28 percent of the market in 1991. Large employers (20 percent of the market) at times self-insure and offer company health plans, sometimes contracting out the administrative functions to financial intermediaries. Finally, only a very small proportion of those covered by private insurance (4 percent) enrolls in indemnity plans (that is, reimbursement insurance).

Regulation of the private insurance market was virtually nonexistent until 1998. Exclusions and restrictions are common, financial soundness is unchecked, and fraudulent practices are perceived to be prevalent. In fact, the poor reputation of prepaid group insurance...
practices has been the main factor contributing to the relative growth of medical cooperatives and company health plans over the past 10 years, although prepaid group practices still capture the largest portion of the market. Lately, however, a number of consumer protection and financial regulation initiatives have arisen.

Chile

Chile's health system is one of the most closely and thoroughly studied in Latin America (Bitran 1998; Sapelli 1999). The country undertook wide-ranging and innovative reforms of its health and social security systems starting in the early 1980s, partly in response to failures of the bureaucratic centralized regimes that had prevailed previously. Similar to the arrangements in Argentina, the reformed Chilean system of compulsory health insurance pairs private provision of insurance for some (in the hope of improving the efficiency of risk sharing) with public insurance for others (to satisfy an equity objective). Funding is primarily through a proportional payroll and pensions tax, although there is no equivalent of Argentina's redistribution fund. Colombia’s reforms have been more ambitious, more fully implementing a redistributive allocation mechanism while privatizing or at least corporatizing insurance provision to a greater degree. The pattern of insurance coverage by income and risk category that has resulted has been criticized by some commentators. However, it may be possible to argue that this pattern, though not first best, could be the best that can be achieved under certain assumptions about the redistributive capacity of the tax and transfer system.

Under the reformed mixed system with public and private insurance options, formal sector workers and pensioners are required to contribute 7 percent of their incomes (up to a cap) to finance health insurance. Each individual has the choice of allocating contributions to 1 of about 35 private insurance companies, known as Instituciones de Salud Previsional (ISAPRES), or to the Fundo Nacional de Salud (FONASA, National Health Fund). The designated recipient of the funds then provides insurance coverage for the individual and his or her dependents. Individuals who are unemployed or who work in the informal sector are automatically covered by the public insurance system.

In 1995, ISAPRES covered about 31 percent of contributors and accounted for about half of insured medical care spending. ISAPRES can offer multiple policies and are free to charge corresponding premiums. These premiums can vary on the basis of age, gender, and the number of insured and on the quality and extent of insurance. (ISAPRES cannot discontinue insurance and can impose at most an 18-month waiting period on insureds for preexisting conditions.) Individuals are permitted to increase their contributions above 7 percent of income to purchase a higher-cost policy. In 1995, ISAPRES offered close to 9,000 policies, reflecting a near continuum of vertically differentiated insurance products matching the distribution of wages.
One important characteristic of contracts offered by ISAPRES is that they often include stop-loss components, limiting the financial risk of the insurer to a certain amount (above which the individual receives no reimbursement). This feature limits the extent of insurance actually offered and means that the high-risk clients—especially the elderly—choose not to participate. Table 1 reports the shares of each age group enrolled in FONASA and the ISAPRES.

Of course, those with low incomes also tend to choose not to participate in the ISAPRE system because the premiums are unaffordable and because the public system provides free insurance. Figure 3 confirms this pattern of demand, although it has changed over time. FONASA classifies beneficiaries into four income groups, labeled A (the poorest) through D (the richest). (It also has a fifth classification, E, for nonbeneficiaries, that is, those enrolled with an ISAPRE or not formally registered with FONASA.) In 1994, half of FONASA's beneficiaries came from the lowest income group, 25 percent from the next group, and around 12–13 percent from each of the two richer groups, C and D. In 1996, this picture started to change, with individuals from the poorest group representing just one-third of FONASA enrollees, whereas the share of group D enrollees doubled to about 25 percent. The shares of groups B and C did not conform with this convergent tendency.

This general pattern of coverage—in which those with high risks and low incomes use the public system and others use the ISAPRES—is fragmentary, although this criticism does implicitly assume that uniform coverage is optimal. However, in a second-best world in which the government wishes to redistribute from the rich to the poor and from those with low risk to those with high risk, such a pattern may not be unreasonable. What is perhaps questionable is the extent to which higher-income individuals are effectively required to purchase superfluous insurance or at least insurance of low marginal value. This feature tends to exacerbate incentives to select low-risk/high-income consumers.

A comparison with Colombia's more focused system is useful in this respect. Under current practice, Colombian users of the contributory regimen (corresponding to Chilean ISAPRES) receive a more or less standard insurance package independent of income, implying an in-kind redistribution from high-income contributors to lower-

| Table 1. Chilean Health Insurance: Distribution of Coverage by Age |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Age group (years) | 0-1 | 2-14 | 15-24 | 25-54 | 55-64 | 65+ | Total |
| FONASA | 65.8 | 64.7 | 58.3 | 57.0 | 68.3 | 79.9 | 61.8 |
| ISAPRES | 24.9 | 23.9 | 22.4 | 26.7 | 16.5 | 6.9 | 23.1 |
| Other | 9.3 | 11.4 | 19.3 | 16.3 | 15.2 | 13.2 | 15.1 |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

income contributors, which results in two tiers of more or less uniform health insurance. In the ISAPRE system, however, above the public level of insurance there is a continuum of insurance qualities that increases with income. The Colombian arrangement therefore allows for some redistribution within the group of relatively higher-income individuals. Perhaps more important, because the average value of the implicit voucher used in the ISAPRE system is equal to the average contribution across all ISAPRE users, there is limited redistribution (through the health system at least) between the two groups in Chile. In contrast, because the cost of the standard insurance package under the Colombian contributory regimen need not (in fact, does not) equal the average contribution by users, there is scope for further intergroup redistribution.

Of course, to make definitive statements about the extent of redistribution among groups requires knowing the full structure not only of taxes but also of other government expenditures. Focus on the earmarked taxes and the provision of insurance is useful for predicting the effects on distribution of marginal parameter changes. For example, an increase in the payroll tax rate in Colombia would increase the resources available for both the low-quality and high-quality insurance packages. Doing the same in Chile would necessarily increase the quality of the high-quality packages, with probably little effect on that of the publicly provided services. However, Bitran (1998) finds that within the FONASA-financed public insurance, the incidence of net benefits is reasonably progressive, suggesting a degree of within-group redistribution through the public system.

A final characteristic of insurance contracts generated under the ISAPRE system is that because they are tied closely to current wages, quality tends to follow the life-
cycle pattern of wages, which may be somewhat different from the time profile of an individual's demand for insurance. It is not just the case that some individuals with high lifetime earnings will be induced to purchase more insurance over their life spans than they would desire, but that the pattern of coverage may not match their pattern of needs. Even if capital markets work well, this problem will persist unless insurance companies can write long-term contracts in which high contributions at ages of high earnings but relatively low risk are made in exchange for lower contributions later in life when earnings are lower and risks higher.

In terms of the organization of medical care, about 70 percent of Chile's population is covered by the public insurance system, which finances care delivered mainly through public hospitals and a mix of public and private ambulatory care. Medical services financed through the public budget are funded primarily from FONASA and the general budget (for the indigent), with additional resources deriving from co-payments by public patients, payments from ISAPRES that use public facilities, and other sources. FONASA acts primarily as a decentralized financing agency, collecting contributions and distributing funds to providers through a network of 26 health services. Following a contraction in public health spending in the 1980s (reflecting in part the growth of ISAPRES during that period), real public health expenditures more than doubled between 1990 and 1996. Despite this increase, the perceived quality and adequacy of public services remained stagnant, leading to a recent debate about reform of the internal organization of the public system. Some have argued in favor of moving toward internal markets and introducing competitive pressures into the delivery system, and others—particularly some labor unions and parts of the medical profession—have resisted what they see as a move toward privatization.

Although a decentralized organizational infrastructure exists in the form of the 26 health services, central bureaucratic control of some crucial decisions remains in the hands of the central authorities, including the Ministry of Finance. In particular, labor inputs and compensation are dictated from the center and funded directly from the budget. Some central control of staff allocation is likely to be necessary in the health system, especially one in which providers are not necessarily profit maximizers and in which service provision has a redistributive role. However, centrally made allocations tend to be unresponsive to changing cost structures and needs and provide little incentive for innovation.

One part of the financial apparatus that does provide incentives for cost consciousness is the use of a prospective payment system for financing the use of drugs and material supplies. Less common services are funded on a reimbursement basis from a global budget under a separate mechanism. A movement toward full prospective funding could also generate a degree of competition similar to that envisioned in the quasi-market reforms of, for example, the United Kingdom. Whether this competition is socially beneficial depends, as usual, on the responsiveness of consumer demand to quality, as determined by both geographical constraints and limitations on
information. When it is not socially beneficial, which is especially likely in poor and/or rural areas, direct monitoring and regulation of quality are necessary.

Conclusions

Despite the well-known failures of insurance and health care markets associated with imperfect information, the primary motivation for large-scale public intervention in the sector has been equity. Most often this derives implicitly from a concern for good-specific equity, but it can also be justified as part of a more general second-best redistributive mechanism. Within the context of public intervention in the pursuit of equity goals, it is reasonable to assert that the reforms in Latin America of the past two decades have sought not so much to improve the efficiency of private markets but to improve the efficiency of public provision, either through direct use or mimicking of such markets. This has sometimes been achieved by altering the focus and function of preexisting institutions—for example, the obras sociales in Argentina—or by encouraging the growth of new institutions, such as the ISAPRES in Chile.

Coupled with the reforms of the ways insurance and care are organized and delivered, countries have attempted to various degrees to extend formal coverage to previously marginalized groups and to finance this extension fairly. Colombia can be seen as instituting an implicit two-tier voucher scheme, financed through a proportional wage tax. Chile’s system has a similar financing mechanism, but the distribution of benefits is less progressive, so that the net effect is in principle less redistributive. Argentina’s remodeled obras system is something of a halfway house: the financing base is similar, but the distribution of benefits in terms of the quality of insurance increases with income, although there is some implicit redistribution from richer to poorer obras. On the face of it, Brazil’s health insurance system serves less of a redistributive function than those of the other countries, to the extent that there is no earmarked tax dedicated to financing health insurance. However, this highlights the limitations of examining the health sector independently of the general tax and transfer system. The taxes paid by higher-income persons in Brazil are not reduced when individuals opt for private insurance. An important issue, which I do not address in this article, would be to analyze the extent of redistribution generated by the general tax base.

Note

William Jack is an Assistant Professor at the Department of Economics at Georgetown University. His e-mail address is wgj@georgetown.edu. This article is a shortened version of a background paper prepared for a World Bank regional analysis of social risk management in Latin America and the
Caribbean. The author thanks the following individuals for discussions about the article: Girindre Beeharry, Chico Ferreira, Charlie Griffin, Jeff Hammer, Jeni Klugman, Jerry LaForgia, Maureen Lewis, Andre Medici, and Miryan Sato.

References

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


Urbanization in Developing Countries

Vernon Henderson

The rapid urbanization in many developing countries over the past half century seems to have been accompanied by excessively high levels of concentration of the urban population in very large cities. Some degree of urban concentration may be desirable initially to reduce inter- and intraregional infrastructure expenditures. But in a mature system of cities, economic activity is more spread out. Standardized manufacturing production tends to be deconcentrated into smaller and medium-size metropolitan areas, whereas production in large metropolitan areas focuses on services, research and development, and nonstandardized manufacturing. The costs of excessive concentration (traffic accidents, health costs from exposure to high levels of air and water pollution, and time lost to long commutes) stem from the large size of megacities and underdeveloped institutions and human resources for urban planning and management. Alleviating excessively high urban concentration requires investments in interregional transport and telecommunications to facilitate deconcentration of industry. It also requires fiscal deconcentration, so that interior cities can raise the fiscal resources and provide the services needed to compete with primate cities for industry and population.

A high degree of concentration—the share of large metropolitan areas in total urban population—characterizes the rapid urbanization in many developing countries today. That concentration is also reflected in urban primacy—the share of the largest city in national urban population. And it involves megacities of unparalleled size, presenting major problems in health and quality of life, international industrial competitiveness, management and institution building, social cohesion, and stability. What has fostered such high degrees of urban concentration in developing countries, a pattern not seen in many developed areas? Has concentration become overconcentration? If so, what are the key institutional and policy issues for alleviating overconcentration and promoting more orderly development and urbanization?

Developing countries today face greater urbanization challenges than developed countries faced. Developed countries urbanized at a comparatively leisurely pace. The United States was 40 percent urbanized in 1990, 70 percent in 1960, and 75+ percent in 1990. This gradual pace is in marked contrast with that in many developing countries. For example, the Republic of Korea was 40 percent urbanized in 1970 and
78 percent urbanized by 1990. What took the United States 90 years to accomplish took Korea 20 years and Brazil 30 years.

That gradual pace, combined with relatively high gross domestic product (GDP) and education per capita at the beginning of the century, allowed time for the development of the political and economic institutions and market instruments essential for an efficient form of urbanization and a reasonable quality of urban life. These included mechanisms for the internal governance and financing of cities, intergovernmental arrangements, regulatory and financial instruments for intercity communications and transport networks, a civil service with technical expertise in urban and regional planning and service provision, and institutions for efficient functioning of national and local land markets. For rapidly urbanizing developing countries, the societal learning required to adapt rural institutions and governance to urban ones became a crash course, leaving little room for timely experimentation and adjustment.

This article looks at the form urbanization takes—the degree of concentration within a typical institutional and policy context—rather than at urbanization itself—the share of the national population residing in urban areas. It looks first at the basic theoretical concepts and empirical patterns governing a mature urbanized economy. What does a system of cities look like in a medium- or high-income country with well-developed urban institutions and governance and well-functioning land and capital markets? That provides a benchmark. In looking at the urbanization process in developing countries, the article then explores why urban concentration sometimes increases strongly in developing countries. When is concentration excessive, and what are the costs of excessive concentration? What do inter- and intracountry data show? The final section looks at the policies and institutions that help shape urbanization, in particular economic liberalization, fiscal deconcentration, intercity infrastructure investments, globalization, and urban institutions. What are the key elements for achieving reasonably efficient urbanization that benefits most segments of the national population?

Cities in a Mature Urban System

Urbanization is a natural part of development. Labor-saving technologies in agriculture and shifts in the composition of national output away from agriculture release labor for industrialization. Industrialization occurs disproportionately in urban areas because of opportunities to exploit scale economies of local agglomeration. About 70 percent of the cross-country variation in urbanization is explained by variations in GDP per capita. Figure 1 illustrates the log-linear relationship for 1990.1

It helps to have a framework to examine the link between urbanization and development and to deal with the fundamentals of urban allocation in a country. A framework can help structure thinking about the productivity of cities—is big good?—and to understand phenomena such as product cycles and new information technologies.
It can help answer questions about why cities form, what determines or limits their size at various points in time, why they grow, how they specialize in different types of production, and what determines the size distribution of cities in a country. This discussion sets the stage for examining the role of institutions and infrastructure investments in urban development and for analyzing the effects of economic liberalizations, globalization, and political decentralization on urbanization.

