The Evolution of Poverty and Inequality in Indian Villages
Rajshri Jayaraman and Peter Lanjouw

Appraising Workfare
Martin Ravallion

Public Social Spending in Africa: Do the Poor Benefit?
Florencia Castro-Leal, Julia Dayton, Lionel Demery,
and Kalpana Mehra

Rethinking the Causes of Deforestation: Lessons
from Economic Models
Arild Angelsen and David Kaimowitz

Formal and Informal Markets for Water:
Institutions, Performance, and Constraints
K. William Easter, Mark W. Rosegrant, and Ariel Dinar

Judicial Reform and Economic Development:
A Survey of the Issues
Richard E. Messick

The Costs and Benefits of Regulation:
Implications for Developing Countries
J. Luis Guasch and Robert W. Hahn
The Evolution of Poverty and Inequality in Indian Villages

Rajshri Jayaraman • Peter Lanjouw

This paper examines the evolution of poverty and inequality in rural India by reviewing longitudinal village studies. It explores the main forces of economic change—agricultural intensification, changing land relations, and occupational diversification—from a wide range of disciplinary perspectives, and it considers the roles of various institutions as conduits of change. Although most village studies support the survey-based judgment that rural poverty declined in India during the 1970s and 1980s, they find that progress has been slow and irregular and that inequalities within villages have persisted. These continued inequalities may constrain both the scope for further poverty reduction from economic growth and the impact of policy interventions.

Information on rural living conditions in India is abundant, compared with that for most other countries. Sample survey and census data, collected regularly since independence, have been used to trace the evolution of consumption levels and poverty rates, demographic and occupational trends, educational levels, and health and nutritional status. These data have been analyzed at the national as well as district levels (for recent examples, see Drèze and Sen 1995; Datt and Ravallion 1996; Drèze and Srinivasan 1996; and Sen 1996).

There is also a rich tradition of village studies in India, reflecting a wide range of disciplinary backgrounds and methodologies. These studies bring context and perspective to our understanding of Indian rural life, highlighting relationships between households and their surrounding community and illustrating the roles played by village institutions.

Although survey and census data have been widely analyzed, there have been only a few attempts to bring together the findings of the numerous village studies (but see Lipton with Longhurst 1989; Harriss-White 1992; and for comparisons of survey and village studies, Bardhan 1989 and Harriss-White 1996). This paper attempts such a review, focusing on the subset of Indian village studies that analyze change over time. Most of the studies reviewed involved at least one revisit to the village.
few record either single, very long-term visits or projects in which the researcher made a particular effort to record change over time. The paper also draws on several studies that take a broader, regional, perspective.

There is no statistical basis for generalizing beyond the village studies reviewed. Although we examine a large number of studies, covering a wide range of locations, we do not have a random sample. The study villages were selected for various specific reasons. If they share one common characteristic, it may be that they are, on average, less remotely located than villages are in general. The data collected, moreover, are dissimilar, limiting ready comparisons. One of the challenges in undertaking our review, in fact, has been to interpret the findings of investigators using varying methodologies and wearing differently tinted spectacles. The degree to which their findings can be extrapolated to rural areas in general is thus ultimately a question of judgment and personal inclination.

Table 1 lists the 35 longitudinal village studies reviewed in this paper. The regional coverage is fairly broad, although it is clear that Tamil Nadu and, to a lesser extent, Uttar Pradesh, are the most heavily represented. Important omissions include Madhya Pradesh, Haryana, Kashmir, Himachal Pradesh, and the northeastern states. In addition, the information available for Orissa, Karnataka, and Kerala is relatively old.

<table>
<thead>
<tr>
<th>State</th>
<th>District or region</th>
<th>Name of village*</th>
<th>Period studied</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andhra Pradesh</td>
<td>Mahbubnagar</td>
<td>Aurepalle, Dokur</td>
<td>1975–85</td>
<td>Walker and Ryan (1990)</td>
</tr>
<tr>
<td></td>
<td>Kathihar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karnataka</td>
<td>Mysore</td>
<td>Wangala, Dalena</td>
<td>1954–70</td>
<td>Epstein (1973)</td>
</tr>
<tr>
<td>Karnataka</td>
<td>Mysore</td>
<td>Rampura</td>
<td>1948–70</td>
<td>Srinivas (1976)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maharashtra</td>
<td>Vidarbha</td>
<td>Olegao</td>
<td>1920–70</td>
<td>Arrwood (1979)</td>
</tr>
<tr>
<td>Maharashtra</td>
<td>Vidarbha</td>
<td>Palso</td>
<td>1947–87</td>
<td>Gadre, Wahle, and Galgalikar (1987)</td>
</tr>
<tr>
<td>Orissa</td>
<td>Phulbani</td>
<td>Bispur</td>
<td>1950s</td>
<td>Bailey (1957)</td>
</tr>
<tr>
<td>Rajasthan</td>
<td>Nagaur, Jodhpur</td>
<td>Two unnamed villages</td>
<td>1963–84</td>
<td>Jodha (1989)</td>
</tr>
<tr>
<td>State</td>
<td>District or region</td>
<td>Name of village*</td>
<td>Period studied</td>
<td>Author</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------</td>
<td>-----------------------------------</td>
<td>----------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Tamil Nadu</td>
<td>South Arcot</td>
<td>Iruvelpattu</td>
<td>1916–81</td>
<td>Guhan and Mencher (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nangavaram, Nanganur, Kalladai,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peripatti</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tamil Nadu</td>
<td>Coimbatore</td>
<td>Three unnamed villages</td>
<td>1981–82</td>
<td>Heyer (1992)</td>
</tr>
<tr>
<td></td>
<td>Sambuvarayar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uttar Pradesh</td>
<td>Mainpuri</td>
<td>Karimpur</td>
<td>1925–84</td>
<td>Wiser and Wiser (1971); Wadley and Derr (1989)</td>
</tr>
<tr>
<td>Uttar Pradesh</td>
<td>Agra</td>
<td>Sunari</td>
<td>1964–85</td>
<td>Fuhs (1988)</td>
</tr>
<tr>
<td>West Bengal</td>
<td>Bardhaman, Birbhum, Purulia</td>
<td>Seventy-two unnamed villages</td>
<td>1972–86</td>
<td>Bhattacharya, Chattopadhyay, and Rudra (1987a, 1987b)</td>
</tr>
<tr>
<td>West Bengal</td>
<td>Bardhaman</td>
<td>Seven unnamed villages</td>
<td>1960s–1980s</td>
<td>Chattopadhyay (1992)</td>
</tr>
</tbody>
</table>

a. In many of these studies, village names have been changed to preserve confidentiality.

Source: Authors’ compilation.
We are concerned, in this review, with the evolution of material living standards in rural India. We trace not only changes in poverty or wealth but also the forces that may have governed these changes, such as agricultural intensification, changes in land relations, and occupational diversification. The first section gives a brief account of agricultural intensification and then examines land and labor—two key determinants of rural income—and the evolution of their links to living standards. The second section looks at changes in levels of poverty. Village studies support the common view that poverty has been declining in rural areas but also cite cases of clear impoverishment over time. Villagers’ own perceptions of poverty are discussed in this section. The third section examines changes in inequality, the nature and degree of which are difficult to gauge from large-scale statistical data. Village studies suggest that there has been little or no decline in the economic stratification of Indian villages. The fourth section summarizes the conclusions from our review and discusses policy implications. Because relatively few village studies set out explicitly to evaluate policy, they offer little guidance for the design of specific interventions. An important lesson that emerges from our review, however, is that inequality at the village level can significantly affect the way policies operate in practice.

The Sources of Income

In most of India’s half million or so villages, agriculture remains central to the local economy. The intensification of agriculture that began with the green revolution during the 1960s—the introduction of new fertilizers and seeds and expanded application of existing inputs, such as irrigation and labor—has had a profound impact in many parts of the country (Griffin 1979; Lipton and Longhurst 1989; Singh 1990; Hazell and Ramasamy 1991). The degree to which agricultural practices have changed has varied markedly by region and agroclimatic zone, but few rural areas have remained unaffected.

During the early years of the green revolution, there was much debate about its distributional impact. Many observers thought it could lead to rising income inequality. They reasoned that large farmers, with lower risk aversion and better access to credit, would make more extensive use of the new technologies and that the greater economies of scale in the new farming methods would enable them to increase their agricultural income far more than small farmers could, thus widening the income gap between the two groups. It was further suggested that inequality in land ownership, and even landlessness, would increase as the early beneficiaries of the green revolution, the large farmers, bought land from smaller farmers. We consider below the extent to which village studies support these concerns.
Agricultural Intensification

Virtually all of the studies surveyed find some expansion of agricultural production since the late 1950s. In Palanpur, Uttar Pradesh, productivity improved significantly between 1957–58 and 1993, and particularly after the mid-1960s. Output per acre of wheat, the main food crop in Palanpur, increased by about four times during the period, an annualized per capita growth rate of about 2 percent (Bliss, Lanjouw, and Stern 1998; see also Saith and Tankha 1992, for another village in Uttar Pradesh, and Leaf 1983, for Punjab). Although the most significant growth in output is seen in the northern states of Haryana, Punjab, and western Uttar Pradesh, substantial growth occurred elsewhere as well (for Tamil Nadu, see Gough 1981; Guhan and Mencher 1983; Athreya, Djurfeldt, and Lindberg 1990; Ramachandran 1990; Harriss 1991; and Hazell and Ramasamy 1991). Even some of the semi-arid regions in the ICRISAT studies—Mahbubnagar, in Andhra Pradesh, and Sholapur and Akola, in Maharashtra—showed increasing agricultural intensification between 1975–76 and 1983–84 (see Walker and Ryan 1990, for a review of these remarkable studies sponsored by the International Crop Research Institute for the Semi-Arid Tropics).

Irrigation, which permits multiple cropping and a shift to high-value crops, appears to have made the greatest contribution to agricultural intensification. Saith and Tankha (1992) note that in Parhil, Uttar Pradesh, farmers began harvesting three crops a year by 1987 and that high-value crops such as vegetables had replaced mixed crops. In Palanpur, where virtually all cultivated land was irrigated by the early 1980s, mixed crops and coarse cereals yielded to wheat, rice, and sugarcane (Bliss, Lanjouw, and Stern 1998). Similarly, in Iruvelpattu, Tamil Nadu, villagers were harvesting three crops a year by the early 1980s, including two paddy crops (Guhan and Mencher 1983).

Although relatively few studies explicitly compare output per hectare and farm size, those that do find little evidence of economies of scale. In Palanpur, output per hectare was found to be unrelated to farm size in any of the survey years (Bliss, Lanjouw, and Stern 1998), and in North Arcot district, Tamil Nadu, the new technologies appear to have been broadly “scale-neutral” (Hazell and Ramasamy 1991).

Land Ownership and Tenancy Relations

Incidence of Landlessness. Statistical evidence on trends in landlessness in rural India is scant and sometimes inconsistent (see Raj 1976). The number of village studies showing a drop in landlessness, however, is surprising in light of the popular presumption that agricultural intensification has caused it to rise.

Moreover, village studies suggest that impoverishment is only one of the causes of landlessness. Others include changes in household structure, population growth,
migration, and occupational change. Thus, increased landlessness, even where it is observed, need not be associated with worsening economic conditions.

Rodgers (1983) reports net losses in land owned and sharecropped in 46 small and marginal farming households in northeast Bihar between 1971 and 1981. Yet landholdings and local wages were not the sole determinants of household incomes, which actually rose slightly over the period. This was in part attributable to the growth in remittances from family members in Assam. In such cases, a “decline” into landlessness might indicate that households can afford to withdraw from agriculture altogether and might therefore imply an improvement in living standards.

Attwood (1979), using retrospective data for a village in Maharashtra’s famine tract, finds that the proportion of landless households increased between 1920 and 1970 but that this increase was caused mainly by in-migration. The availability of nonfarm employment in the local cooperative sugar factory made it possible for these landless immigrants to enjoy a living standard comparable to that of the landed population.

Drèze, Lanjouw, and Sharma (1998) show that landlessness in Palanpur was about 14 percent between 1957–58 and 1974–75, rose to 23 percent by 1993, and fluctuated considerably over time across households. Very few households became landless because the land was sold, however. More commonly, this occurred when sons left their father’s household before his death, prior to inheriting their share of the land. Much of the movement out of landlessness in Palanpur then occurred when these sons subsequently acquired their inheritance. In Gokilapuram, Tamil Nadu, Swaminathan (1991) notes high immobility in land ownership at either end of the landholding scale. Once a household is landless, it tends to remain so, but in between the extremes, and particularly among holders of 2.5 to 5 acres of land, there is considerable upward and downward mobility.

In the ICRISAT villages described by Walker and Ryan (1990), the general trend from 1950 to 1982 has been toward decreased landlessness in areas of reliable rainfall, increased equality of landholdings, and a decline in average farm size (see also Gadre, Wahile, and Galgalikar 1987). Athreya, Djurfeldt, and Lindberg (1990) note that in Thiruchirapalli district, Tamil Nadu, smallholder cultivation increased, landlessness decreased, and inequality in landholdings declined over the last generation. In irrigated areas, the incidence of landlessness decreased from 64 to 55 percent, and the land area occupied by landholdings of 25 or more acres decreased from one-half to one-third of the total land area. In the dry areas, half of those who were landless acquired some land. In Karimpur, Uttar Pradesh, the proportion of landless families decreased from 1925 to 1975, with a further sharp drop between 1975 and 1984 (Wadley and Derr 1989). In North Arcot, Tamil Nadu, Hazell and Ramasamy (1991) find no evidence of increased loss of land by smallholders over time.

**LAND TRANSACTIONS.** Srinivas (1976:82) reports that in Mysore, “parting with ancestral land was a serious matter under any circumstances.” The low levels of activity
in land markets in most of the villages studied suggest that this attitude may well
generalize to rural India more broadly (and to other developing countries as well: see
Binswanger and Rosenzweig 1986a).

In Palanpur, Drèze, Lanjouw, and Sharma (1998) find that the land market is
quite inactive in the sense that sales and purchases of land occur rarely. From 1957
to 1993, the average amount of land sold in Palanpur was barely 0.5 percent a
year. Analogous results are reported for Parhil (0.5 percent) by Saith and Tankha
(1992), for the ICRISAT villages (0.7 percent) by Walker and Ryan (1990), and for
Fonogram (less than 1 percent) by Beck (1994). Bliss, Lanjouw, and Stern (1998)
suggest that typical motives for land transactions in industrial countries, such as
life-cycle changes, moving to a job, and changes in perceived returns to different
forms of assets, are muted in rural Indian villages. In addition, the “thinness” of
the land market in villages such as Palanpur can be linked to a number of different
kinds of transaction costs.

Although inactivity in the land market challenges the notion of rapid land polar-
ization in rural areas, activity need not imply increased polarization. Harriss (1991)
finds that small landowners in North Arcot district frequently added to their hold-
ings, whereas large landowners generally sold land. In Palanpur, most of the land
sold between 1957 and 1993 was sold by households of the Thakur caste, which was
the dominant landowning caste in the village, but one whose traditional occupation
was not cultivation (Drèze, Lanjouw, and Sharma 1998). A similar pattern of sales
has been noted by Jha (1994) in Bihar.

Some active land markets do appear. Epstein (1973) reports that the conversion
of dry land to wet (through canal irrigation) in rural Karnataka led to an “overnight”
increase in land prices, presumably because the sudden arrival of irrigation led to a
rapid change in land productivity and the expected returns to land ownership. Early
land sales proved to be economically disadvantageous for the sellers, and those who
had better access to credit and were less risk-averse were able to buy this land and
profit in the long term.

Irrigation is often accompanied by a shift to cash crops which, although poten-
tially lucrative and possibly less exposed to harvest fluctuations than nonirrigated
crops, are vulnerable to the vagaries of the market. Booms and busts in the sugar
market were a driving force behind distress sales in the Maharashtra village studied
by Attwood (1979) and were an important cause of downward mobility among the
largest cane growers and their creditors. In this particular case, the sales had an equal-
izing effect on the distribution of land ownership.

LAND LEGISLATION. Village studies suggest that land legislation directed toward lim-
iting the size of landholdings or securing sharecroppers’ tenure has had mixed re-

results. Most studies find that the ultimate changes induced by land legislation are not
those intended by its architects.

Rajshri Jayaraman and Peter Lanjouw
Athreya, Djurfeldt, and Lindberg (1990) conclude that in Thiruchirapalli, land-reform legislation compelling large landlords to sell holdings resulted in a more equal distribution of land. They attribute this result to the particularly active tenants’ movement in the area, combined with a high incidence of absentee landlords. The experience elsewhere in Tamil Nadu was not so positive, however. In Tanjavur district, Gough (1987) finds that land reform led to the widespread evasion of land ceilings and to the eviction of tenants. Kapadia (1993) reports a similar pattern among the Pallars in Poovaloor.

In neighboring Kerala, Mencher (1980) finds that the 1970s land legislation giving permanent tenure and ownership rights to former tenants had two effects. Landlords who had failed to rid themselves of tenants prior to the legislation lost land because of it, but the only benefit most agricultural laborers gained was the permanent right to their house sites and to a tiny piece of the immediately surrounding land.

In the semi-arid ICRISAT villages, little land changed hands directly because of land-ceiling and tenancy legislation, but because the threat of confiscation was perceived as real, large farmers saw an increased risk in land accumulation (Walker and Ryan 1990). This perceived risk, coupled with the advent of irrigation, caused a shift away from the acquisition of rainfed land to investment in irrigation. Rather than evicting tenants, large farmers in the ICRISAT villages seem to have developed an aversion to long-term tenancy—a development that Walker and Ryan claim has led to more equitable access to land among prospective tenants.

In Karimpur, Uttar Pradesh, two developments changed the pattern of land distribution (Wadley and Derr 1989). The first, the Uttar Pradesh Zamindari Abolition and Land Reforms Act of 1950, combined with land ceilings to eliminate absentee land agents (zamindars) and give land ownership to the tillers and also to redistribute land from Brahmans to non-Brahmans. The results were by and large progressive. The subsequent land consolidation during 1967–70, however, although meant to regroup the holdings of small farmers, seems to have benefited richer farmers disproportionately. Drèze, Lanjouw, and Sharma (1998) note that in Palanpur, an attempt at very modest land redistribution in 1976 (six households were allotted one acre each of government-owned Palanpur land) involved considerable corruption and did not benefit the poor. A land-consolidation exercise carried out in Palanpur in 1985–86 was broadly successful in reducing land fragmentation, but influential farmers exploited their connections with the headman to get the best land.

In West Bengal, Chattopadhyay (1992) claims that land-ceiling laws had no direct impact on the village of Rajoor because no land was declared as surplus and vested with the government for redistribution. He finds, however, two important indirect effects. First, large joint families, in an attempt to evade the land ceilings, separated into smaller proprietary units, an action that led to fragmentation of large families and the erosion of their dominance in village politics. Second, as
observed elsewhere, the land-ceiling legislation led to the large-scale eviction of tenants.

**TENANCY ARRANGEMENTS.** Village studies provide ample evidence of the resilience of tenancy as an institution. They point to three broad developments over time: a move from sharecropping to fixed-rent contracts, including agreements for payment in kind (which might be misconstrued as sharecropping); a greater involvement by landlords in cost-sharing and in the supervision of day-to-day operations; and an emergence of "reverse tenancy," that is, the leasing of land by larger landholders from smaller landholders.

In Palanpur, the tenancy market, which is very active, has several notable features. First, with the exception of the poor, who may be more excluded than in the past, both tenants and landlords appear remarkably similar on average and are widely distributed along the scales of per capita income, land ownership, and caste status. Second, there is considerable turnover each year, with landlords and tenants re-sorting themselves and sometimes changing roles (Dreze and Sharma, 1996).

The evolution of tenancy contracts in Palanpur has been closely associated with the importance of nonlabor inputs (irrigation, fertilizers, seeds) accompanying the intensification and increased market orientation of agriculture. Cost-sharing of inputs, which has become more common in sharecropping contracts, has led to a reduction in economic differentiation between landlords and tenants, as poor households lose the ability to contribute their share of the cultivation costs. A similar trend is noted in Tamil Nadu by Ramachandran (1990) and in Uttar Pradesh by Srivastava (1995). Furthermore, ownership of indivisible productive assets such as pumpsets or tractors induces some of the larger landowners to lease-in land, giving rise to the phenomenon of reverse tenancy (Walker and Ryan, 1990; Janakarajan, 1996). The emergence of tenants who are less risk-averse and credit-constrained than in the past also explains the shift away from sharecropping contracts to fixed-rent tenancy contracts in Palanpur.

**Occupational Change**

**DECLINE OF TRADITIONAL LABOR SERVICES.** Many village studies observe a decline in traditional caste occupations. Wadley and Derr (1989) note that the Hindu jajmani system, in which customary payments are received in return for regular service to a patron, became virtually extinct in Karimpur, Uttar Pradesh, between 1925 and 1984. They attribute the loss of demand for these services to three main factors: behavioral change, technological change and mechanization, and increased monetization of transactions.

Athreya, Djurfeldt, and Lindberg (1990) comment on the extinction of some services and increased casualization of others in Tamil Nadu. They report that even
those traditional artisans who continue to be regulated by the old jajmani system often earn cash from some customers and work as agricultural wage laborers. Leaf (1983) finds that, in Ludhiana, Punjab, most harijans, who were formerly weavers, are now full-time agricultural workers. Some traditional occupations, however, such as carpentry, continue to be in strong demand (Drèze, Lanjouw, and Sharma 1998).

Ramachandran (1990) points out that barriers to employment in agriculture are negligible. In Gokilapuram, Tamil Nadu, two-thirds of the service-caste members have moved out of their traditional activities into agriculture. Indeed, agricultural labor has become the most caste-heterogeneous activity in the village.

THE CASUALIZATION OF LABOR CONTRACTS. Village studies also note a decline in traditional farm-labor arrangements. Long-term relationships between employers and laborers have declined in favor of casual, nonpersonalized, contracts, and the sub-contracting of specific cultivation tasks to labor “gangs” has increased.

Decline is not the same as demise, however. In some cases, the traditional farm-servant arrangement has been replaced by the “right of first call,” whereby workers first check at their patron’s house to see if their services are needed before seeking employment elsewhere. For Ramachandran (1990), the right of first call remains a manifestation of “unfreedom” for agricultural laborers. Epstein (1973), however, recounts that casual farm workers in unirrigated Dalena expressed envy at the continuation of the permanent farm-servant institution, however “unfree,” in neighboring Wangala.

In Thanjavur, Tamil Nadu, Gough (1987) observes a movement over time away from in-kind payments. In the 1950s, attached laborers were generally given clothing, life-cycle-rites goods, and a plot of land; in the 1980s, they were given no land and received more of their income in cash. Walker and Ryan (1990) find a similar decline of in-kind payments to farm servants in several of the ICRISAT villages. Only in Sholapur, which is drought-prone and characterized by high income variability, is payment in kind still important.

The increase in casualized labor contracts and cash payments suggests that wages may reflect valuable distributional information. Because agricultural daily-wage rates are remarkably uniform for all laborers (of the same gender) within a village, wage trends should show whether laborers’ incomes are rising over time—assuming no offsetting changes in days of employment (Drèze and Mukherjee 1989; Datt 1996).

Two additional trends are observed in village studies. First, permanent servants are being replaced by gang (often migrant) labor, and second, piece-rate contracts have increased over time. Indeed, gang labor, in which job payment is shared among members of the gang, is a collective piece-rate system. Breman (1993) claims that the movement after World War II from cane cultivation (which is highly labor intensive) to fruit tree cultivation (which requires hired labor only during the picking
season) was the primary cause for the expansion of gang labor in Gujarat. Athreya, Djurfeldt, and Lindberg (1990) find that in Thiruchirapalli, Tamil Nadu, contract gang labor has been gradually replacing wage labor, even for labor-intensive crops. Leaf (1983) notes that in Punjab, the green revolution has led to a marked rise in “periodic labor needs,” which has been met by an increased use of gang labor (see also Desai 1983, and Saith and Tankha 1992).

Drèze and Sharma (1996) note that in Palanpur, farmers have a clear view of the respective merits of piece-rate and daily-wage contracts. Piece-rate contracts dispense with the need for close supervision and encourage a timely completion of tasks. They are well suited to activities such as harvesting but may be less appropriate for tasks requiring quality control. They may also be appealing to highly productive workers, who can raise their earnings above the daily-wage rate through piece-rate work, as well as to the least productive workers, who may be excluded from the daily-wage labor market (Baland, Drèze, and Leruth 1996).

**EX�んAGNNT OF NONAGRICULTURAL EMPLOYMENT.** Discussions of occupational change in village India tend to concentrate on the decline of traditional labor services. This trend is often viewed negatively, on the assumption that workers are being “pushed” out of traditional occupations. A relevant but little discussed topic is the expansion and diversification of nonagricultural employment.

Many village studies note an increase in nonagricultural employment. Wiser and Wiser (1971) mention a bus stand tea stall and new bicycle and tractor repair shops. Epstein (1973) reports on the movement of entrepreneurs in 1970 to cafés, shops, cattle-trading posts, cane crushers, and rice mills, none of which had existed in 1955. Srinivas (1976) notes investment in bus lines, and Saith and Tankha (1992) comment on band-playing as a specialty of growing importance in Parhil, Uttar Pradesh. Although these are all instances of self-employment, evidence suggests that nonagricultural wage employment has expanded even more rapidly than nonagricultural self-employment (Visaria and Basant 1994). Because these new employment opportunities tend to be caste heterogeneous, they compensate in part for the reduced market for traditional caste-specific labor services.

Nonagricultural labor-market opportunities appear to be an important means of offsetting declines or high variances in village incomes. In North Arcot, Tamil Nadu, new agricultural machinery displaced hired labor in paddy cultivation, but real-wage rates in agriculture rose (at least for some activities) as a result of increased nonagricultural employment and the consequent tightening of village labor markets (Hazell and Ramasamy 1991). In the ICRISAT villages, nonagricultural earnings became increasingly important sources of income in the 1980s, raising mean household income and dampening variability (Walker and Ryan 1990). Decreased variability, in particular, seems important in raising rural living standards.
In Palanpur, both regular and casual employment outside the village has expanded (Bliss, Lanjouw, and Stern 1998; Dréze, Lanjouw, and Sharma 1998). The demand for employment in this sector, where wage rates and work conditions are attractive relative to agricultural work, exceeds the supply of jobs available. The process through which these jobs are obtained appears to be governed by both personal connections and the ability to pay a bribe. Regular nonagricultural jobs tend to cluster in a small number of establishments where an initial employee has helped others to gain employment. Those who follow the first entrant are frequently of the same caste or are otherwise related.

The role of personal contacts and influence in job search might explain the large gaps between agricultural and regular nonagricultural wages, the low turnover of regular nonagricultural jobs, and the apparent disadvantage that persons of low social status have in competing for regular nonagricultural jobs, even given comparable skills and endowments (Unni 1997). Although the better educated, or otherwise privileged, may have more opportunities for nonagricultural employment, the poor, with lower reservation wages, have in the past been most active in pursuing nonagricultural opportunities. This appears to be changing, however; the village elites are now aggressively seeking nonagricultural employment. In Palanpur, there has been a clear shift over time in the distribution of nonagricultural earnings, with the better-off in the village acquiring an increasing share. This gradual reduction in the share for disadvantaged groups has also been observed by Leaf (1983) in the Punjab and by Wadley and Derr (1989) in western Uttar Pradesh (for which, see also Sharma and Poleman 1993 and Ranjan 1994).