**Why Urbanization: The Economics of Agglomeration**

Economic activity agglomerates in cities because of local external economies of scale in production—a plant’s productivity is enhanced if other plants are located nearby. Why is this? Urban models of the micro-foundations of urban externalities follow Alfred Marshall’s (1890) description of spillover benefits for a plant from others in the neighborhood—information spillovers about technology, suppliers, purchasers, and market conditions (Fujita and Ogawa 1982); scope for local intraindustry specialization of plants in specific activities (Becker and Henderson 2000); increased diversity of local suppliers to local export producers (Dixit and Stiglitz 1977; Abdel-Rahman and Fujita 1990); and search for and matching improvements between workers and firms in local labor markets (Helsley and Strange 1990). In addition,
the proximity of buyers and sellers reduces transport costs of trade (Krugman 1991) and search costs in retailing (Caplin and Leahy 1998).

Research suggests that plants producing standardized products (textiles, steel, ceramics, food processing) learn mostly from other plants in the same industry locally, so that they benefit primarily from what are called localization economies, as first documented by Hoover (1948) for industries such as footwear in the United States. Recent econometric evidence for Korea shows similar benefits (box 1). In contrast, urbanization economies, deriving from the overall local scale of a metropolitan area, benefit plants engaged in nonstandardized production (special-order machinery, high-fashion ap-

---

**Box 1. Evidence on Agglomeration Economies**

Econometric work has quantified the degree of local external economies, distinguishing between localization and urbanization economies. Industries subject to localization economies, which depend on local own-industry scale, tend to locate in small and medium-size specialized cities. Industries more subject to urbanization economies are drawn to large diverse metropolitan areas. For Japan, Nakamura (1985) found strong evidence of localization economies in iron and steel, nonferrous metals, nonelectrical machinery, transport equipment, and precision instruments and evidence of urbanization economies for such industries as publishing and furniture. For Indonesia, Henderson and Kuncoro (1996) found strong localization economies for apparel (including textiles), nonmetallic minerals, machinery (including transport equipment and electrical machinery), and miscellaneous manufacturing (toys, jewelry, musical instruments) and evidence of strong urbanization economies for wood, furniture, publishing, and apparel. Results for the United States are similar (Henderson 1988).

For the Republic of Korea, Henderson, Lee, and Lee (2001) estimate urban scale economies using city industry data for 1983, 1989, and 1991–93. They examine the determinants of value added per production worker across cities by industry, controlling for capital per worker and accounting for time, city, and subindustry fixed effects. Their results on the magnitude of localization economies are reported in the table below. A coefficient of 0.06–0.08 means that a 1 percent increase in local own-industry employment results in a 0.06–0.08 percent increase in plant output. So a plant in a city with 1,000 workers in other firms in the same industry would, without changing its own inputs, increase its output by over 70 percent by moving to another city with 10,000 workers in the same industry. That is a big incentive. Henderson, Lee, and Lee (2001) show that the ranking of the magnitude of externalities exactly matches the ranking of industries by extent of spatial concentration across cities. So heavy and transport industries tend to be concentrated in a few highly specialized cities, whereas traditional industries with low scale externalities are more dispersed.

In searching for urbanization economies, Henderson, Lee, and Lee (2001) found that measures of metropolitan scale had no effect on productivity in any industry (see table 1). Drawing on Jacobs (1969) and Glaeser and others (1992), Henderson, Lee, and Lee created a diversity measure for each city based on the sum over all manufacturing industries of the squared deviations in each industry's share of local employment from its share of national employment. In perfectly diverse cities employment shares match national shares. In high-tech industries (only), Henderson, Lee, and Lee (2001) find strong significant diversity economies, where a one-standard-deviation increase in diversity increases productivity by 60 percent. This accords with the intuition that the bright lights of diverse large metropolitan areas, with cross-industry fertilization of ideas, are important for evolving high-tech industries but not standard ones.
Table 1. Magnitude of External Economies of Scale for the Republic of Korea

<table>
<thead>
<tr>
<th>Industry</th>
<th>Localization economies (elasticity)</th>
<th>Urbanization economies (percentage increase in productivity for one standard deviation increase in diversity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional (food, textile, apparel, wood and paper products and furniture)</td>
<td>0.021*</td>
<td>NA</td>
</tr>
<tr>
<td>Heavy (basic and fabricated metals, chemicals, and plastics)</td>
<td>0.082*</td>
<td>NA</td>
</tr>
<tr>
<td>Transport</td>
<td>0.096*</td>
<td>NA</td>
</tr>
<tr>
<td>Machinery and electrical machinery</td>
<td>0.053*</td>
<td>NA</td>
</tr>
<tr>
<td>High tech (computers, communication, television, radio and scientific instruments)</td>
<td>0.056* 0.599*</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 5 percent level.


parel, entertainment services, publishing), firms in interactive services with centralized markets (certain financial services, including stock and commodity markets), and high-tech firms in research and development (R&D) and new products (electronics, software development) (Jacobs 1969). For example, plants engaged in product development need to get ideas and specialized employees from beyond their own industry by interacting with the scientific consulting industry, universities, the software industry, a diversity of suppliers—whatever tweaks the inventive process. They also need large local markets to test products and ideas locally and to get immediate feedback.

There may also be dynamic components to externalities. One is local knowledge accumulation (Rauch 1993), as in endogenous growth models (Romer 1986). Black and Henderson (1999) develop an urban growth model that suggests city sizes will grow over time, as local human capital development enhances economies of agglomeration. They document the strong empirical relation between human capital accumulation in individual cities and growth of city size.

Local knowledge accumulation implies that history matters. Cities with a history of employment in a particular industry will have built up a stock of local “trade secrets,” which makes locating there much more attractive than in cities with no experience in the product. This presents an obstacle that newly industrializing cities have to overcome if they are to attract such an industry, although there is debate about the strength of this force (Holmes 1997; Rauch 1993; Glaeser and others 1992; Henderson, Kuncoro, and Turner 1995).

**Factors Limiting City Sizes over Time**

Although bigger cities offer higher productivity because of scale economies, residents of bigger cities are burdened with higher costs of living—for housing, food, public

Vernon Henderson
utilities, commuting, and so on (Muth 1969; Fujita and Ogawa 1982). In reviewing evidence on Latin America (Thomas 1980) and the United States, Henderson (1988) finds a doubling or more of costs of living in moving from a small city to a large metropolitan area. Rousseau (1995) calculates that living costs in Paris are 89–94 percent higher than the average for the rest of France. Richardson (1987) estimates that the marginal investment cost (including utility, public service, and housing costs) of absorbing an extra family in average urban areas of Bangladesh, Egypt, Indonesia, and Pakistan is three times that in rural areas—higher still for the primate city of a country.

In a market economy, city size reflects a tradeoff between the scale productivity benefits of local employment growth and the cost-of-living increments of local population growth. Improvements in commuting technology, which reduce local costs of living, and local knowledge accumulation, which accentuates scale economies, suggest that efficient city sizes increase over time. As commuting technology continues to improve and human capital investments rise in developing countries, the sizes of cities may continue to increase well into the 21st century, just as they did throughout much of the 20th century. At the same time, national population growth means that the number of cities may need to grow to accommodate population increases. Generally, both the size and number of cities have grown with national population growth in most countries over the past century.

**Urban Specialization, Diversification, and Productivity**

Most employment in a city (55–60 percent, often more) is engaged in nontraded goods production, such as housing, local retailing, local services, and local government. But as a result of the factors already discussed, mature medium-size and smaller cities will be somewhat specialized in the products and to some extent the services they produce for export (to other cities or abroad). Why is this? For products experiencing localization economies, cities specialize because adding a different industry does not help own-industry productivity, though it causes the cost of living to escalate. Such products include primary metals, pulp and paper, textiles, and machinery. Some of these products involve natural resource use and weight reduction in production, suggesting hinterland location. Specialization also extends to the service sector for market and transport node towns, state government and education centers, and agricultural services towns, as well as certain areas of health, recreation, and standardized business services.

Theory and evidence suggest that the size of specialization cities is positively related to the degree of localization economies of the good in which the city specializes. So steel and auto cities tend to be larger than textile cities and traditional service centers, which in turn are larger than pulp and paper and food-processing cities. As reported in Henderson (1988), this phenomenon of specialization has been studied for Brazil, India, Korea, and the United States (see box 2), among other countries.
Box 2. Specialization among U.S. Cities

Bergsman, Greenston, and Healy (1972); Henderson (1988); and Black and Henderson (1998) use cluster analysis to classify cities by production patterns. Black and Henderson distinguish 55 city types, or clusters, based on 80 industrial categories covering all private employment in the United States in 1990. Production patterns of cities differ significantly across city types. Average size and educational composition also tend to vary by city type. They divide the 55 city types into eight broad categories: clothing and food (49 cities); wood products (14 cities); electronics (20 cities); heavy manufacturing (48 cities); oil and chemicals (19 cities); national and regional market centers (52 cities); health, hotel, state government, and recreation services (72 cities); and business, finance, education, and transport services (44 cities).

Out of 318 total cities, 35 focus on health services; these cities average 220,000 people. There are 6 highly specialized food processing cities and 13 pulp and paper and lumber cities, both averaging about 165,000 people. There are three highly specialized primary metals cities and nine former primary metals centers whose production composition is changing, both averaging 260,000–290,000 people. There are two highly specialized insurance cities averaging 627,000 and three recreation centers (Reno, Las Vegas, and Cape May) averaging 440,000. All the smaller types of manufacturing contain industries subject primarily to localization economies (see box 1). At the other extreme are large market centers—nine with traditional mixed industrial bases ( wholesale, transport, and business services) and averaging 2.7 million people, seven with diverse modern services plus some manufacturing (health, education, engineering and management) and averaging 2.4 million people, and eight with newer mixed industrial bases (high-tech and engineering and management) averaging 1.4 million people.

Medium-size cities (under 750,000 people) play a key role in standardized manufacturing. In 1992 they accounted for 34 percent of the population of all U.S. metropolitan areas and 35 percent of manufacturing employment. They accounted for 44–54 percent of all employment in food, textiles, pulp and paper, wood products, petroleum, nonmetallic minerals, and primary metals industries. Medium-size cities are underrepresented (under 27 percent) in high-tech instruments and headquarters activity, activities subject to urbanization economies. In contrast, certain services cluster in large metropolitan areas. Though the primate cities of New York, Chicago, and Los Angeles have about 12 percent of the national population, they account for 40 percent of investment security services, 19 percent of banking services, 20 percent of holding and other investment services, and 19 percent of legal services.

City size and diversity are closely related. The normalized Hirschman-Herfindahl index of lack of diversity—the sum over 20 manufacturing industries of squared deviations of each subindustry’s local share in manufacturing employment from its national share—has a sharp negative slope declining from an average of about 0.15 for the smallest cities (with a maximum observed value of 0.36) to almost zero for the largest cities (with an observed minimum of 0.0090) (based on an ordinary least squares regression of $g = 0.4868 - 0.03119 \log \text{population}$, where the standard error on the slope coefficient is 0.00381, the number of observations = 317, and $R^2 = 0.18$.)
However, many products require a diversified and hence large local environment to thrive. These include high-tech products, R&D, high-fashion apparel, publishing, special-order machinery, financial services, company headquarters, and such business services as upscale advertising and placement services. Large metropolitan areas produce these products, but there is also relative specialization—some emphasize financial and international export market services, and others may emphasize high-tech environments. But it is even more complicated than that. For example, diverse metropolitan areas are rich in large social networks, so they can become migration points, attracting migrants from the hinterlands especially in times of crisis, such as famine and civil unrest. These low skill–low wage migrants are an attractive labor pool for producers such as apparel manufacturers, who want both big local product markets and cheap labor.

Across the range of products and city types, the relative size distribution of cities is fairly stable over time (see Becker, Williamson, and Mills 1992 on India and Eaton and Eckstein 1997 on France and Japan). Big and small cities coexist and grow in population size in parallel, keeping the same relative size and offering residents similar standards of living. For example, both wages and costs of living differ by more than 100 percent in a city of 50,000 people and a metropolitan area of 5 million people, so real incomes are equal in both types of cities.

A plant seeking to locate in the bigger city must be able to afford wages that are twice as high, as well as high rents and public service taxes. To remain competitive, the plant must enjoy correspondingly larger output per worker. Rousseau (1995) estimates that productivity for the types of firms located in Paris is 240 percent higher than in the rest of France, but wage, tax, and rent differentials raise costs 245 percent. Most industries cannot afford these costs because they do not experience large enough productivity differentials, and so they locate outside Paris. Experience has been similar for the types of industries found in specialized medium-size and smaller U.S. cities and those subject to localization economies in Korea (see boxes 1 and 2).

Policymakers tend to focus on the greater output per worker in larger cities and ignore the higher opportunity costs of labor and land, creating a policy bias toward locating heavy, often polluting industries in larger cities. Traditional policy toward state-owned industry in China clearly demonstrates this bias: plants in big cities received favored access to the newest technology, best managers, appropriate funding, and export-import licenses. Shanghai’s 1981 share of national industrial output (especially in heavy primary metals and machinery) was several times the share in New York or Chicago in 1914 (early period of U.S. industrialization and urbanization) and in 1977 (Henderson 1988), despite Shanghai’s smaller relative share of national urban population. It is not surprising that many township industries today outcompete many state industries in Shanghai. They are not burdened with high costs of housing and feeding mega-city workers.
For new products or plants, development appears to be accelerated by location in large metropolitan areas, in centers of information exchange (see box 3). New plants, which by definition are inexperienced, improve their chances of surviving by locating in an environment rich with external information about technology, suppliers, and markets. If these plants learn, grow, mature, and become more self-sufficient in information, they can move further out to satellite towns, where land and wage costs are lower.

New plants, especially smaller, informal sector plants, also play a key role in fostering local productive environments. In the Silicon Valley computer-electronics center around San José, California, most plants are small—in the informal, less regulated sector. In high-tech industries, small plants that rapidly appear, die, or split off play a key role in innovation, representing new ideas and ventures. They also play an important role in traditional industries by filling special orders and supplying key components to large-scale plants in the formal sector. Their role in developing countries is no different. For example, Mukherjee’s (1990) work on Calcutta shows that the informal sector is a vibrant, productive sector that interacts strongly with the formal sector, meeting specific demands for components and repairs (see also Banerjee 1985).

The story for products is similar. Radio, television, and personal computer technology has typically developed in intense, usually metropolitan environments, such as

---

**Box 3. Product Cycles and Stages**

*Plant Cycles from the Seoul Metropolitan Area*

Central cities are viewed as incubators for new plants. Once surviving plants mature and have established lines of communications, many decentralize to suburbs or more distant satellite cities. Table 2 shows patterns for the greater Seoul metropolitan area. In the early 1980s, new plants outnumbered relocators 5.5 to 1 in the central business district but made up just 60 percent of the number of relocators in satellite cities. So new plants locate mostly in the core and then, if successful, relocate later to the suburbs. Among relocators, bigger plants move furthest, perhaps because they are less dependent on outside information and more concerned about wage and land costs.

*Product Cycles for Japanese Electronic Firms*

Fujita and Ishii (1994) examine the plant location patterns of the nine major Japanese electronic firms as of the early 1990s. They divide plants hierarchically into headquarters units, basic R&D, developmental R&D, trial production, and mass production. At the top are the most sophisticated, information-driven activities, with all information coming from the top down within the corporation.

A strict locational hierarchy accompanies this hierarchy of activities. Within Japan, basic R&D plants all go to major metropolitan areas, mostly Tokyo (see table 3). Mass production plants are mostly outside metropolitan areas, and those in metropolitan areas are in suburbs, not the central city (in contrast to headquarters and basic R&D). Similarly, outside Japan, more sophisticated activities are located in the industrialized nations, whereas mass production of standardized parts, components, and products is located in the Association of Southeast Asian countries, China, and India.
Table 2. New Plants and Relocating Plants in the Greater Seoul Metropolitan Area

<table>
<thead>
<tr>
<th>Location</th>
<th>Ratio of new plants to relocators</th>
<th>Distance of relocators (percent of plants for each employment category by distance moved)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Central business district</td>
<td>5.5</td>
<td>1–5 km: 30, 25–99 km: 43, 100+ km: 43</td>
<td></td>
</tr>
<tr>
<td>Metropolitan area and suburban</td>
<td>2.6</td>
<td>1–5 km: 30, 25–99 km: 42, 100+ km: 59</td>
<td></td>
</tr>
<tr>
<td>rings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satellite cities</td>
<td>0.6</td>
<td>35+ km: 27, 100%</td>
<td></td>
</tr>
</tbody>
</table>

Source: For new plants to relocators ratio, Lee, Choe, and Park (1985); for distance moved, Chun and Lee (1985).

as Tokyo. As production stages moved from the experimental to the routine or standardized, these industries were dispersed to hinterland locations or abroad. Product development and R&D require the information-rich environments found in large metropolitan areas, whereas efficient, standardized production requires the low land and labor costs found in smaller cities and less developed countries (Duranton and Puga 2001).

Urbanization with Rapid Economic Growth

For developing countries, studies suggest that urban concentration increases in the early stages of economic development as part of increasing regional disparities, and then decreases in later stages of development as part of decreasing regional disparities, or regional convergence (Williamson 1965; see also Parr 1985; El-Shakhs 1992; Table 3.

Table 3. Location Hierarchy of Japanese Plants Inside Japan and Abroad

<table>
<thead>
<tr>
<th>Location</th>
<th>Share of total plants in the major metropolitan areas</th>
<th>Abroad: Share of total plants located abroad in North America, European Union, and newly industrialized countries, as opposed to less developed countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headquarters and basic</td>
<td>100 (70)</td>
<td>100 (13)</td>
</tr>
<tr>
<td>R&amp;D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental R&amp;D</td>
<td>70 (54)</td>
<td>Regional headquarters 100 (13)</td>
</tr>
<tr>
<td>Trial production</td>
<td>91 (21)</td>
<td>R&amp;D 91 (32)</td>
</tr>
<tr>
<td>Mass production</td>
<td>43 (333)</td>
<td>Production 59 (335)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are the total number of plants in all locations in each category (row).

Source: Fujita and Ishii (1994).
and Wheaton and Shishido 1981). Hansen (1990) suggests that in the early stages of development, when resources are scarce (especially infrastructure capital) it is impossible to invest enough in public infrastructure to support widespread urban agglomeration and a system of interconnected cities joined by paved highways, rail transit, and communication.

An efficient allocation of economic infrastructure requires concentration in just one or two cities (perhaps megacities) initially, conserving on spending that would go instead to linking cities or developing other cities. Such concentrated investment induces immigration and industrialization of focal megacities and increases urban concentration. With accompanying economic growth, the country will develop some appropriate institutions and a pool of skilled technocrats. Eventually, the country will be able to invest in hinterland regions, allowing other major urban centers to develop as well as smaller and medium-size cities. That leads to urban deconcentration.

The concentration phase has been observed in many developing countries, but any subsequent deconcentration seems to be modest. There is a very strong cross-country relation between GDP per capita and measures of primacy and concentration: concentration increases up to mean (of log) GDP per capita worldwide and then decreases thereafter (box 4). The relation is significant and robust to specification. However, the magnitudes are tiny—roughly a 2.5 percent change in concentration as GDP per capita grows from low levels ($1,000) to the mean ($5,300).

This analysis shows few signs of population deconcentration. It does find exceptionally high concentration in certain regions, particularly Asia (excluding China), where primacy is about 20–30 percent higher than can be explained by typical measured economic forces. This is the Asian megacity phenomenon. This high concentration seems to stem from political and institutional forces that lead to a high centralization of resources, so that a high concentration of industry accompanies population concentration in these megacities.

Though significant population deconcentration may not be widespread, there appear to be strong forces encouraging industrial deconcentration from megacities, as costs of production escalate and living conditions deteriorate. Countries such as Korea that have encouraged industrial deconcentration have experienced enormous effects (box 5). Industrial deconcentration starts with deconcentration from the core primate city into its suburbs or satellite cities and then to the rest of the country. Deconcentration into the rest of the country enhances specialization across cities, as industrializing cities focus on a specific activity. Korea experienced the first stage of deconcentration into satellite cities before 1980 and the second into the hinterlands over the past 20 years.

Thailand and Indonesia seem to be experiencing some of the first stage of deconcentration of industry into satellite cities—starting in Thailand’s case, certainly, from very high levels of concentration. Consider the Bangkok metropolitan area (see Isarankura 1990 and Lee 1988). Between 1970 and 1986 the urban population of
Box 4. International Comparisons of Urban Concentration

Data for more than 85 countries for 1960, 1970, 1980, and 1990 show clear relationships between urban concentration measures and country characteristics (table 4). The first five columns examine primacy—the population in the largest city. The sixth column examines total population in cities over 750,000.

Basic Patterns

The size of the largest city grows with national urbanization but much less than proportionally (an elasticity well below one). Urban primacy is weaker in larger countries, as economic resources are spread out with geographic size. As GDP per capita grows from low levels up to $5,300 (approximate mean world GDP per capita in 1990), urban primacy increases. After that, it declines. Thus urban primacy is the same at $1,300 as it is at $17,000. But even at its peak, all else remaining unchanged, urban primacy is only 2 percent larger than at $1,300 (or $17,000). International trade also affects primate city size: a one-standard-deviation increase (48.8) in openness (imports plus exports as a percent of GDP) decreases primate city size by 27 percent.

A final major issue concerns regional differences, presumably capturing differences in broad institutional arrangements and attitudes. Regional dummy variables have little effect on other coefficients and captured unmeasured aspects of regions. Primacy is highest in Asia, in most formulations. This appears to capture the emerging Asian megacity phenomenon of development in countries such as Bangladesh, Indonesia, Japan, and Thailand, which is highly focused on one city. In the basic model, urban primacy also is significantly higher in Latin American and Sub-Saharan countries.

Centralization and Infrastructure Investments

Economic models suggest that centralization increases urban concentration. Centralization is hard to measure, but two effects are notable. First, when the primate city is the national capital, it is 25 percent or more larger than otherwise, suggesting resource centralization goes with political centralization. Second, the share of the central government in government consumption (nondefense spending), a crude summary measure of fiscal centralization, has strong, significant effects on primacy beyond the national capital effect on a limited sample covering 1980 and 1990 (column 3, table 4). A 1 standard deviation (0.31) increase in the central government’s share increases primate city size by 9 percent. Government centralization appears to contribute strongly to urban concentration.

Another way to approach this question is to see what effect decentralized infrastructure investment (interregional transport and communications) has on urban deconcentration. Again, this is hard to measure. Data on the density of roads, telephone mainlines, and navigable waterways are available, but each set of data has problems. Telephone mainlines are highly correlated with GDP per capita and concentrated in urban areas, so effects are hard to disentangle. Road density is less correlated with GDP per capita, but the measure has no control for road quality or width. The density of navigable waterways (length of navigable waterways divided by land area) represents both geographic features (rivers and lakes) and investment in infrastructure (canals, dredging, rerouting, flood control). Navigable waterways are, however, a good measure of accessibility of hinterlands to the coast.

These problems aside, analysis shows that increased infrastructure density strongly reduces urban concentration (columns 4 and 5, table 4, with shifting sample sizes). After controlling for population and land area, waterways have enormous impacts—elasticities of five or six—opening up a country and reducing urban concentration. Road and telephone mainlines have negative but insignificant density effects, reducing urban concentration. These crude measures suggest that infrastructure investments help decentralize resources and shape urban deconcentration.
Table 4. Determinants of Urban Primacy

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log national urban population</td>
<td>0.706*</td>
<td>0.707*</td>
<td>0.613*</td>
<td>0.755*</td>
<td>0.753*</td>
<td>0.894*</td>
</tr>
<tr>
<td>(0.022)</td>
<td>(0.026)</td>
<td>(0.038)</td>
<td>(0.025)</td>
<td>(0.044)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>Log national land area</td>
<td>-0.029</td>
<td>-0.037**</td>
<td>0.021</td>
<td>-0.068*</td>
<td>-0.067</td>
<td>0.0028</td>
</tr>
<tr>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.033)</td>
<td>(0.019)</td>
<td>(0.045)</td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>Capital city dummy variable</td>
<td>0.263*</td>
<td>0.263*</td>
<td>0.452*</td>
<td>0.380*</td>
<td>0.390*</td>
<td></td>
</tr>
<tr>
<td>(1 if city is the capital)</td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.096)</td>
<td>(0.064)</td>
<td>(0.089)</td>
<td></td>
</tr>
<tr>
<td>Log per capita GDP</td>
<td>1.28*</td>
<td>0.910*</td>
<td>0.534</td>
<td>1.52*</td>
<td>2.39*</td>
<td>1.16*</td>
</tr>
<tr>
<td>(0.324)</td>
<td>(0.356)</td>
<td>(0.866)</td>
<td>(0.374)</td>
<td>(0.583)</td>
<td>(0.433)</td>
<td></td>
</tr>
<tr>
<td>Log per capita GDP sq.</td>
<td>-0.078*</td>
<td>-0.053*</td>
<td>-0.030</td>
<td>-0.089*</td>
<td>-0.136*</td>
<td>-0.065*</td>
</tr>
<tr>
<td>(0.020)</td>
<td>(0.022)</td>
<td>(0.052)</td>
<td>(0.023)</td>
<td>(0.036)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>-0.0056*</td>
<td>-0.0056*</td>
<td>-0.0092*</td>
<td>-0.0049*</td>
<td>-0.0060*</td>
<td>-0.0065*</td>
</tr>
<tr>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td>(0.0016)</td>
<td>(0.0011)</td>
<td>(0.0014)</td>
<td>(0.0013)</td>
<td></td>
</tr>
<tr>
<td>Share of central government in government consumption</td>
<td>0.306*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.129)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of waterway density (ratio of length of waterways to national land)</td>
<td>-6.61*</td>
<td>-5.98*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.02)</td>
<td>(1.24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of road density (ratio of length of roads to national land)</td>
<td>-0.048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.045)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of telephone main lines (per 1,000 people)</td>
<td>-0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1970</td>
<td>0.147*</td>
<td>0.143*</td>
<td>0.117*</td>
<td></td>
<td>0.089</td>
<td></td>
</tr>
<tr>
<td>(0.060)</td>
<td>(0.059)</td>
<td>(0.069)</td>
<td>(0.069)</td>
<td>(0.074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1980</td>
<td>0.329*</td>
<td>0.315</td>
<td>0.131*</td>
<td>0.246*</td>
<td>0.152*</td>
<td>0.227*</td>
</tr>
<tr>
<td>(0.063)</td>
<td>(0.064)</td>
<td>(0.096)</td>
<td>(0.062)</td>
<td>(0.071)</td>
<td>(0.076)</td>
<td></td>
</tr>
<tr>
<td>Year 1990</td>
<td>0.379*</td>
<td>0.363*</td>
<td>0.166*</td>
<td>0.285*</td>
<td>0.217*</td>
<td>0.272*</td>
</tr>
<tr>
<td>(0.067)</td>
<td>(0.070)</td>
<td>(0.103)</td>
<td>(0.068)</td>
<td>(0.079)</td>
<td>(0.079)</td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>0.084</td>
<td>-0.206</td>
<td>0.207*</td>
<td>0.311*</td>
<td>0.240*</td>
<td></td>
</tr>
<tr>
<td>(0.098)</td>
<td>(0.162)</td>
<td>(0.098)</td>
<td>(0.151)</td>
<td>(0.113)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latin America</td>
<td>0.191*</td>
<td>-0.085</td>
<td>0.120*</td>
<td>0.065</td>
<td>0.194*</td>
<td></td>
</tr>
<tr>
<td>(0.073)</td>
<td>(0.112)</td>
<td>(0.069)</td>
<td>(0.107)</td>
<td>(0.081)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>0.221*</td>
<td>0.302*</td>
<td>0.226*</td>
<td>0.207*</td>
<td>0.347*</td>
<td></td>
</tr>
<tr>
<td>(0.082)</td>
<td>(0.112)</td>
<td>(0.071)</td>
<td>(0.091)</td>
<td>(0.084)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>0.207*</td>
<td>0.240*</td>
<td>0.058</td>
<td>0.024</td>
<td>0.211*</td>
<td></td>
</tr>
<tr>
<td>(0.081)</td>
<td>(0.120)</td>
<td>(0.080)</td>
<td>(0.124)</td>
<td>(0.087)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-3.71</td>
<td>-2.07</td>
<td>0.708</td>
<td>-4.97</td>
<td>-8.67</td>
<td>-5.05</td>
</tr>
<tr>
<td>Number of observations</td>
<td>320</td>
<td>320</td>
<td>127</td>
<td>286</td>
<td>147</td>
<td>237</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.895</td>
<td>0.899</td>
<td>0.889</td>
<td>0.919</td>
<td>0.910</td>
<td>0.922</td>
</tr>
</tbody>
</table>

'Significant at the 5 percent level.
"Significant at the 10 percent level.
Note: Numbers in parentheses are standard errors.
' The dependent variable is population in cities over 750,000.
Sources: Author's calculations based on data from the United Nations and World Bank.
Box 5. Deconcentration in the Republic of Korea

The urban primacy of Seoul peaked in the early 1970s and declined thereafter (table 5), even though Seoul's share of national population continued to rise with rapid national urbanization. Seoul's share of the population of the surrounding Kyonggi Province has not changed much in 25 years. What is dramatic is Seoul's loss of manufacturing employment. With rising suburbanization, Seoul's share declined from 76 percent in 1970 to 45 percent in 1983 and 30 percent in 1993 (table 6), reflecting both natural market forces as firms left high wages and rents in Seoul for nearby cities and policies encouraging suburbanization.