AGRICULTURAL WAGES. The green revolution and the expansion of the nonfarm economy appear to have raised agricultural wages in rural India by increasing the demand for labor. Until recently, secondary data suggested that real wages in rural India showed no significant upward trend (Kurien 1980). Evidence now suggests, however, that wages did rise in most regions of India in the 1970s and 1980s (Acharya 1989; also Guhan and Mencher 1983; Leaf 1983; Harriss 1989; Ramachandran 1990)—although there is no clear trend in the 1990s (Unni 1996).

In Palanpur, real wages have risen fairly steadily since 1974–75, remaining well above the levels that prevailed in either 1957–58 or 1962–63 (Dréze, Lanjouw, and Sharma 1998). This increase is all the more dramatic when expressed in terms of wheat purchasing power. Because the relative price of wheat fell in Palanpur as production rose over time, one day of casual labor could purchase more than 8 kilograms of wheat in 1993, compared with less than 3 kilograms in 1957–58. Similarly, Hazell and Ramasamy (1991) show that, in North Arcot, agricultural earnings for landless laborers, small paddy farmers, and nonagricultural households doubled between 1974–75 and 1983–84, as large farmers withdrew from agricultural labor and employment opportunities in dairying and nonfarm activities expanded.
LABOR MIGRATION. The rise in work-related migration also affects village incomes. Using large-scale survey analysis, Kurien (1980) finds that in rural Tamil Nadu, the relatively poor seem increasingly willing to give up not only their traditional occupations, but also their places of work. Walker and Ryan (1990) observe a tightening of the labor market in Aurepalle, Andhra Pradesh, from 1974 to 1985 and a 60 percent rise in wage rates, which they attribute to temporary migration to nonagricultural jobs in Hyderabad. Saith and Tankha (1992) find that in Parhil, Uttar Pradesh, the incidence of out-migration virtually doubled.

The effects of labor migration are difficult to examine in a closed village study (but see Breman 1985; Walker and Ryan 1990; and Hazell and Ramasamy 1991). Nevertheless, village studies do note an apparent shift from low-caste, low-skilled migration to high-caste, high-skilled migration. In Karimpur, extravillage service jobs had formerly been held solely by the outcaste poor, who migrated to Calcutta to work as sweepers. By 1984, service jobs had become the “desired occupation” of wealthier Jati families (Wadley and Derr 1989). In Punjab, as well, high-caste Jats have migrated out more frequently since the green revolution than have outcaste harijans (Leaf 1983).

PROLETARIANIZATION? Evidence from secondary sources, such as the census, suggests that an increasing proportion of the rural population is working for wages. This increase has sometimes been seen as a trend toward “proletarianization” of the labor force. This term can be interpreted in (at least) two ways. It may simply describe the shift away from self-employment (mainly in agriculture) to wage labor, or it may suggest that smallholder cultivators are being pushed out of agriculture into wage labor, rather than being pulled by new employment opportunities.

Harriss (1991) argues that in North Arcot, Tamil Nadu, proletarianization is mainly a process of farming households supplementing their cultivation incomes with wages from both agricultural and nonagricultural employment. He interprets a sharp increase in the number of occupations outside of cultivation or agricultural labor between 1973 and 1984 as a process of “proletarianization without depeasantization” (Harriss, 1989; but compare Athreya, Djurfeldt, and Lindberg 1990, who describe a process of “peasantization rather than proletarianization” in Thiruchirapalli). Ramachandran (1990) also demonstrates that a rise in agricultural wage employment does not necessarily imply a withdrawal from cultivation. He finds that in Gokilapuram, Tamil Nadu, smallholders supplemented their cultivation activities with agricultural employment but did not sell their own land. He therefore suggests that the expansion of the market for hired labor has helped to preserve the institution of smallholder cultivation.

The degree to which expanded agricultural wage employment is associated with a withdrawal from cultivation seems to be closely linked both to the incidence of tenant evictions accompanying land-to-the-tiller legislation in the 1970s and to dimin-
ishing farm size caused by population pressure on the land. In Thanjavur, Tamil Nadu, the number of landless male agricultural laborers pushed out of cultivation for these reasons increased dramatically (Gough 1987). Workers have also been pulled away from farms by increased nonagricultural opportunities. Whether push or pull influences predominate across rural India varies with the particular local experience.

SOCIAL IMPLICATIONS OF OCCUPATIONAL CHANGE. Occupational change is likely to be accompanied by changes in the prevailing social order. Breman (1993:21), for instance, regards the disintegration of the jajmani system as a product of multiple factors and states that “the relationships lost their local flavor in the process of enlargement of scale. Commercialization of agriculture and continuously increasing government intervention diminished the importance of local autarky and autonomy. The drift of members of the artisan and serving castes to the urban centres, from which they began to serve the surrounding countryside, contributed to the rise of an ever-increasing number of different contacts which went beyond the village. Depending on the accessibility of the region, this development began early or late.”

The degree to which a village is linked to the market economy can significantly affect its occupational and social structure. Gough (1981), in her analysis of societal change in Thanjavur, Tamil Nadu, remarks that coastal regions, which have been more “disturbed” than their inland counterparts by external change, break more easily with traditional caste-based occupational structures. Where an area’s comparative advantage lies in agriculture, however, the traditional social order (and associated occupational structure) is less likely to change. Epstein (1973) describes a similar process in Karnataka. One of the villages studied, Wangala, was served by a recently constructed irrigation canal; the other, Dalena, was not. Irrigation in Wangala enabled farmers to grow cash crops, but because the village’s economy remained almost wholly agricultural, the new cash economy coexisted with the traditional system of hereditary relationships between Wangala farmers and their “functionaries.” The system was, in fact, reinforced by the introduction of labor-intensive cash crops. In neighboring, unirrigated, Dalena, economic diversification led to increased factionalism. Agriculture no longer bound the villagers together, and greater integration into the rural economy led Dalena commuters into the wider world, where differentiation along caste lines and other social institutions were diluted.

Poverty and Living Standards

General Trends

Although aggregate economic statistics in India point to a steady (albeit slow) decline in poverty, absolute levels of deprivation remain high. Few village studies track
absolute poverty, measured in terms of consumption or income. They generally scrutinize, instead, the positions of households relative to one another. Beck (1994) thus argues that poverty rose in three West Bengal villages during the late 1980s, even though the incomes of the poorest households increased. His judgment, reflecting the perceptions of villagers themselves, is based on the observation that the rich in these villages enjoyed significantly larger increases in income than the poor did during the same period.

Village studies that do focus on absolute poverty also observe a decline. In Palanpur, the proportion of the population below any reasonable poverty line clearly decreased between 1957–58 and 1983–84, although poverty rates fluctuated markedly throughout the period, and households moved in and out of poverty in response to price changes, harvest levels, and the partitioning of household lands (Drèze, Lanjouw, and Sharma 1998). In any given year, however, irrespective of the overall level of poverty, households of the low-ranked Jatab caste, and households relying primarily on casual agricultural work for income, were disproportionately represented among the poor. This relatively high and constant risk of poverty among agricultural laborers and low-caste households is stressed in many village studies (Mencher 1980; Rodgers 1983; Ramachandran 1990; Jha 1994). Drèze, Lanjouw, and Sharma (1998) argue that because agricultural wage labor is seen as a last resort by Palanpur villagers, it is a reliable indicator of poverty in any year.

Walker and Ryan (1990) find that two-thirds of the villagers in the ICRISAT studies moved in or out of poverty in at least one of the nine consecutive years between 1975–76 and 1983–84. These were nearly all middle-size cultivators. The nonpoor—those who never crossed the poverty line—were more educated, did not actively participate in the casual labor market, and owned more land than their neighbors. The consistently poor were predominantly landless harijans with high dependency ratios.

Village studies recording income levels for more than two years are rare. The more usual comparisons of income across two periods, corresponding to an initial study and a revisit, are vulnerable to the effects of harvest fluctuations. Where comparisons of income are handicapped in this way, however, other indicators may be scrutinized to assess changes in longer-term living standards. Jodha (1989), using the more “conventional” per capita income measure, finds that 20 percent of the households in his two Rajasthani villages were poorer in 1982–84 than in 1964–65 (although he does recognize that crop yields were better on average in 1964–65). The proportion of households below the poverty line increased from 18 to 23 percent, with both upward and downward mobility across the poverty line. Households that appear to have become poorer according to the income measure did seem to be better off, however, when judged by other, qualitative, indicators of economic well-being. Jodha points to expanded alternatives for employment and borrowing, to an increase in consumption activities with high income
elasticities (travel, slack-season purchases), and to investment in lumpy consumer durables (pukka houses and structures).

Increased ownership of consumer durables has been observed in many village studies. Leaf (1983), for example, finds that in rural Punjab, a substantial improvement in housing and shelter accompanied the green revolution. Bhattacharya, Chattopadhyay, and Rudra (1987a) note that in rural West Bengal, ownership of nontraditional items such as radios, wristwatches, and bicycles increased considerably between 1972–74 and 1985–86. Although such increases have been widely interpreted to imply an improvement in living standards, it is important to note that expanded ownership of “modern” consumer durables (the increase most often cited) is in part a relative-price effect. In Palanpur, for example, modern consumer durables have become cheaper relative to livestock and have increasingly been substituted for livestock in dowries (Drèze, Lanjouw, and Sharma 1998). Similarly, an increased consumption of goat meat in Punjab can be attributed both to an improvement in general welfare and to the pumpset revolution, which led to a shift from owning draft animals to maintaining livestock for food and marketing purposes (Leaf 1983).

Instances of Impoverishment

Although both large-scale surveys and village studies find a general drop in poverty after the 1960s, there are pockets where poverty has increased. These instances of impoverishment are worth scrutinizing.

Rapid population growth, which can offset rising productivity and reduce per capita income, is one factor that can cause impoverishment. Village studies note that growth in the agricultural labor force may follow land-augmenting technological change, particularly when such change induces in-migration of landless laboring households. Ramachandran (1990) observes a decline in wage rates in Gokilapuram between 1948 and 1975, which he attributes to a sharp growth in the relative and absolute size of the agricultural labor force in the Cumbum valley. Walker and Ryan (1990) find that in Kanzara, the influx of landless laboring households from neighboring areas has kept wages low, despite the increased demand for agricultural labor. Where there are factors pushing smallholders into landlessness and agricultural labor, the need for offsetting “pull” factors becomes paramount. As Jha (1994) observes, “push” sometimes dominates. Population growth, tenant eviction, and declining migration opportunities in Bihar have led to a decrease in employment days per worker and to stagnating incomes for daily-wage and attached laborers.

Many parts of rural India, moreover, are vulnerable to drought, and drought can devastate the poor. Hazell and Ramasamy (1991) show that in North Arcot, Tamil Nadu, average incomes in 1982–83 (a severe drought year) were lower than in 1973–74 for all agriculturally dependent households. Many households in North Arcot had no financial reserves with which to cope with droughts, particularly when low
rainfall years ran together. The expansion of irrigation did not exercise a stabilizing effect in North Arcot because irrigation water comes from groundwater reserves, which depend on rainfall for their recharge. Although groundwater irrigation is a useful way of redistributing the monsoon rains in this region, it provides only limited capacity for carrying water from good to bad years. The close association between poverty and access to groundwater is explored by Bhatia (1992) for rural Gujarat and by Saith and Tankha (1992) and Drèze, Lanjouw, and Sharma (1998), who express concern that unregulated and expanding pumpset irrigation is depleting groundwater in the Gangetic plain.

Even against a background of generally declining poverty, certain subgroups of the population face a high risk of impoverishment through illness, accidents, or lifecycle events. Drèze (1990) notes, for example, that in most parts of India, women, who are usually younger than their husbands, face a high probability of becoming widowed during their lifetime. Whether widowhood translates into a sharp decline in their living standards depends on local inheritance laws, whether women are allowed to work for wages, and whether the widow has children (in particular, sons) who can provide support.

The Perceptions of the Poor

Income- or consumption-based measures of well-being are often suspected of failing to capture local perceptions about living standards. Village studies are particularly useful in this regard because the authors are close to their subjects. Bhattacharya, Chattopadhyay, and Rudra (1987b) find that 60 to 80 percent of the households they canvassed in rural West Bengal felt that their standard of living had improved with respect to social consumption (drinking water, medical care, education, roads and transport, and recreation and culture), and less than 10 percent felt it had deteriorated. More generally, Bhattacharya, Chattopadhyay, and Rudra (1987b) find that about half of the respondents felt their overall living standards had improved, while about a quarter thought they had deteriorated. The main reason for their judgments was a perceived increase or decrease in real income.

In his study in western Rajasthan, Jodha (1989) points to reduced reliance on traditional patrons as a reason for a perceived improvement in well-being. The importance attached to personal freedom and dignity is a recurrent theme, even among the very poor, and a desire to secure such freedom can be a driving force behind the acquisition of assets. Srinivas (1976:111) writes that "landownership and wealth were occasionally able to mitigate if not overcome the effects of birth in a ritually low caste. . . . No wonder then, that there was a general scramble for land . . . for it meant freedom from hunger and bondage to patron, and also self-respect." Wadley and Derr (1989) find that in Uttar Pradesh, those who perceived an improvement in their conditions spoke of it in terms of increased personal freedom rather than re-
duced hunger or poverty. Beck (1994) reports that in West Bengal, 49 out of 58 respondents claimed they valued self-respect more than food.

Some think otherwise, however. Epstein (1973) quotes a Dalena villager as saying, “you cannot eat social acceptance.” Breman (1993) claims that in Gujarat, both landowners and agricultural laborers preferred long-term labor contracts. Such contracts guaranteed employment for laborers during the slack season and laborers for landowners during the peak season. That concerns about personal freedom can be something of a luxury is starkly pointed out by Mencher (1980), who warns against complacency regarding rural Kerala, where the remarkable inroads achieved in health and literacy might suggest that poverty has been eliminated. Mencher argues that the “miracle of Kerala” has not been a miracle for agricultural laborers, many of them women, who are still daily confronted with the uncertainty of how they are going to feed their children.

Inequality

Changes in inequality are closely related to the manner in which rural poverty has evolved. Understanding the distribution of living standards in village communities can therefore provide important insights for poverty-reduction initiatives. There have been many efforts to strengthen local decisionmaking power in rural India, fueled by the impression that local control will contribute to poverty reduction. The Jawahar Rozgar Yojana (JRY) employment program, for instance, introduced by the central government in 1989, relies on decentralization. The JRY scheme is implemented by village panchayats and promotes the creation of durable community assets. How much one can expect from such a scheme depends, to some extent, on how well represented the poor are in local decisionmaking. This, and factors such as village solidarity, are likely to be affected by the degree of polarization in village living standards.

Because village studies typically take the entire village population as their domain, they are better suited than large-scale surveys for studying village-level inequality. Large-scale surveys rarely sample enough households within a given village to yield reliable measures of village inequality. Moreover, because of factors such as aging, it is risky to make inferences about the evolution of inequality by following a specific sample of households surveyed over time.

Land Ownership

Village study findings on changes in land distribution challenge several common clichés. First, contrary to conventional wisdom, there seems to be considerable movement in the distribution of land, movement not necessarily linked to a highly active
land market. Second, there is no uniform trend toward increased inequality in land distribution. Whether land inequality has increased or decreased varies across the villages studied and often depends on whether the area in question is wet or dry.

Demographic change is particularly relevant to changes in land distribution. Srinivas (1976:112) writes that “while a man may have had his descendents in mind when buying land he also knew that it would be divided after his death. Big estates were more usually built up through the accident of single sons in more than one generation.” In Palanpur, demographic factors account for the bulk of observed changes in the distribution of land (Dreze, Lanjouw, and Sharma 1998). Removing the effect of household partitioning by aggregating the land ownership of each “dynasty” (defined as the union of all households descending from a particular household in the first survey year, 1957–58) reveals that inequality in per capita land distribution remained fairly constant between 1957–58 and 1993.

Swaminathan (1991), however, finds high and rising inequality in landholdings in Gokilapuram between 1977 and 1985 (despite a decline in the already lower inequality in nonland assets). Hazell and Ramasamy (1991) find no evidence of increased inequality in landholdings in North Arcot following population growth and partitioning but note that average farm size declined slightly across all quartiles between 1973 and 1983. In the ICRISAT villages, Walker and Ryan (1990) actually find increased equality in land distribution, with large landowners losing ground to small landowners in both relative and absolute terms.

The positions of smallholders and the landless appear to be relatively static. Athreya, Djurfeldt, and Lindberg (1990) report that in Thiruchirapalli, the most immobile groups in their respective areas are the landless in the wet area and those who began with very small holdings in the dry area. Similar patterns are noted by Rao (1972); Attwood (1979); Cain (1981); Caldwell, Reddy, and Caldwell (1982); Harriss (1991); and Swaminathan (1991).

Athreya, Djurfeldt, and Lindberg (1990) note that the ownership structure is often extremely polarized in wet areas but shows relative equality in dry areas. In Thiruchirapalli, this pattern has social origins: in the wet area, lower castes were traditionally not allowed to own land, whereas in the dry area, land commanded no price at all until the 1900s and is still quite cheap. Hazell and Ramasamy (1991) note that the only evidence of a mild worsening in land distribution among their Tamil Nadu villages was in Duli, which is poorly irrigated. In dry areas, rising inequality need not increase poverty because the productive value of the land is already so low.

**Income and Wealth**

In Palanpur, there is little evidence that the green revolution led to a marked widening in the distribution of income (Lanjouw and Stern 1998a). In fact, income inequality appears to have declined somewhat between 1957 and 1984. The sharp
expansion of irrigation during the green revolution years to farms other than the wealthiest (from 50 percent of the land in 1957–58 to 96 percent by 1974–75) had an important equalizing effect on incomes. Sharma and Poleman (1993) find a similar decline in Walidpur, in western Uttar Pradesh, between 1963–64 and 1988–89. Epstein (1973), however, notes that in Karnataka, there was a sharp concentration of income between 1957–58 and 1970. The key factor in the villages Epstein studied was access to irrigated land, and this remained highly unequal throughout the interval. Households with good land endowments became much richer, while agricultural wage earners competed with a sizable pool of migrant workers for low wages.

In Palanpur, Drèze, Lanjouw, and Sharma (1998) find both lower income inequality across the green revolution period and a nonnegligible widening in the distribution of wealth between 1962–63 and 1983–84. These findings need not be contradictory. Inequality in wealth may rise, even in the presence of an unchanged distribution of income, if the savings function is convex at low levels of income, that is, the marginal propensity to save rises with income. An equiproportionate increase in income for all households could thus easily lead to some polarization of new wealth. In Palanpur, the rise in net wealth inequality also reflects a highly uneven accumulation of liabilities to credit institutions, as well as the particular vulnerability of the poor to corrupt accounting practices within formal credit institutions (Bell and Srinivasan 1985; Bhende 1986; Binswanger and Rosenzweig 1986b; Iqbal 1988; Bouman 1989; Krishnan 1990; and Banik 1992).

Swaminathan (1988) also finds, for Gokilapuram between 1977 and 1985, an immense disparity in, and perceptible worsening of, the distribution of wealth, defined as the households’ own estimation of the current value of their assets. It is worth noting, however, that although the poor in Gokilapuram become relatively worse-off, the per capita wealth of even the poorest households increased in real terms by a factor of about 2 over this eight-year interval.

Whether inequality has increased or decreased over time in rural India is, perhaps, of secondary importance to the well-documented observation that levels of inequality still generally remain high. Although various village studies trace alternative paths of income or wealth inequality, all have started from positions of considerable disparity in living standards. Even where inequality seems to have fallen, only a small fraction of the total has been eliminated over time.

Caste

Just as land and income distribution in rural India has changed in response to population growth, technological change, and occupational diversification, so caste relations have changed. Many studies comment on the shift in local perceptions of, and attitudes toward, caste-based social rankings. Drèze, Lanjouw, and Sharma (1998) point out that in Palanpur, people of all castes can now sit together on the same
string cots (charpai), and the dominant Thakurs can no longer exercise arbitrary force over the lowest-ranked Jatabs.

The disappearance of many traditional occupations has inevitably undermined the differentiation of castes by behavior and associations. Walker and Ryan (1990) ascribe the erosion of the caste hierarchy in the ICRISAT villages to the tendency of traditional service-related castes to supplement their livelihood with agricultural labor. Epstein (1973) also notes the role of occupational diversification. She writes that in the unirrigated village of Dalena, economic diversification diluted caste lines, whereas in irrigated Wangala, where the agrarian base remained largely intact, the all-India antiuntouchability policies were largely ignored.

The presence of upwardly mobile households clearly affects the social hierarchy in a village. In Palanpur, an ongoing rivalry has pitted the Thakurs, the highest-ranked, traditionally noncultivating and nonlaboring caste in the village, against the Muraos, a traditionally cultivating caste (Drèze, Lanjouw, and Sharma 1998). The economic status of the Muraos has risen substantially since the 1950s and has surpassed that of the Thakurs, resulting in a gradually improved social status for the Muraos and an escalating political rivalry with the Thakurs.

Caste mobility is most marked at the middle and upper ends of the caste hierarchy. Kapadia (1993) notes that in Poovaloor, Tamil Nadu, the relative economic power of the Brahmans has declined. Wadley and Derr (1989) find a similar decline in Karimpur, and Fuhs (1988) observes that in Sunari, Uttar Pradesh, many Brahmans now even work as laborers for Jat farmers. Breman (1993) finds considerable upward mobility among the middle caste Kolis and Dhodias in two villages in Gujarat. Jha (1994) finds similar mobility in rural Bihar, as does Da Corta (1993) in Andhra Pradesh.

Upward mobility of low-caste households is more unusual. Drèze, Lanjouw, and Sharma (1998) find no change in the position of Jatabs in Palanpur society, where they represent about 12 percent of the village population. Jatabs have experienced slower per capita income growth than other groups, almost stagnant levels of education (100 percent female illiteracy in 1993; 88 percent male illiteracy), virtual exclusion from regular, nonagricultural, wage employment, and unaltered land endowments. This is despite their being targeted by various government programs. Guhan and Mencher (1983) find that in Iruvelpatti, Tamil Nadu, the combined effects on the harijans of economic inequality, social discrimination, and physical segregation are persistent and readily visible. Epstein (1973) paints a similar picture for Wangala and Dalena in Karnataka.

In some villages, low-caste households have improved their relative standing, through either the influence of (often external) political parties or the forum of collective action. Chattopadhyay (1992) attributes the disintegration of the structure of “dominance and subordination” of landless laborers and sharecroppers by three “big-men” families in Bardhaman district, West Bengal, to an awareness of new possibil-
ties fostered by political parties. Gough (1987) attributes the increase in real wages for farm servants in Thanjavur between 1952 and 1976 to the efforts of the communists to ensure good treatment for farm workers, as well as to the rise of labor unions. Sreekumar (1995) cites the importance of collective action in Changel, Bihar, where the consolidation of Yadavs, Noniyas, and Dhanuks, all "backward" castes, led to the break-up of the Khayasth-Brahman monopoly on village politics.

Collective action may not be easily achieved, however. Breman (1993) finds that intercaste tensions among the lower castes in his two Gujarat villages, as well as effective resistance from the higher caste Anavils, have prevented solidarity among the lower castes. Athreya, Djurfeldt, and Lindberg (1990) suggest that in Thiruchirapalli, factionalism among the higher castes was a key to reductions in poverty and inequality among the lower castes.

An interesting countertrend to the erosion of caste distinctions is the process called "sanskritization" by Srinivas (1966). Srinivas (p. 28) notes that "when a caste or section of a caste achieved secular power, it usually also tried to acquire the traditional symbols of high status, namely the customs, ritual, ideas, beliefs, and life-style of the locally highest castes." Upwardly mobile castes therefore often come to adhere more, rather than less, stringently to caste norms over time, but they adopt the norms of a caste that is higher than the one from which they originate.

**Gender**

Although women may have benefited from the general improvement of living standards, their gains have been small compared with the persistent inequalities between men and women in many parts of rural India. In some cases, their relative position has actually declined. This decline is linked in part to the process of sanskritization and the institution of dowry, and in part to the reduced participation of women in the labor force following occupational diversification and technological change.

One of the more alarming findings of village studies has been a decline in the female-male population ratio. In Karimpur, Uttar Pradesh, Wadley and Derr (1989) interpret such a decline among the Jati caste as a growing negative valuation of women linked to the rising incidence in nonagricultural employment outside the village by Jati men. Because female farm workers in Karimpur are generally not hired independently of their husbands, the Jati women now have fewer income-earning opportunities (see also Epstein 1973, for Dalena). Drèze, Lanjouw, and Sharma (1998) link a similar decline in the female-male ratio among Jatads in Palanpur to the absence of any expansion in female labor-force participation and to a growing identification of disadvantaged castes with the patriarchal norms of privileged castes. Drèze and Sen (1995) record a decline in the female-male ratio among scheduled castes in Uttar Pradesh more generally (from 0.94 in 1901 to 0.88 in 1991).
An important aspect of sanskritization has been the shift from brideprice to dowry. In a study of rural propertied elites northeast of Coimbatore, Tamil Nadu, Heyer (1992) notes a movement from brideprice in the 1930s to dowries in the 1950s and a doubling of the real value of dowries paid from the 1950s to 1970s. Bailey (1957), Epstein (1973), and Wadley and Derr (1989) find similar trends. As Heyer (1992) explains, the growing importance of dowries has led to the redistribution of capital from households with higher daughter-son ratios to households with lower ratios. Maintaining wealth from generation to generation therefore depends on keeping the number of surviving daughters relatively small.

Summary and Policy Implications

Agricultural Production, Land, and Employment

All of the studies reviewed find that agricultural production has grown significantly during the past 20 to 30 years. The broad package of new inputs that accompanied the green revolution has had at least some impact almost everywhere in India. Irrigation, in particular, has had a profound effect.

The new technologies do not appear to have favored large farmers disproportionately. Although large farmers have better access to credit for the purchase of new inputs, the expansion of irrigation may be quite progressive if, prior to expansion, only large farmers had irrigated land.

Landlessness has not obviously increased over time and in some cases has decreased. Because increased destitution is only one possible cause of landlessness, loss of land is not necessarily a symptom of impoverishment.

In most of the villages studied, the land market was found to be relatively inactive. There were exceptions, however. In some instances, canal irrigation, which sharply and dramatically alters returns to land, led to an increase in land transactions. Even in such cases, however, the small and vulnerable farmers were not necessarily the ones losing their land. Land sales, when they occurred, often involved transfers of land from formerly well-endowed, but noncultivating, households to cultivating castes.

Land legislation has had both positive and negative effects on land distribution. The beneficial effects have usually been indirect, such as encouraging farmers to invest in new technologies, rather than to expand their landholdings. A very common response to land-to-the-tiller legislation, however, has been the eviction of tenants. This has increased both landlessness and the size of the agricultural wage-labor force.

Tenancy, where it remains, appears to be evolving along with the new cultivation practices. Many studies point to the frequency of fixed-rent contracts, the greater involvement of landlords in cost-sharing, and the phenomenon of "reverse tenancy."