However, a key aspect of deconcentration in Korea has been the spread of industry throughout the country following the massive intercity infrastructure investments of the late 1970s and early 1980s and the economic and political liberalization of the 1980s. The employment shares of Seoul and the other two major metropolitan areas of Pusan and Taegu declined in the 1980s, while satellite cities (over 50,000 urban residents in 1983) in the provinces of Kyonggi and South and North Kyongsang surrounding the three major metropolitan areas held constant (table 7). Where did industry go? Mostly to rural areas and cities in other provinces.

This deconcentration occurred in all manufacturing industries and was comprehensive. Korea's 10 major manufacturing industries (over 130,000 employees nationally in 1993) are traditional food processing, textile, apparel, nonmetallic minerals, modern chemicals, rubber, fabricated metal, motor vehicles, machinery, and communications equipment. For each of the 10 industries, Lee (1998) calculates an Ellison and Glaeser (1997) or normalized Hirschman-Herfindahl index of the degree of spatial concentration across provinces and 50 cities in 1983. The index for industry $j$ is

$$G_j = \sum_{i=1}^{n} (S_{ij} - S_j)^2$$

where $S_{ij}$ is region $i$'s share of total national employment in industry $j$, and $S_j$ is region $i$'s share of national manufacturing employment. With perfect deconcentration, $G_j = 0$, because each region's share of industry $j$'s employment mimics its share of national manufacturing employment. With perfect concentration $G_j$ exceeds one, where one region has all of industry $j$'s employment ($S_j = 1$ for that region) and little of national employment.

Across the nine provinces from 1983 to 1993, $G$ drops dramatically for all six modern industries—by 95 percent or more in most cases and a minimum drop of 42 percent. Nationally, modern industries deconcentrated rapidly throughout the country. For traditional industries, $G$ changes only modestly. Traditional industries, especially food processing and nonmetallic minerals, were already deconcentrated and remained that way.

In contrast to deconcentration across the nine provinces, concentration increases across the 50 cities for all industries except textiles, for which concentration remains unchanged. The rise is dramatic for some modern industries, such as chemicals and rubber, and in most cases it is strong. Is this a paradox—declining regional concentration in modern manufacturing industries, but rising urban concentration in the same industries? Not at all. Cities specialize in production. So with liberalization, infrastructure investments, and decentralization inducements, there is national deconcentration of industry spatially. However, these same policy changes allow cities to become more specialized as they industrialize. Korea is well on the path to having a mature system of cities.
the core Bangkok metropolitan area grew 4.3 percent a year, while key satellite sub-
urban areas (Samut Prakarn, Nonthaburi, and Pathum Thani) grew 5.7–8.0 per-
cent a year. Manufacturing employment grew at average annual rates of 8.2–12.6
percent in all five suburban areas defined by Isarankura (1990), but at 7.2 percent
in the Bangkok area, which was increasingly focusing on service production. Lee
(1988) reports that in 1982 the start-up rate for new manufacturing establishments
was 7.4 percent in Bangkok and 8.9–16.9 percent in the suburban areas. Despite
these differentials strongly favoring suburban areas, the absolute population growth
rate in the core metropolitan area remains very high.

Although industry may eventually deconcentrate from primate cities, deconcen-
tration may be delayed too long for many developing countries, which appear to
undergo strong initial sustained increases in population concentration. Are there
compelling reasons to believe that this concentration is excessive? If so, what are the
costs of excessive concentration?

Is Concentration Excessive?

Many economic policies inadvertently promote inefficient concentration, delaying
the provision of appropriate infrastructure and services outside the central city until
long after it is efficient to do so. Urban development in the primate cities may be
significantly underpriced, considering that air pollution, noise pollution, contami-
nation from waste disposal, and traffic congestion remain unpriced or underpriced
externalities. These externalities are much lower in small cities than in large cities,
so that they are relatively more underpriced in big cities, which confers an advan-

| Table 5. Metropolitan Seoul's Share of Population (percent) |
|----------------|---------|---------|---------|---------|
| National urban population | 34      | 41      | 38      | 33      |
| National population       | 9       | 17      | 22      | 25      |

Source: Lee (1997).

| Table 6. Metropolitan Seoul's Share of Population and Manufacturing in Kyonggi Province (percent) |
|----------------|--------|--------|--------|--------|
| Population     | 62     | 63     | 67     | 61     |
| Manufacturing  | 76     | 61     | 45     | 30     |

Note: Data excludes Inchon metropolitan area.

Source: Lee (1997).
<table>
<thead>
<tr>
<th>Table 7. Cities’ Share of National Manufacturing Employment (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Seoul</td>
</tr>
<tr>
<td>Pusan and Taegu</td>
</tr>
<tr>
<td>Satellite metropolitan areas</td>
</tr>
<tr>
<td>Other cities, other rural areas</td>
</tr>
<tr>
<td>Rural areas of satellite city provinces</td>
</tr>
</tbody>
</table>

Source: Lee (1997).

The advantage is accentuated in megacities, where negative externalities are particularly severe.

In some countries, export and import licensing and financial markets are centralized. People seeking licenses and loans need to locate in the primate city so that they can apply political, social, and financial resources to the pursuit of their applications. In Korea initial deconcentration of industry from Seoul to industrial parks in satellite cities in the 1970s was impeded by the bureaucratic bottlenecks in capital and international markets. Plants needed to be within half to three-quarters of an hour commuting time from downtown Seoul (Kwon 1985; Chun and Lee 1988) to make access feasible. These bottlenecks were eased in the early 1980s. In Indonesia, even after liberalization, financial markets for large loans remain highly centralized. Prospective borrowers are required to apply in person in Jakarta. Similar impediments surround export-import permissions, nonmarket allocations of raw materials, location permits, and monopoly licensing (Henderson and Kuncoro 1996).

Why are licensing procedures, financial markets, and allocation mechanisms more centralized in some countries? One reason is rents. If the authorities seek rents on licenses, permits, or loans, then a central authority would not wish to see decentralization, which would create potential competitors for the rents. Another reason is the large city bias of many policies, which are designed to favor the centralized development of large cities at the expense of more rural hinterlands. O (1993) and others illustrate this for China.

**Costs of Excessive Deconcentration**

What are the costs of excessive urban concentration? Strong and robust econometric evidence in Henderson (1999) suggests that excessive concentration significantly reduces economic growth rates. Growth losses in countries with excessive primacy may be as great as those resulting from significantly deficient human and physical capital investment. Why? Because wage, land, and infrastructure costs escalate with
megacity size relative to scale benefits for standardized products, making it more difficult to attract foreign investment and to remain competitive in world markets. Also, excessive concentration leads to degradation of the quality of life as population concentration rises. Problems with congestion, contamination of ground water, gastrointestinal and infectious diseases, and poor air quality are severe in many cities (see box 6 on Jakarta). The quality of life and public service issues are exacerbated by a lack of technical expertise in planning and managing megacities. As environmental quality in megacities declines, infrastructure investment becomes increasingly focused on trying to alleviate the effects of these unpriced negative externalities.

Influencing Deconcentration

As discussed, better pricing of goods and negative externalities in large metropolitan areas and liberalization and decentralization of financial markets, licensing, and permits can help countries avoid excessive concentration. So can intercity infrastructure investments, economic openness, and appropriate government structure and institutions, the subject of this section.

Infrastructure Investment

A key condition for urban deconcentration seems to be investment in modern intercity transport and communications. Conceptually the point is straightforward. Manufacturers cannot move from the main city (typically a port city) of a country to hinterland locations unless they are linked to the port and national and international markets by good transport and communication. Containerization and ready movement of freight are essential for purchasing inputs and getting products to market at reasonable delivered prices in a timely fashion. In a globalized world of just-in-time production and inventories slashed to reduce costs, reliable transport that is reasonably fast and cheap is essential. Good communications are also critical for interaction with buyers, suppliers, parent firms, and government regulators.

What is the evidence? As already noted, urban concentration is negatively related to crude measures of infrastructure—navigable waterways (requiring interconnecting canals, dredging, and maintenance) and road and telephone density (see box 4). That evidence is supported by econometric modeling using cross-country panel data (Henderson 1999). For example, with the development of major toll roads east and west from Jakarta into Botabek, Bekasi, and Tangerang (known collectively as Jabotabek), metropolitan (dki) Jakarta's share of Jabotabek's formal sector manufacturing employment fell from 57 percent in 1986 to 44 percent in 1991, and its share of start-ups dropped dramatically, especially for the corporate portion of the formal sector (Henderson, Kuncoro, and Nasution 1996).
Box 6. Unpriced Externalities of Mega Cities: The Example of Jabotbek

Sewer and Water in dkj Jakarta

Only 2.1 percent of the population of Jakarta in 1990 was served by an off-site sewerage connection, such as a conventional trunk sewer. Most of the rest relied on pit latrines and septic systems, which are unsuitable for high-density areas. Septic system failures are common, resulting in soil and water contamination. The sewerage problem is compounded by the use of groundwater by half the population (only 10 percent of households had piped water in 1992, and 20 percent purchased water from vendors; World Bank 1993).

These problems were most intense in the low-income areas of Jakarta Utara (North Jakarta) around the coastal port area, where households have no trunk sewers or piped water connections. Overpumping of groundwater has caused heavy and growing salination: 47 percent of low-income residents buy water from vendors (just 6 percent of high-income residents do). Disease incidence was higher in Jakarta Utara as well. In 1986–88 the incidence of gastroenteritis was 37.4 per 1,000 residents; of typhoid, 8.6; and of dysentery, 11.5, compared with 23.7, 3.0, and 5.2 for the rest of dkj Jakarta.

These high health costs have arisen because of unpriced externalities—both sewage disposal and groundwater pumping. Not only older neighborhoods are affected. New developments are poorly served as well because of poor land use planning and development policies, as the land development process in the suburban Botabek area shows.

Land Development in Botabek

Three national ministries control land use planning in Indonesia—Bappenas (planning), Public Works, and Home Affairs—but without any clear delineation of responsibilities. Local officials issue the land development permits. Aksoro (1994) argues that local officials in rural suburbanizing areas are uninformed about land use, often even lacking maps showing current use. “Planning” is simply a cataloging of evolving land use as dictated by private developers. In a typical scenario, Public Works builds an arterial highway from Jakarta into the rural hinterland with little consultation with local (rural) officials. Large-scale, sophisticated developers approach local officials for development permits. Negotiated developer permits give each developer complete monopsony power for three years to buy land in the designated area for future development.

The most important failure is the lack of required land improvements. Developers “strip-develop” in shallow, unconnected pockets along arterial roads, to avoid having to invest in side and feeder roads. That creates urban sprawl, with long strips of development along highways, requiring even more highway construction to encourage suburbanization. Proper planning would require deep-pocket development, with proper side and feeder roads linking developments to one another. Additionally, developers are not required to invest in proper infrastructure. Requirements for piped water connections are not enforced, so that groundwater pumping (with unauthorized unmetered deep wells) occurs, contributing to metropolitan Jakarta’s water problems. Installation of sewage lines and connections is incomplete. Even where lines do exist, untreated sewage simply flows into canals and unauthorized ponds.

All these negative externalities of population growth in metropolitan Jakarta are unpriced. Because these externalities are much worse in Jakarta than in other cities, the lack of pricing or other regulation means that immigration into Jakarta is underpriced, or subsidized, relative to its true cost and relative to its cost in other cities. That encourages overpopulation. Simply saying “price the externalities” is not going to change the process. Institutional change is also required.
rapid industrial deconcentration (see box 5) can be directly linked to its heavy nationwide investment in telecommunications since the late 1970s. From enormous disparities in the early 1970s, telecommunications capital stock per capita converged to nearly regional equality by the early 1990s, leading the deconcentration process.

**Globalization and Openness**

Simple cross-country regressions, as well as more complex time series regressions, suggest a strong negative correlation between international trade and urban concentration. A one standard deviation increase in openness is associated with a 25 percent decline in primacy and similar declines in other measures of concentration (see box 4). The new economic geography models predict that a country’s exposure to trade may encourage hinterland development (Krugman and Venables 1995; Krugman and Livas 1996; Puga and Venables 1996).

The link is not obvious. In a country with a dominant primate city and poor intercity infrastructure, greater trade associated with industrialization could promote enhanced primacy. So, if a Nike or Sony wants to locate assembly of standardized products or production of low-tech components in a cheap-labor country, it would initially focus on the primate, coastal city if the hinterlands are cut off by poor infrastructure. However, the relatively high land, labor, and congestion costs of operating in an increasingly oversized megacity would provide strong incentives to deconcentrate. The new economic geography literature suggests that products destined for international markets are less tied to focal points of national population and local demand and are freer to seek hinterland locations. If local cost conditions deteriorate, the threat of international producers moving to other countries may spur domestic infrastructure investments to permit deconcentration. Exporters facing stiff international competition also may be more eager to move to cheaper hinterland locations.

The deconcentration of manufacturing from metropolitan Jakarta to its hinterland region was spurred in part by foreign producers. In 1989, firms with more than 50 percent foreign ownership accounted for only about 9 percent of manufacturing employment in medium-size and large plants in both the Bobatek suburb of Jakarta and DKI Jakarta. By 1993 these firms accounted for 27 percent of such employment in the Botabek suburb but only 14 percent in DKI Jakarta, despite the development of free export zones in DKI Jakarta.

**Role of Government Structure and Institutions**

Government structure and institutions may play a key role in determining the deconcentration of urban activity. Many economists believe that decentralized governments and institutions lead to greater deconcentration. The evidence in Henderson (1999) suggests significant but surprisingly small effects on deconcentration, as coun-
tries shift to decentralized governmental structures. However, the cross-sectional evidence on historical institutions is strong. Henderson (1988) finds, other things being equal, that a federal system of government reduces a common Hirschman-Herfindahl index of urban concentration by more than 40 percent. Ades and Glaeser (1995) find a similar reduction in a primacy index for a similar switch in government regime (see also Petrakos and Brado 1989).

The concentration regressions in box 4 find that the mere fact that the primate city is the national capital increases its size by more than 25 percent, indicating that resource centralization is not uncorrelated with the seat of power. A crude measure of the share of the central government in overall government consumption has a very strong positive effect on concentration, over and above the national capital effect. Contrast the high urban concentration and high government decentralization in Bangladesh, Indonesia, or Thailand, for example, with the low urban concentration and low government centralization in China or India.

Economists look to the Tiebout (1956) model to explain the relation between urban concentration and government centralization. This model calls for local fiscal autonomy in the provision of services that are local in nature. The key to a successful local political process is local participation in political processes (local democratic voting outside the influence of the central government and reasonably unconstrained by national-level party officials), local determination of local taxes and revenues at the margin, and local determination of expenditure levels.

Local fiscal autonomy is viewed as efficient for two interrelated reasons. First, each locality's industrial needs and tastes for services differ, and there is asymmetric information—the national government will be unaware of specific local needs and tastes, which can be better ascertained through local political processes. If localities pay for public services, at least at the margin, they will make expenditure decisions based on (true) marginal costs, so allocations may be more efficient. Second, unitary governments may favor the national capital, creating a primate city bias in public services, infrastructure investments, and allocations from capital and import-export markets. This favoritism draws in immigrants. A decentralized situation creates and empowers hinterland cities and local governments to compete freely for industry and residents with the national capital, through tailored investments in infrastructure and public service offerings.

A key element in decentralization is revenue sources for the operating and capital accounts for autonomous local (or state) governments. Cities such as Bombay and Karachi testify to the fact that cities can operate with almost entirely local revenues, despite the common impression that local governments in low-income countries cannot raise sufficient local revenues (Bahl and Linn 1992:table 13.1). But financing local infrastructure investment and funding the capital account remain thorny issues. Efficient levels of long-term (intergenerational) investment cannot readily be financed from current taxes; they need to be financed over the long term, generally
by floating bonds and establishing sinking funds to pay them off. That requires a sophisticated capital market to monitor local government creditworthiness, good communications and information gathering, and probably an adequate level of economic development.

Low-income and high-income countries alike also rely heavily on intergovernmental transfers. India, for example, allocates funds to states and then to localities from the national plan, on a 30 percent/70 percent grant-to-loan ratio, using a formula based on population, income, special circumstances, and so on. Several problems typically attend this system. The loans are often forgiven, so the loan costs to localities are uncertain. Funding is forced, automatic, and unresponsive to local needs in the level and composition of expenditures. Funding is also inadequate. For large metropolitan areas in India and Pakistan, inadequate funding is addressed by creating metropolitan development authorities, which can float bonds, thereby having some control over financing levels at the margin. Indonesia has national development funds for which localities can apply. These are supposed to be revolving funds (new loans are financed out of repayments), but, as in India, loans are not repaid.

Conclusion

The rapid urbanization in many developing countries over the past half century seems to have been accompanied by excessively high levels of concentration of the urban population in very large cities. This pattern is seen most prominently in the recent development of Asian megacities.

Some degree of urban concentration may be desirable initially to reduce inter- and intraregional infrastructure expenditures. But in a mature system of cities, standardized manufacturing production tends to be deconcentrated into smaller and medium-size metropolitan areas, whereas production in large metropolitan areas focuses on services, R&D, and nonstandardized manufacturing. The problem in today’s developing countries is that there appear to be strong biases toward excessive concentration.

Such concentration reflects several forces. Externalities of congestion and pollution are relatively underpriced in megacities relative to smaller cities, encouraging overpopulation of larger cities. Rent seeking and urban bias by central government bureaucrats result in the centralization, often in capital cities, of capital markets and licensing for exports, imports, plant production, and material allocations, forcing a centralized location of production. Finally, infrastructure investments and public services provision outside of main cities and interregional linkages seem to remain at too low levels for too long.

The costs of excessive concentration stem from the large sizes of megacities and underdeveloped institutions and human resources for urban planning and manage-
ment. These costs include traffic accidents, health costs from exposure to high levels of air and water pollution, and time lost to inordinately long commutes.

Alleviating excessively high urban concentration requires investments in interregional transport and telecommunications to facilitate deconcentration of industry. It also requires fiscal deconcentration, so that interior cities can raise the fiscal resources and provide the services needed to compete with primate cities for industry and population.

Notes

Vernon Henderson is Eastman Professor of Political Economy at Brown University. Vernon Henderson can be reached via e-mail at Vernon_Henderson@brown.edu.

1. The correlation is from a regression of percent urbanized on GDP per capita pooling cross-country data for 1960, 1970, 1980, and 1990 for a total of 508 observations. Fixed-effects formulations also show a similar critical role of GDP per capita.

2. But there is now spatial balance overall in Jabotabek. Plant-level wage regressions for 1993 indicate that wages in Botabek suburbs in medium-size and large plants for production workers are 21 percent lower than in central Jakarta, controlling for industrial composition, foreign ownership, and plant legal organization. Productivity regressions indicate that plant productivity (controlling for industry, capital-to-labor ratio, and foreign ownership) is 14 percent lower in Botabek. For a capital share coefficient of 0.35 that represents a perfect balance on average, where the impact of the productivity loss is just offset by the reduction in labor costs. Assume a constant cost of capital with capital rental equal to its value of marginal product in a Cobb-Douglas production function with an estimated capital share of 0.35 and a labor share of 0.65. Then, if producers are in equilibrium, the ratio of wages in central Jakarta to those in Botabek (> 1.208) should equal the productivity gap (1.14)^1/0.65 or 1.22.

References


Developing Countries and a New Round of WTO Negotiations

Thomas W. Hertel • Bernard M. Hoekman • Will Martin

This article summarizes some of the results and findings emerging from an ongoing World Bank research and capacity-building project that focuses on the World Trade Organization (WTO) negotiating agenda from a developing country perspective. Recent research suggests that the potential gains from further multilateral liberalization of trade remain very large. The payoffs associated with attempts to introduce substantive disciplines in the WTO on domestic regulatory regimes are much less certain. This suggests that the focus of current and future negotiations should be primarily on the bread and butter of the multilateral trading system—the progressive liberalization of barriers to trade in goods and services on a nondiscriminatory basis. In addition, priority should be given to ensuring that rules are consistent with the development needs of poorer countries and to helping developing countries implement WTO obligations.

The Uruguay Round of multilateral trade talks, concluded in 1994 after eight years of often confrontational negotiations, was a landmark in the history of the trading system. Agriculture and textiles and clothing—two sectors that for all intents and purposes had been removed from the ambit of the General Agreement on Tariffs and Trade (GATT)—were brought back into the fold. The system of multilateral rules was extended to include intellectual property rights and services and, because of the single undertaking rule, all countries desiring to become members of the new World Trade Organization (WTO) were required to accept a variety of disciplines in areas ranging from customs valuation to subsidies.

Reflecting the very limited liberalization that had occurred in agriculture and services, the two agreements on these subjects included provisions calling for new negotiations within five years of the entry into force of the WTO. Other WTO agreements contained review provisions. To increase the scope for beneficial tradeoffs across issues, the 1998 WTO ministerial meeting called for the development of an agenda for “further liberalization sufficiently broad-based to respond to the range of interests and concerns of all members.” In the lead-up to the subsequent ministerial meet-
that was expected to launch a new round, WTO members submitted numerous proposals regarding the issues that should be included on a negotiating agenda. A number of influential voices called for such a round to be a "development round" and to seek greater balance in addressing issues of concern to developing countries.\textsuperscript{1} In the event, the November 1999 ministerial meeting in Seattle turned out to be a fiasco, failing to launch a round.

Domestic politics in the United States played a key role in the failure to attain consensus on a broad negotiating agenda, greatly reducing the willingness of the U.S. administration to agree to put items on the table that were opposed by domestic lobbies. Strong differences on the scope of agricultural liberalization between the European Union, on the one hand, and the United States and other agricultural exporters, on the other hand, were also important.\textsuperscript{2} Another major factor was the active and full-fledged participation by developing countries, many of which refused to accept the agenda being pushed by a number of high-income countries in some areas—most notably the United States on labor standards—and raised concerns about implementation problems associated with the Uruguay Round. Many also expressed general dissatisfaction concerning the process through which a negotiating agenda was being set. Small countries in particular perceived themselves to be left out completely, not having access to the forums where potential agenda-setting compromises were being crafted.

In the two years following the Seattle ministerial meeting, a great deal of effort at the WTO was focused on dealing with the implementation concerns of developing countries and on building the confidence of the smaller and poorer members in the trading system. Many of the implementation concerns deal with the trade and development issues surveyed herein. An important part of the confidence-building agenda has been proposals for higher-income countries to grant unrestricted market access to the least developed countries.

The Doha Development Agenda that emerged from the 2001 WTO ministerial meeting in Doha, Qatar, launched a broader set of negotiations. The Doha agenda gives great prominence to development concerns, reflecting proactive participation by developing countries in the process. Negotiations are to take place on market access for manufactures, dispute settlement, WTO rules, disciplines on regional integration, environment, and Trade Related Aspects of Intellectual Property Rights (TRIPS) (geographical indications). These negotiations will complement the ongoing negotiations on agriculture and services mandated by the Uruguay Round agreements.\textsuperscript{3} Negotiations may be launched on the four so-called Singapore issues—trade facilitation, transparency in government procurement, competition policy, and investment policy—at the 2003 WTO ministerial meeting if consensus exists on the modalities of such talks. The Doha agenda explicitly deals with key concerns of developing countries, including the implementation issues from the Uruguay Round, the need for technical cooperation and capacity building in developing countries, and market
access for the least developed countries. The meeting also dealt with the concerns of developing countries about intellectual property rights and public health.

This article summarizes some of the results emerging from a collaborative research and capacity-building project involving scholars in developing countries, international experts, and World Bank staff. The aim of the project is to generate both cross-country and country-specific analysis on the costs and benefits of further multilateral rule making and liberalization. The article focuses first on market access issues (the potential gains from further liberalization of trade in agriculture, manufacturing, and services) and then on topics that are of particular concern to developing countries—policies for investment and export development; rule making and implementation; labor, environment, and related standards issues; and participation of developing countries in the WTO. A message emerging from much of this research is that if the objective is economic development, the focus of attention should not be on efforts to graft a development dimension onto the trade agenda. Rather, the focus should be on making trade a more effective instrument of development. In many cases, that implies a need to strengthen significantly the supply side of the economy through investment in infrastructure and strengthening of trade-related institutions.

Agricultural Liberalization

Barriers to trade tend to be highest in agriculture and services. Average tariffs on agricultural imports are in the 15–20 percent range, with peaks for some commodities exceeding 100 percent in many countries, both developing and industrial. By contrast, average manufacturing barriers are quite low in Organisation for Economic Co-operation and Development (OECD) countries, but significantly higher in developing nations. However, certain manufacturers (such as clothing) continue to confront high tariffs in many high-income countries. Tariff barriers faced by developing countries on their exports of agricultural products are estimated to average 15.6 percent in rich countries and 20.1 percent in developing countries (table 1). The rates for the industrial countries are much lower. Estimates of the implied tariffs paid (constructed by multiplying the marginal tariffs levied on the relevant trade flows by the value of the corresponding trade flow) suggest that more than half of the levies charged on developing country exports are associated with their exports to industrial countries. These prevailing patterns of protection imply that many developing countries have a large stake in achieving significant agricultural liberalization. Hertel and others (1999) built a model of the world economy in 2005—at which time Uruguay Round commitments will have been fully phased in. They estimate that a 40 percent reduction in post–Uruguay Round agricultural tariffs and export subsidies will cause an increase in global real income of about US$60 billion per year. This figure increases...
by $10 billion if domestic support is also reduced by 40 percent, although the uncertainty in the degree to which such producer payments are linked to production decisions makes such analysis difficult (ABARE 1999).

Measured in dollar amounts, industrial countries capture the largest gains from liberalization, reflecting the reduction in the cost of agricultural support policies for OECD consumers. However, the percentage real income gains—reported in the first set of bars in figure 1—are largest in developing regions such as South Asia (other than India) and Southeast Asia (other than Indonesia). Virtually all developing regions, except the net food importing Other Middle East region, experience overall gains from these multilateral reductions in agricultural protection. The bulk of these gains derive from efficiency improvements generated in the developing countries themselves (the second set of bars in figure 1). This finding parallels the conclusion that in the Uruguay Round, the largest gains accrued to the countries that undertook the most liberalization of their own trade barriers (Martin and Winters 1996).

**Modalities for Agricultural Negotiations**

The precise outcome of agricultural negotiations will depend heavily on the specific modalities used. Anderson, Hoekman, and Strutt (2001) identify the priorities for further progress in agriculture as reducing import barriers, disciplining domestic support, and eliminating of export subsidies. They make clear that it would require substantial reductions in import barriers even to begin to approach parity with the treatment of manufactures trade. Nothing short of elimination of export subsidies would be sufficient to do so. They also note the large overlap between the agricultural reform agenda and the “second-generation” regulatory issues—competition
The Uruguay Round led to virtually complete tariffication of agricultural border protection. Unfortunately, the process that achieved this allowed substantial "dirty tariffication" in industrial countries—setting tariff bindings far above the tariff implied by prevailing nontariff barriers—and very high ceiling bindings in developing countries (Hathaway and Ingco 1995). The gap between applied tariff rates and tariff bindings in agriculture is particularly large in many developing countries (Abbott and Morse 1999), implying that substantial reductions in tariff bindings are required to achieve any liberalization of applied rates (Francois 1999). One approach to dealing with the gap between bound and applied rates is to make applied rates the basis for future negotiations, in effect requiring all countries to bind at applied rates. This is unlikely to be feasible because it would also create perverse incentives for countries to keep applied tariff rates high to conserve bargaining chips for future negotiations. A better approach would probably be to devise a formula that would impose the largest reductions in the highest tariff bindings. Josling and Rae (1999) suggest a "cocktail" approach that uses a formula to reduce the very highest tariffs, which are likely to contain a good deal of "water"; subjects moderate tariffs to a uniform percentage cut; and abolishes nuisance tariffs.
De Gorter (1999) shows that since the Uruguay Round was completed, OECD countries have become intensive users of tariff rate quotas (TRQs). TRQs can be important market access barriers. Under a TRQ, there is an out-of-quota tariff that applies to imports above a specified quota quantity. Volumes below the quota limit pay a lower in-quota tariff. Elbehri and others (1999) provide indicators of the extent to which TRQs are binding access constraints for a number of sensitive agricultural commodities, such as sugar, dairy, meats, and grains. They conclude that for the United States, the European Union, Canada, and Japan, imports exceeded the quota volume in 13 out of 16 cases. The allocation of associated quota rents is uneven, with many countries allocating a substantial share of the quota rents to exporters. By contrast, in the Philippines and the Republic of Korea, the two developing countries where Abbott and Morse (1999) find evidence of binding TRQ regimes, it appears that importers retain the quota rents.