Rajshri Jayaraman and Peter Lanjouw
This evolution of tenancy corresponds to expectations, given the increased intensity of agriculture and the spread of modern practices and technologies. The persistence of tenancy as an institution suggests that it fulfills a useful purpose—permitting the cultivation of land by those who can put it to most productive use, without requiring a full transfer of property rights.

Traditional artisanal occupations have generally declined in number and importance, but this decline has usually occurred against a background of rising real wages. As a result, although most traditional artisans have moved into manual-labor occupations, it is not clear that their living standards have declined. Daily-wage and piece-rate contracts are increasingly replacing long-term agreements. Although such casualization reduces the dependence of laborers on their employers, it can also increase their insecurity.

Rural households benefit from nonagricultural employment not only from the incomes received, but also from reduced exposure to agricultural fluctuations. These benefits have not been shared equally across households, however. Although village households without assets have not been excluded from nonagricultural employment, regular nonfarm employment appears to be linked to personal contacts, as well as to payment of bribes. These factors have probably favored the more affluent and highly ranked villagers, and evidence suggests that this group has become more active in the nonagricultural sector over time.

The expansion of nonagricultural employment opportunities has "tightened" rural agricultural labor markets and raised agricultural wages in many of the villages studied. This increase is less marked in regions where in-migration has risen sharply.

Although proletarianization has been widely observed, it is not clear how many smallholder cultivators are being pushed out of agriculture by declining returns and land polarization, and how many are being pulled away by new opportunities in the nonagricultural sector. The fact that proletarianization often occurs against a background of rising real agricultural wages suggests that pull factors may be just as important as push factors.

In villages having a strong comparative advantage in agriculture, traditional social and occupational structures have generally persisted, even in the face of dramatic changes in terms of trade or technology. In villages that are more heterogeneous economically, however, traditional occupational structures, such as caste-based trades, have become more diffuse, as villagers have acquired links with the outside world and have diversified their activities.

**Poverty**

Village studies point to a slow decline in rural poverty but note considerable movement in and out of poverty. Some of this movement can be attributed to fluctuations...
in harvest quality and to personal calamities. Low caste status and agricultural wage labor are associated with long-term poverty.

Even where poverty does not appear to have fallen, increased ownership of consumer durables, shifts toward consumption of higher quality goods, and self-reported perceptions point to gradually improving living conditions. In addition, the rural poor often single out reduced dependence on patrons as an important improvement in their living conditions.

Although the direction of change is encouraging, poverty remains extremely high in many villages. Perhaps as important to understanding which forces have reduced poverty is the question of why poverty has fallen so slowly in the face of often dramatic growth.

**Inequality**

In contrast to declining poverty in rural areas, it is difficult to find reduced inequality. Although there has been considerable movement in the distribution of welfare over time, no clear trend emerges, particularly one favoring the very poor.

Land distribution in rural areas is less static than often suggested, but much of the movement may be more apparent than real. Demographic change associated with household partitioning alters the distribution of land but is not likely to reflect fundamental changes in the distribution of wealth. Even where significant changes in land distribution occur, the high degree of immobility among the landless and smallholders suggests that the relatively poor are not the primary beneficiaries. The green revolution may have improved the distribution of welfare, but the scope for improvement depends closely on local conditions.

Rural growth, and in particular the integration of the village economy into the wider rural economy, has eroded the traditional caste structure in many villages. Several studies point to a discernable breakdown in the correlation between ritual status and economic status. In many villages, the traditionally dominant, non-cultivating, caste is being overtaken economically by households of other castes, usually traditional cultivators who have taken greatest advantage of the new agricultural opportunities. This process leads, on occasion, to sanskritization, whereby high-caste practices are emulated by other upwardly mobile castes. Although such changes are significant, they do not affect the lives of all castes in the same way. The lowest castes in Indian villages often remain easily distinguishable from the rest of village society by their very low material well-being and the limited opportunities they have to improve their living standards.

In addition, there has been little evidence of improvement in gender relations. The decline in the (already extraordinarily low) female-male population ratio suggests that gender-based inequality may be increasing. Sanskritization may play a
role in this, in that women are often required to observe behaviors associated with higher castes, such as full purdah or withdrawal from wage-labor activity. The shift from brideprice to dowries and the trend toward higher dowries, moreover, further strengthen the perception among households that daughters drain family wealth.

Policy

Conspicuous by its absence from this review is an evaluation of the various public policy measures directed toward village living standards and community life. Although longitudinal village studies seem particularly well-suited to making such an assessment, most of them are silent on this issue.

Drèze, Lanjouw, and Sharma (1998) review every instance of public service provisioning in Palanpur between 1957–58 and 1993—including the building of public schooling facilities, the implementation of the Integrated Rural Development Programme and other credit programs, the JRY employment program, and the provision of widows’ pensions. Their conclusion is that, except for the modest success of a program providing two water handpumps near the low-caste quarters, the programs have been extremely disappointing. A recurrent observation is that privileged individuals or groups direct the benefits of programs to themselves at the expense of other village members.

It seems worth asking whether high inequality within villages, and in particular the lack of strong village cohesion, present major obstacles to the successful implementation of policies in Indian villages. The answer could have important implications for the design of policies, especially policies that seek to avoid the recognized pitfalls of centralized delivery by exploiting the potential of decentralized mechanisms. Unless great care is taken to ensure accountability, such decentralized schemes may accomplish no more than the previous efforts.

Notes

Rajshri Jayaraman is a doctoral student in the Department of Economics at Cornell University. Peter Lanjouw is an economist in the Development Economics Research Group of the World Bank and Professor of Economics at the Free University of Amsterdam, Netherlands, where he is currently on leave.

The authors are very grateful to Jean Drèze for guidance in the preparation of this paper, although they do not wish to implicate him for any remaining errors and omissions. They also thank four reviewers for their suggestions, as well as Harold Alderman, Zoubida Allaoua, James Fairhead, Barbara Harriss-White, Jesko Hentschel, Ravi Kanbur, Valerie Kozel, Takashi Kurusawa, Jenny Lanjouw, Michael Lipton, Lant Pritchett, Martin Ravallion, Naresh Sharma, and participants at a seminar held at Queen Elizabeth House, Oxford University, for valuable comments.
References

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


Harriss, John. 1989. “Knowing about Rural Economic Change.” In Bardhan, ed.
hold Mobility, and the Politics of an ‘Awkward Class.’” In Hazell and Ramasamy, eds.
Harriss-White, Barbara. 1992. “Rural Poverty in India: Micro-Level Evidence.” In Barbara Harriss-
University Press.
the ODA-funded project between Queen Elizabeth House, Oxford, and Madras Institute of De-
Distribution of Capital in a South Indian Community.” Journal of International Development
Iqbal, Farrakh. 1988. “The Determinants of Moneylender Interest Rates: Evidence from Rural In-
Villages in Northern Tamil Nadu.” Paper presented at a dissemination workshop for the ODA-
funded project between Queen Elizabeth House, Oxford, and Madras Institute of Development
Case Study from Bihar.” Ph.D. dissertation, Jawaharlal Nehru University, Centre for Economic
Studies and Planning, School of Social Sciences, New Delhi. Processed.
Kapadia, Karin. 1993. “Mutuality and Competition: Female Landless Labour and Wage Rates in
35.
and Political Weekly, Annual Number (February):365–90.
Lanjouw, Peter, and Nicholas Stern, eds. 1998. Economic Development in Palanpur over Five De-
University Press.
erty.” Economic and Political Weekly, Special Number (October):1781–1802.
Economic and Political Weekly (September):A133–44.


Appraising Workfare

Martin Ravallion

Workfare programs aim to reduce poverty by providing low-wage work to those who need it. They are often turned to in a crisis. This article offers some simple analytical tools that can be used to rapidly appraise the cost-effectiveness of an existing workfare operation as a basis for deciding whether the program should be expanded. For pedagogic purposes, two stylized versions of a range of programs found in practice are analyzed: one for a middle-income country, the other for a low-income country. The cost of a given gain to the poor is about the same for both programs, although the components of that cost are very different, with implications for the timing of benefits. The author points to program design changes that could enhance the impact on poverty.

Workfare programs require participants to work in order to obtain benefits. Such programs have been widely used for fighting poverty in crises such as macroeconomic or agroclimatic shocks in which large numbers of poor, able-bodied people have become unemployed. Such interventions are relatively complex and difficult to evaluate. Other things being equal, the appraisal and design should be well informed and rigorous. But time is short in a crisis, and data are often far from ideal. What can be done to obtain a reasonably credible and yet rapid assessment of the likely gains to the poor from a given outlay on a workfare program?

This article offers a mini-manual for the rapid appraisal of an existing workfare program to determine if its expansion would be cost-effective. By “rapid appraisal” I mean that the work can be done by two people in about two weeks with the sort of data normally (though not invariably) available at short notice. Box 1 summarizes the data requirements. I assume that the appraisal must address two main questions: How much impact on poverty can be expected from outlays on the existing program? How might the program be modified to enhance the gains to the poor?

To illustrate, I consider two stylized versions of the programs found in practice. The first is in a middle-income country in which unemployment has risen sharply in the wake of a macroeconomic stabilization and reform program. The other is in a
Box 1. Data for a Rapid Appraisal

The rapid appraisal method proposed here requires information on

- The poverty rate in the relevant country or region
- The wage rate of unskilled (informal sector) labor
- The unemployment rate among the poor
- The labor intensity of current workfare projects, their (financial) benefit-cost ratio, their cost recovery rate, and the extent to which the projects are targeted to poor areas.

The likely sources of this information are household or labor force surveys, the project’s administrators (both central and local), and interviews with participants in the program.

A rapid appraisal cannot normally quantify the distribution of benefits among the poor, so only the aggregate transfer from the budget is estimated here. I note, however, some of the qualitative ways in which sensitivity to distribution among the poor (notably by putting higher weight on gains to the poorest) can bear on the appraisal and design of the program. I also note the implications of attaching a high value to reducing poverty in terms of current incomes, as is often the case in a crisis.

The program’s budget is taken to be predetermined. The issues are how cost-effective the program is in raising the incomes of the poor, and how performance might be improved. Such cost-effectiveness calculations can be deceptive if the budget is not fixed. This can happen if the design of the program also affects the resources available, by influencing how generously the nonpoor support the program. External benefits that accrue to the nonpoor from the assets created can help mobilize broad public support. For example, insurance benefits to nonpoor participants were a factor in public support for workfare schemes in rural areas of South Asia (Ravallion 1991). However, it is unclear how important such considerations are in a crisis. I return to this point later.

A further limitation (in common with other estimates of cost-effectiveness) is that I largely ignore benefits that come in the form of better insurance. This limitation could well be serious because insurance against income losses is thought to be a significant benefit from workfare programs in practice (Ravallion 1991). I do, however, consider ways in which program design can enhance risk benefits.

Any public program must be assessed relative to the best alternative use of the same resources. The best alternative will vary with country circumstances, including administrative capabilities. One option that is probably feasible everywhere is a uniform distribution of the program budget to every household (whether poor or not). If the transfer to the poor as a percentage of total spending on the program is less
than the percentage of households that are poor, then the uniform allocation is preferable. Of course, a workfare scheme that passes this test may still be inferior to some other option; in highly industrialized countries and some transition economies, for example, a well-designed unemployment insurance scheme might be feasible and more cost-effective (Wilson and Fretwell 1996).

An Overview of the Programs

A workfare program can reduce poverty in two ways: by providing paid work for the unemployed from poor households, and by producing goods or services that poor families value. Workfare will naturally be more labor intensive than if the government simply maximized the present value of the assets created because the workfare program attaches positive value to employing poor people, independent of the gains to society as a whole from the outputs obtained. So a workfare program will tend to operate at a point where there is a tradeoff between the value of the assets created and employment (figure 1). The program will operate to the right of the point that maximizes the present value of the assets created.

Figure 1. The Tradeoff in a Workfare Program
This tradeoff poses a difficult question: How much emphasis should be given to immediate employment versus creation of durable assets? The stylized program in MINC puts relatively more emphasis on the assets created than does the LINC scheme. Municipal governments in MINC appear to use the scheme as an extra source of funds for maintaining or upgrading minor roads, sanitation facilities, and so on, using roughly the same combination of labor and nonlabor inputs as is customary. For example, a number of subprojects entailed connecting new dwellings under construction (clearly for well-to-do households) to the sewerage system. In others, pavements were repaired in well-to-do neighborhoods. In LINC the subprojects are mainly minor roads, soil conservation, reforestation, and irrigation, and the technology used tends to be somewhat more labor intensive than that used in similar projects outside the workfare program.

In neither country are the subprojects targeted to poor communities per se. In MINC the projects are just as likely to be in nonpoor neighborhoods, and in LINC the beneficiaries of rural development projects are often relatively well-off local landowners. In both cases, the work done clearly has some value to the community at large, although the projects in LINC seem unlikely to pass a conventional cost-benefit test. The municipal or provincial government usually provides cofinancing to cover the nonwage cost; local residents or nongovernmental organizations usually do not provide cofinancing. Cost recovery is rare, even from well-off beneficiaries.

The Arithmetic of Cost-Effectiveness

The share of the government's outlay that benefits the poor—the cost-effectiveness ratio—can be decomposed into various components that are either estimated from the available data or calibrated from plausible assumptions. Here I suggest a decomposition that I have found useful in practice, although there are other possibilities; my aim is to provide an example that can be adapted to specific circumstances.

The cost-effectiveness ratio can be decomposed exactly into five other variables:

(i) **Budget leverage.** The government can require cofinancing from nonpoor neighborhoods for subprojects that will benefit them. Let government (central plus local) spending be $G$, and let this spending be leveraged up to result in a total budget of $G + C$, including private cofinancing ($C$).

(ii) **Labor intensity.** Some of the participants may not be poor, so let the share of all wages paid in total operating cost be $(W + L)/(G + C)$, where $W$ is the wage received by the poor and $L$ denotes leakage to the nonpoor.

(iii) **Targeted labor earnings.** This is the proportion of the wages paid out to poor workers, $W/(W + L)$.

(iv) **Net wage gain.** This is the share of the gross wage received by the poor after subtracting all costs of participation, including income forgone from other work.

The net wage gain is \( NW/W \), where \( NW \) stands for wages net of forgone income or other costs of participation.

(v) *Indirect benefit.* Let \( IB \) denote the indirect benefits to the poor, such as when the assets created are local public goods in poor neighborhoods.

The total gain to the poor is \( B = NW + IB \), which, as a proportion of public spending on the program, gives the cost-effectiveness ratio:

\[
\frac{B}{G} = \frac{G + C}{G} \cdot \frac{W + L}{W} \cdot \frac{W}{W + L} \cdot \frac{NW}{W} \left( 1 + \frac{IB}{NW} \right).
\]

It is useful to further decompose the indirect benefit to net wage ratio \( (IB/NW) \) as:

\[
\frac{IB}{NW} = \frac{SB}{G + C} \cdot \frac{NW}{G + C}.
\]

This gives the net wage ratio as the product of a further three ratios:

(vi) *Targeted indirect benefits.* Let the social benefits (to the whole population) from the assets created be \( SB \). The share going to the poor is then \( IB/SB \).

(vii) *Benefit-to-cost ratio for the project.* This is simply the ratio of \( SB \) to total cost, \( G + C \).

(viii) *The share of net wage gains in total cost.* This can also be written in terms of three of the ratios in equation (1):

\[
\frac{NW}{G + C} = \frac{NW}{W} \cdot \frac{W}{W + L} \cdot \frac{W + L}{G + C}.
\]

in which the labels iv, iii, and ii correspond to the ratios in equation (1).

Some benefits, particularly the bulk of the indirect benefits, accrue in the future. One can also measure the cost-effectiveness ratio in terms of current benefits \( (CB) \) by replacing all values in these formulas with current values, or values within some specified period. I define the “current period” as the period during which indirect benefits are negligible.

In the above formulation, cost recovery from the nonpoor will increase the budget leverage ratio, \( (G + C)/G \), but will not change other variables. One can explicitly introduce the cost recovery rate, \( k = C/(SB - IB) \), that is, the ratio of the privately financed component of the total cost to the amount of the total benefit that does not accrue to the poor. One can then obtain the following formula for the budget leverage ratio:
\[
\frac{G + C}{G} = \left[ 1 - k \left( \frac{IB}{SB} \right) \cdot \frac{SB}{G + C} \right]^{-1}
\]

in which the labels vi and vii correspond to the ratios in equation (2).

**Impact on Labor Earnings**

In discussing how best to estimate plausible values for the cost-effectiveness ratio, I focus initially on the workfare scheme in MINC; with the basic ideas in place, the application to LINC will be straightforward. Let the data sources be the MINC Statistical Bureau (MSB) and the LINC Statistical Bureau (LSB).

There is a strong association between poverty and unemployment in MINC. Tabulations from the MSB’s recent national sample survey indicate an unemployment rate of 40 percent in the poorest decile of households ranked by income per person, compared with 20 percent for all households. The unemployment rate falls steadily as income per person rises.

**The Wage Rate**

As is typically the case in workfare programs, no means test is applied. The scheme aims to self-select the poor and provide work for as many people as possible without undermining their incentive to take regular jobs when they are available. This approach will work if the workfare wage rate is low enough. (It should not be forgotten that the program has a budget constraint; increasing the wage rate for poor workers means that fewer poor people would benefit from the program.) A relatively low wage rate reduces the need to ration the number of workfare jobs and so enhances the risk benefits to poor people by providing a reliable fallback in times of need (Ravallion 1991; Ravallion, Datt, and Chaudhuri 1993).

The statutory minimum wage in MINC is $250 a month, well above the current monthly workfare rate of $200. (As an emergency employment program targeted to the poor, workfare is exempt from the minimum wage rate statute.) The minimum wage, however, may also be above the market wage, given that enforcement is difficult in most developing countries. So how does the workfare wage compare with the market wages received by the poor?

In the poorest 10 percent of MINC households (ranked by household income per person), the average monthly earnings for the principal job (when it entailed at least 35 hours of work a week) were $330, well above the workfare wage. The poorest
decile received the lowest average wage among all deciles. Average wages for the second poorest and all higher deciles were more than double the workfare wage. On the basis of these data, it is reasonable to assume that the prevailing workfare wage will be unattractive to anyone who is not considered poor in MINC, and it is unlikely to attract poor workers out of their current job.

**Net Wage Gain**

The net wage gain (NW/W) is probably the most difficult variable to estimate in equation (1) and, possibly for this reason, it is often set to unity. This would be justified if labor supply for a workfare scheme came only from unemployment and if no other participation costs were incurred by the poor. But this assumption is difficult to accept. Poor people cannot afford to be idle. Some time will be devoted to informal, often family-run, farms or other enterprises. (This is less work than needed, but it still creates some income.) Even if all workers were unemployed at the time they joined the scheme, they would not necessarily have remained unemployed had the program not existed. Even a worker who has been unemployed for some time typically faces a positive probability of finding extra work, including self-employment or some informal sector activity. Joining the program will leave less time for search. So the net income gain will be lower than the gross wage rate paid. How much lower?

Consider a typical unemployed poor worker who is searching for a job at the time the program is opened. Without the program, the worker faces a probability $P'$ of finding extra work of some sort, at a wage $W'$. So expected earnings without the program are $P'W'$. (One can interpret $P'$ as the proportion of time in which work would otherwise be found during the workfare period.)

Now introduce the workfare program. Let the probability of finding alternative work while working on the program be $P$ (which may not be the same as $P'$). The workfare wage is $W$. The expected income gain when the program becomes available is then $PW' + (1 - P)W$. So the expected net wage gain (NW) to workers from introducing the scheme is:

$$NW = (1 - P)W - (P' - P)W'.$$

Suppose, for example, that joining the scheme means that the worker can no longer search for a regular job and hence has zero chance of getting one ($P = 0$). Then the expected gain is $W - P'W'$, that is, the program's wage minus expected earnings from finding a regular job. The example does not seem plausible in this setting, however. The worker can still search in nonwork hours, and participation in workfare may help in getting a regular job (by the extra experience and possibly the extra knowledge of work opportunities) sufficiently to compensate for the lost search time.
So suppose instead that joining the scheme has no effect one way or the other on the probability of finding regular work; \( P' = P \). Then the expected gain is \( (1 - P)W \), that is, the proportion of the worker's time that would otherwise be unemployed times the wage rate.

In one special case, the calculation of net wage gain is greatly simplified. This occurs when no extra nonworkfare employment is available to the poor with or without the program, that is, \( P' = P = 0 \). Then any income forgone by a workfare participant will be made up by an equal gain to a poor nonparticipant. Employment for the poor is then a zero-sum game. Because poor workers as a whole will forgo no income, \( NW/W = 1 \). It appears that this special case of zero forgone income is often (at least implicitly) assumed in discussions of workfare schemes. But zero forgone income does not seem plausible on a priori grounds, even for the poor as a whole, as discussed earlier.

What are reasonable assumptions for MINC? As already noted, the MSB survey indicated that 40 percent of those in the poorest decile were unemployed. If a worker in the poorest decile who is choosing between the program and the labor market does not accept a workfare job, he will no doubt find some work. Assuming that he has the average probability of being employed at the average wage received by workers in that decile, he will be employed 60 percent of the time at a monthly rate of $330. His expected wage if he does not accept employment through workfare is then about $200. So the current wage rate in the program of $200 turns out to be the minimum expected wage needed to attract the average worker in the poorest decile out of unemployment.

This calculation is based on averages. Actual gains to participating workers will be distributed around these averages; some workers will face relatively low chances of finding a full-time job or even casual part-time work while searching for a full-time job. Such workers will find the workfare wage rate more attractive. There are also regional differences; the same wage will be more (less) attractive in low-cost (high-cost) regions, and unemployment rates will differ from region to region. The gains from workfare will be found among those who face below-average prospects of other employment, or below-average wage rates—or both—or who live in areas where the cost of living is relatively low.

With such a high average unemployment rate in the poorest deciles, it is not unreasonable to presume that participants face unusually low prospects of finding full-time work during their spell of workfare employment. How much lower is hard to say. If program participants in MINC face a 50 percent higher unemployment rate than the poorest decile, then \( P = 0.4 \). Also assume that joining the program has no effect on the probability of finding a regular job. So the expected net benefit for those joining the workfare program will be 60 percent of the program wage rate. This is close to data-based estimates on the net wage ratio for an Argentinean workfare program that was implemented in 1997 (Jalan and Ravallion 1998).
The Cost-Effectiveness of the MINC Program

One can now make a rapid appraisal of the cost-effectiveness of the workfare program in MINC. Because private cofinancing is negligible, I set \( C = 0 \). The central government’s accounts indicate that its own contributions (entirely for workfare jobs) represented one-third of total cost, so \((W + L)/(G + C) = 1/3\). Because the MINC wage rate discussed earlier is very unlikely to be attractive to people who are not poor, I set \( L = 0 \). From the data and assumptions discussed in the previous section, I have assumed that \( NW/W = 0.6 \). Then \( NW/(G + C) = 0.2 \).

Because MINC makes no explicit attempt to target poor areas, the poor are as likely as the nonpoor to benefit indirectly from the projects; so \( IB/SB = 0.2 \) (the poverty rate in MINC). The projects in MINC produce benefits sufficient to cover their cost; \( SB/(G + C) = 1 \). (This reflects the fact that the labor intensity is about the average for similar public works projects.) Together, these assumptions yield \( BI/NW = 2 \). Later I consider alternative assumptions.

Combining these numbers, the value of \( B/G \) implied by equation (1) is 0.40. Equivalently, it takes $2.50 to increase incomes of the poor by $1. Assuming that all of the indirect benefit is in the future, the \( CB/G \) ratio is 0.20, so it takes $5 to transfer $1 to the poor today.

Because the poverty rate in MINC is 20 percent, the \( B/G \) estimate of 0.40 is double the share that poor people would obtain from a uniform, untargeted allocation of the same budget across the whole population (in which everyone gets about the same amount, whether poor or not). In terms of its impact on the current incomes of the poor, however, the workfare scheme does no better or worse than a uniform lump-sum transfer to all households, whether poor or not.

The LINC Program

The LINC scheme operates primarily in rural areas. Unlike in MINC, poverty and unemployment (at least as conventionally measured) are not strongly correlated in LINC. Indeed, data from the LSB indicate that unemployment rates rise as income rises, starting with the poorest, peaking at about the middle of the distribution, and falling thereafter. Substantial underemployment is thought to exist among the poorest families, however; a worker might be classified as employed, yet work for only half the week.

As in MINC, there is no private cost recovery (\( C = 0 \)). The wage rate in the LINC program is tied to a statutory minimum wage rate for agricultural labor that is well above the prevailing wage rate for casual unskilled agricultural labor. The high workfare wage rate thus attracts participants who are not poor or unemployed. As a result, workfare jobs are heavily rationed. Anecdotal evidence from field trips suggests that

Martin Ravallion
when deciding who gets work, local program administrators do not always favor the poorest, either deliberately or because they do not know who is poor. So, unlike MINC, there is definite leakage to the nonpoor in LINC, although the forgone income is probably lower than in MINC. I assume 0.75 for both the targeting of earnings \([W/(W+L)]\) and the net wage gain \([NW/W]\). This is consistent with an estimate of forgone income in an Indian workfare program (Datt and Ravallion 1994).

The indirect benefits to the poor are clearly smaller in LINC than in MINC; the nonpoor landowners capture the bulk of the benefits from the assets created. The poor do receive some indirect benefits, however, notably through second-round effects on employment from higher farm productivity. I assume that the poor obtain one-fourth of the indirect benefits from the program. The high labor intensity means, however, that the social benefits are only sufficient to cover one-half of the cost.

On plugging these numbers into equation (1), LINC’s value of \(B/G\) is almost identical to that for MINC, 0.41, and the cost of transferring $1 to the poor is also about $2.50 under LINC’s program. As in MINC, it is unlikely that any of the indirect benefits will raise current incomes (within a few months, say). The current benefit ratio is 0.28 (this is \(CB/G\), as given by the value of \(B/G\) when \(IB = 0\)). So it costs $3.55 to increase the current earnings of the poor by $1 with LINC’s program. Recall that the poverty rate in LINC is 50 percent. So the absolute gain to the poor from an untargeted allocation of the same gross budget is higher than the gain from the program.

Table 1 summarizes the cost-effectiveness calculations for these two stylized programs under the base-case assumptions discussed above. (Costs are rounded off to the nearest $0.10.)