Understanding the impact of TRQs is critical to predicting the outcome of attempts to liberalize trade. For example, reducing out-of-quota tariffs would increase imports only if the demand for imports exceeded the quota amount such that the out-of-quota tariff were operational. If imports were less than the quota level, reductions in out-of-quota tariffs would be ineffective. However, marginal expansion of the TRQ would be ineffective if imports were greater than the TRQ; the only effect would be to increase the volume of imports on which scarcity rents are earned. If imports were less than the TRQ, expanding the quota would also be ineffective. Only reductions in in-quota-tariffs would stimulate greater imports in this case. Thus, reductions in out-of-quota tariffs would be the most effective instrument for achieving market liberalization in most cases. However, it could be desirable to accompany such cuts with expansion of the quotas.7

Agricultural liberalization (especially movement toward eliminating export subsidies) may increase world prices of food products and thus have a negative effect on developing countries' net food imports. However, any such impact would be offset to some degree by the increase in domestic supply that higher prices would stimulate. Current policies result in large global price swings that are highly detrimental to developing countries, and farmers in many developing nations suffer from a significant antiagricultural policy bias. Even if the prices of imports rise, complementary reforms at home can make net food importers better off: they initially lose welfare by unnecessarily stimulating food imports and the price rise curtails that stimulus (Wang and Winters 2000). However, mechanisms are needed to ensure that any price-increasing effects of reforms do not reduce the real income/consumption of the poorest in society. Such social safety nets may not exist or may not function adequately in many countries. Multilateral trade liberalization generally takes a long time to negotiate and implement. This provides an opportunity for governments to develop and/or strengthen safety net programs and complementary policies to maintain the real incomes of the poorest in society.
It should be emphasized that the need for such mechanisms is quite general and not specific to agricultural reform. All types of policy reforms that are beneficial for the economy as a whole, as well as exogenous shocks of various kinds, may have detrimental consequences for the poor. Mechanisms to offset the negative impact of shocks should therefore be broad in scope and not conditional on changes in trade policies only.

Industrial Tariffs

There has been a sweeping change in the structure of international trade in the past two decades. In the mid-1960s, manufactured exports accounted for only around one-quarter of developing country exports. By the early 1980s, this share had risen to around one-third. Since then, growth has accelerated; as of the mid-1990s, the share was around three-quarters, and it is projected to go on rising (figure 2). Much of the increase in developing country exports during the past three decades has not followed a north-south pattern. The share of developing country exports going to other developing countries has risen sharply as the importance of developing countries in the world economy has risen and barriers to trade have declined in both de-

Figure 2. The Increasing Share of Manufactures in Developing Country Exports

Source: Hertel and Martin (2000).
veloping and industrial countries. Developing countries therefore have a strong interest in including industrial products in WTO negotiations. Although industrial countries impose low average tariffs on their imports of manufactured items, the average tariff on imports from developing countries is four times higher than average tariffs originating in the OECD (table 2). This is primarily because of the relatively high tariffs on such products as textiles and clothing. Estimates of the implied tariffs paid suggest that the barriers developing countries face in other developing countries account for more than 70 percent of the total tariffs levied on their industrial exports. This situation contrasts sharply with that in agriculture (see table 1).

A computable general equilibrium (CGE) analysis of the impact of a 40 percent cut in applied tariffs on manufactures by all countries suggests that global trade volume would expand by some $380 billion in 2005, or about 4.7 percent of projected merchandise and nonfactor service trade (Hertel and Martin 2001). This increase is reflected in almost all products, including nonmanufactures. The largest increase is for wearing apparel. Even after the phase-out of quotas agreed in the Uruguay Round, trade volume in this sector rises by a further 20 percent, reflecting the heavy tariff protection in high-income countries. Textiles and autos follow in importance. Figure 3 reports real income and efficiency gains as a share of 2005 income by region. The difference between these two variables reflects terms-of-trade effects. (If the real income gain exceeds the efficiency gain, then the terms-of-trade effect is positive; the opposite is also true.) Efficiency gains depend on the degree to which a country liberalizes its markets. Sharp tariff cuts give rise to increased access to cheaper imported goods and generate gains in consumption as well as improvements in the efficiency of use of domestic resources.

The largest efficiency gains (as a share of income) occur in developing economies, with countries or regions where tariffs are highest in the 2005 base gaining the most

<table>
<thead>
<tr>
<th>Importing region</th>
<th>High income</th>
<th>Developing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exporting region</td>
<td>High income</td>
<td>Developing</td>
</tr>
<tr>
<td>Import-weighted average tariffs (percent)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High income</td>
<td>0.8</td>
<td>10.9</td>
</tr>
<tr>
<td>Developing</td>
<td>3.4</td>
<td>12.8</td>
</tr>
<tr>
<td>World</td>
<td>1.5</td>
<td>11.5</td>
</tr>
<tr>
<td>Implied tariff paid (US$ billions)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High income</td>
<td>16</td>
<td>93</td>
</tr>
<tr>
<td>Developing</td>
<td>23</td>
<td>57</td>
</tr>
<tr>
<td>World</td>
<td>40</td>
<td>150</td>
</tr>
</tbody>
</table>

Source: Hertel and Martin (2000).
Figure 3. The Welfare Impact of a 40 Percent Cut in Manufactures Tariffs

![Graph showing the welfare impact of a 40 percent cut in manufactures tariffs across different regions.](image)

Source: Hertel and Martin (2000).

(China, Other South Asia, and India). China's greater gains relative to India (which is projected to have higher protection levels in 2005) are due to the fact that the manufacturing sector in China is larger and more trade oriented. Tariff cuts in the industrial economies of Japan, Western Europe, Australia/New Zealand, and North America generate almost no efficiency gains because tariffs are already extremely low. However, the bulk of the gains go to the developing countries, which are estimated to receive three-quarters of the total gains from liberalizing manufacturing trade.

These results suggest that there are strong economic and political economy reasons for developing countries to support the inclusion of industrial products in any multilateral round of negotiations. From a political perspective, industrial products make up a very large share of exports; frequently these products are produced by a relatively small number of producers that can provide active support for the politically difficult reforms required by a trade negotiation. From an economic perspective, the substantial static welfare gains outlined above are a good reason to support their inclusion, as are the potential dynamic gains associated with moving to a more outward-looking manufacturing sector.

The quantitative analysis of liberalization of trade in manufacturing and agriculture is highly stylized and simplified. The use of a uniform percentage cut in applied...
rates of protection provides only a rough guide to the potential benefits from a broad-based liberalization. In practice, the policy instruments on which negotiations focus are tariff bindings, which may be higher than applied rates. The actual outcome will depend heavily on the precise approach to liberalization used. Theory predicts that the gains are likely to be larger than those indicated if negotiators choose a top-down approach that reduces the variance of protection by more than a uniform cut. The gains will be smaller if less reduction in the variance of protection is achieved—that is, if politically sensitive tariff peaks are preserved. Further research is necessary to take these potential differences into account.

The research summarized uses static models and does not consider the dynamic effects of liberalization or the fact that many industries are imperfectly competitive. Current CGE techniques allow such factors to be incorporated into analyses. This will affect the magnitude of the predicted net gains, however, it will not affect the basic message that emerges: developing countries have a major stake in the attainment of further reductions in barriers to trade in both agriculture and manufacturing.

Services

In contrast with agricultural and industrial tariffs, it is much more difficult to employ numerical general equilibrium techniques to assess the potential gains from alternative liberalization options for trade in services. The required information on prevailing barriers to trade and investment simply does not exist. In the case of merchandise trade, the main barrier is the tariff. Although differences between tariff bindings and applied rates and accounting for preferential trade agreements and subsidies certainly complicate analysis, the prevailing policies are relatively straightforward to characterize. This is not the case with services. Frequently market access barriers are enforced "behind the border" and are embodied in regulations that control entry and/or operations, impose limitations on foreign equity holdings or nationality constraints, or require professionals to recertify as a condition for operating in a market. Because services often are not tradable, firms and/or providers have to move to the location of the buyer/consumer of a service (or vice versa). This implies that regulatory regimes pertaining to the temporary entry (visa restrictions or economic needs tests) or longer-term entry (foreign direct investment policies) of service suppliers and consumers must be considered in determining the overall policy stance of a country toward trade in services.

Ongoing work seeks to improve tariff-equivalent estimates of the effect of service policies and to use this information in CGE modeling (for example, Brown and Stern 2001). This research also involves efforts to construct openness indicators for modes of supply, especially foreign direct investment, and for specific sectors using qualitative assessments of the extent to which actual policies raise the costs of entry and/or
The staff of the Australian Productivity Commission has made a noteworthy attempt, identifying existing policies affecting foreign direct investment, assigning each a weight, and summing across weights to obtain an overall restrictiveness index. Their results suggest that across Asia-Pacific Economic Cooperation countries, communications, financial services, and transport are subject to the greatest barriers to foreign direct investment, reflecting the existence of ownership limits or an outright ban on foreign ownership. The most restrictive countries include Korea, Indonesia, Thailand, and China—all countries that appear to have restrictive service sectors using a variety of other measures (Francois and Hoekman 2000).

Developing countries have a large stake in enhancing the efficiency of domestic service providers and improving their ability to contest foreign service markets. Although the large OECD countries dominate global trade in services, developing countries dominate the list of countries that are most specialized in (dependent on) service exports as a source of foreign exchange. Often this reflects the importance of tourism and/or transportation services. But developing countries have also become large exporters of transactions processing, back-office services (Jamaica), and information and software development services (India). There is enormous potential to exploit recent and emerging technological developments—such as e-commerce—that facilitate cross-border trade in services and provide firms with incentives to slice up the value chain geographically.

Recent research suggests the emphasis in the next set of General Agreement on Trade in Services (GATS) negotiations should be on three issues: expanding the coverage of specific commitments; increasing the transparency of prevailing policies; and improving multilateral disciplines.9

**Expanding the Coverage of the GATS**

The sectoral coverage of specific commitments on national treatment and market access is limited for many countries. By one measure, high-income countries made commitments on only about half of all services, of which only one-half involved commitments of free access. That is, governments committed to imposing no restrictions on market access or national treatment for only 25 percent of all service activities. Developing countries made even fewer commitments. In the case of major developing countries, on average, free access commitments were made for only 15 percent of the service sector (Hoekman 1996). Subsequently, successful negotiations expanded the coverage of specific commitments for basic telecoms and financial services. These negotiations were important both for keeping momentum going and because the services involved are vital intermediate inputs. Despite the success in concluding these agreements, they have not led to a significant increase in the coverage of the GATS, as the Hoekman (1996) compilation included commitments made as of 1994 in both sectors.10
Thus, many governments have refrained from even binding the status quo. More commitments have tended to be made with respect to foreign direct investment (mode 3) than other modes. A number of countries' commitments favor infusions of foreign equity into existing firms over entry by new firms. As noted by Mattoo (2000), such protection of incumbents and/or existing market structures is difficult to rationalize and must be carefully monitored because it can easily result in a transfer of rents to foreign firms rather than a socially desirable increase in competition and lower prices/higher quality output.

A strong case can be made that the GATS should cover all services. There is no rationale for excluding certain sectors or modes of supply from the national treatment and market access disciplines, given that the GATS allows for derogation of both principles. One way of moving toward this would be to apply a formula approach to expanded coverage in the next round of negotiations, setting minimum coverage targets for GATS members to be attained by a specified date (which may vary depending on per capita income level to allow for a transition period). This could include agreement that a specified share of all commitments must involve full binding of status quo policies. A more ambitious approach would be to seek agreement on a deadline for full coverage to be reached. This should include politically sensitive but economically important sectors, such as air and maritime transport. Weak (managed) competition on many international transport routes—both sea and air—imposes large costs on developing countries (Francois and Wooton 2001).

From a market access perspective, developing countries have a great interest in ensuring that substantially more commitments are made on mode 4—supply of services through temporary entry by service providers—and cross-border trade (mode 1), which is of great importance for e-commerce. Although the tradability of services has been increasing rapidly due to technological developments, in many cases it remains imperative that service providers be able to work on the premises of their clients. Currently, virtually all GATS members maintain restrictions on such trade, usually through the application of economic needs tests and other requirements imposed on requests for entry visas. Achieving concrete agreements to liberalize access to service markets through mode 4 would go far toward making the GATS a more balanced market access liberalization instrument. The difficulties of making progress in this area are obvious, as there is vigorous opposition on the part of unions and sectoral interest groups in industrial countries. However, there are also industries in the importing countries that have a large stake in being able to employ foreign service providers. The software industry in the United States is just one prominent example; it was the major force behind a temporary expansion in the number of H-1B visas allocated by the U.S. government for temporary movement of service professionals (Chandra 1999). The challenge confronting developing countries is to build coalitions with domestic industries in large markets to achieve a permanent increase in the number of visas for service providers that may be issued (that is, a quota expansion).
Toward Greater Transparency and Better Rules

A major weakness of the GATS is that it does not force members to come clean regarding the measures that are used to restrict the ability of foreigners to contest domestic service markets. It is very unlikely that negotiators will be willing to reopen the issue of scheduling commitments, and efforts to adopt a "negative list" approach are likely to be counterproductive. But a negative list-reporting exercise for transparency purposes deserves serious consideration. This would involve all members reporting information on all measures that affect market access and national treatment in all services sectors. It would result in a comprehensive database of status quo policies and provide a focal point for reform efforts.

Market access and national treatment only apply on a sector-mode basis under the GATS and are subject to exceptions if governments schedule them. The ability to make commitments by modes of supply can distort the incentives to use alternative modes or to make a commitment in one mode irrelevant because providers need to have access to more than one mode. One way to reduce potential inconsistencies is to require one-to-one mappings between commitments relating to different modes (non-discrimination across modes). Such modal neutrality is an objective worth pursuing because, as the literature often emphasizes, trade and investment have increasingly become complementary. It is also frequently noted that it will become increasingly difficult to maintain a clear distinction between trade in goods and trade in services, as technology may give producers the choice of delivering their products in tangible or disembodied (digitized) form.

Ideally, scheduling of liberalization commitments should shift from the sectoral (specific) to the horizontal (general). This would allow negotiating efforts to center more on developing disciplines that make sense from a long-term growth and economic development perspective. In general, these effects are likely to focus on safeguarding the contestability of markets while maintaining national sovereignty to regulate activities to attain health, safety, and prudential and related objectives. In this perspective, it may be useful to consider generalizing the appropriate parts of the so-called Reference Paper for telecoms to other infrastructure network services to establish a horizontal set of procompetitive disciplines. Useful work could also strengthen the reach of the most favored nation principle and extend it to the area of standards and certification to ensure that (mutual) recognition agreements minimize discrimination.

The Uruguay Round left open a number of outstanding issues, including whether to adopt rules on procurement, subsidies and safeguards. Evenett and Hoekman (2000) argue GATS-specific disciplines on procurement should not be sought, as any disciplines should cover both goods and services. Moreover, what really matters for foreign firms is to have access to service procurement markets, and frequently this can only be achieved if they have a commercial presence in a country. In such cases,
the binding constraint is not a policy of discrimination, but the ability of foreign firms to establish (enter). This suggests the focus of attention should be on expanding market access commitments under the GATS.

Multilateral disciplines on subsidies might help avoid policies that are mutually destructive from the viewpoint of developing countries—for example, seeking to attract foreign direct investment through the use of incentives. Subsidies are an important source of distortions in OECD markets for some services (for example, transport). However, to be effective in disciplining the use of firm-specific fiscal incentives, subsidy rules will have to be quite comprehensive to ensure that countries cannot sidestep them through the use of alternative policies. Here again, the same conclusion arises as for procurement—any disciplines should be general, not sector specific.

One area where a more compelling case can be made for service-specific rules concerns safeguards. The limited nature of liberalization commitments in mode 4—temporary movement of service providers—may in part be due to the nonexistence of safeguard instruments. Given that this mode of supply is of major interest to developing countries, a safeguard instrument could be tied to mode 4 liberalization commitments. These safeguards could be explicitly aimed at providing country governments with an insurance mechanism that can be invoked if liberalization should have unexpected detrimental impacts on their societies (Hoekman 2000).

Industrial, Investment, and Export Development Policies

Many developing countries pursue a variety of programs for industrial development, agricultural extension and export promotion. These may involve assistance with adopting new technologies, penetrating new markets, and general advertising campaigns that aim at “selling” the country and enhancing the visibility of export products. During the 1990s, an increasing number of countries also implemented so-called matching grant schemes that subsidize a proportion of the cost of improving production facilities, obtaining ISO 9000 certification of management systems, and exploring new export markets.

Many developing country governments have expressed concerns regarding their ability to pursue industrial policies without running afoul of WTO rules and disciplines. Multilateral rules on subsidies and related industrial policies were tightened substantially for developing countries in the Uruguay Round. Export subsidies became prohibited (except for the least developed countries) and trade-related investment measures (TRIMs)—for example, local content schemes that imply discrimination against imports—were outlawed. The TRIMS agreement prohibits both mandatory measures and the more common policies with which compliance is necessary to obtain an advantage (such as a tax concession, import duty exemption, or subsidy). Matching grant schemes—an instrument often used in developing countries—may
be regarded as export subsidies insofar as the provision of the grant element is made conditional on exports.

In almost all these cases, the rationale for activist policies is the existence of distortions created by market failures or other government policies. It is well known that if the source of the problem is policy induced, the case for a subsidy is only a second-best one (Bora, Lloyd, and Pangestu 2000). A good case can be made that WTO disciplines on export subsidies and TRIMS are generally likely to be beneficial. Export subsidies are distortionary for the world as a whole and can easily be captured by private interests seeking rents. In practice, they are very difficult to justify on the basis of distortions or market failure. In contrast, production subsidies (and taxes) can be an efficient way to offset externalities, and are allowed under WTO rules (although the effect of direct subsidies may be countervailed by importing countries if they can be shown to materially injure domestic competitors). The adoption of a “green box” approach toward subsidies in the Uruguay Round allows substantial freedom for governments to use subsidy instruments in cases where this is called for on economic grounds and reduces the scope for other countries to second-guess the motivation underlying the use of such instruments. The types of subsidies that are defined to be acceptable (but in principle countervailable) are subsidies where economic theory suggests intervention can help offset market failures (that is, support for research and development and certain types of agricultural input subsidies for developing countries). A good case can be made that disciplines should not be tightened in this area, that is, the range of “acceptable” subsidies should not be tightened.

Many developing countries have resisted the requirement to abolish TRIMS, arguing that they need such instruments to encourage industrialization. In the Uruguay Round, it was decided that the TRIMS agreement would be reviewed in 2000, at which time it would be complemented by provisions on competition and investment policy (Low and Subramanian 1996). Because the five-year review deadline coincided with the built-in negotiating mandate on such topics as services and agriculture, TRIMS became a potential negotiating topic for a millennium round. Some countries argued in favor of wrapping the TRIMS discussion into a more general WTO negotiation on foreign direct investment policies. This position can be motivated on the basis that TRIMS are just part of the relevant policy landscape: investment measures are often general, not trade related. Many countries apply licensing and approval regimes and impose related red-tape costs on foreign investors. They may also prohibit entry through foreign direct investment altogether or impose equity ownership restrictions. The TRIMS agreement does not apply to such nontrade-related policies, nor does it affect service industries.

Neither the economics nor the political economy of seeking WTO disciplines on foreign direct investment is straightforward (Markusen 2001; Moran 1998). Restrictive policies may reflect welfare-enhancing attempts to shift foreign profits to the domestic economy or welfare-reducing rent-seeking activities by bureaucrats and...
their constituents. Sometimes the effect of policies is simply to waste real resources (so-called frictional costs). A key question concerns the value added offered by multilateral rules in this area to developing countries, given that much can (and should) be done through unilateral reform to attract investment. This is an underresearched topic. Hoekman and Saggi (2000) argue that there are potential payoffs but that these may be difficult to realize. They also argue that in the area where foreign direct investment matters most as a mechanism to contest markets—services—a WTO instrument already exists. The GATS extends to foreign direct investment policies as countries can make specific market access and national treatment commitments for this mode of supply for any or all services. Thus, a lot can already be achieved using existing structures.

As far as more traditional TRIMS are concerned, the available empirical evidence suggests that local content and related policies are generally ineffective or costly to the economy. Often they do not achieve the desired backward and forward linkages, and they encourage inefficient foreign entry and create potential problems for future liberalization as those who enter lobby against a change in regime (Moran 1998; Pursell 2001). The major policy question is implementing the agreement—that is, phasing out illegal TRIMS. Governments may be constrained in eliminating costly status quo TRIMS because protected industries are politically powerful. Although the Uruguay Round agreement incorporated transition periods, some countries may need extensions of transition periods as well as assistance in designing effective and credible transition paths (Hoekman 2001).

**Achieving Balance: Rule Making and Implementation**

Resource constraints impede the ability of many developing countries to identify and defend their interests in multilateral negotiations and to participate in WTO activities (Blackhurst, Lyakurwa, and Oyejide 2000). Even if countries are able to influence the set of subjects to be negotiated so that notional symmetry prevails in terms of defining the agenda, outcomes can easily be asymmetric, reflecting differences in negotiating power. Under the GATT, this asymmetry was exemplified by the exclusion of agriculture and textiles and clothing from many multilateral disciplines and the use of various instruments of contingent protection by OECD countries, including some that were illegal under the GATT—for example, voluntary export restraint agreements. In the Uruguay Round, negotiating power asymmetries were illustrated by the implicit bargain in which the Multi-Fibre Arrangement, voluntary export restraints, and some agricultural barriers of concern to developing countries were abolished in return for introduction of a TRIPS agreement proposed by OECD countries. This involved a “payment” for the elimination of practices that violated the spirit (if not the letter) of the GATT.
Asymmetry under the Uruguay Round was also reflected in the fact that developing countries became subject to a large number of disciplines in areas that were voluntary under the GATT—including rules on customs valuation, antidumping, subsidies, technical product standards, and sanitary and phytosanitary measures. In these areas, it is difficult (if not impossible) to trade concessions. Negotiators focused instead on the identification of specific rules that should be adopted by all. In practice the norms chosen were those that were (are) applied in industrial countries (Finger and Schuler 2000). In contrast to traditional trade liberalization, a “one-size-fits-all” approach may not be optimal. Nonetheless, one size fits all was a central pillar of the Uruguay Round—developing countries were only granted additional time in which to implement obligations. In many areas, these periods were five years and expired at the end of 1999. Many developing countries remain far from compliance, making extension of time limits for implementation—and possibly renegotiation or granting of waivers—priority issues for them. Extensions of the time limits for implementation of the TRIMS agreement were finally agreed in August 2001 in the run-up to the Doha ministerial conference.

Implementation became an issue partly because the costs associated with complying with some WTO agreements can be significant. As noted by Finger and Schuler (2000), such costs can easily exceed the entire development budget of a least developed country. This is not because the rules themselves are necessarily onerous but because of the ancillary investments needed to make the agreements work successfully. It is not at all clear from a development perspective that the resources required for implementation of WTO agreements, whatever the amount might be, would not be better used to build schools or improve infrastructure. Ensuring that WTO agreements are conducive to (consistent with) attainment of development objectives should therefore be a major objective of the next round. A necessary condition for this is full ownership of the WTO agreements, which, in turn, is predicated on participation in the development of the rules.

At the time of the Uruguay Round, the negotiators could draw on only limited developing country experience in the “new areas.” Poor countries have, for instance, yet to attempt to create intellectual property regimes that make traditional knowledge or cultural products into negotiable and defensible assets. Nor have they identified the alternative options that can be used to upgrade and enforce national product, health, and safety standards or to regulate service sectors that are subject to market failures. In many of these areas, the trial and error experience—the assessments of the real-world impacts of alternative policy options—that can inform the effective incorporation of the development dimension into multilateral rules does not exist. It seems desirable that future rule-making attempts should focus on the goals to be achieved, rather than on specifying the approaches to be used. WTO rules should allow for experimentation and learning, and should be complemented by adequate assistance to help countries develop appropriate regulatory instruments.
Intellectual Property Rights and Competition Policy

The Uruguay Round TRIPS agreement obliged all WTO members to enforce intellectual property rights (IPRs), although with transition periods for developing countries. Whether developing countries will gain from stronger protection of IPRs is a matter of vigorous debate. Those in favor argue that dynamic benefits—operating through foreign direct investment, technology transfers and licensing, and innovation within and for the domestic market—will more than offset any static losses. Those against note that dynamic benefits are uncertain, although on balance the short-run impact of the TRIPS regime—which will operate through higher prices, lower domestic output, and more imports—is likely to cause a transfer of income from poor to rich countries, with at best marginal impacts on economic efficiency, resulting in net transfers to firms in high-income countries.13

The scale of the transfer depends on the market structures that prevail and the closeness of available substitutes. For example, estimates by Maskus (2000a) suggest that because Lebanon is a net importer of pharmaceutical products and technologies and currently has relatively little inventive capability in the sector, the static impact of stronger patents is likely to be negative, increasing average prices by some 10 percent and resulting in lower output and fewer firms. But in other sectors where Lebanon produces IPR-sensitive products (such as printing and publishing, music, and film/video), stronger protection would be beneficial. The static net impact is therefore unclear. Dynamic effects are even more uncertain.

An important factor determining the impact of the TRIPS agreement is the ability of governments to intervene to offset socially detrimental outcomes. The agreement has a number of provisions that authorize the use of policy measures against abuses of IPRs. Competition law has an important role to play in this connection. For example, because rights holders will frequently use their IPRs to segment markets, developing countries may have a strong interest in applying an international exhaustion rule (allowing parallel imports). This would imply that domestic buyers could purchase patented and branded products wherever they find the most favorable prices. This is fully compatible with the TRIPS agreement, although both the European Union and the United States are active proponents of a national/regional approach to exhaustion. Whether to adopt an international exhaustion rule should be a matter for national authorities to decide independently. Of course, this is just one aspect of the interrelation among IPR and competition regimes, but it illustrates that a one-size-fits-all rule should be avoided (Maskus 2000b).

Developing countries have an interest in adopting strong competition policies, the main pillar of which should be a liberal trade and foreign direct investment policy stance. Competition law is required to ensure markets are contestable, especially in nontradable sectors. Antitrust legislation may also be required to maximize the benefits (or minimize the costs) of certain WTO agreements, the TRIPS agreement being one.
example. What matters from the point of view of WTO negotiations is whether attempts to agree on binding multilateral disciplines would be beneficial for developing countries. In principle, a commitment to more thorough policing by industrial countries of their own exporters, of global strategic alliances (for example, in telecoms), and import-competing industries' use of instruments of contingent protection could have very general benefits for developing countries (Maskus and Lahouel 2000). Hoekman and Holmes (1999) argue that developing countries should use the occasion of a trade round to put their interests on the table, recognizing that the quid pro quo they can expect will depend importantly on what they are willing to offer. Seeking modifications in antidumping laws and commitments by OECD competition authorities to provide assistance to developing country competition authorities are examples of the type of quid pro quo that could be sought. However, realism suggests that the primary focus should be on the design of appropriate national policies. After more experience has been obtained with the design and implementation of national regulations, countries will be better able to judge the appropriate type of multilateral agreement in this area.

Labor, Environmental, and Related Standards Issues

The collapse of the Seattle ministerial meeting in 1999 was in part due to differences in views between industrial and developing countries on whether the WTO should regulate and enforce labor and environmental standards. Determining the extent to which multilateral disciplines should extend "behind the border" and cover more areas of domestic regulation is one of the key challenges facing WTO members. Although the case for domestic regulation to address market failures or pursue non-economic objectives is indisputable, the case for international harmonization is not. Efforts to impose standards on all WTO members and enforce them via the threat of trade sanctions threaten to embroil the WTO in issues in which it has neither the technical ability nor the political legitimacy to act effectively (Rollo and Winters 2000).

The TRIPS agreement reflects the first clear attempt at harmonizing domestic policy and setting certain minimum absolute standards of IPR protection to which all members must adhere. The GATT was largely an instrument of negative integration—it revolved around agreements not to do certain things (discriminate in trade policy or raise tariffs above bound levels). The TRIPS agreement is the first major example at the multilateral level of what has been called positive integration (Tinbergen 1954). One result is that for the first time, failure to implement a certain type of regulatory regime (in this case to prevent the use of production processes by domestic firms that violate IPRS) can give rise to dispute settlement and possibly trade sanctions.

A question that has arisen concerns the implications of TRIPS on the scope of the WTO. If IPRS can be brought into the WTO, why not other areas of domestic regulation.
as well? In a useful approach to answering this question, Maskus (2000b) identifies a series of screens or criteria that can be used to determine if there is a good economic case for bringing a regulatory area under the WTO. These criteria are that the issue: (1) is strongly trade related; (2) gives rise to international externalities; (3) is associated with policy coordination failures that can be addressed effectively through WTO dispute settlement; and (4) has the potential to strengthen the trading system. In the case of labor standards, Maskus surveys the literature and concludes that, first, there is little credible evidence that deficient enforcement of core labor standards has a major impact on trade. Second, he finds that both the theoretical and empirical basis for arguing that lax labor standards in developing countries suppresses wages of low-skilled workers in OECD nations is very weak. Third, he argues that WTO-type enforcement (relying on trade sanctions) will worsen labor outcomes. Finally, he points out that linkages between core labor standards and existing WTO disciplines are nonexistent (Maskus 1997, 2000b).

There is considerable international agreement that certain core labor rights should be globally recognized and protected. Development and encouragement of implementation of such rights is the task of the International Labour Organisation (ILO). One of the principal arguments for inclusion of labor standards in the WTO is to provide an enforcement mechanism for ILO conventions. However, trade remedies to enforce labor standards would worsen the problems they were aimed to solve (by forcing workers in targeted countries into informal or illegal activities) and burden the trading system (by increasing the likelihood of controversial disputes). Account should also be taken of the nonnegligible danger that such instruments would be captured by protectionist interests seeking to limit imports from labor-abundant developing countries.