Table 1. Cost-Effectiveness of the Two Workfare Programs under the Base-Case Assumptions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Middle-income country</th>
<th>Low-income country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget leverage: ((G+C)/G)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Labor intensity: ((W+L)/(G+C))</td>
<td>0.33</td>
<td>0.5</td>
</tr>
<tr>
<td>Targeting: (W/(W+L))</td>
<td>1.0</td>
<td>0.75</td>
</tr>
<tr>
<td>Net wage gain: (NW/W)</td>
<td>0.6</td>
<td>0.75</td>
</tr>
<tr>
<td>Poor people’s share of total benefits: (IB/SB)</td>
<td>0.2</td>
<td>0.25</td>
</tr>
<tr>
<td>Benefit/cost ratio: (SB/(G+C))</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Current + future gains to the poor per $1 of spending: (B/G)</td>
<td>0.40</td>
<td>0.41</td>
</tr>
<tr>
<td>Cost of $1 gain to the poor (dollars)</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>Current earnings gain per $1 of program spending: (CB/G)</td>
<td>0.20</td>
<td>0.28</td>
</tr>
<tr>
<td>Cost of $1 extra current earnings (dollars)</td>
<td>5.00</td>
<td>3.60</td>
</tr>
</tbody>
</table>

Note: The poverty rate is assumed to be 20 percent in the middle-income country and 50 percent in the low-income country.
Comparisons with Other Safety Net Operations

One must be cautious in comparing these estimates with cost-effectiveness ratios for other programs because the numbers are often not strictly comparable. For example, the numbers in table 1 include forgone incomes, but these are often ignored in other estimates of cost-effectiveness ratios. There are also systematic differences in the target group; for example, workfare reaches able-bodied adults, while child nutrition programs do not directly do so; rather than choosing between them, a government may need both types of program to provide a comprehensive safety net.

Assessments of safety net programs in several middle- and low-income countries show a wide range of cost-effectiveness. For cash transfer programs in Eastern Europe, Subbarao and others (1997: table 3.5) present estimates of the proportion of the public transfer going to the poor ranging from 19 to 58 percent. The same source (table 4.2) also estimates the leakage to the nonpoor from targeted food programs for several developing countries. The proportion of the total transfer to the poor ranges from 19 to 93 percent. The latter figure is an outlier; excluding it, the range is 19 to 69 percent. For food subsidy programs in India, Radhakrishna and others (1997) estimate that the share of expenditure reaching the poor is 16 to 19 percent. For housing subsidies in various countries, Subbarao and others (1997: table 4.5) estimate the share going to households below the median income at between 10 and 50 percent.

This wide range of experience makes generalizations difficult. But in terms of the impact of my stylized workfare programs on current incomes, one might well do worse with other instruments, particularly subsidies to goods for which demand rises with income. One could probably do better, however, or at least no worse, with an untargeted lump-sum transfer. Factoring in the estimated future income gains to the poor, the workfare programs start to look better than many other safety net operations, including untargeted lump-sum transfers, in MINC but not in LINC.

Risk Benefits

The benefits from lowering the risk of reduced income are rarely factored into calculations of cost-effectiveness. How would their inclusion affect these comparisons? The risk benefits from a good workfare program can be large, as has often been demonstrated in famines (Ravallion 1997a). Even in normal times, existing (market and nonmarket) arrangements for insurance leave poor people exposed to risk (see, for example, Jalan and Ravallion, forthcoming). The risk benefits depend on the degree of risk aversion and the effect of the safety net on the riskiness of incomes, which will depend in turn on how flexibly the program responds to changing household circumstances. In this respect, some safety net programs are quite unresponsive.
and therefore are as ineffective as insurance; ration cards for subsidized foods, for example, are often held for long periods and are hard to get quickly. Workfare schemes are more responsive to income risk provided the work is easily obtained when needed. That will depend on wage rates and the budget. If the wage is set so high that jobs are heavily rationed (given the available budget), the scheme will not provide reliable insurance for the poor.

This is more of a concern in LINC, yet its risk-prone rural economy is a setting in which protection is greatly needed. Insurance is difficult for the poor to obtain without incurring unacceptably high costs (including lost opportunities for escaping poverty in the longer term through potentially risky investments in human and physical capital). A full accounting of the risk benefits would probably make the MINC scheme look better but would have less effect on the calculations for LINC.

Options for Enhancing Performance

Rapid appraisals can also indicate ways in which the program’s impact on poverty might be improved. Box 2 provides a checklist of recommendations for a cost-effective workfare scheme.

---

**Box 2. Elements of a Good Workfare Program**

- The wage rate should be no higher than the market wage for unskilled manual labor in agriculture or the informal sector during a normal year in the setting in which the program is introduced.
- Restrictions on eligibility or other forms of rationing should be avoided; ideally the only requirement for eligibility should be the fact that work is wanted at this wage rate.
- If rationing is unavoidable (because demand for work exceeds the budget available at the wage set), then the program should be targeted to poor areas and confined to the time periods in which hardship appears to be greatest. Flexibility should be allowed in future budget allocations, however, to reflect any revealed differences in demand for the scheme.
- The labor intensity (share of wage bill in total cost) should be higher than normal for similar projects in the same setting. How much higher will depend on the relative importance attached to immediate income gains versus (income and other) gains to the poor from the assets created. This proportion will vary from setting to setting.
- The subprojects should be targeted to poor areas to ensure that the assets created are of maximum value to poor people. Any exceptions—in which the assets largely benefit the nonpoor—should require co-financing from the beneficiaries, and this money should go back into the workfare budget.
- Performance should be monitored using careful evaluation.
The Wage Rate

The wage rate for the MINC program seems about right, but the LINC wage is too high given the budget. As a result the LINC program has fewer jobs to offer than it could have, and the jobs that are available often do not go to the poor. A wage rate that is no higher than the market wage for unskilled agricultural labor in a normal agricultural year will reduce leakage, provide wider coverage of the poor with the current budget, provide better insurance, and protect incentives to take up normal work when it becomes available. It can be safely assumed that anyone who is willing to do unskilled manual labor for that wage in LINC is poor. So (assuming that the other ratios are unchanged), this alteration to the LINC program would bring the benefit ratio up to 0.50 and reduce the cost of a $1 gain to the poor to $2; LINC's workfare program would then do as well as a uniform lump-sum transfer. The current benefit ratio \((CB/G)\) would rise to 0.375—still less than a lump-sum transfer.

Organized labor is likely to resist this reform, arguing that the government cannot undercut its own statutory minimum wage rate (although this wage is not enforceable). Labor unions in MINC initially took the same position, but the counterargument—that an exception should be made for emergency programs—won the day there.

Cost Recovery

Another way to enhance the scheme's impact on poverty is to introduce cost recovery for the benefits accruing to the nonpoor. If costs are recovered at a rate of 25 percent (but all other assumptions of the base case are retained), the value of \(B/G\) in MINC rises to 0.50, bringing the cost of transferring $1 to the poor down to $2. For LINC, \(B/G = 0.45\). At a cost recovery rate of 50 percent, the value of \(B/G\) in MINC reaches 0.67, bringing the cost of an extra $1 to the poor down to $1.50. In LINC, a cost recovery rate of 50 percent is enough for \(B/G\) to reach 0.50. At a cost recovery rate of 75 percent, the cost of transferring $1 to the poor falls to $1 in MINC and to about $1.75 in LINC. Clearly, such initiatives could greatly improve program performance in reaching the poor in both countries.

Labor Intensity

Increasing the share of outlays spent on labor can greatly enhance the effectiveness of workfare programs in raising current incomes of the poor. The labor share for some projects in MINC is very low; about one-fifth are electricity and gas projects, with an average labor share of 10 percent. If these were entirely privately cofinanced, the low labor share would not be a concern. More typically, however, such projects, to be justified, would have to yield large indirect benefits to the poor, and that they should
do so seems unlikely. Data for MINC indicate that by dropping these projects, the overall labor share would be 0.40, implying that $B/G = 0.44$; this would mean that it would cost $2.27$ (instead of $2.50$) to transfer $1$ to the poor through the scheme. Indeed, if the MINC program had the labor intensity of the LINC program, and all other characteristics were the same as in table 1, the value of $B/G$ would rise to 0.50, again reducing the cost of a $1$ transfer to the poor to $2$.

Recall, however, that there is a tradeoff between higher labor intensity and the indirect benefits from the program in both countries. As illustrated in figure 1, a workfare scheme will operate at labor intensities that entail a tradeoff between $(W + L)$ and $SB$. So raising the labor intensity will lower the social benefits. For the sake of argument, suppose that a labor share of two-thirds in MINC was enough to drive the social benefits from the projects down to zero. Then $B/NW = 1$. With the other ratios unchanged, the value of $B/G$ would be 0.40, exactly what it is in the current scheme. So as long as it is possible to cover at least some of the cost of the scheme from the outputs generated with a labor intensity as high as that in LINC, it would be better to switch to a more labor-intensive scheme.

Even given the seemingly steep tradeoff with indirect benefits, a more labor-intensive program could be more effective in reducing poverty. The case is even stronger when aiming for high current transfers to the poor.

**Restrictions on Eligibility**

The MINC program restricts eligibility to heads of households. By constraining the family’s own adjustment, however, such restrictions may actually reduce the program’s effectiveness. For example, insisting that only the head of the household can join the program will reduce the net gain to poor families to the extent that other household members have less attractive labor market options and therefore lower forgone income. The best way to raise the net transfer benefit is to let poor households rearrange their own activities so as to take advantage of the workfare scheme.

**Designing a Program to Enhance Indirect Benefits to the Poor**

Field trips to poor areas and discussions with local residents revealed plenty of scope for worthwhile community infrastructure projects in both MINC and LINC. It is not unreasonable to expect (pecuniary and nonpecuniary) benefits to poor people from such projects.

To illustrate the implications for the cost-effectiveness calculations, suppose that the value of the indirect benefits to poor communities from the subprojects in MINC rose to half the total benefits. Keeping all other assumptions the same, $B/NW$ would rise to 2.25, and the overall $B/G$ ratio would rise to 0.70. Thus the cost of transfer-
ring $1 to the poor would fall by more than 40 percent—from $2.50 to about $1.40.
Under the same assumption, the cost of a $1 transfer to the poor in LINC would fall to $1.90. If at the same time the social benefits could be increased to cover three-fourths of the cost (still a conventional benefit-to-cost ratio under 1), a $1 gain to the poor in LINC would cost about $1.50.

Sound project selection is crucial to achieving higher indirect benefits from the assets created by workfare. Technical corroboration of the subproject’s viability helps; is it likely to work on purely technical, engineering grounds? The ex ante appraisal should, however, consider other factors likely to make the indirect benefits more pro-poor. Appraisers can identify subprojects likely to be of value to poor people and reward these using a point system, allowing subprojects to compete for funds. Census-based poverty indicators, for instance, are available at the local level in both middle- and low-income countries. Higher points should then be given to subprojects from poor areas. A municipality’s past success in completing subprojects can also be rewarded with extra points. The point system could also be used to give an incentive to municipal governments in nonpoor areas that are willing to fully finance subprojects in their areas.

Direct community involvement in determining what project is to be funded and how it is to be implemented is highly desirable. This input can serve as a source of information on the likely benefits to the area and as an indicator of the subproject’s longer-term sustainability. It can also help avoid “program capture” by local elites. If a bona fide local community group confirms that the subproject is valuable (even aside from the direct employment benefits), extra points should be allocated.

Other design features can enhance the value of the assets created. Relying as far as possible on workers in the same community as the project will probably improve the quality of the work because the beneficiaries have a personal longer-term interest. A bonus, contingent on successful completion to a standard that can be verified, would also help.

In all cases, the central government must rely on the existing fiscal structure. Systemic factors are likely to influence the ability of some provinces to reach poor areas; for example, budget constraints can mean that poorer provinces will have a harder time targeting their own poor areas (Ravallion 1997b). Central government incentives to improve performance and provide technical assistance in proposing viable projects at the local level could help. Progress in placing subprojects in poor areas is not difficult to monitor, although this will be easier in MINC, where the statistical system is better developed.

When a points system is used and both project selection and the budget allocation are decentralized, a check should be made for horizontal inequality between areas in the minimum number of points for a subproject to be accepted. If there are large differences, then reallocations of the budget may be called for, with more
money going to areas where the minimum points needed for project acceptance are higher. A good information system for monitoring projects can help in all these respects.

Conclusion

This article shows how to conduct rapid appraisals of workfare programs. For illustrative purposes, two programs are considered, one in a middle-income country and one in a low-income country. The programs are stylized versions of those found in practice. The cost of a $1 gain to the poor using the program is about $2.50 in both cases, although the components of that cost are quite different; the poor obtain higher gains in current earnings in the low-income country, reflecting the program’s higher labor intensity. The cost of a $1 gain in current earnings is $5 for the middle-income country and $3.50 for the low-income country. The amount received by the poor from a given public outlay under the program is double what they would receive from a uniform (untargeted) transfer in the middle-income country, but the poor would receive more with such a lump-sum transfer than with workfare in the low-income country. A comparison with the cost-effectiveness ratios of other types of safety net operations suggests that workfare schemes are more effective than poorly targeted food and housing subsidies but not as effective as other options. These comparisons may be deceptive, however; the same costs are not always considered, and the same options are not always feasible.

Selected reforms could enhance the benefits to the poor. It should be possible to switch to more labor-intensive production methods for subprojects in middle-income countries. There is also scope in both middle- and low-income countries for enhancing the indirect benefits to the poor from the assets created. Redistributive cofinancing—whereby cost recovery is applied only to asset creation in nonpoor areas—could also greatly improve cost-effectiveness.

Tradeoffs between some of these options are likely. In particular, too high a labor intensity will mean that the projects yield negligible indirect benefits. Circumstances will no doubt influence the choices made with respect to such a tradeoff. In a crisis situation, it is understandable that officials will opt for high labor intensity. In more normal times, where the political sustainability of the safety net is also an issue, indirect benefits will tend to get greater weight. The present calculations suggest that by any one of these routes—greater cost recovery from the nonpoor, higher labor intensity, or greater indirect benefits—design changes should make it possible to enhance appreciably the gains to the poor from a given outlay.

It cannot be denied that these calculations are rough. Naturally, the more rapid the appraisal, the more assumptions will be needed to make up for missing data. This type of appraisal is no substitute for a rigorous evaluation of a program after it
has been implemented, but it can help inform public choice and program design. It can also help identify key areas where further data and analysis would have a high return.

Notes

Martin Ravallion is with the Development Research Group of the World Bank. For their helpful comments, the author thanks Harold Alderman, Polly Jones, Nora Lustig, W. R. Savedoff, K. Subbarao, Dominique van de Walle, Ann Velenchik, Michael Walton, and participants at seminars at the World Bank.

1. On the arguments for and against workfare programs, see Ravallion 1991; Besley and Coate 1992; Ravallion and Datt 1995; Lipton and Ravallion 1995, section 6; Mukherjee 1997; and Subbarao 1997.

2. An example is the Sri Lankan food stamp scheme: a better-targeted program (with the poor getting a higher share of the budget) was introduced, but this subsequently undermined political support from the middle class, and the poor ended up with less than they had before the reforms. For further discussion of these and other issues of targeting, see Besley and Kanbur (1993) and van de Walle (1998).

References

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


Public Social Spending in Africa: Do the Poor Benefit?

Florencia Castro-Leal • Julia Dayton • Lionel Demery • Kalpana Mehra

Education and health care are basic services essential in any effort to combat poverty and are often subsidized with public funds to help achieve that purpose. This paper examines the effectiveness of public social spending on education and health care in several African countries and finds that these programs favor not the poor, but those who are better-off. It concludes that this targeting problem cannot be solved simply by adjusting the subsidy program. The constraints that prevent the poor from taking advantage of these services must also be addressed if the public subsidies are to be effective.

Public subsidies for social services such as education and health care rest on two basic policy objectives—efficiency and equity. Efficiency gains can be achieved when the subsidies produce external benefits or correct for a market failure. Equity is also an important objective of public spending. Education and health care, in particular, are understood to be basic services that are essential in any fight against poverty. The World Bank’s strategy for poverty reduction, for example, combines broad-based growth with human capital development (World Bank 1990). And for that, public subsidies on investments that enhance human capital must benefit the poor.

To what extent has public social spending in Africa been effective in reaching the poor? To answer this question, this article reviews the benefit incidence of government spending. It finds that government subsidies in education and health care are poorly targeted to the poor and indeed favor those who are better-off. Improving targeting to the poor involves not simply rearranging the public subsidies but also addressing the constraints that prevent the poor from accessing these services. The article examines these issues by reviewing the evidence on the benefit incidence of health spending in seven African countries and education funding in nine countries in the region.
What Is Benefit Incidence?

Measuring the benefits of publicly provided goods to individuals is a matter of long-standing concern in the economics literature. For market-based goods and services, the prices consumers pay can be taken as reflecting underlying values and can be used to yield measures of welfare that can be compared across individuals and over time. But when governments subsidize the provision of private goods (such as health care, education, and many infrastructure services), the supply is usually rationed, and the price paid (if any) does not necessarily reflect the marginal value to individual consumers.

Two broad approaches have been pursued to measure the value to the beneficiaries of government-subsidized goods and services. The first, based on the Aaron and McGuire (1970) methodology, emphasizes the individual’s own valuation of the good (that is, the demand, or virtual, price). The difficulties inherent in estimating these prices led to the development of a less demanding approach (reviewed by de Wulf 1975 and Cornes 1995) that values publicly provided goods at their marginal cost (Brennan 1976). This second approach is called benefit incidence; it combines the cost of providing public services with information on their use to show how the benefits of government spending are distributed across the population (Meerman 1979, Selowsky 1979, van de Walle and Nead 1995). Until recently no such studies had been undertaken in Africa. This article fills that empirical gap.

A benefit incidence analysis involves three steps:

1. Estimating the unit cost per person, or unit subsidy (in current expenditures), of providing a service.
2. Imputing the unit subsidy to households or individuals who are identified (usually from household surveys) as users of the service. Individuals who use a subsidized public service in effect gain an in-kind transfer. Benefit incidence measures the distribution of this transfer across the population.
3. Aggregating individuals (or households) into subgroups of the population to compare distribution of the subsidy among different groups. The most common grouping is by income or a related measure of the welfare of the individual. The studies reported here group individuals according to the total expenditure per capita of the households to which they belong.

Health Spending in Africa

Recent improvements in household survey data that provide information on the welfare of households and their use of public services offer an opportunity to estimate the distribution of government subsidies in the social sectors. This section re-

**Health Care Delivery Systems**

The public health systems in the seven countries are very similar (with the exception of South Africa, which has a much more developed private sector). Typically, public facilities provide more than two-thirds of the medical care in these countries. Private nonprofit (mostly charitable) organizations provide the remaining one-third. In Tanzania, for example, private nonprofit hospitals account for about half of all hospitals and about 3 percent of all health centers. Private for-profit medical care is increasing in most of the countries, but from a low base.

Table 1 shows how households respond to an injury or illness. (Country coverage varies in the tables that follow depending on the availability of data. Thus, for example, Kenya is not included in table 1, and Tanzania is not included in table 2.) These responses reflect the availability, cost, and quality of health services, as well as the circumstances of the individual households. The results are not strictly comparable across countries because the design of the survey instruments is not standardized. Moreover, the results suggest a bias: poorer households are less inclined to report illness than are their better-off counterparts. Perhaps that is because the poor accept illness as a normal feature of life and do not consider it an event. Lower reporting could also occur because poorly educated respondents are less likely to recognize untreated illnesses, a problem that is discussed by Chernikovsky and Meesook (1986) and van de Walle (1995).

Evidence from these countries shows that patterns of treatment are strikingly different across household groups:

- The poor are more inclined to self-treat than are the rich, and they are less likely to seek private modern care.
- The richest groups rely heavily on publicly provided care, particularly in Côte d’Ivoire, Guinea, and Tanzania. Only in South Africa is there evidence of the richest groups opting out of the state system in favor of private care.
- The poor rely mainly on the public system, but the private sector is important for both the poor and the nonpoor in Ghana, South Africa, and Tanzania.
- Interestingly, with the exception of Guinea, there is little reliance on traditional health providers.²

These countries have three-tiered public health systems, with basic clinics and dispensaries at the first level, district-level hospitals at the secondary level, and referral and specialty hospitals at the tertiary level. Resources (and hence services) are generally concentrated at the tertiary level; typically, less than 25 percent of recurrent expenditures accrue to the primary level. The public systems are traditionally
<table>
<thead>
<tr>
<th>Country, year, and quintile</th>
<th>Ill during previous four weeks</th>
<th>Ill during previous four weeks</th>
<th>Ill during previous four weeks</th>
<th>Ill during previous four weeks</th>
<th>Ill during previous four weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Côte d'Ivoire, 1995&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Côte d'Ivoire, 1995&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Côte d'Ivoire, 1995&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Côte d'Ivoire, 1995&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Côte d'Ivoire, 1995&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>30</td>
<td>73</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>50</td>
<td>35</td>
<td>55</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Ghana, 1992&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ghana, 1992&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ghana, 1992&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ghana, 1992&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ghana, 1992&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>33</td>
<td>59</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>58</td>
<td>43</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>24</td>
<td>60</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>32</td>
<td>31</td>
<td>52</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Madagascar, 1993&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Madagascar, 1993&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Madagascar, 1993&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Madagascar, 1993&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Madagascar, 1993&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>20</td>
<td>72</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>34</td>
<td>52</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>South Africa, 1994&lt;sup&gt;a&lt;/sup&gt;</td>
<td>South Africa, 1994&lt;sup&gt;a&lt;/sup&gt;</td>
<td>South Africa, 1994&lt;sup&gt;a&lt;/sup&gt;</td>
<td>South Africa, 1994&lt;sup&gt;a&lt;/sup&gt;</td>
<td>South Africa, 1994&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>12</td>
<td>25</td>
<td>46</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>26</td>
<td>14</td>
<td>9</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>12</td>
<td>42</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>22</td>
<td>27</td>
<td>32</td>
<td>39</td>
</tr>
</tbody>
</table>

— Not available.

<sup>a</sup> The reference period was two weeks, so proportions were multiplied by 2 to make estimates approximately comparable.

<sup>b</sup> Refers to private care received at home.

<sup>c</sup> Includes all other providers.


Subsidized from general revenues, although recently each country has implemented cost recovery at most public health care facilities to help finance services and to improve quality. In almost all countries, health care personnel (particularly physicians) are concentrated in urban areas, where they provide tertiary-level care, and are comparatively scarce in rural areas.

Although resources and services are heavily focused on specialized health care, the main causes of illness and death in all seven countries are preventable and easily treated diseases, such as acute respiratory illness, diarrhea, and malaria. In Madagascar it is estimated that 90 percent of illnesses could be prevented or treated at the primary level, provided the services are of good quality and accessible to the majority of the population (World Bank 1996a:79). In an effort to provide better primary and preventive care, most of these countries have begun to decentralize public health...
care systems. Several have recently modified the structure of their health care systems, but few have actually made major resource reallocations. In Côte d’Ivoire the share of total recurrent expenditures devoted to primary care was scheduled to increase from 35 percent in 1991 to 42 percent in 1995. Instead, the share declined in that year to 32 percent.

Unit Subsidies in Health

Estimates of the unit subsidies for public health care in six African countries are given in table 2. The unit subsidy represents the net current cost to the government of an individual visit to a health facility. It is computed as total recurrent spending on facilities, less any revenue from cost recovery (the amount that is returned to the treasury), normalized by the number of visits. Typically, this figure is obtained from government accounts. In some cases, visits are estimated from the household survey used to identify users of the facility. In others, health ministry data are used. The subsidy for a visit to a health center or primary health clinic is generally less costly to the government than a visit to a hospital, and outpatient visits are substantially less costly than inpatient visits. In Ghana an outpatient visit is one-tenth the cost of an inpatient visit, and in Guinea the ratio is 1 to 7.

Unit cost data are limited in several respects. First, only in some cases—Ghana and South Africa—do the data refer to actual recurrent spending on health facilities; in the other countries, they are based on budgeted expenditures, which may differ significantly from outcomes. Second, there is little disaggregation by type of facility, type of consultation, or region of the country, masking variations in the costs of consultation. The unit costs were generally averaged into two groups—visits to health centers and visits to hospitals. No distinction was made between different types of hospital care (such as secondary and tertiary hospitals). And making a distinction between outpatient and inpatient visits was feasible in only two countries, Ghana

<table>
<thead>
<tr>
<th>Country/year</th>
<th>Monetary unit</th>
<th>Health center</th>
<th>Outpatient</th>
<th>Inpatient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Côte d’Ivoire, 1995</td>
<td>CFAF</td>
<td>1,252</td>
<td>1,787a</td>
<td></td>
</tr>
<tr>
<td>Ghana–Accra, 1992</td>
<td>Cedis</td>
<td>6,489</td>
<td>4,044</td>
<td>49,553</td>
</tr>
<tr>
<td>Ghana–other, 1992</td>
<td>Cedis</td>
<td>1,129</td>
<td>1,275</td>
<td>14,427</td>
</tr>
<tr>
<td>Guinea, 1994</td>
<td>GNF</td>
<td>902</td>
<td>1,321</td>
<td>7,926</td>
</tr>
<tr>
<td>Kenya (rural only), 1992/93</td>
<td>K Shs.</td>
<td>15</td>
<td>151a</td>
<td></td>
</tr>
<tr>
<td>Madagascar, 1994</td>
<td>FMG</td>
<td>1,413</td>
<td>2,136a</td>
<td></td>
</tr>
<tr>
<td>South Africa, 1992/93</td>
<td>DBSA</td>
<td>98</td>
<td>516a</td>
<td></td>
</tr>
</tbody>
</table>

*Average cost of all hospital visits.

Source: For Kenya, see Dayton and Demery (1994); for other countries, see table 1.

Florencia Castro-Leal, Julia Dayton, Lionel Demery, and Kalpana Mehra
Table 3. Benefit Incidence of Public Spending on Health for the Poorest and Richest Quintiles in Selected African Countries

<table>
<thead>
<tr>
<th>Country/year</th>
<th>Primary facilities</th>
<th>Hospital outpatient</th>
<th>Hospital inpatient</th>
<th>All health</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poorest</td>
<td>Richest</td>
<td>Poorest</td>
<td>Richest</td>
</tr>
<tr>
<td>Côte d'Ivoire, 1995</td>
<td>14</td>
<td>22</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td>Ghana, 1992</td>
<td>10</td>
<td>31</td>
<td>13</td>
<td>35</td>
</tr>
<tr>
<td>Guinea, 1994</td>
<td>10</td>
<td>36</td>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>Kenya (rural only), 1992</td>
<td>22</td>
<td>14</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>Madagascar, 1993</td>
<td>10</td>
<td>29</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>South Africa, 1994</td>
<td>18</td>
<td>10</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>Tanzania, 1992/93</td>
<td>18</td>
<td>21</td>
<td>11</td>
<td>37</td>
</tr>
</tbody>
</table>

— Not available.

Note: Hospital subsidies combine in- and outpatient spending in Côte d'Ivoire, Guinea, Kenya, Madagascar, and South Africa.

Source: See tables 1 and 2.
and Guinea. Further, a lack of data on regional health expenditures means that unit subsidies were generally computed at the national level. In Ghana and Madagascar, however, where regional data were available, differences among regions were significant. For example, spending per visit to a primary health care facility in Accra was almost six times that for other areas of Ghana. Such inequalities may also hold in other countries but were masked in the aggregate data to hand. It should be emphasized that the data for South Africa are for 1992–93, to correspond to the household survey year. A great deal has changed since then, with the election of the Government of National Unity. And these changes will undoubtedly influence the benefit incidence of health (and education) spending.

**Who Benefits from Health Subsidies?**

By combining the unit costs of health care delivery with the use of publicly funded health facilities, we can estimate the benefit incidence of government spending on health. For convenience, we report here the benefit incidence of spending to the poorest quintile (that is, the poorest 20 percent of the population, ranked by expenditure per capita) and the richest quintile (table 3).