Attainment of core labor standards can be pursued more effectively through instruments that are targeted directly at improving outcomes. For example, efforts could be made to improve the quality of and access to primary education for poor children to reduce child labor exploitation, for example, through programs to subsidize the purchase of school supplies, provide transportation, and reduce the costs of schooling. It is important to avoid a confrontational approach to this issue and to pursue collaborative solutions that help developing countries improve labor standards. Attempts to force countries to adopt standards that do not reflect national preferences and conditions should be rejected (Bhagwati and Srinivasan 1996). Gains from trade arise in large part because countries differ, and national social or environmental policies are simply one determinant of these differences. They do not constitute barriers to trade or give rise to unfair trade.\(^{14}\)

In the case of the environment, the argument for adopting substantive rules in the WTO is also weak. However, an important difference compared with labor standards is that trade policies may have adverse environmental consequences. An example would be subsidies and trade protection for the coal industry, which may discourage
a shift to cleaner-burning fuels. There may also be potentially important cross-border environmental spillovers, which give rise to a need for international cooperation. This has been reflected in various multilateral environmental agreements. Such agreements might include trade sanctions as enforcement instruments. Sanctions can give rise to WTO dispute settlements if a nonsignatory is targeted by signatories; therefore, the WTO membership may need to devise procedures that set out the conditions under which they would permit such sanctions. Rollo and Winters (2000) suggest the following as necessary conditions to ensure legitimacy and reduce the chances of protectionist capture:

- Sanctions must be genuinely a last resort.
- Decisions to sanction a transgressor should be collective, with a very substantial majority of WTO membership.
- Provision must be made for frequent review to determine if sanctions can be removed.
- Sanctions should be applied by all signatories of the multilateral environmental agreement.
- WTO members should have no discretion about what products to restrict.

Developing Country Participation and Market Access

Many of the contributions that have emerged from the ongoing research on which we have drawn emphasize that participation constraints are a general problem, in particular for the least developed countries; that lack of information and limited cross-country experience greatly constrain the ability of countries to exploit the "wiggle room" that is embodied in many WTO agreements; that there is a need to ensure that governments have the scope to pursue policies in a manner that makes sense from a development perspective; that fulfillment of offers of financial and technical assistance by high-income countries have proven to be disappointing; and that provisions requiring such countries to take into account the interests of developing countries have proven to be meaningless (Blackhurst, Lyakurwa, and Oyejide 2000; Finger and Schuler 2000; Michalopoulos 1999; Hoekman 2001).

The end result has been limited developing country ownership of many agreements, and a general suspicion of the WTO in large segments of civil society. This can be remedied only if the next round results in agreements that are supported by strong constituencies in developing countries. The preconditions for achieving greater balance appear to be there—developing countries have demonstrated a willingness to participate actively and constructively in the WTO. This was reflected in the run-up to the Seattle ministerial and the role played in the process of defining a negotiating agenda. The inability (unwillingness) of the industrial countries to accept the neces-
sary compromises helped scuttle the talks, but arguably helped set the stage for a more balanced agenda to be crafted at Doha. That said, to paraphrase Wang and Winters (2000), much will have to be done to “put Humpty Dumpty back together again” and repair the damage manifested in Seattle. Many developing country delegations left Seattle frustrated with WTO procedures that excluded them from deliberations. This compounded prior resentment about the outcome of the Uruguay Round and concerns about the asymmetric burden of implementing WTO obligations.

The process in Doha was much more inclusive—all meetings were open to all delegations, and chairs of working groups reported regularly to meetings of heads of delegations. The exclusionary “green room” approach to hammering out deals was largely avoided. Although successful in generating a final agreement, the Doha process is not something that can be applied on a day-to-day basis in Geneva.

In contrast to past negotiating rounds, modalities should be developed to ensure that agreements are consistent with development strategies and priorities, thereby avoiding the types of implementation-related tensions that arose after the Uruguay Round. In areas involving positive integration measures, this requires both greater flexibility in terms of time frame and modalities of implementation and binding commitments by high-income countries to provide the financial and technical assistance that is required to assist countries in implementing WTO obligations in ways that support the economic development process. Political will and leadership are required to refrain from overburdening the institution by expanding its reach into areas that are not trade related, where it has no expertise, or where it is clear that interests among members diverge too much for agreement to be feasible.

One response to the concerns of the poorest countries has been a number of initiatives for richer countries to grant unilaterally improved market access for the least developed countries. Although such preferential market access has well-known disadvantages, including the fact that it tends to divert export opportunities away from other developing countries, it could potentially play a catalytic role in stimulating exports from the poorest countries. A careful examination of tariff peaks in the Quad (Canada, the European Union, Japan, and the United States) countries finds that, though average tariffs are low, tariff peaks and tariff escalation have a disproportionate negative impact on the exports of the least developed countries (Hoekman, Ng, and Olarreaga 2001). A related study by Ianchovichina, Mattoo, and Olarreaga (2001) estimates that opening Quad markets to Sub-Saharan African countries would raise their real incomes by $1.8 billion. Improved market access by itself is not enough. Such unilateral efforts need to be complemented by supply-side reforms to strengthen the export response (World Bank 2001a, 2001b).

There are good reasons to believe that the agenda that was adopted at the Doha ministerial meeting is a good one from a development perspective. It is consistent with the conclusions of research revealing that there is still a large market access agenda and that dealing with this matters most from a development perspective (Finger and
Research also suggests that much greater care is required to determine the development relevance and payoffs of extending the WTO into domestic regulatory areas (Finger and Schuler 2000; Hoekman and Kostecki 2001). The work program that was adopted for investment, competition policy, trade facilitation, and government procurement provides countries with an additional two years to determine what might make sense to discuss in the WTO. Furthermore, there is a strong emphasis on the need for technical assistance and capacity building for developing countries.

Concluding Remarks

This article has surveyed a range of recent research regarding the interests of developing countries going into a new round of WTO negotiations. It is impossible to do full justice to the complexity of the issues that arise in many of the areas that could figure on the WTO negotiating agenda. Accordingly, readers are referred to the papers provided in the references for a more thorough treatment of these issues.\textsuperscript{16} It is also impossible to generalize regarding the interests of developing countries. Nations are very diverse, reflecting differences in per capita incomes, initial conditions, and endowments.

Since the late 1970s, assessments of trade rounds have relied heavily on CGE modeling techniques. These allow simulations of the economy-wide effects of policy changes. Such models are particularly well suited to assessing the impact across industries and countries of reductions in tariffs, quotas, and subsidies. We have summarized some key findings from one such model with respect to agricultural and manufacturing liberalization in the wake of the Uruguay Round. We conclude that developing countries are likely to gain relatively more than rich countries from further liberalization. However, CGE models are much less amenable to assessing the implications of rule making, and, as tariffs have fallen, this aspect of the negotiations has become increasingly prominent. It is therefore very difficult (if not impossible) to assess quantitatively the overall impact of multilateral trade talks, either ex ante or ex post. Many dimensions of WTO agreements simply cannot be quantified. Other instruments—ranging from the use of economic theory to cross-country econometric analyses and case studies—must be used to assess the likely impact of rules on regulatory policies.

It is important to separate initiatives that aim at directly reducing barriers to trade from initiatives to further expand and deepen WTO provisions pertaining to regulatory regimes. The available research strongly suggests that the potential benefits from vigorous pursuit of a market access agenda are significant. This conclusion spans agriculture, trade in manufactured items, and services. Future efforts in the WTO should give priority to reducing barriers to trade that, in the case of services, include policies that restrict the ability of foreign firms to contest markets through a variety of modes. Developing countries have a strong interest in taking up the banner of
manufacturing tariff cuts, given that manufactures account for some three-quarters of their merchandise exports. Efforts to further discipline the ability of governments to abuse instruments of contingent protection minimize the trade-restricting impact of product standards (especially sanitary and phytosanitary measures) and facilitate trade more generally. Although not discussed in this article due to space constraints, these efforts are also of great importance (see UNCTAD 1999; Finger, Ng, and Sonam 2000; Messerlin and Zarrouk 2000).

A good case can be made for greater scrutiny of attempts to introduce substantive rules on domestic regulatory and legal regimes that imply harmonization to rich-country standards. Efforts to agree on procedural disciplines that increase transparency of policies and measures that focus on achieving objectives, rather than specifying approaches, are less likely to create implementation problems.

Greater efforts to consider the development relevance and impact of proposed rules will help create a trading environment that allows developing countries to help themselves by specializing in sectors where they have a comparative advantage. Doubts remain about the introduction of substantive disciplines in the WTO on domestic regulatory regimes if this entails harmonization to OECD norms that have not been determined to be in the interests of developing countries. However, such doubts certainly do not extend to the bread and butter of the multilateral trading system—the progressive liberalization of barriers to trade in goods and services on a nondiscriminatory basis. This is an area where there still remains much to be done, and where traditional GATT negotiating modalities can be an effective mechanism to overcome resistance to reform in both developing and industrial countries.

Notes

Thomas W. Hertel is with the Department of Agricultural Economics at Purdue University; his e-mail address is hertel@agecon.purdue.edu. Bernard M. Hoekman is with the Development Research Group at the World Bank and the Centre for Economic Policy Research in London; his e-mail address is bhoekman@worldbank.org. Will Martin is with the Development Research Group at the World Bank; his e-mail is wmartin1@worldbank.org. Output from this project can be downloaded from the Web, at http://www.worldbank.org/trade. Earlier versions of this article were presented at the annual meetings of the Latin American Economic Association (October 20–23, 1999), the American Economic Association (January 7–9, 2000), and the Annual Bank Conference on Development Economics (April 19, 2000). Parts of the article draw on collaborative work with Kym Anderson, Mike Finger, Joe Francois, and Aaditya Mattoo. The authors are grateful to them and to Philip English, Costas Michalopoulos, and Silvia Saborio for helpful comments and suggestions. The section on agricultural liberalization draws on the results of a parallel World Bank project led by Merlinda Ingco. The project is supported by a generous grant from the Department for International Development, United Kingdom; the governments of Italy and the Netherlands; Société Generale de Surveillance; the World Bank Institute; and the World Bank’s Research Support Budget.

1. For example, Clare Short, U.K. minister of Development; Mike Moore, director-general of the WTO; and Joseph Stiglitz, former chief economist, World Bank (Stiglitz 2000).

3. In the case of agriculture, extensive discussions in Doha centered on the objectives of the talks, in particular the insistence of exporters that elimination of export subsidies be an objective. However, this was in the context of an already existing mandate for negotiations.

4. In addition to researchers based in national think tanks, the project draws on the work of a number of research networks, including the Latin American Trade Network; the Economic Research Forum for the Arab Countries, Iran, and Turkey; the African Economic Research Consortium; the Coordinated African Program of Assistance on Services; and the Trade Policy Forum of the Pacific Economic Co-operation Council.

5. These estimates exclude the effects of preferences on tariffs paid, which clearly results in some over-estimation of these barriers. However, Hoekman, Ng, and Olarreaga (2001:1) find that the beneficial effects of these preferences tend to be diminished by the exclusion of "sensitive" products.

6. A number of studies of complete, global liberalization of post–Uruguay Round trade barriers have been undertaken. Anderson and others (forthcoming) analyze the benefits of trade reform by and to industrial and developing countries and review the range of estimates. The 40 percent liberalization scenarios were chosen as a degree of liberalization likely to be feasible in a round of negotiations.

7. Elbehri and others (1999) find that reducing by one-third the over-quota tariffs on sugar imported into the United States and the European Union results in losses for almost half of the exporting countries, a consequence of the reductions in quota rents. When the out-of-quota tariff reduction is paired with a 50 percent increase in the TRQs, most countries experience gains, and the remaining losses fall to negligible levels. Thus, there may be some grounds for liberalizing on both price and quantity margins simultaneously to secure acceptance of the overall liberalization program by importing and exporting nations alike.

8. See Warren and Findlay (2000) for an excellent survey of recent work. Stern (2001) collects a number of recent country studies on services by developing country analysts.

9. For a comprehensive discussion, see Mattoo (2001) and the contributions in Sauvé and Stern (2000).

10. However, the quality of the commitments, especially in basic telecoms, improved substantially. Of particular importance was the adoption of the so-called Reference Paper on regulatory principles. This is discussed further later.

11. See Chadha (2001), Chandra (1999), and Mattoo (2000) for more detailed analyses and proposals in the area of trade in services.

12. Implementation of these agreements did not cost high-income countries anything, given that the rules basically codified existing practices in these countries.


14. The focus here should be on policies, not outcomes. Thus, industries may shift production facilities abroad because environmental policies are less stringent in the host location, but as long as policies are appropriate to local conditions and reflect national (or local) preferences, this is desirable.

15. Hoekman (2001) suggests creation of a mechanism to allow development concerns and capacity constraints to be taken into account when defining implementation horizons for WTO agreements.

16. Many of these articles are available online at http://www.worldbank.org/trade.

References

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


Introduce Your Library to
THE WORLD BANK
RESEARCH OBSERVER

EDITOR: Shantayanan Devarajan, The World Bank
WBRO is intended for anyone who has a professional interest in development. Articles debating such key issues as development economics and developmental policy are accessible to non-specialist readers. For more information, visit our website at:

www.wbro.oupjournals.org

If you feel that a subscription would benefit students and colleagues, photocopy this page, fill in the name and address, and Oxford University Press will send a FREE sample copy of The World Bank Research Observer without obligation!

Institution:

Librarian:

Address:

City: State: Zip:

Country: 0104intlib/inwbro

Oxford University Press
Great Clarendon Street, Oxford OX2 6DP, UK
Tel: +44 (0) 1865 267907 Fax: +44 (0) 1865 267485
E-mail: jnl.orders@oup.co.uk
Journals in Economics from Oxford University Press

Journal of International Economic Law
Dedicated to encouraging thoughtful and scholarly attention to the relation of law to international economic activity.
www.jiel.oupjournals.org

Contributions to Political Economy
A forum for academic discussion of original ideas and arguments drawn from lines of thought associated with the works of classical political economists Marx, Keynes, and Sraffa.
www.cpe.oupjournals.org

Contemporary Economic Policy
A journal of Western Economic Association International, CEP publishes quality research and analysis on policy issues of widespread concern.
www.cep.oupjournals.org

Oxford Economic Papers
A general economics journal, OEP publishes refereed papers in economic theory, development and history, as well as applied economics and econometrics.
www.oep.oupjournals.org

Economic Inquiry
A journal of Western Economic Association International, this peer-reviewed, general economics journal publishes original research from top scholars in the field.
www.ei.oupjournals.org

The Review of Financial Studies
Provides significant new research in financial economics, striving to establish a balance between theoretical and empirical studies.
www.rfs.oupjournals.org

Contact Oxford University Press

In the Americas Elsewhere
Oxford University Press Oxford University Press
2001 Evans Road, Cary, NC 27513 Great Clarendon Street, Oxford, OX2 6DP
Tel: (919) 677-0977 Fax: (919) 677-1714 Tel: +44 (0) 1865 267907 Fax: +44 (0) 1865 267485
E-mail: jnl-orders@oup-usa.org E-mail: jnl.orders@oup.co.uk