Two clear messages emerge. First, health spending in Africa is not well targeted to the poorest. Typically the share of the subsidy to the poorest quintile was significantly less than that to the richest 20 percent. The inequality was greater in some countries (notably Côte d'Ivoire, Ghana, Guinea, and Madagascar) than in others, but overall, the poorest 20 percent of the population received less than 20 percent of the subsidy. Moreover, the share received by the richest quintile was far in excess of 20 percent (except in South Africa, where the richer households rely on private care; see table 1). The second message is that health spending is reasonably progressive; the subsidy to the poorest quintile amounts to a higher share of that group’s total household expenditures than did the subsidy to the richest quintile (see table 3). This progressiveness was particularly striking in South Africa but was also true of the other countries. This finding suggests that if the government gave all households an annual income transfer, rather than subsidized health care, income-expenditure distribution would improve, other things being constant.

**Understanding the Benefit Incidence of Health Subsidies**

To understand why health spending is not targeted to the poor in Africa, it is helpful to look at the allocation of health budgets to different levels of service—notably, hospital and nonhospital care—and the poorest quintile’s share of total visits for each level of service. These two measures are obviously related: as governments change the allocations of spending across subsectors, they influence the way households choose
among treatment options, which would in all probability change the quintile shares of health visits. For convenience we examine each in turn.

The allocation of spending across services within the health sector is not favorable to the poor. Governments allocate significant shares of their health budgets to hospital-based services, which the poor generally do not use. In Ghana, for example, two-thirds of the health budget was spent on hospital services; a major portion went to one large teaching hospital in Accra. In South Africa the share allocated to hospitals was 89 percent. And in both Madagascar and Kenya more than half of the health budget was devoted to hospitals. It is safe to say that targeting health spending to the poor in Africa would require spending less on hospitals and more on primary facilities.

Spending on hospital-based health care, however, can be justified to some extent because many large hospitals train medical personnel for lower levels of care. Moreover, one of the reasons why governments subsidize tertiary health services is that there is no insurance market. Households in developing countries cannot insure themselves against the risk of serious illness or injury and the consequent need for very expensive treatment. As the data show, this allocation of the health subsidy can be at the expense of the equity objective, because the poor tend not to use hospital services.

In Kenya, South Africa, and Tanzania, budget reallocations toward primary care would in themselves improve the targeting of spending to the poor. There, the poorest quintiles use primary facilities in good measure, gaining about one-fifth of the primary subsidy—a pattern similar to that found elsewhere in the developing world (Demery 1997). But in the other African countries, budget reallocations alone would not necessarily fix the targeting problem. In Côte d’Ivoire, Ghana, Guinea, and Madagascar the share of the subsidy received by the poorest quintile was low at all levels of health care, including primary facilities. Given the costs and benefits involved, household decisions about using publicly subsidized health care services result in far fewer visits to primary facilities from poor households than from better-off ones. The point is that budget reallocations must be accompanied by increased use of primary facilities by poor households. To identify the interventions that would have this effect, it is necessary to understand why the poor limit their use of publicly funded health facilities.

We consider here five principal factors that affect the use of health services by the poor: income, service quality, access, direct user charges, and gender.

**INCOME.** Health care is a normal good, which means that household spending on health—and the use of health facilities—increases with income (table 4). But as table 1 shows, the richest groups use mainly publicly subsidized health care (except in South Africa). In Ghana the richest quintile directed almost 60 percent of its health spending to the public sector, much of it on hospital consultations (Demery and others 1995). This means that health spending is very unlikely to be targeted to
the poor. Given the fundamental influence of income on the demand for health care, the only way in which public subsidies can be well targeted to the poorest is by diverting the demand for health care by those who are better-off to the private sector. That is no easy task in countries where private health care is generally poorly developed, largely because of the dominance of the public sector. This change has to be considered a long-term objective. In the meantime, are there other factors amenable to shorter-term policy interventions that might mitigate this powerful income effect?

QUALITY DIFFERENCES. Alderman and Lavy (1996) report that the demand for health care is sensitive to the quality of the service provided. Even the poor limit their demand for health care when services are of poor quality. But the poor are less sensitive to changes in quality of service (Lavy and Germain 1994). Thus uniformly poor-quality service would discourage demand more among the rich than the poor, which would be inconsistent with the observed share of each quintile’s participation in health care services. The observed pattern can therefore be explained only by significant differences in the quality of service offered to the rich and the poor. So, for instance, drug availability, staff skills, and the quality of health facilities may vary considerably and to the disadvantage of the poor. Is there any evidence of such variations in quality? A special survey of health facilities designed to accompany the Ghana Living Standards Survey of 1989 suggests that there is (Lavy and Germain, 1994:13). And the earlier discussion of unit subsidies in Ghana implied the presence of large variations in quality (table 2). Similarly, in Antananarivo, the region in Madagascar where most (23 percent) of the poor live, the government unit subsidy for basic

<table>
<thead>
<tr>
<th>Country, year, and quintile</th>
<th>Nonfood expenditure</th>
<th>Total expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Côte d’Ivoire, 1988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>3,347</td>
<td>13.4</td>
</tr>
<tr>
<td>Richest</td>
<td>14,407</td>
<td>6.3</td>
</tr>
<tr>
<td>Ghana, 1992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>2,964</td>
<td>12.7</td>
</tr>
<tr>
<td>Richest</td>
<td>12,452</td>
<td>7.5</td>
</tr>
<tr>
<td>Madagascar, 1993/94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>1,133</td>
<td>6.9</td>
</tr>
<tr>
<td>Richest</td>
<td>4,581</td>
<td>1.5</td>
</tr>
</tbody>
</table>

a. Spending figures for Côte d’Ivoire are in CFA; for Ghana in cedis; and for Madagascar in FMG.

Source: See table 1.
health care was just 41 percent of the subsidy going to the richest region, Antsiranana, with a total poverty headcount of only 7 percent (World Bank 1996a). These comparisons suggest that there may well be differences in the care provided at different health facilities, to the disadvantage of poorer households.

ACCESS AND OPPORTUNITY COSTS. Poor households, which are often some distance from government health facilities, typically face long journeys and high opportunity costs to obtain health care. In South Africa, for example, those in the poorest quintile must travel almost two hours on average to obtain medical attention, compared with an average of 34 minutes for the richest quintile (Castro-Leal 1996a). The Ghana Living Standards Survey of 1992 also recorded longer travel and treatment time for poorer households. Time spent away from economic activity represents much greater private opportunity costs for the poor, who, unlike their salaried counterparts, have to forgo income in order to obtain medical care. These costs can dominate the decision to seek care.

Lavy and Germain (1994) found that halving the distance to public health facilities in Ghana would increase their use among the population at large by an estimated 96 percent. In Kenya distance was also a significant factor in the demand for health care, although not as dramatic as in Ghana (Mwabu, Ainsworth, and Nyamete 1993). Lavy and Germain (1994) found that the poor were willing to pay less than the nonpoor in absolute terms, but more relative to their income, to reduce the distance traveled. Gertler and van der Gaag (1990) found that individuals at the lower end of the income distribution in Côte d’Ivoire were far more sensitive to changes in the time required to obtain care than were those at the upper end. Time, in effect, rations the market. These studies are based on cross-sectional evidence, however, so direction of causation is uncertain; the relationship between use and distance might be capturing the effect of geographic variations in health care utilization on government decisions about placement of health facilities rather than the other way around.

PRICE. The cost of a medical consultation is far more of a burden for the poor. And ample evidence suggests that when prices are raised through cost recovery schemes, the poor are more likely than the nonpoor to cut back on their use of health services (Gertler and van der Gaag 1990; Lavy and Germain 1994). Longitudinal studies based on controlled experiments such as those by Litvack and Bodart (1993) in the Cameroon and by Gertler and Molyneaux (1997) in Indonesia confirm that price increases without compensating improvements in quality discourage utilization by the poor. Increasing user charges, other things being equal, lowers the share of the poor in total visits to health facilities. Charges must therefore be introduced carefully; they must be targeted to services used mainly by the nonpoor; and if applied to services used by the poor, they should be accompanied by improvements in both access and quality.
Table 5. Benefit Incidence of Health Spending by Gender for Selected African Countries

<table>
<thead>
<tr>
<th>Country/year</th>
<th>Quintile</th>
<th>Percentage share</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Côte d’Ivoire, 1995</td>
<td>Poorest</td>
<td>52.9</td>
<td>47.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>55.0</td>
<td>45.0</td>
<td></td>
</tr>
<tr>
<td>Ghana, 1992</td>
<td>Poorest</td>
<td>44.3</td>
<td>55.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Richest</td>
<td>65.0</td>
<td>35.0</td>
<td></td>
</tr>
</tbody>
</table>

Source: See table 1.

GENDER.

Income, quality, access, and price interact with social relationships to produce sharp inequalities in the distribution of health benefits by gender. Females in the top quintiles in Côte d’Ivoire and Ghana, for example, typically use publicly funded health facilities more than do their male counterparts (table 5). But this advantage changes markedly for the poorest quintiles. The gender advantage is largely eroded in Côte d’Ivoire, although poor females still use facilities more than do males, and in Ghana poor females use health facilities less than do males in the same quintile. The reasons for this are unclear. Supply-side effects may account for the difference—the facilities available to the poor may not provide the perinatal care used by their richer counterparts. Or the difference may arise from demand-side household preferences. Poor households may decide that females should not use health facilities, either because of underlying social values favoring males over females or because of higher opportunity costs of female time. Either way, poor households behave differently from rich households, and this difference explains to some extent the weak targeting of the health subsidy to the poor in Africa.

Education Spending in Africa

Education has long attracted government subsidies in Africa, in part because of the expected high social externalities involved, but also because of equity considerations. The case for subsidizing primary education is particularly strong, given the wide benefits it brings. Literacy and numeracy are critical to sustaining modern democracies. And a growing weight of evidence from the endogenous growth literature highlights the favorable growth effects of education (Bruno, Ravallion, and Squire 1996; Demery, Sen, and Vishwanath 1995).

Characteristics of the Education Systems

Formal education in all nine countries for which we have data (Côte d’Ivoire, Ghana, Guinea, Kenya, Madagascar, Malawi, South Africa, Tanzania, and Uganda) includes...
six years of primary school (seven in Tanzania), three years of lower secondary school, three years of senior secondary school, and four years of university education. Most public systems have vocational, technical, and teacher-training programs parallel to the university system. Movement through the educational system is generally determined by student performance in national examinations. The government is the main provider of education in all nine countries, although the size of the private sector varies substantially. At one extreme, for-profit provision of primary education is prohibited in Tanzania (although the number of private secondary schools is increasing dramatically). And at the other, almost 30 percent of primary and secondary students in Accra attend private schools.

Public education is financed by both governments and households. Of total recurrent spending on education in Ghana, the government contributes about 65 percent and households about 35 percent (Demery and others 1995). Household out-of-pocket contributions include school fees, uniforms, books, supplies, and the like. Households also incur opportunity costs (of the time forgone while attending school), as well as transaction costs (mainly transportation to and from school). Attendance fees vary: in Côte d'Ivoire and Tanzania primary schooling is free, but in Ghana and Guinea nominal fees are charged at all levels.

Although all governments consider primary education to be the highest priority, the degree to which the budget reflects this priority varies. The share of the education budget allocated to primary schooling ranged from just 40 percent in Guinea in 1994 to more than 70 percent in Malawi in 1994-95. Wages and salaries dominate the functional categories in the budget. In Ghana and Malawi wages accounted for 94 and 97 percent, respectively, of total costs, and in Tanzania the share was 81 percent. Elsewhere, the distribution between salaries and supplies was not as skewed, with salary expenditures accounting for between one-half and two-thirds of recurrent expenditures.

Enrollment rates vary by education level and household income in the following ways (table 6):

- Enrollment rates in primary schools are generally lower than the average for all low-income countries, although variation among them is substantial.
- Enrollment rates are extremely low at the secondary level, at around 10 to 40 percent, substantially lower than the average for low-income countries worldwide (42 percent for girls and 55 percent for boys in 1993, according to World Bank 1996c). The exception is South Africa, where secondary education is almost universal.
- Enrollment rates are significantly lower for the poor at all levels, and particularly at the secondary level. Again, South Africa is an exception, with both primary and secondary rates among the poorest quintile being close to the national average.
- The overall enrollment rates for boys and girls at the primary level is about the same in many of these countries, but a gap emerges among poorer quintiles. A
large gender gap in enrollments is evident at the secondary level for most income groups.

Low enrollment rates are not the only indicator of poor performance; repetition rates are uniformly high—more than 30 percent—and completion rates are correspondingly very low. In addition, most of the nine countries have a problem with late starting age. In Tanzania more than 80 percent of all primary school students were late in enrolling; the average starting age was 9 for girls and 10 for boys (Mason and Khandker 1997:5).

**Unit Subsidies in Education**

Unit subsidies for education are computed as net recurrent spending (total government recurrent spending less cost recovery to the treasury) per student. In most of the studies reported here, unit subsidies are based on government expenditure data and enrollment estimates from household surveys; in some cases, tertiary enrollments are based on government statistics. For most countries the unit subsidies apply to the country as a whole, taking into account only differences between the levels of education. But for Madagascar and South Africa, it is possible to disaggregate further (box 1). Unit subsidies increase with the level of education, markedly so in Guinea and Malawi (table 7). Typically the outlays for secondary schools are about

---

**Table 6. Gross Enrollment Rates in Primary and Secondary Education for the Poorest and Richest Quintiles in Selected African Countries**

<table>
<thead>
<tr>
<th>Country/year</th>
<th>Primary</th>
<th></th>
<th>Secondary</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poorest</td>
<td>Richest</td>
<td>All</td>
<td>Poorest</td>
</tr>
<tr>
<td>Côte d'Ivoire, 1995</td>
<td>51</td>
<td>99</td>
<td>75</td>
<td>12</td>
</tr>
<tr>
<td>Ghana, 1992</td>
<td>75</td>
<td>101</td>
<td>88</td>
<td>27</td>
</tr>
<tr>
<td>Guinea, 1994</td>
<td>16</td>
<td>84</td>
<td>44</td>
<td>2</td>
</tr>
<tr>
<td>Kenya, 1992</td>
<td>100</td>
<td>108</td>
<td>105</td>
<td>9</td>
</tr>
<tr>
<td>Madagascar, 1993</td>
<td>48</td>
<td>113</td>
<td>83</td>
<td>2</td>
</tr>
<tr>
<td>Malawi, 1994, 1990b</td>
<td>74</td>
<td>133</td>
<td>108</td>
<td>4</td>
</tr>
<tr>
<td>South Africa, 1994</td>
<td>112</td>
<td>97</td>
<td>106</td>
<td>81</td>
</tr>
<tr>
<td>Tanzania, 1993/94c</td>
<td>77</td>
<td>87</td>
<td>81</td>
<td>3</td>
</tr>
<tr>
<td>Uganda, 1992</td>
<td>72</td>
<td>116</td>
<td>93</td>
<td>4</td>
</tr>
</tbody>
</table>

a. Decile averages.
b. Primary enrollment rates are for 1994, and secondary enrollment rates for 1990.
c. Unweighted average of male and female enrollment rates.

Box 1. Disaggregating Unit Subsidies and Education Spending

Where spending is unevenly distributed geographically (or in other ways), the use of aggregate unit subsidies can mask inequality in public spending. In South Africa Castro-Leal (1996b) obtained five levels of unit subsidy based on the budgets of the different “houses” of government, which were divided along racial lines. Unit subsidies varied enormously, but enrollment rates were high, even among the poorest groups receiving the lowest subsidy. In Madagascar unit subsidies were obtained for the six main regions of the country (World Bank 1996b). The subsidies did not vary as much as in South Africa, but enrollment rates declined sharply at low income levels.

Two estimates of the benefit incidence of education spending are reported in the table below. One is based on the disaggregated unit subsidies, while the other is computed using an average unit subsidy at each of the three education levels. In South Africa the disaggregation of unit subsidies makes a significant difference to benefit incidence. For education spending as a whole, the use of average subsidies makes it appear as though each quintile received roughly its proportionate share of the education budget. But in fact, the poorest quintile gained only 14 percent and the richest 35 percent of total education spending because of unit cost variations between the races. The Madagascar estimates tell a quite different story. Here, the use of national average unit subsidies (at each level of schooling) changes the benefit incidence estimates only marginally compared with the use of region-specific unit subsidies. The differences are literally a matter of tenths of a percentage point.

Why the difference between South Africa and Madagascar? Three factors explain this outcome. First, the unit subsidies were far more variable in South Africa than they were in Madagascar, reflecting as they did the years of the apartheid regime. Second, the population within the quintiles was distributed across regions in Madagascar, so that there was some variability in the unit subsidies within quintiles. In South Africa the population in the poorest quintile was almost entirely black, and it was the black population that received the lowest unit subsidy. Third, enrollment rates were uniformly high in South Africa, whereas enrollment rates in Madagascar varied significantly across income groups. The lower enrollment rates among the poorer groups in Madagascar were probably caused in part by the lower unit subsidies allocated to them. When national average unit subsidies are used, variations in the unit subsidy are missed, but their effects on the enrollment patterns across income are captured and are reflected to some extent in the benefit incidence estimates.

<table>
<thead>
<tr>
<th></th>
<th>Share of quintile in total subsidy</th>
<th>South Africa, 1994</th>
<th>Madagascar, 1993</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Primary spending</td>
<td>Secondary spending</td>
<td>Tertiary spending</td>
</tr>
<tr>
<td></td>
<td>Disaggregated Mean</td>
<td>Disaggregated Mean</td>
<td>Disaggregated Mean</td>
</tr>
<tr>
<td><strong>Population quintile</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>18.9</td>
<td>25.8</td>
<td>11.5</td>
</tr>
<tr>
<td>Richest</td>
<td>27.8</td>
<td>13.5</td>
<td>38.6</td>
</tr>
</tbody>
</table>

| Poorest             | 16.8 | 17.2 | 1.9 | 2.0 | 0.0 | 0.0 | 8.2 | 8.3 |
| Richest             | 14.4 | 14.0 | 41.8 | 41.5 | 88.6 | 88.6 | 41.2 | 41.0 |

Source: Castro-Leal (1996b); World Bank (1996a).
Table 7. Unit Education Subsidies by Level in Selected African Countries

<table>
<thead>
<tr>
<th>Country/year</th>
<th>Monetary unit</th>
<th>Primary</th>
<th>Secondary</th>
<th>Tertiary as ratio of primary</th>
<th>Tertiary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costa d'Ivoire, 1995</td>
<td>CFAF</td>
<td>64,840</td>
<td>117,462</td>
<td>1.8</td>
<td>348,453</td>
</tr>
<tr>
<td>Ghana, 1992</td>
<td>Cedis</td>
<td>24,824</td>
<td>65,275</td>
<td>2.6</td>
<td>392,707</td>
</tr>
<tr>
<td>Guinea, 1994</td>
<td>GNF</td>
<td>47,625</td>
<td>116,812</td>
<td>2.5</td>
<td>2,595,705</td>
</tr>
<tr>
<td>Kenya, 1992/93</td>
<td>K Shs.</td>
<td>1,368</td>
<td>3,868</td>
<td>2.8</td>
<td>42,050</td>
</tr>
<tr>
<td>Madagascar, 1994</td>
<td>FMG</td>
<td>50,504</td>
<td>192,491</td>
<td>3.8</td>
<td>1,140,000</td>
</tr>
<tr>
<td>Malawi, 1994/95</td>
<td>Kwachas</td>
<td>220</td>
<td>909</td>
<td>4.1</td>
<td>15,523</td>
</tr>
<tr>
<td>South Africa, 1994</td>
<td>DBSA</td>
<td>1,124</td>
<td>2,055</td>
<td>1.8</td>
<td>5,657</td>
</tr>
<tr>
<td>Tanzania, 1993/94</td>
<td>T Shs.</td>
<td>6,600</td>
<td>7,500</td>
<td>1.1</td>
<td>—</td>
</tr>
<tr>
<td>Uganda, 1992/93</td>
<td>U Shs.</td>
<td>11,667</td>
<td>37,352</td>
<td>3.2</td>
<td>373,525</td>
</tr>
</tbody>
</table>

— Not available.

Source: See table 6.

twice the amount spent on primary schools. Tertiary unit subsidies were significantly greater than other levels.

Who Benefits from Education Subsidies?

Combining the unit cost data with information on the use of publicly subsidized education from household surveys yields estimates of the benefit incidence of government education spending. The subsidy for education, like that for health, is generally progressive but poorly targeted (table 8). In absolute terms, the poorest quintile gains less than 20 percent of the subsidy—significantly less in most cases (Costa d'Ivoire, Guinea, Madagascar, South Africa, Tanzania, and Uganda). The richest quintile gains far more, especially in those same five countries. Yet the subsidy for public education is more equally distributed than household income or expenditure. The monetary benefit to the poor, as a share of total household expenditure, is more than the benefit to the rich, particularly in Kenya and South Africa. Generally education subsidies represent a greater gain to poor households in these countries than do health subsidies.

The high share of the primary school subsidy imputed to the poorest quintile, shown in table 8, is misleading because the education needs of this group are so much greater than those of other groups. In both Costa d'Ivoire and Ghana, the share of primary-school-age children in the poorest quintile is much greater than the share of subsidies that quintile receives (table 9). The contrast between needs and benefits is even more striking in the case of secondary school subsidies. The poorest quintile in Costa d'Ivoire accounts for 21 percent of secondary-school-age children but receives only 7 percent of the subsidy.
Table 8. Benefit Incidence of Public Spending on Education in Selected African Countries  
(percent)

<table>
<thead>
<tr>
<th>Country/year</th>
<th>Primary subsidy</th>
<th>Secondary subsidy</th>
<th>Tertiary subsidy</th>
<th>Total subsidy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poorest</td>
<td>Richest</td>
<td>Poorest</td>
<td>Richest</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>Richest</td>
<td>Poorest</td>
<td>Richest</td>
</tr>
<tr>
<td></td>
<td>Poorest</td>
<td>Richest</td>
<td>Poorest</td>
<td>Richest</td>
</tr>
<tr>
<td>Côte d'Ivoire, 1995</td>
<td>19</td>
<td>14</td>
<td>7</td>
<td>37</td>
</tr>
<tr>
<td>Ghana, 1992</td>
<td>22</td>
<td>14</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Guinea, 1994</td>
<td>11</td>
<td>21</td>
<td>4</td>
<td>39</td>
</tr>
<tr>
<td>Kenya, 1992</td>
<td>22</td>
<td>15</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Malawi, 1994</td>
<td>20</td>
<td>16</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>Madagascar, 1993</td>
<td>17</td>
<td>14</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>South Africa, 1994</td>
<td>19</td>
<td>28</td>
<td>11</td>
<td>39</td>
</tr>
<tr>
<td>Tanzania, 1993/94</td>
<td>20</td>
<td>19</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>Uganda, 1992</td>
<td>19</td>
<td>18</td>
<td>4</td>
<td>49</td>
</tr>
</tbody>
</table>

- Not available.

Source: See table 6.
Table 9. Benefit Incidence and Education Needs in Côte d'Ivoire and Ghana

<table>
<thead>
<tr>
<th>Country/quintile</th>
<th>Primary</th>
<th></th>
<th>Secondary</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Share of subsidy</td>
<td>Share of school-age population</td>
<td>Share of subsidy</td>
<td>Share of school-age population</td>
</tr>
<tr>
<td>Côte d'Ivoire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>19.1</td>
<td>23.8</td>
<td>6.8</td>
<td>20.9</td>
</tr>
<tr>
<td>Richest</td>
<td>13.9</td>
<td>13.5</td>
<td>37.2</td>
<td>20.9</td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>21.8</td>
<td>24.3</td>
<td>14.9</td>
<td>20.4</td>
</tr>
<tr>
<td>Richest</td>
<td>14.1</td>
<td>13.7</td>
<td>18.6</td>
<td>16.8</td>
</tr>
</tbody>
</table>

Source: See table 6.

These demographic differences across the quintiles arise in part because of the use of per capita household expenditures as the welfare measure; as a result of that measure, poor households are both larger and have more children than better-off households. Because of this, they gain a significant proportion of the primary education subsidy. If, instead, per adult equivalent expenditures were used, these demographic differences might disappear (Lanjouw and Ravallion 1994). To see whether their estimates were sensitive to the measures of welfare used, Demery and others (1995) normalized household expenditures on both household size and adult equivalence (using a scale proposed in Deaton and Muellbauer 1986) and found that spending was significantly less targeted to the poorest under the revised welfare measure and that a larger share went to the richest quintile (table 10). The exercise confirmed the sensitivity of the benefit incidence results to the welfare measure (see also van de Walle, Ravallion, and Gautam 1994; Jarvis and Micklewright 1995).

<table>
<thead>
<tr>
<th>Level/quintile</th>
<th>Adult equivalent expenditures</th>
<th>Per capita expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>17.4</td>
<td>21.8</td>
</tr>
<tr>
<td>Richest</td>
<td>19.7</td>
<td>14.0</td>
</tr>
<tr>
<td>Secondary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>18.6</td>
<td>14.9</td>
</tr>
<tr>
<td>Richest</td>
<td>16.6</td>
<td>18.6</td>
</tr>
<tr>
<td>Tertiary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>9.5</td>
<td>6.0</td>
</tr>
<tr>
<td>Richest</td>
<td>29.8</td>
<td>45.2</td>
</tr>
</tbody>
</table>

Source: Demery and others (1995).
Understanding the Benefit Incidence of Education Subsidies

Unlike health, the share of the education subsidy accruing to the poorest varies noticeably by level of service. The poorest quintile is seen to gain far more from spending at the primary level—typically about one-fifth of the subsidy, compared with about one-tenth of the subsidy at the secondary level and almost nothing from the subsidy at the tertiary level. Thus the more governments spend on primary education, the more the poor will benefit.

Yet, understanding the differential enrollment rates remains important if governments are to improve the targeting of education subsidies to the poor. Although a rich and growing literature exists on the constraints facing the poor in accessing health services, far less evidence is available on the demand for education in Africa. Again, the major determinants of demand are income, quality, and costs (opportunity and direct costs).

Table 11 shows how much the poorest and richest income groups in three countries spend on education. Typically, rich households spend more than the poor. Unless better-off groups can be encouraged to use private service providers, especially at the secondary and tertiary levels, it is difficult to envisage how government education subsidies can be better targeted to the poor.

Education systems in most of the countries reviewed here need to be improved. Less well documented is variation of service quality within a country and the extent to which the poor are disadvantaged. Unit cost variations suggest that the services provided in poorer rural communities are inferior to those extended to urban-based

<table>
<thead>
<tr>
<th>Country/quintile</th>
<th>Year</th>
<th>Education spending per capita</th>
<th>Education spending as percentage of Nonfood expenditures</th>
<th>Total expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Côte d'Ivoire</td>
<td>1995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td></td>
<td>2,083</td>
<td>5.1</td>
<td>2.6</td>
</tr>
<tr>
<td>Richest</td>
<td></td>
<td>23,964</td>
<td>6.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Ghana</td>
<td>1992</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td></td>
<td>1,924</td>
<td>8.3</td>
<td>3.0</td>
</tr>
<tr>
<td>Richest</td>
<td></td>
<td>6,872</td>
<td>4.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Madagascar</td>
<td>1993/94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td></td>
<td>338</td>
<td>2.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Richest</td>
<td></td>
<td>3,000</td>
<td>1.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

a. Spending figures for Côte d'Ivoire are in CFAF; for Ghana in cedis; and for Madagascar in FMG.
Source: See table 6.
schools. Case and Deaton (1998) reported that pupil-teacher ratios in black schools were more than twice those in white schools under the former regime in South Africa, and their econometric results suggest that this policy discouraged school enrollment and educational attainment among black households. Evidence from Uganda shows that the amount that actually reaches rural schools is much less than aggregate data would suggest. Ablo and Reinikka (1998) found that for each dollar the government spent on primary education, only 36 cents actually reached the rural schools. This means lower-quality education in such areas.

In Ghana Lavy (1992) found that leaking, unusable classrooms and lack of electricity had significantly negative effects on decisions to enroll in primary schools, and Glewwe and Jacoby (1992) reported that other variables (for example, no desks) also influenced primary enrollment. The government has increasingly required local communities to meet capital and nonwage recurrent costs, which means that educational services in poor communities are likely to be inferior because their resources are so constrained.

Households that enroll children in school encounter costs, such as the costs of supplies and transportation and the loss of the children’s work. What effect do these costs have on education? Mason and Khandker (1997) could find little evidence in Tanzania that out-of-pocket costs had a negative effect on enrollments. But work on Ghana by Lavy (1992), by Norton and others (1995), and more recently by Chao and Alper (1998) suggests that these costs do reduce enrollments. Tan, Lee, and Mingat (1984) estimated the elasticity of school enrollment with respect to changes in direct costs at about -0.5 in Malawi. But all these studies failed to establish whether the poor are more sensitive to these costs than the population at large, thus explaining the observed pattern of school enrollment across income groups. Intuition would suggest that they are, but evidence from Africa is not available. (For evidence from Indonesia, see King 1995; on Peru, see Gertler and Glewwe 1989.)

The distance to the nearest school might also explain why enrollments are so low among poorer households. For example, Lavy (1992) found distance an important constraint in Ghana. But interestingly, it is not the distance to the primary school that influences primary school enrollments, but the distance to postprimary schooling. Parents are clearly making decisions based on the whole education investment profile. The decision to enroll girls was more sensitive to the access costs of postprimary education than was the decision to enroll boys. Because distances to secondary schools are longer than those to primary schools for poorer households, enrollments are likely to be low at all levels. Although primary schools are more widespread and accessible, access is still a problem for many poor rural communities in Ghana: Chao and Alper (1998) found that enrollment was almost 70 percent in communities with at least one primary school in 1992, compared with only 43 percent in those with no primary school. They estimated that reducing the distance to a primary school by one mile would raise enrollment by 3 percentage points. In contrast, Mason and
Khandker (1997) report that distance is important only in secondary school enrollment in Tanzania, largely because most communities in that country are served by primary schools.

In most of the countries covered here, fewer girls than boys from low-income groups attend primary school, and fewer girls than boys from all income groups attend secondary school. This means that gender bias is an important explanation for the poor targeting of education spending in Africa (see Demery 1997).

**Concluding Observations**

This review does not present a particularly encouraging picture. Although spending on social services is usually justified on equity and efficiency grounds, most health and education subsidies in the region are not particularly well targeted to the poorest. Subsidies to primary education are an exception, but even here, they appear inequitable when judged against the numbers of school-age children in the poorest groups and when alternative measures of economic welfare are used. There are grounds for considering that the inequality results shown here underestimate the true inequality. Regional variations in unit subsidies and in the quality of services provided—largely ignored in the results reported here—are likely to further disadvantage the poor. Moreover, the poor are less able than the better-off to augment government subsidies by contributing to the services obtained (table 12).

One of the most fundamental factors responsible for weak targeting is the positive income elasticity of demand for these services. In the long run, the strategy must be to encourage private providers so that the public subsidy can be directed more effectively at services used mostly by the poor. But there are instruments that could improve targeting in the short and medium term. The first involves reallocating public

**Table 12. Social Sector Spending on Poorest and Richest Quintiles in Ghana, 1992**

<table>
<thead>
<tr>
<th>Sector/Quintile</th>
<th>Government Subsidy</th>
<th>Household Spending</th>
<th>Total Spending</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Share</td>
<td>Mean</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>2,296</td>
<td>12</td>
<td>1,998</td>
</tr>
<tr>
<td>Richest</td>
<td>6,515</td>
<td>33</td>
<td>7,099</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorest</td>
<td>8,731</td>
<td>16</td>
<td>1,761</td>
</tr>
<tr>
<td>Richest</td>
<td>11,067</td>
<td>21</td>
<td>5,072</td>
</tr>
</tbody>
</table>

*Note:* Mean is in cedis per capita; share is percentage of spending for all quintiles.

subsidies toward services used primarily by the poor. On this, the evidence is mixed: in health, many of the poor do not use any services very much, even primary services. Expenditure reallocations would improve targeting only if they led to a significant increase in the use of such services by the poor. In education, there is somewhat more scope to channel resources to the poor through primary education, but even here, enrollment rates are low, especially among girls.

Changes in household behavior, therefore, are critical. Two factors appear to be important: quality of service and access to facilities. The poor are not well served by the public provision of health services. Such services that are available are costly to access. The evidence suggests that the poor would be willing to increase their use of health services if both quality and access could be improved. It also points to the need for increased attention to the infrequent use of health services by poor women. Improving quality and reducing cost would also seem to be critical for raising the demand for education among poor communities in Africa, although here the evidence is somewhat patchy.

A well-designed user-fee policy could potentially improve the benefit incidence of health and education spending, but the decision to impose such fees should be undertaken with care. Fees should be applied to services where total demand (for private and public services) is price inelastic and where good-quality private services are available. They should not be applied where good-quality private services are unavailable or where the demand is very price elastic (those services used mainly by the poor, for example). If user fees are combined with significant improvements in both access and quality, there is growing evidence that the poor will increase their use of the service.

Benefit incidence has provided important insights into the problems facing governments in Africa that are struggling to deliver essential social services to poor communities. But although it highlights the problems, it is short on answers. For Africa, at least, the message is that reallocations of public expenditures are not sufficient; policies must be based on a sound understanding of the factors that govern household decisions about health care and schooling and of the means by which subsidized services can lead to better outcomes for the poor.

Notes

Florencia Castro-Leal is an economist and Kalpana Mehra is a research analyst in the Poverty Reduction and Economic Management Network at the World Bank, Julia Dayton is a Ph.D. candidate in health economics at Yale University, and Lionel Demery is lead poverty specialist in the Poverty Reduction and Social Development Group in the Africa Regional Office of the World Bank.

1. Current expenditures are used because they benefit current beneficiaries. Capital spending may well have a quite different incidence, but it benefits future beneficiaries.
2. This welfare measure, now an established one for poverty analysis, is described in Ravallion (1993). Usually, household expenditures include imputed values for own-produced consumption and take into account regional variations in prices. In most cases, the welfare measure normalizes household expenditure on household size (the exception here being the study of South Africa, which uses total household expenditure per adult equivalent as its welfare measure). As shown here, the results are sensitive to the welfare measure used. Our reliance on per capita expenditure comes from the use of this measure in the studies that we draw on. But for future work, benefit incidence should explore ranking households by other measures. The effects of different assumptions about economies of scale in household consumption, for example, should be investigated. Given the effects of random variations in observed expenditures, an alternative approach would be to use instrumented or predicted values of the welfare measure (see Behrman and Knowles 1997).

3. These data are not comparable across countries. The Guinea estimate, for example, assumes that all private treatment provided at the home of the respondent is traditional (which is not true of the other countries). There is some suggestion in these data that respondents report visits to private pharmacists and traditional caregivers as "self-treatment," which would explain the apparently low use of traditional care. Such underreporting of traditional care would leave reported use of modern health providers (and the analysis that follows) unaffected. There may well be indirect effects of government health spending that affect such services, but without strong empirical evidence about these effects in the studies reviewed here, we assume that such care is unaffected by public subsidies.

References

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


van de Walle, Dominique. 1995. “The Distribution of Subsidies through Public Health Services in Indonesia.” In van de Walle and Nead, eds.


Rethinking the Causes of Deforestation: Lessons from Economic Models

Arild Angelsen • David Kaimowitz

This article, which synthesizes the results of more than 140 economic models analyzing the causes of tropical deforestation, raises significant doubts about many conventional hypotheses in the debate about deforestation. More roads, higher agricultural prices, lower wages, and a shortage of off-farm employment generally lead to more deforestation. How technical change, agricultural input prices, household income levels, and tenure security affect deforestation—if at all—is unknown. The role of macroeconomic factors such as population growth, poverty reduction, national income, economic growth, and foreign debt is also ambiguous. This review, however, finds that policy reforms included in current economic liberalization and adjustment efforts may increase the pressure on forests. Although the boom in deforestation modeling has yielded new insights, weak methodology and poor-quality data make the results of many models questionable.

Concern is rising about the adverse consequences of tropical deforestation. The loss of forest cover influences the climate and contributes to a loss of biodiversity. Reduced timber supplies, siltation, flooding, and soil degradation affect economic activity and threaten the livelihoods and cultural integrity of forest-dependent people. Tropical rain forests, which constitute about 41 percent of the total tropical forest cover, are considered the richest and most valuable ecosystem on the earth’s land surface. During the 1980s about 15.4 million hectares of tropical forests were lost each year, according to estimates by the United Nations Food and Agriculture Organization (FAO 1992). From 1990 to 1995 the annual loss was estimated at 12.7 million hectares (FAO 1997), but it is unclear whether this reduction represents a slowdown in actual forest clearance or new definitions and better data.

This concern has led economists to expand their efforts to model why, where, and to what extent forests are being converted to other land uses. Kaimowitz and Angelsen
(1998), in a comprehensive review of more than 140 models, describe why landholders behave the way they do and examine the links between the larger economy and decisions to clear—or to protect—the forest. The models vary with regard to the precise definition of forest, if indeed they provide any definition at all. In most instances in this paper, the term **deforestation** describes the complete long-term removal of tree cover. Like all social science models, those discussed here simplify complex multidimensional processes and highlight only a few of the many variables and causal relations involved in changing patterns of land use. These models, however, do allow one to think about deforestation more systematically and to explore the possible effects of policy or other exogenous changes on land use.

### A Framework for Analyzing Deforestation

The conceptual framework used here is helpful both in understanding deforestation processes and in classifying modeling approaches. Five types of variables are used in models of deforestation:

- **The magnitude and location of deforestation**—the main dependent variable
- **The agents of deforestation**—those individuals, households, or companies involved in land use change and their characteristics
- **The choice variables**—those decisions about land allocation that determine the overall level of deforestation for the particular agent or group of agents
- **Agents' decision parameters**—those variables that directly influence agents' decisions but are external to them
- **The macroeconomic variables and policy instruments**—those variables that affect forest clearing indirectly through their influence on the decision parameters

Figure 1 illustrates the relations among the main types of variables and provides a simple, logical approach to analyzing deforestation at three different levels: sources, immediate causes, and underlying causes. This schematic varies somewhat from the existing literature, which is rather inconsistent in its use of these terms. The starting point is to identify the agents of deforestation (small farmers, ranchers, loggers, plantation companies) and their relative importance. These agents' actions are the sources of deforestation. Theoretically at least, the magnitude of various sources can be directly measured—although it may be difficult to do so—and no economic analysis is required.

Next one might focus on agents' decisions, which are based on their own characteristics (background, preferences, and resources) and on decision parameters such as prices, technology, institutions, new information, and access to services and infrastructure. Together, these factors determine the set of available choices and the in-
Figure 1. Variables Affecting Deforestation

<table>
<thead>
<tr>
<th>Underlying causes of deforestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macroeconomic-level variables and policy instruments</td>
</tr>
</tbody>
</table>

↓

<table>
<thead>
<tr>
<th>Immediate causes of deforestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision parameters</td>
</tr>
<tr>
<td>Institutions</td>
</tr>
</tbody>
</table>

↓

<table>
<thead>
<tr>
<th>Sources of deforestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agents of deforestation: choice variables</td>
</tr>
</tbody>
</table>

↓

| Deforestation |

Source: Authors' construction.

centives for different choices. The decision parameters may be seen as the immediate causes of deforestation.

Finally, the agents’ characteristics and decision parameters are themselves determined by broader forces. These underlying causes of deforestation influence agents’ decisions through several channels—the market; the dissemination of new technologies and information; the development of infrastructure; and institutions, particularly the property regime.

For the sake of simplicity, figure 1 and the discussion so far imply that causal relations go in only one direction. But important effects also go in the opposite direction; for example, the decisions agents make will have important feedback effects on market prices (general equilibrium effects). Agents’ collective actions, political pressures, and demographic behavior also affect underlying causes.

A clear distinction among the three levels is necessary for several reasons. First, it is useful to single out the parameters that directly affect decisionmakers. Second, the levels of variables are related to the type of model used: microeconomic models focus on immediate causes, whereas macroeconomic models tend to deal with underlying causes. Third, because the underlying causes determine the immediate causes, which in turn influence the agents who are the sources of deforestation, mixing these levels...
confuses the causal relations involved and leads to serious misspecifications in regression models. And fourth, the results regarding the sources and immediate causes are more conclusive than those for the underlying causes.

Although the article focuses on the immediate and underlying causes of deforestation, some comments on the sources are in order. There is a broad consensus that the expansion of cropped areas and pastures is a major source of deforestation and that the expansion of pastures is especially important in Latin America (Kant and Redantz 1997). No similar consensus has formed about logging, although it seems to be a direct source of deforestation in some contexts and an indirect source in others. Logging roads, for example, facilitate access to the forest by farmers (Burgess 1993). Southeast Asia is one region in which logging has contributed significantly to deforestation. Evidence regarding fuelwood collection and open-pit mining is weak, although it implicates them as the sources of some significant deforestation, particularly for fuelwood in Africa.

Surprisingly little is known about how the characteristics of agents affect their behavior. Researchers know that subsistence-type households are less responsive to market signals than families who are more market oriented, but existing models say little about the prevalence of such behavior. Nothing significant can be generalized from available information about the role of farm size, farmer background, or timber company characteristics. The conventional poverty-environment argument is that poorer families are more likely to clear the forest, either to grow crops or to cut wood, because they have shorter time horizons (higher discount rates); the counterargument says such families are less likely to do so because they lack the necessary capital to put additional land into production (see, for example, Rudel 1993). Existing models provide little evidence on this issue.

Analytical and empirical models suggest that time preferences and risk aversion are important to farmers and loggers. But their practical effect depends on what the relevant investment decisions are assumed to be. High discount rates and risk aversion are both likely to reduce investment, but that investment could be either to clear the forest or to conserve it (Southgate 1990; Mendelsohn 1994; von Amsberg 1994). These two types of investment are not symmetrical, however: whereas forest clearing requires action, forest conservation requires little more than leaving the forest alone. This suggests that forest clearing is more likely to be considered the relevant investment decision.

The Models Reviewed

We have reviewed more than 140 papers containing economic models that represent key processes associated with deforestation. The exclusive focus on formal models does not imply that these models are necessarily more useful or more accurate than
informal studies based solely on descriptive statistics. Such "soft" analyses and studies complement formal models and offer important insights that are difficult to capture in formal models.

Quantitative models have many clear limitations. They focus on variables for which quantified data are available. They do not, typically, explicitly address issues related to market failure (users do not capture the full value of preserving tropical forests), which one could argue are the "real" underlying causes of deforestation (Pearce 1996). Further, institutional factors are rarely included. One could argue, however, that the variables related to market failure or institutional arrangements are fairly stable over time compared with prices, for example, and are therefore less relevant to changes in rates of deforestation.

Global reviews such as this one inevitably emphasize the similarities between countries and regions, rather than their differences. The factors affecting deforestation, the interactions between them, and the magnitude of their effects all vary significantly from one location to another. Models based on data from distinct locations can reach conflicting conclusions not only because they use distinct definitions, variables, or methodologies, but also because the processes themselves differ.

In this context, it should be noted that most of the empirical and simulation models that analyze a single country or region focus on a handful of countries: Brazil, Cameroon, Costa Rica, Indonesia, Mexico, Thailand, and, to a lesser extent, Ecuador, the Philippines, and Tanzania. Most of these countries are medium or large in size and population, are relatively politically stable, and have large areas of tropical rain forest, so the results presented here may be more applicable to countries with those characteristics.

Categories of Models Reviewed

Table 1 shows the distribution of the models analyzed here. We have classified the models based on two criteria: scale (household and firm, or microeconomic, level;
regional level; and national, or macroeconomic, level) and methodology—analytical, simulation, and empirical.¹

Analytical models are abstract, theoretical constructs. They include no empirical data but rather clarify the implications of different assumptions about how agents behave and how the economy operates, which may not be obvious. Simulation models use parameters based on stylized facts drawn from various sources to assess scenarios. Most simulation models at the microeconomic level are whole-farm analyses using (linear) programming techniques, whereas the most common macroeconomic simulation models are computable general equilibrium (CGE) models. Empirical models quantify the relations between variables based on empirical data. Almost all empirical models use regression analysis, usually the standard ordinary least squares (OLS) method. Below we describe the three types of model.

**Microeconomic Models**

As the name suggests, these models seek to explain how individuals allocate their resources, using standard economic variables such as background and preferences, prices, institutions, access to infrastructure and services, and technological alternatives. A major distinction is between models that assume all prices are market determined and farmers are fully integrated into perfect markets (Southgate 1990; Mendelsohn 1994; Bluffstone 1995; Angelsen 1999) and those that do not (Dvorak 1992; Holden 1993; Angelsen 1999). Within the former category, production decisions are guided by market prices (including off-farm wages) and can be studied as a profit-maximizing problem. When farmers are not fully market integrated, decisions are based, in part, on farmers’ subjective (and endogenously determined) shadow prices. Factors such as resource endowments (poverty) and household composition are important, and the consumption side must be included when making the production decision. This distinction turns out to be critical for how model makers predict land use will change in response to changes in population, agricultural prices, and income. Analytical models have been very useful in highlighting the role played by the underlying market and behavioral assumptions (Angelsen 1999).

The main strength of farm-level simulation (programming) and regression models reviewed here lies in their use of generally good-quality survey data regarding the magnitude of deforestation and description of farmers’ behavior. Strictly speaking, however, these conclusions apply only to the area studied. Some of the conclusions of the simulation models depend heavily on the market assumptions discussed above, which the models normally do not test. Farm-level regression models often say little about farmers’ response to price changes because the price variation within the area is normally too small to allow such analysis.
Regional Models

The coverage in such models is limited to a region or area with a distinct and characteristic ecology, agrarian structure, institutional and political history, set of trade networks, and pattern of settlement and land use (Lambin 1994:16). Analytical or simulation models rarely focus on a region, although there are a few exceptions. Although deforestation is inherently a spatial phenomenon, most models lack an explicit spatial dimension; thus they cannot answer the where question.

Most regional models are thus regression models, which may be spatial or nonspatial. Spatial models measure the impact on land use of variables such as how far the forest is from markets and roads, topography, soil quality, precipitation, population density, and zoning categories. This type of analysis has become more popular since the advent of digitized land use data and geographic information systems that have made it easier to manipulate the data. Nonspatial models, however, are more common. These models use data obtained at a provincial or regional level in a manner similar to multicountry regression models, but the regional models generally have better data on forest cover: about half the models reviewed here used satellite data, either alone or in combination with land surveys.

Decisions affecting the rate of deforestation are taken at the household level, but the most interesting consequences affecting biodiversity and watersheds often occur at the district or regional level. Accounting for behavioral changes of farmers and other agents is difficult in spatial models. It should soon be possible in some cases to use panel data for spatial regression models, which will facilitate the inclusion of price variables (Foster, Rosenzweig, and Behrman 1997). It should also be possible to incorporate agricultural census and survey data into a geographic information systems framework, which would allow modelers to take into account many additional variables.

Macroeconomic Models

National and multicountry models emphasize the relations among underlying variables, decision parameters, and deforestation. Analytical, simulation, and regression models are all well represented at this level.

To model complex macroeconomic processes in a strictly analytical framework and still reach interesting conclusions, model makers have generally had to place strict limits on the number of variables and make some strong assumptions. Both analytical and computable general equilibrium (simulation) models at the national level add two important dimensions to the analysis that are absent in household- and firm-level models. First, they make some prices endogenous. Thus they move beyond simply asking how decision parameters influence agents and look at how the underlying variables determine one particular set of decision parameters (prices).
This provides an important link to macroeconomic variables and policy instruments. Second, most models include the interactions among different sectors, for example, (subsectors of) agriculture, forestry, and manufacturing, which makes them useful in analyzing the underlying causes of deforestation.

Some computable general equilibrium (CGE) models take a conventional approach and assume that land is a factor of production and that forest is cleared up to the point where the current land rent is zero (Coxhead and Jayasuriya 1994; Aune and others 1997). Others pay particular attention to the property regime (Persson and Munasinghe 1995; Unemo 1995). A third group applies a forest rotation (Faustmann) approach (Thiele and Wiebelt 1994; Thiele 1995). CGE models can be criticized for the poor quality of their data and the parameters commonly used, their questionable assumptions about perfect markets, and (particularly in the case of the forest rotation approach) their descriptions of farmers’ or loggers’ behavior. In such models the conclusions depend heavily on the responsiveness of the variables to changes in prices and income, and these elasticities are often chosen rather arbitrarily.

Multicountry (global) regression models comprise the single largest category of deforestation models. They rely on national data to make global generalizations on the major processes affecting tropical deforestation. But problems with the method and the data make their usefulness and validity questionable. First, most researchers use deforestation data from the Food and Agriculture Organization Forest Resource Assessments (1981, 1992) or from the FAO production yearbooks. We agree with Rudel and Roper (1997: 54) that neither is “acceptable for empirical analysis of the causes of deforestation” because they are based largely on dubious data sources or are mere extrapolations based on forest cover data from a single point in time. For example, in the 1990 assessment (FAO 1992), only 21 of the estimates for the 90 countries were based on two or more national forestry inventories. For the remaining countries, deforestation rates were extrapolated from a single data point using a model with population density and ecological classes as its only explanatory variables. Three countries had no forest inventory at all; of the 66 countries with one inventory, 39 inventories were taken before 1981. The data for African countries are particularly poor.

Because of the difficulty of obtaining reliable data, many multicountry regression models use the percentage of forest land as a proxy for deforestation. Kummer and Sham (1994) argue persuasively, however, that forest cover depends on the percentage of land originally in forests and the total amount of forest cleared throughout human history and is not related in any simple way to recent deforestation. Moreover, many models mix sources, immediate causes, and underlying causes in their independent variables. (The work of Kant and Redantz 1997 is an exception to this general picture.) Besides potential statistical problems of multicollinearity and biased estimates, this mixing will also distort the interpretation of cause and effect.
And finally, to produce meaningful cross-country results, it is important that the variables included affect deforestation in roughly the same manner across countries. This is obviously a strong assumption because studies indicate that the effect of and interaction among economic growth, foreign debt, population, and other variables may differ greatly from one country to the next. In principle, this problem could be overcome by adding interaction terms among the independent variables, but in practice, the degrees of freedom are too small to do that.\(^3\)

In sum, most of the existing multicountry regression models do not accurately estimate the direction and size of the effects that different variables have on deforestation. Because of these weaknesses, we have given less weight to these models in the discussion.

### The Immediate Causes of Deforestation

The main source of deforestation is clearing by households or companies for agriculture or timber. The question is: what factors make farmers and loggers decide to clear more forests? Table 2 gives an overview of the main results of the models.

#### Agricultural Prices

Substantial evidence supports the assertion that higher prices for agricultural products stimulate forest clearing. As frontier agriculture becomes more profitable, both the existing population and migrants from other areas begin to shift resources into forest clearing. Higher prices also provide capital to put additional land into agricultural production.

On the theoretical level, there is only one reason why higher agricultural prices might not increase deforestation: when farmers exhibit a preference for subsistence-type farming, they will opt for leisure once they have reached some minimal consumption level. In this case they will produce less when prices are higher because they can meet their basic consumption needs without clearing more land. Microeconomic simulation models that assume subsistence behavior, such as Ruben, Kruseman, and Hengsdijk (1994) and Angelsen (1999) find less deforestation when agricultural prices are higher, while models that assume farmers are profit maximizers show the opposite (Monela 1995).

Although it is possible that some households might respond to higher agricultural prices by reducing the amount of land farmed, there is no evidence for this at more aggregated levels. Regional regression models on Mexico by Barbier and Burgess (1996) and Deininger and Minten (forthcoming), on Sudan by Elnagheeb and Bromley (1994), on Tanzania by Angelsen, Shitindi, and Aarrestad (1998), and on
Table 2. Major Results on Immediate Causes of Deforestation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Analytical</th>
<th>Simulation and empirical</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural output prices</td>
<td>Increase</td>
<td>Increase</td>
<td>Farm-level analytical models predict increase, unless there are strong income effects (subsistence models).</td>
</tr>
<tr>
<td>Agricultural input prices</td>
<td>Indeterminate</td>
<td>Mixed</td>
<td>Fertilizer price increases may induce shift to more land-extensive systems.</td>
</tr>
<tr>
<td>Off-farm wages and employment</td>
<td>Reduce</td>
<td>Reduce</td>
<td>Among the most significant findings.</td>
</tr>
<tr>
<td>Credit availability</td>
<td>Indeterminate</td>
<td>Increase</td>
<td>Depends on whether the relevant investment is forest clearing or forest management and agricultural intensification; most studies find that credit finances deforestation.</td>
</tr>
<tr>
<td>Technological progress</td>
<td>Indeterminate</td>
<td>Little</td>
<td>Similar to price increase; new labor-intensive technologies may reduce deforestation if labor supply is inelastic.</td>
</tr>
<tr>
<td>on frontier farms (direct effects)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessibility (roads)</td>
<td>Increase</td>
<td>Increase</td>
<td>Among the most significant findings, although roads are partly endogenous.</td>
</tr>
<tr>
<td>Homesteading property regime</td>
<td>Increase</td>
<td>Little</td>
<td>Claims to future land rents give farmers an additional incentive to clear land.</td>
</tr>
<tr>
<td>Land tenure security</td>
<td>Indeterminate</td>
<td>Increase</td>
<td>Empirical evidence is relatively weak.</td>
</tr>
<tr>
<td>Timber prices</td>
<td>Indeterminate</td>
<td>Increase</td>
<td>Empirical findings are weak but tend to find a positive link.</td>
</tr>
</tbody>
</table>

a. Data may not be reliable.

Source: Authors' analysis.

Thailand by Panayotou and Sungsuwan (1994) all find a positive correlation between higher agricultural prices and deforestation. Binswanger and others (1987) found a positive correlation between total cropped area and agricultural prices in a cross-country analysis of 58 countries. All the analytical macroeconomic and computable general equilibrium models also show that increased agricultural prices boost deforestation, although this result is as much a product of their initial assumption that farmers are profit maximizers as it is of the empirical evidence.

It should be emphasized that this discussion refers only to changes in the aggregate terms of trade for agriculture with respect to other sectors. Changes that affect the relative prices of different crops and livestock products may have quite different effects. Thus it is impossible to predict how specific policies will affect forest clearing.
without looking at their impact on prices for specific products and the pressure each product puts on forests. For example, Gockowski (1997) shows that deforestation increased in Cameroon after relative prices shifted in favor of plantains, the production of which requires substantial forest clearing, from cocoa, which requires less land.

Although the prices of agricultural products and other decision parameters can be taken as given by the individual farmer, they are not truly exogenous in the models (as is also the case for many of the other variables in table 2). Output prices are a function of total supply. The regression models reviewed do not attempt to separate out predetermined (exogenous) changes (taxes, exchange rates, and so on) from the response to these changes. The response to an exogenous price increase will dampen the initial increase, but this effect is likely to be small because output from recently cleared land often has a small market share.

Prices of Agricultural Inputs and Credit

The theory of how changes in agricultural input prices affect forest clearing leads to indeterminate conclusions, and the empirical evidence is mixed, particularly for fertilizers. Analytical models point to two conflicting effects. On the one hand, higher fertilizer prices lead farmers to adopt more extensive production systems that use more land and less fertilizer. On the other hand, the higher costs associated with increased fertilizers make agriculture in general less profitable and can lead to a reduction in the amount of land devoted to crops.

Attempts to resolve the issue empirically have been only partially successful. Linear programming and regression models suggest that fertilizer price increases in southern Africa provoke greater deforestation or have little impact (Monela 1995; Aune and others 1997; Holden 1997; Mwanawina and Sankhayan 1996), whereas in some Latin American contexts such price increases may reduce deforestation (Barbier and Burgess 1996). Higher fertilizer prices seem most likely to induce greater forest clearing when farmers are wavering between intensive sedentary agriculture and more extensive shifting cultivation systems. This finding adds a cautionary note about the possible negative impact of current policies aimed at reducing fertilizer subsidies in Sub-Saharan Africa (Holden 1997).

The evidence regarding the prices of other agricultural inputs, such as seeds, pesticides, and hand tools, suggests that higher prices reduce forest clearing (Ruben, Kruseman, and Hengsdijk 1994; Ozório de Almeida and Campari 1995; Monela 1995). In these cases the reduced profitability of agriculture appears to outweigh any shift toward more extensive production.

In theory, credit expansion could reduce the pressure on forests if it were used for more intensive agriculture or for forest management investments. It will, however, increase the pressure if used to finance activities associated with forest clear-
ing, such as extensive cattle ranching. Most empirical evidence on credit comes from farm- and regional-level regression analysis in tropical Latin America and concludes that credit availability is positively correlated with deforestation (Ozorio de Almeida and Campari 1995; Barbier and Burgess 1996; Andersen 1997; Pfaff 1997). The only significant exceptions are two studies of indigenous farmers in Bolivia and Honduras, which found that farmers who used credit deforested less (Godoy and others 1996, 1997). In these cases, families with credit may be less dependent on forest-based activities or may choose to engage in off-farm work to repay their loans. Modeling work in Africa and Asia has largely ignored the issue of credit availability (with the exception of Monela 1995, who finds a positive relationship between credit availability and forest clearing in Tanzania), perhaps because it is less important there.

Wages and Off-Farm Employment

All types of microeconomic models strongly suggest that higher rural wages reduce deforestation by making agricultural and forestry activities more costly. They also suggest that, at the individual household level, greater off-farm employment opportunities produce a similar effect by competing with such activities for labor (Holden 1993; Ruben, Kruseman, and Hengsdijk 1994; Bluffstone 1995; Godoy and others 1996, 1997; Pichón 1997).

Regional and national analytical and simulation models also support these conclusions, although the hypotheses have yet to be successfully validated in macroeconomic empirical models because of limited data on wages and off-farm labor. One has, therefore, strong reasons to believe that policies that favor rural wage increases and generate off-farm employment opportunities for rural people should reduce deforestation. Such policies should simultaneously conserve forests and diminish poverty.

Technological Progress in Agriculture

Technology has both a direct effect on farmers’ behavior and an indirect effect resulting from its impact on product and factor prices (including wages). We focus here on the first set of effects, leaving the second for a later section.

Technological changes that increase yields without significantly altering labor or capital requirements can be expected to increase deforestation. The extent of forest clearance is likely to be even greater if technological changes are labor- or capital-saving, or both, since this will free up resources for farming additional land (Southgate 1990). Conversely, if the new technology is more labor- or capital-intensive and if farmers find it difficult, expensive, or inconvenient to hire wage labor or obtain credit, then such changes can lead farmers to devote more labor and capital to their
existing farms, leaving them with fewer resources for expansion. Under these circumstances the net effect is indeterminate (Larson 1991). More generally, technologies that make more intensive production systems more profitable reduce the need for clearing additional forest land for agriculture, according to linear programming models by Nghiep (1986) and Holden (1993).

These findings imply that agricultural research and extension policies designed to limit deforestation should focus on promoting profitable technologies that are labor- and capital-intensive and more easily applicable to land already under cultivation. The empirical evidence, however, is still limited, and this is clearly an important area for future research.

**Accessibility and Roads**

Analytical and empirical models and studies find that greater access to forests and markets accelerates deforestation. Roads, rivers, and railroads all facilitate access. Forest fragments are more accessible than large compact forests, and forests in coastal countries and islands are more accessible than those in continental countries (Krutilla, Hyde, and Barnes 1995; Rudel and Roper 1996).

Spatial regression models are well suited for studying the effects of access. Models of this type for Belize (Chomitz and Gray 1996), Cameroon (Mertens and Lambin 1997), Costa Rica (Sader and Joyce 1988; Rosero-Bixby and Palloni 1996), Honduras (Ludeke, Maggio, and Reid 1990), Mexico (Nelson and Hellerstein 1997), and the Philippines (Liu, Iverson, and Brown 1993) all show a strong relation between roads and deforestation. Several find a similar result between proximity to markets and forest edges. Most studies show that forest clearing declines rapidly beyond distances of 2 or 3 kilometers from a road, although Liu, Iverson, and Brown (1993) report significant forest clearing up to around 15 kilometers from the nearest road. These results are also supported by nonspatial regression models from Brazil (Andersen 1997; Pfaff 1997), Ecuador (Southgate, Sierra, and Brown 1991), the Philippines (Kummer and Sham 1994), and Thailand (Panayotou and Sugisawa 1994; Cropper, Griffiths, and Mani 1997).

The simple correlation between distance to roads and deforestation found in regression models tends to overstate the causality, since some roads are built precisely because an area has been cleared and settled, rather than vice versa. And both the land and the roads can be simultaneously influenced by a third set of factors, such as soil quality or population density. Model makers have attempted to account for this alternative by including some of those factors as separate independent variables, using road density, say, and analyzing only forest clearings that occur after roads are built. These attempts have been only partially successful, but no policy intended to influence deforestation can be considered comprehensive unless it provides clear guidelines on investments in transportation infrastructure.
Property Regime and Tenure Security

In the absence of well-defined and secure property rights, forest clearing often becomes a way to claim property rights to land (homesteading). Such strategic behavior has been reported by Anderson and Hill (1990), Mendelsohn (1994), and Angelsen (1999). Under these circumstances, there are at least three reasons why forests may be cleared beyond the point where the current net benefits are zero. First, even though profits may be negative in the first few years, technological progress, new roads, and so on will make cultivation profitable in the future, and farmers need to act now so that others do not claim the land before they do. Second, in many cases land prices may reflect not agricultural potential but rather speculation that the purchaser will profit from selling the land at some future date (Clark, Fulton, and Scott 1993). And finally, in situations where users compete for forest land, such as in conflicts between communities and government agencies, deforestation by one agent is costly to the other. Hence there may be incentives to clear the land oneself in order to squeeze out the competitor (Angelsen 1997).

Some empirical evidence suggests that where farmers can obtain property rights by clearing forests, land-titling projects can encourage them to clear larger areas (Kaimowitz 1996). Secure tenure encourages investment by making it less risky, and if the investment involves clearing land in the forest, deforestation should increase as a result. Nevertheless, household- and regional-level regression models from Latin America show that deforestation is lower in areas with secure land tenure (Southgate, Sierra, and Brown 1991; Godoy and others 1996; Pichón 1997). Thus a conclusion is premature at this time.

Timber Prices

The literature on the effect of logging on deforestation is smaller, and the results are less conclusive. The effect of higher timber prices remains particularly controversial. Higher prices for timber are likely to promote deforestation by making logging more profitable (Capistrano 1990; Gullison and Losos 1993; von Amsberg 1994; Barbier and others 1995; Deacon 1995; Mxstad 1995). Higher timber values also increase the net benefits of clearing land (assuming the timber is sold) and encourage deforestation (Southgate 1990; Deininger and Minten forthcoming).

Using a traditional supply-demand framework, trade restrictions, such as log export taxes and import bans, would reduce total demand for timber by lowering prices and production even if lower prices increased domestic demand. Other authors suggest, however, that in the medium term, low timber prices discourage efficient harvesting and processing techniques, leading in turn to more logging (Barbier and others 1995). Low timber prices may also discourage efforts to prevent farmers from clearing logged areas (van Soest 1996).
The Underlying Causes of Deforestation

It is harder to establish clear links between underlying causes and deforestation. Macroeconomic variables influence decisions through complex paths, and many of the causal relations are indirect. Further, such studies typically require data that often do not exist or are of poor quality. Table 3 summarizes the major findings on the underlying causes of deforestation, with these reservations in mind.

**Population Pressures**

Deforestation rates may increase because the population is growing and needs more land for food, fuelwood, timber, or other forest products. Growing populations also affect labor markets, as an abundant supply of labor pushes down wage rates. But

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect of increase in variable, by model type</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>Increase</td>
<td>Increase; The empirical results suggest that population density is positively correlated with deforestation, but the evidence is weaker than often believed; regional population should be considered endogenous.</td>
</tr>
<tr>
<td>Income level</td>
<td>Indeterminate</td>
<td>Increase; Higher income increases demand for agricultural and tropical products and access to markets but also increases off-farm employment.</td>
</tr>
<tr>
<td>Economic growth</td>
<td>Indeterminate</td>
<td>Mixed; Should induce downward pressure on agricultural prices and upward pressure on wages and interest rates (unless the changes reduce labor and/or capital intensity).</td>
</tr>
<tr>
<td>Technological progress</td>
<td>Reduce</td>
<td>Limited evidence; Same as above.</td>
</tr>
<tr>
<td>(general equilibrium effects)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign debt</td>
<td>Indeterminate</td>
<td>Mixed; Theory weak; empirical evidence weak and contradictory.</td>
</tr>
<tr>
<td>Trade liberalization and devaluation</td>
<td>Indeterminate</td>
<td>Increase; Higher agricultural and timber prices increase clearing, but income declines may offset this in the short run; relative prices also matter.</td>
</tr>
</tbody>
</table>

---

*Table 3. Major Results on Underlying Causes of Deforestation*

*a. Data may not be reliable.*

*Source: Authors' analysis.*
population growth may also induce technological progress and institutional changes that contribute to reduced pressures on forests.

Analytical models that consider the labor supply to be exogenous give quite different results from those that assume it to be highly elastic with respect to wages. In the former, deforestation rates tend to be much more sensitive to agricultural price changes, and agricultural intensification is more likely to diminish forest clearing (Angelsen 1999).

Several multicountry regression models show a positive correlation between population density and deforestation (such as Palo 1994; Rock 1996). Many of their results are spurious, however, because they rely on the FAO Forest Resource Assessments, which are themselves based on population data. As Rudel and Roper (1997: 54) note, "a variable which FAO used to construct the dependent variable is now being used to predict the value of that variable!" At the regional level, studies from Brazil (Andersen 1996; Pfaff 1997), Ecuador (Southgate, Sierra, and Brown 1991), Mexico (Barbier and Burgess 1996), the Philippines (Kummer and Sham 1994), and Thailand (Katila 1995; Cropper, Griffiths, and Mani 1997) also find a positive correlation between population density and deforestation. In the multicountry and regional studies, this correlation disappears when additional independent variables are added, implying that population may be acting as a proxy for some other factors in these models (Capistrano 1990; Deacon 1994; Harrison 1991).


Few models focus specifically on the relation between population and the demand for agricultural and forest products. Economic liberalization and globalization are likely to make this aspect less important at the national and regional levels because global demand is increasingly likely to determine prices and demand. New prospects for agricultural and forestry exports may lead to rapid deforestation in countries where small domestic markets previously limited deforestation.

At the local and regional levels, population is endogenous and is determined by infrastructure availability, soil quality, distance to markets, off-farm employment opportunities, and other factors. Several studies show that population growth in previously forested, low-population areas occurs in response to road construction, available high-quality soils, and growing demand for agricultural products (Harrison 1991; Southgate, Sierra, and Brown 1991; van Soest 1995; Andersen 1997). Government policies that affect migration (and hence population) at this level include road construction, colonization policies, agricultural subsidies and tax incentives, and gasoline prices. This implies that the latter factors, rather than population growth per se, are the causes of deforestation in these areas. People migrate to forested areas
because clearing forest for agriculture is economically attractive, and so the size of the population in those areas cannot be considered an independent variable in models of deforestation.

Income Level and Economic Growth

Higher national income and economic growth can be expected to reduce the pressure on forests by improving off-farm employment opportunities, but to increase it by stimulating demand for agricultural and forest products and improving access to virgin forests and markets. Countries with higher incomes may also demand that forests be protected rather than depleted. Forest depletion may contribute to economic growth, implying a causal relation in the opposite direction.

Many studies of developing countries associate higher national per capita income with greater deforestation (Capistrano 1990; Burgess 1993; Krutilla, Hyde, and Barnes 1995; Barbier and Burgess 1996; Mainardi 1996). Again, these models have significant data and methodological weaknesses and should be regarded with caution. Evidence on the impact of income growth rates is even weaker. Because there is no strong short- or medium-term relation between economic growth rates and average per capita national income, the fact that higher incomes are associated with more deforestation does not necessarily imply that higher growth rates will be.

The models are also not very clear about whether deforestation declines or is even reversed beyond certain income levels as countries become richer, a possibility noted by the “forest transition” hypotheses (Mather 1992; Grainger 1995) and by the environmental Kuznetz curve literature (for example, Stern, Common, and Barbier 1996). Based on the dubious FAO data, several authors claim to have found an environmental Kuznetz curve for deforestation; that is, at low levels of income, an increase in income will accelerate the rate of deforestation, but higher income beyond a certain level reduces deforestation. But the levels of per capita income they estimate must be reached before deforestation declines vary considerably (Panayotou 1993; Cropper and Griffiths 1994; Rock 1996). In addition, the driving forces behind such a possible transition are still unclear. They could be economic forces (the attraction of off-farm employment, a higher value placed on pristine forest by the public and the government, or expanded state capacity to enforce forest protection). Even if such a relationship does exist, income levels in most tropical countries are well below the level at which deforestation begins to decline.

External Debt, Trade, and Structural Adjustment

Some studies find a positive correlation between external indebtedness and deforestation (Burgess 1991; Kahn and McDonald 1994; Mainardi 1996; Kant and Redantz 1997), while others find no clear connection (Capistrano 1990; Kimsey 1991; Inman
The empirical studies are based on poor-quality data; the analytical models make very simplistic assumptions about government objectives and policy formation that limit their empirical relevance.

According to analytical models, policies to improve the terms of trade for agriculture tend to raise the prices received by farmers and hence increase deforestation (Jones and O'Neill 1994, 1995). Thus structural adjustment policies of this type may potentially increase pressure on forests, and policies such as overvalued exchange rates, industrial protectionism, and urban-biased spending may actually be good for forest conservation—although obviously not necessarily for other parts of the economy.

Market characteristics and general equilibrium effects can either strengthen or dampen these policy effects. Increases in agricultural and timber prices will generate more deforestation when labor supply is relatively elastic. If it is not, the initial effect of price increases will be dampened as rural wages rise in response to greater demand for labor. Conversely, higher rural wages could potentially generate more demand for agricultural and forest products.

Structural adjustment and trade liberalization policies designed to increase the terms of trade in favor of agriculture may have short- or medium-term recessionary consequences that reduce urban food demand, which could lead to lower, rather than higher, agricultural prices and thus to less deforestation. But a recession might also lower urban employment, putting downward pressure on rural wages and consequently stimulating deforestation (Jones and O'Neill 1995).

Policies designed to increase agricultural and forest product exports are likely to affect deforestation more than policies that promote production for the domestic market (since the latter are more likely to exert downward pressure on prices). Similarly, pro-agricultural policies can be expected to have stronger deforestation effects in the contexts of globalized agricultural markets and trade liberalization.

The previous findings are supported by several analytical macroeconomic and computable general equilibrium models, which show that currency devaluation, trade liberalization, and agricultural subsidies increase deforestation (Cruz and Repetto 1992; Jones and O'Neill 1994, 1995; Wiebelt 1994; Barbier and Burgess 1996; Mwanawina and Sankhayan 1996). It should be remembered, however, that these models depend heavily on more or less arbitrary assumptions about price elasticities and use generally poor data. Moreover, all of them tend to look at the agricultural and forestry sectors at a very aggregated level. Changes in relative prices within these sectors may have a greater impact on deforestation than the overall sectoral terms of trade, and to date these models have shed little light on this subject.

These findings suggest the difficulties of evaluating the effects of macroeconomic policies; important effects are not included. For example, will increased public revenues give officials the leverage they need for better regulatory intervention? Or will affluence mean additional investments that increase forest clearance? One lesson is
that any general claims about the relations among economic liberalization, structural adjustment, and deforestation are misleading. In particular, claims that structural adjustment programs will “generally contribute to both economic and environmental gains” (Munasinghe and Cruz 1995) seem unjustified based on the evidence. If anything, the findings support the opposite claim because higher agricultural output and timber prices lead to increased pressure on forests.

The Indirect Effects of Technological Change

Technological inputs also have indirect (general equilibrium) effects on product, labor, and factor markets. Technologies that increase aggregate supply and lower prices should reduce pressures to clear additional forest land. In some cases this may even offset the initial effects of technology on deforestation, as is possible in the case of maize production in the Philippines reported by Coxhead and Shively (1995). Technological changes that affect products with inelastic demands are more likely to reduce deforestation. Labor-intensive technologies will raise rural wages and should dampen—and even reverse—the deforestation associated with the increased profitability of agriculture. In fact, the more labor-intensive the technology, the more rigid the labor supply, and the more prices of agricultural products respond to changes in labor costs, the greater will be the effect. Similarly capital-intensive technologies might have the same effect if farmers have limited access to capital.

Technologies such as irrigation that require substantial infrastructure and that benefit farmers with access to markets are particularly likely to reduce pressure on forests; they will tend to push down agricultural prices and bid up wages without increasing the profitability of frontier farming. At the empirical level, some studies conclude that technological progress leads to more deforestation (Katila 1995), while others find the opposite (Panayotou and Sungsuwan 1994; Southgate 1994; Deininger and Minten forthcoming).

Rethinking the Causes of Deforestation

This review raises serious questions concerning the conventional wisdom about the causes of deforestation, either by providing contrary evidence or by showing the weakness of the supporting evidence. In particular, the models raise significant doubts about the following hypotheses:

- The population thesis. The models offer only weak support for the explanation that population growth is a driving force of deforestation. The correlations are largely based on flawed data or incorrectly specified models. At the local and regional levels, population should be considered endogenous, particularly in the medium to long term.
The poverty thesis.
There is little empirical evidence on the link between deforestation and poverty. If forest clearing requires investment, rich people may in fact be in a better position to clear new forest land. Moreover, off-farm employment opportunities simultaneously affect both poverty and deforestation, and any apparent relation between poverty and deforestation may actually be reflecting the off-farm employment-deforestation connection. Poverty (and discount rates) should therefore be considered endogenous variables.

The win-win thesis.
The thesis advocated by the World Bank and others, that economic growth and the removal of market distortions are good for people and forests, finds limited support in this review. Economic liberalization and currency devaluations tend to yield higher agricultural and timber prices that, in general, will promote deforestation. Moreover, higher incomes, within the relevant range of income found in developing countries, is likely to increase the pressure on forest resources.

The making-the-forest-valuable thesis.
Those who oppose boycotts of tropical timber and other timber market restrictions often claim that lower timber prices will discourage sound forest management. This review of the literature suggests that lower timber prices should both reduce logging activities and restrain agricultural encroachment stimulated by logging.

The tenure security thesis.
Land titles and more secure tenure have contradictory effects. Where forest clearing gives farmers a claim to the land, increasing the security of such claims may lead to greater forest clearing. This finding contradicts the conventional thesis of resource and environmental economics that more secure property rights are good for the environment.

The intensification thesis.
How improvements in agricultural technology affect forest clearing cannot be determined a priori, without information regarding the type of technology and the output and factor market elasticities. On the one hand, intensification programs targeted at farmers living near the forest frontier make farming more profitable and may shift resources to forest clearing and attract new migrants, although this effect may be at least partially outweighed by the resulting downward pressure on agricultural prices and upward push on wages. On the other hand, new technologies for nonfrontier agriculture should reduce pressure on the agricultural frontier. Labor-intensive technological changes are more likely to reduce pressure on forests than general yield-augmenting productivity increases and labor-saving technologies.

Although the evidence is not sufficient to reject all of these hypotheses, it does at least raise significant doubts. It is time to rethink the causes of deforestation and redirect research to focus more on issues such as the impact of credit markets, technological change, poverty reduction, and land tenure.
Notes

Arild Angelsen is a senior scientist and David Kaimowitz is a principal scientist at the Center for International Forestry Research, Bogor, Indonesia. The authors would like to thank Dennis Dykstra, Stein Holden, Ottar Mastad, David Pearce, and William Sunderlin, who provided comments on a draft version of this paper.

1. Models also vary with regard to their temporal nature (static-dynamic), type of data used (cross-section, time series, panel), spatial-nonspatial, and specific methods used.

2. The exceptions include Wiebelt’s (1994) regional CGE model for the Brazilian Amazon and the dynamic ecological-land tenure analysis (DELTa) model built for Rondônia in Brazil by scientists from Oak Ridge National Laboratory (that is, Dale and others 1994).

3. Many studies include regional dummies, but this approach allows only point intercepts to vary across regions, rather than the slopes (coefficients). This problem can be solved by multiplying regional dummy variables by the global variables to create separate explanatory variables, but only at the expense of considerable degrees of freedom (Mainardi 1996; Kant and Redantz 1997). Another potentially useful approach in the case of panel data, suggested by one of the reviewers and yet to be explored in analysis of deforestation, is to run regressions separately for each country. Then the estimates are averaged over countries, and these averages are much more precise than the individual country estimates. It should be noted that OLS on the pooled data may not converge to the country average effects.

4. Local labor supply is likely to be much more elastic in the long run because of the possibility of migration.

References

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


94 *The World Bank Research Observer*, vol. 14, no. 1 (February 1999)


Formal and Informal Markets for Water: Institutions, Performance, and Constraints

K. William Easter • Mark W. Rosegrant • Ariel Dinar

Water markets—either formal or informal—can be an efficient method for reallocating scarce water supplies. At the same time certain constraints can raise the transaction costs of trading water. This paper reviews the conditions necessary to establish successful water markets, identifies potential problems, and offers mitigating strategies. It also uses examples of several informal and formal water markets already in operation to illustrate these problems and the solutions to them.

One response to growing demands for limited supplies of water is to reallocate available supplies through water marketing strategies. Although marketing is not a new concept, what is new is the growing recognition that the policy of developing new water supplies to meet future needs is no longer viable. Thus government officials are more open than ever before to new ideas for improving water management. In this context estimates of the economic benefits from trading water within and among sectors illustrate the potential for relatively substantial gains from trade (Vaux and Howitt 1984; Colby 1990; Chang and Griffin 1992; Weinberg, Kling, and Wilen 1993; Rosegrant and Binswanger 1994; Hearne and Easter 1997; Thobani 1997; Diao and Roe 1998). As a result countries such as Mexico have taken the plunge and adopted water use rights and water trading strategies as part of their new policy for managing water resources. Others, such as Peru and Pakistan, are considering or have considered implementing such programs (World Bank 1995, 1997).

This article proposes that countries facing water shortages under their current water pricing systems consider water marketing as a way to reallocate water resources. We illustrate the importance of understanding a country’s institutional framework before embarking on a comprehensive overhaul of water policies and review the conditions required for effective water markets. Recent studies of formal and infor-
mal markets highlight the gains from the efficient allocation of water as well as the
constraints that raise the transaction costs of trading water. As we point out, water
markets can provide the appropriate economic incentives to improve the efficiency
of water use and encourage the reallocation of water to higher-valued uses without
encountering the traditional opposition of existing water users.

The Institutional Setting

If a country has little experience with private markets for allocating scarce goods and
services, water is unlikely to be one of the first goods exposed to private market
forces. In contrast, in a country that is exploring new ways to use the private market
to improve the allocation of publicly managed resources, scarce water may be a good
candidate for market trading, depending on one’s view of the requirements for mar-
ket exchanges.

There are two distinctly different opinions about the institutional setting required
for efficient market exchanges: The neoclassical view posits that a legal system is
required; a more pragmatic view emphasizes the importance of informal contract
enforcement. Greif (1997:239–40), for example, observes:

This neoclassical view that places the legal system at the center of contract
enforcement in market economies has recently been criticized on the basis
of evidence indicating that many contemporary exchange relations in the
West and elsewhere are informal. The associated contract enforceability is
not provided by the legal system but is based on reputation, general morality,
and personal trust within social networks. Empirical evidence indicates the
importance of two distinct systems of informal contract enforcement: the
individualistic system of informal contracts enforcement prevalent in the
West, under which the reputation and morality of the individuals are key,
and the collectivist system of contract enforcement prevalent in most other
societies, under which personal trust within the social network is critical.

Cooter (1997) comes to a similar conclusion in reviewing the problem of con-
tracting and establishing a rule of law that is consistent with a country’s social norms.

At least in the case of markets for irrigation water, both the formal neoclassical
legal system and the informal system appear to be at work. The transfer of perma-
nent water rights seems to require the certainty that is provided by a legally based
approach in which water rights are recorded and can be defended in court. Water
transfers among districts are likely to change the amount of water that is returned to
streams and rivers, called the return flow, and a formal market may be required to
prevent losses of return flows to downstream users. If, however, the sales are tempo-
rary, that is, for one season or less, and do not affect return flows, informal markets
based on informal water rights can suffice. These informal sales will likely be among farmers in the same water district and in many cases among farmers served by the same canal. In addition, these sales are not likely to be anonymous, and enforcement of the contracts will not be provided by the legal system but rather will be based on reputation and personal trust. This suggests that to obtain more interdistrict or interjurisdictional water trades, a country will have to develop legal water rights that can be verified and defended in court at a reasonable cost.

Informal water markets work fairly well for groundwater as long as recharge to streams is adequate and the market has a sufficient number of sellers (Palanisami and Easter 1991; Shah 1993; Saleth 1998). The “tit for tat” game-theory enforcement strategy appears to work; if farmers do not pay, their future supplies will be cut, and if a seller does not deliver, the buyer can use another supplier. In one area of Gujarat, India, farmers have pipelines from three or four different suppliers coming to their fields (Shah 1993), and they can buy from the supplier who offers the best price and service. Shah found that “while the main beneficiaries of private investments in pipelines have been the buyers of water, early operators in the water business were motivated mainly by the desire to establish monopoly positions and to overcome topographical constraints in supplying water to a large command” (pp. 61–62). In other words, although sellers are motivated by profits to sell water, the buyers may be the big beneficiaries.

If a country decides to establish a formal water market, the community of users needs to support the concept as fair and beneficial. Thus the law must be written so that the resulting allocation of rights is equitable. If the economic rents from water trading are concentrated in the hands of a few individuals or the negative effects on third-party users are large and unmitigated, the community is not likely to obey the law. Cooter (1997:214) notes that “a modern economy needs effective laws to promote cooperation among people. Yet, states enact many laws that few people obey. People tend to disobey, or obey out of fear, laws that are not consistent with social norms and to obey laws that reflect social norms.” In Pakistan, for example, farmers tend to disobey the law against trading canal water. In contrast, the 1981 Chilean water law that establishes private water use rights is widely obeyed because Chile not only has a long record of private water development but also allocates water rights based on past use (Hearne 1998b).

Conditions for Effective Water Markets

Thus effective formal markets are dependent on some basic institutional and organizational arrangements to overcome a number of potential market constraints and to prevent other associated problems from developing (see Garrido 1998a:tables 1 and 2). For example, tradable water rights or water use rights need to be separated from
land rights. In many cases, institutional arrangements will also be needed to deal with third-party effects that result from changes in return flows or declining economic activity in the region in which water sales originate. Adequate management and infrastructure will be needed for trades that are not in the immediate vicinity, such as trades between users on different canals. Countries must consider establishing mechanisms to prevent monopoly control over water and to avoid the overexploitation of groundwater. In both cases, however, these problems can be dealt with through the manner in which water rights are designed, quantified, allocated, monitored, and enforced (box 1). How to do this effectively for groundwater is a particularly vexing problem in many developing and industrial countries.

Box 1. Constraints on Effective Trades in Unregulated Water Markets

<table>
<thead>
<tr>
<th>Potential problem</th>
<th>Frequency</th>
<th>Mitigating strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-party effects from a decline in output and employment in the water exporting area (1, 2, 3, 6, 7)</td>
<td>F</td>
<td>1. Require review and approval of transactions by public agency.</td>
</tr>
<tr>
<td>Reduction or changes in return flows along with any changes in water quality (1, 3, 4, 6, 8, 9)</td>
<td>F</td>
<td>2. Establish a fund to compensate third parties damaged in trading, financed by levies on water transactions.</td>
</tr>
<tr>
<td>Added incentive to overdraft open-access groundwater stocks, damage the aquifer, and increase pumping costs (3, 4, 10, 12, 13, 14)</td>
<td>I</td>
<td>3. Limit trades to a percentage of water rights in a given area or community.</td>
</tr>
<tr>
<td>Increased costs of irrigation system for the remaining farmers (3, 10)</td>
<td>I</td>
<td>4. Revise water rights downward.</td>
</tr>
<tr>
<td>Drop in land values (3, 11)</td>
<td>I</td>
<td>5. Grant water rights to those using return flows.</td>
</tr>
<tr>
<td>Market power for large-scale buyers or sellers (1, 3, 15, 16)</td>
<td>I</td>
<td>6. Limit trading outside the river basin or sector to consumptive water use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. Open litigation to nonholders of water rights.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. Tax or ban trading from upstream to downstream users.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. Set minimum instream flows to maintain aquatic ecosystems.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10. Require buyers to pay a fee for the costs imposed on the irrigation system from which water is transferred.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. Require lending agency to clear permanent water sales.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13. Tax groundwater sales based on their scarcity value.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14. Limit trading in areas with rapidly declining groundwater stocks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15. Provide for regulation of monopolies or expand supply options.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16. Aid small rights holders with legal fees and registration.</td>
</tr>
</tbody>
</table>

a. Numbers in parentheses refer to appropriate mitigating strategies.
b. I, infrequent; F, frequent.
In many cases, governments will also need to reserve or buy some of the water rights to preserve instream uses that have strong “public good” characteristics, such as recreation, fish production, and the preservation of aquatic environments. As Howe (1998) points out, preservation has become a growing concern in the western United States as the demand for recreation and environmental services has grown. California has even reallocated water from irrigation to improve instream flows into the San Joaquin–Sacramento delta that will help preserve aquatic environments in the delta (Archibald and Renwick 1998; Howitt 1998).

Adequate information concerning water supplies and demands is a basic requirement for the efficient operation of markets. In Chile, for example, water user associations were essential in providing such information (Hearne 1998b). The central government generally has a comparative advantage in obtaining water supply data, although the users are better able to determine their own demand. User associations with access to such data can be important conduits for information. Asymmetric information is much less of a problem for water markets with strong user associations, but farmers may withhold information about their willingness to buy or sell water in order to obtain more favorable prices.

Informal water markets may be a good alternative, particularly if water allocation at the local level is a problem and the transaction costs of establishing formal markets are high (meaning the costs of enacting legislation, establishing institutional and organizational arrangements for markets, implementing trade arrangements, and monitoring and enforcing trades). Besides allowing water to be sold to the most productive farmers, informal markets would give all farmers an incentive to use their water more efficiently.

Gains from Water Markets and Organizational Constraints

The informal water markets that have evolved suggest that water users will buy and sell water even if such transactions are illegal or discouraged by governments (Renfro and Sparling 1986; Shah 1993). Problems arise when governments are asked to help develop formal markets or allow informal markets to develop, particularly within government-constructed irrigation projects. Because these are subsidized projects, many government officials maintain that the users should not be able to sell the water at a profit and that poor farmers will be disadvantaged because they will lose access to the water unless they pay higher prices. (Both Meinzen-Dick 1998 and Saleth 1998 dispute this claim, however.) Thus, even though water markets can change the incentives for water users and improve allocation, organizational constraints may prevent their introduction (box 2).

One prerequisite for water marketing is some type of water or use right that can be bought or sold separately from ownership of the land. Such rights may be difficult to establish and are likely to be resisted by public water agencies that fear they will lose a
Box 2. Constraints That Raise Transaction Costs in Unregulated Markets

<table>
<thead>
<tr>
<th>Potential problem</th>
<th>Frequency</th>
<th>Mitigating strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete or poorly defined water rights that are not separate from land</td>
<td>W</td>
<td>1. Register and secure water rights.</td>
</tr>
<tr>
<td>(1, 2, 8, 10)</td>
<td></td>
<td>2. Use proxies of water use (land area and use) to define water rights.</td>
</tr>
<tr>
<td>Inadequate infrastructure, including conveyance and storage systems (3)</td>
<td>F</td>
<td>3. Invest in infrastructure.</td>
</tr>
<tr>
<td>Inadequate water management, lack of water user associations (WUAs), or both (4, 5)</td>
<td>F</td>
<td>4. Provide management training for irrigation agencies.</td>
</tr>
<tr>
<td>Imperfect or asymmetric information about trading (5, 9, 10)</td>
<td>I</td>
<td>5. Provide water users incentives to organize WUAs.</td>
</tr>
<tr>
<td>The granting of more water rights than warranted by existing supplies (11, 12, 13)</td>
<td>I</td>
<td>6. Carry out education program explaining the benefits of water markets.</td>
</tr>
<tr>
<td>Sleeper or inactive water rights that might be sold and activated by water markets (7)</td>
<td>I</td>
<td>7. Tax unused water rights.</td>
</tr>
<tr>
<td>Inappropriate initial allocation of water rights that causes conflicts among water users (14)</td>
<td>F</td>
<td>8. Keep up-to-date single basinwide water rights registries.</td>
</tr>
<tr>
<td>Reallocation by government agencies of water among and within sectors without compensating original users (1, 4, 16)</td>
<td>F</td>
<td>9. Use public agencies or WUAs as clearing-houses for trades.</td>
</tr>
<tr>
<td>Opposition from farmers and environmental groups (6, 10, 15; also strategies 3, 6, 7, 8, 9, 10, 15 from box 1)</td>
<td>F</td>
<td>10. Aid small water rights holders with free legal protection and information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. Encourage spot and option markets for water.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12. Revise all water rights downward.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13. Define two types of rights with one senior to the other.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14. Base allocation of water rights on past use and conduct an auction for any surplus.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15. Use part of the water traded to enhance stream flows.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16. Base water rights on shares of the water supply rather than on absolute volumes.</td>
</tr>
</tbody>
</table>

a. Numbers in parentheses refer to appropriate mitigating strategies.
b. l, infrequent; F, frequent; W, widespread.

great deal of power if they allocate water rights to users. Giving users water rights means that system operators (the government officials) have the responsibility to deliver water more or less when the users want it. In contrast, a government agency that retains the water rights can dictate the conditions under which farmers will receive water, including (in some cases) the necessary side payments. Making the water rights tradable creates an even greater dilemma for government agencies. To prevent loss of control over tradable water rights, the National Water Commission in Mexico and some of the water districts in the western United States limit trading among water districts. In Mexico a water user must obtain special government approval to sell water outside the district or jurisdiction, and any profits from the sale must accrue to the district and not to the seller. This regulation discourages interdistrict trading, but it also reduces the chance that trades will have deleterious third-party effects.
Even if the users set up and hire the management unit that allocates water, the unit may have an incentive to use monopolistic power and discourage water trades with other jurisdictions. If too much water is transferred out of the district, the resulting shortage may reduce economic activity and make the irrigation system—or parts of the system—difficult to operate effectively. Once most of the farmers along a canal have sold their water, the few remaining farmers who own water rights on the canal may be difficult and expensive to serve. Thus, although water markets may change user incentives and encourage efficiency, the management of the system may prevent trades or raise the transaction costs of interdistrict or interjurisdictional trades.

Cooter (1997) argues that an organization seeking to maximize the wealth of its members will behave monopolistically toward outsiders but efficiently toward insiders. By fixing prices, establishing jurisdictional territories, and withholding information from the public, the organization will seek to create monopoly power for its members in dealing with nonmembers. These motivations may affect water user organizations and other water entities that can use their infrastructure as a monopolistic tool to block trades to outsiders. Water user organizations can either say that their canals are used to capacity or charge such high transmission fees that the trades become unprofitable.

The organizational problem appears to involve two important aspects: the resistance to trading water between or among districts or jurisdictions, and the problem of establishing water rights and giving the users more control. Other problems that raise transaction costs include legal challenges by third parties claiming they might be damaged by a transfer, the lack of sufficient infrastructure to transfer water among potential buyers, and the lack of an effective means for verifying and enforcing water rights. The question is whether taking action to reduce these transaction costs is in the best interests of a country. If the answer is even a tentative “yes,” then the second question is how to lower these costs.

Potential Problems and Mitigating Strategies

Clearly, one should not go to the expense of establishing water markets if water is not scarce or likely to become scarce in the foreseeable future. In Chile, for example, where water markets have been encouraged, there is no trading in the southern region because water is not scarce in that area. Water markets will be active only in regions where water is scarce.

Formal water markets are less likely to develop if they are constrained by the high cost of institutional development and by the absence of the management and infrastructure needed to implement trades. In areas where it is costly to establish, allocate, and enforce water rights, markets will be slow to develop. This has been an important factor in South Asia, where the presence of many small, fragmented farms makes it difficult to establish individual water rights. In such cases, water rights may have to
be allocated to water user associations, as they are in in Mexico, or to villages. If water trading is really to expand, these countries may also need to take steps to improve both their infrastructure and their system management. Yet most of these latter improvements are needed even without the desire to establish markets.

In South Asia there has also been a general aversion to markets, especially for allocating basic resources such as water and land, because of the belief that markets will disadvantage low-income farmers. In practice, as Saleth (1998) and Meinzen-Dick (1998) show, even “poor” farmers benefit from market exchanges. Clearly, if there are too few sellers, buyers may be disadvantaged, but Saleth and Meinzen-Dick point out that social conditions in India and Pakistan, for example, tend to mitigate against such exploitation. Saleth finds little or no price discrimination in mature water markets or in situations where kinship and social relationships are strong.

One of the biggest problems in establishing water markets in irrigated areas concerns issues associated with changes in return flows and water-related economic activities. These issues are important in California, as discussed in Archibald and Renwick (1998) and Howitt (1998). When water is transferred into other sectors or out of a river basin, governments need to have mechanisms in place that require the traders to take these third-party effects into account. Archibald and Renwick note, however, that these mechanisms must be carefully designed or they will foreclose many socially beneficial water trading opportunities. Thus the government must ensure that the appropriate institutional and organizational arrangements are in place. Clearly, a strong legal system is an asset in establishing such arrangements. Other key assets are a history of private irrigation development, strong water user associations, and appropriate conveyance and storage systems. Hearne and Easter (1997) illustrate how investments in storage capacity and a flexible infrastructure lower the transaction costs of water trading in Chile.

In contrast, Hearne (1998a) shows how too much government involvement in reallocation decisions can prevent water markets from developing. In Mexico the National Water Commission has foreclosed any possibility of intersectoral market exchanges in several regions. A better strategy may be for the commission to act as a broker in facilitating water trades. That approach would mean a major change in the commission’s function and is not likely to occur without strong political pressure from farmers and other interested parties.

Because the physical, institutional, organizational, and technical conditions that affect the performance of potential water markets vary so much, it is difficult to predict what will happen if water markets are introduced in a new area. The use of experimental markets, as suggested in Dinar and others (1998), may be a low-cost first step toward developing and evaluating alternative institutional arrangements. Although not widely used for water resources, experimental markets can capture some of the complexities involved in water markets.
Experience with Water Markets

Formal water markets specify the volume and share of water to be sold, either for a set period of time or permanently. Informal markets usually involve the sale of unmeasured flows of surface water from a canal for a set period of time or of water pumped from a well for a set number of hours. Although the units sold in informal markets may not be metered, both the buyer and the seller have good information about the volume transferred. The key difference between the two markets is the way in which the trade is enforced. If the users must self-enforce trades because no formal water rights exist that can be enforced through the legal or administrative system, the market is informal. Formal water markets are usually found in North and South America, whereas informal markets are prevalent in the irrigated areas of South Asia.

Informal Markets

Groundwater markets are important for agricultural production and the distribution of water throughout the irrigated areas of South Asia. Saleth (1998) estimates that 20 percent of the owners of the 14.2 million pumpsets in India are likely to be involved in water trading. This means that water markets are providing water for about 6 million hectares, or 15 percent of the total area irrigated by groundwater. In Pakistan a survey reported that 21 percent of well owners sold water (NESPAK 1991). In areas where dependable precipitation recharges the groundwater, the benefits of buying and selling water from tubewells have increased farmers’ income and production. The economic gains from groundwater markets reflect improved efficiency in pump management, in reducing conveyance losses, and in farm-level water use. These markets also increase access to irrigation, especially for smaller-scale farmers who do not own tubewells and cannot afford to invest in a well without a market for their water.

Meinzen-Dick (1998), in one of the few studies estimating the economic returns from access to water markets, found that water markets increased the availability and reliability of water supplies. Both yields and income rose for those who purchased water, particularly for those who also had access to canal water supplies. The highest yields and income, however, were still found among farmers who owned their own tubewells and had access to canal water.

Preventing Overdrafts. Given that markets for the sale of groundwater draw on an open-access resource (that is, one that is available for capture to anyone who has access), it is not surprising that problems arise in areas with high demands and limited supplies. Farmers have an incentive to ignore the scarcity and buffer stock value of the groundwater and pump until their cost of pumping equals the market
price of water (Ramasamy 1996). Over time, the cost of pumping and the price of water rise as the groundwater level declines. For example, the overdraft (that is, water use in excess of recharge) in the Coimbatore District of India is almost 5,000 cubic meters a year. Ramasamy estimates that if the overpumping continues, it will mean a drop in total net returns to farmers of between $42 million and $69 million, a result of the increased costs of power necessitated by increased pumping and additional investment to deepen wells. Here is a case where informal markets may exacerbate the problem, and formal markets may not work any better unless water rights can be established and enforced in strict quantity terms. The problem is not the water markets but the lack of exclusive property rights for groundwater. To establish such rights, the number of wells and the amount of water to be pumped would have to be agreed on and restricted. Such restrictions are probably unrealistic without strong support in the irrigation community. If exclusive water rights can be established, however, the water market should reflect the scarcity value of water and help restrain overpumping.

Blomquist (1995) reports on one case where the demand for water is increasing and the community of water users has been able to stop the overdraft. In the dry Los Angeles metropolitan area in southern California, pumping is metered and taxed so that users have an incentive to shift from local groundwater to more expensive but more plentiful imported water. Surface and imported water are stored and used to recharge the groundwater in the basin. One result has been a halt in saltwater intrusion from the ocean in the area's coastal groundwater basins. In some of these basins, pumping rights have been defined, limited to the basin's average recharge, and made transferable to other users through sales.

A more typical case, reported by Shah (1993), is in coastal Gujarat, India. Here, the overdraft of coastal aquifers has caused a decline in groundwater supplies in some areas and saltwater intrusion in others. Shah argues that any effective reduction in this overdraft is unlikely without good local leadership and the involvement of water user groups. He argues that “legal, quasi-legal, and organizational instruments of public policy will not, on their own, succeed in securing the compliance of farmers unless they are accompanied by measures aimed at affecting private returns to irrigation . . . or unless the structure of property rights on the water resource itself is drastically reformed (p. 147).” Similarly in Pakistan, Meinzen-Dick (1998:218) doubts “whether government would have the institutional capacity to regulate sales among hundreds of thousands of private tubewells, and if it had such capacity, it is unclear what such direct intervention could achieve.”

Yet in both India and Pakistan, any effect that water markets might have on the overdrafting of groundwater is much less than the effect of subsidized electricity. The zero or near-zero marginal cost of pumping means that farmers have an incentive to use water to the point where the marginal value of production is close to zero. This, of course, encourages farmers who can sell water to use their wells at close to
full capacity. The low power rates not only create overdrafting problems but also waste electricity in countries without adequate power.

As noted above, water markets can actually help solve the overdrafting problem by increasing the incentives for efficient water use and making it possible to purchase water from areas where water is abundant. The ability to find another source of water, but at a higher marginal cost, can help promote community action for self-regulation and demand management. Shah (1993) cites a case in coastal Gujarat where self-regulation became possible when additional new supplies were piped into the area.

Overdrafting tends to be concentrated in coastal areas of India and Pakistan and in the hard rock areas of southern India. In many of the northern areas, pumping actually improves growing conditions by lowering the water table below the root zone (Shah 1993; Meinzen-Dick 1998). In cases where water tables are high or recharge rates are rapid, water markets are not likely to cause negative externalities except possibly temporarily if neighboring wells are too close or deep tubewells interfere with shallow wells. Where these externalities are small, personal trust and reputation may be enough to foster competitive informal water markets. This is particularly true where farmers own a number of separate plots that cannot be served by the same well. In such cases, most water sellers are also buyers because most farmers who own a well are able to irrigate only their large plots and must purchase water to irrigate other plots (Shah 1993; Meinzen-Dick 1998; Saleth 1998). In addition, their wells are likely to be underutilized unless they can sell water. Yet because of the costs of conveying water and the need for cooperation from neighboring farmers when water is to be conveyed any distance, high transaction costs can restrict trades in areas with only a few wells and prevent water markets from being competitive.

**COUNTERING MONOPOLY PRICING.** This raises the other concern about water markets, the potential for monopoly pricing and discrimination. Groundwater markets are somewhat confined by the physical limits of the location and supply of groundwater. Still, pipelines can extend markets, and the investment costs of new wells should put a limit on monopoly power. An abusive monopolist who raises prices too high will find others investing in wells and undercutting the price. Shah (1993) notes a lack of balance between the numbers of buyers and sellers in areas with high-capacity wells, where one seller may serve as many as 70 or 80 buyers. He fails to say how many sellers the average individual buyer can access. Monopoly pricing may be avoided if the buyers can purchase water from three or four sellers—so long as the sellers do not collude.

The evidence on monopoly pricing is mixed. In a 1991–92 survey in Pakistan, Meinzen-Dick (1998) found that sellers were pricing water at little more than the cost of pumping. The two most common ways of charging for groundwater are a flat charge per hour of pumping (ranging from $0.57 to $3.27 an hour, depending on...
the pump type, capacity, and location) and arrangements whereby the buyer supplies the diesel and motor oil for the pump and pays an additional fee of $0.16 to $0.24 an hour to the well owner to cover the wear and tear on the engine. Sellers with diesel pumps were just recovering their own costs under either type of contract. 

In contrast, Saleth (1998) suggests that in some areas of India, monopoly rents may be extractive. He cites as evidence the variation in water charges compared with pumping costs in different areas. For example, water charges are 1.3 to 2 times higher than operating costs in the Indo-Gangetic region but 2.5 to 3.5 times higher in the water-scarce hard rock regions of southern India. The difference in rates, however, might be explained in part by the difference in water scarcity and in the value of water in those two regions.

The degree of monopoly power may also be related to the terms of the transaction or contract for water. Not surprisingly, some of the contracts for water are quite similar to contracts for land. Water contracts include crop sharing, crop and input sharing, and labor arrangements. If the payment is cash-based, buyers have more freedom to take their business to another well owner anytime during the season. When the transaction is a contract in kind, especially one based on crop sharing or on crop and input sharing, the buyer is tied to the seller for at least one season, if not longer. Similarly, if buyers contract to pay for the water with their labor, they may find it difficult to change suppliers until they have fulfilled the contract. Yet in the villages, informal markets do not appear to face extreme cases of monopoly rents. In fact, monopoly power that restrains trading in areas with serious problems of declining groundwater levels may help reduce overextraction. In contrast, when suppliers are taking advantage of their monopoly position and there are adequate groundwater supplies, the best strategy is to encourage (legalize) trading and increase competition through community and private well development (Palanisami and Easter 1991).

Thus informal water markets can improve water use and incomes in irrigated areas where water rights are not well defined or recorded. They also may be a good option if formal water markets are likely to produce third-party challenges and result in excessively high transaction costs. Finally, informal markets would work well in traditional irrigation systems where the farmers manage the irrigation system and would be able to maintain a relatively modest level of transaction costs.

Formal Markets

In situations where informal markets can work well, it may not be necessary to incur the extra expense of establishing formal water markets. Formal markets will be required, however, to provide the certainty necessary for permanent water transfers or transactions between different sectors and jurisdictions. Because the need for permanent trades and interjurisdictional water exchanges is likely to become more important as nonagricultural demands for water grow, formal water markets are likely to
become more common. The growing demand in water-scarce regions has been one of the driving forces behind the new interest in water markets. Several studies have illustrated the benefits that are possible from interjurisdictional trading in permanent water rights for short-term use.

In Texas 99 percent of the water traded has been transferred out of the agricultural sector in the Rio Grande Valley to nonagricultural users (Griffin 1998). Of the municipal water rights in the valley that existed in 1990, 45 percent had been purchased since 1970. Although water markets are not active in other areas in Texas, Griffin notes that the surface water law has evolved to the stage where trading will be more widespread in the future. In contrast, the groundwater law is just beginning to evolve.

ECONOMIC GAINS. In a study of the Guadalquivir Basin of southern Spain, Garrido (1998b) finds that the economic gains of trading within an individual water district or community may be relatively modest. In contrast, if permitted, trades among communities subject to different supply constraints and drought conditions could produce substantial gains. Garrido estimates the total welfare gain at no more than 10 percent over the current water allocation for four communities where trades were only intracommunity. Intercommunity trading, however, could produce estimated economic gains in one of the older irrigation communities of almost 50 percent. Garrido also shows that both types of trades are very sensitive to the level of transaction costs. If those costs exceed 8 to 12 percent of the market price, trading and the gains from trading would be too small to justify the expense of establishing formal markets. Yet Garrido may underestimate the potential gains because he considers only the crops traditionally grown in the region (cotton, wheat, corn, oilseed, and sugar beets) and excludes any transfers to nonirrigation uses. Evidence from Chile found significant changes in cropping as a result of water trading (Hearne and Easter 1997).

In contrast, Horbulyk and Lo (1998) found that most potential gains from introducing water markets in Canada’s Alberta Province were likely to come from trades within a subbasin. They considered four subbasins and compared the current water allocation situation with the allocation under four separate markets (one in each subbasin), as well as with a market encompassing the total basin. The four separate market scenarios created 90 percent of the welfare gains that were obtained when unrestricted trading was allowed among the four subbasins. The urban sectors purchased most of the water, except on the South Saskatchewan River, where the agricultural sector purchased additional water when market trading was allowed among the subbasins.

TRADING PATTERNS AND TRANSACTION COSTS. In their analysis of selected water markets in Chile, Hearne and Easter (1997) found trading both within and between sectors. In the case of permanent transactions either within or between sectors, well-established water use rights that were recorded and recognized by the government
were critical in fostering trade. Several trades between farmers and the city of La Serena were not consummated because of uncertainty regarding ownership of the water rights. La Serena is a growing vacation destination located on the coast in a dry region some 400 kilometers north of Santiago. Rapid growth in demand has strained the city’s water supply, particularly during the summer tourist season. The opening of water markets allowed the city to purchase water and delay development of new water sources. Starting in 1992, the city’s water company purchased enough water to increase its water supply by 28 percent. Additional purchases were made by upstream households for domestic uses and by farmers.

Elsewhere in Chile, significant trading occurs in the Limari Valley for agricultural purposes (the urban sector has adequate water). A survey of 37 farmers selling water and 19 farmers buying it reported transfers of rights to 9.2 million cubic meters. The gains from trade (measured as the difference between the value of water to the seller before the sale and the value to the buyer after the sale) were, on average, $2.47 a cubic meter ($3,045 an acre-foot), with a transaction cost of $0.069 a cubic meter ($86 an acre-foot). This sample was neither random nor complete, but the numbers surveyed were large enough to show that the water market was very active and had created significant gains from trade. The largest gains accrued to three grape producers who purchased 5.8 million cubic meters of water (63 percent of the total amount traded in the sample). In these active water markets, transaction costs were low and did not seem to constrain trading. In other areas, such as the upper section of the Maipo River that supplies the southwestern Santiago area and irrigates 100,000 hectares, the transaction costs are high, and trading is quite limited. The Maipo River is divided into three sections for management and water trading. As a result, water rights are uncertain, and the lack of adjustable control structures raises transaction costs and therefore limits trading (Hearne 1998b).

Similarly, Archibald and Renwick (1998) found that high transaction costs in California limited a large number of potentially profitable trades. Two types of transaction costs were identified: administrative-induced costs, which are explicit and included in the price of water sold through the California Water Bank, and policy-induced transaction costs, which stem from existing legal requirements designed to avoid injuring owners of water rights, damaging fish and wildlife, and creating negative third-party effects in exporting areas. Administrative-induced transaction costs, including the costs of locating buyers and sellers and negotiating quantities, timing, and other terms of transfer, were $0.041 a cubic meter ($50 an acre-foot) in 1991 and $0.014 a cubic meter ($17.50 an acre-foot) in 1992 and 1994. Policy-induced transaction costs in the West range from $0.152 a cubic meter ($187 an acre-foot) in Colorado to $0.044 a cubic meter ($54 an acre-foot) in New Mexico, all states with less stringent state and federal transfer requirements than California. Policy-induced transaction costs in this range would be as much as or more than the potential gains from trading in the California Water Bank (Archibald and Renwick 1998).
Because of high transaction costs in Colorado, Howe (1998) recommends shifting the administrative responsibility for water transfers from the water courts to the State Engineer's Office. He also recommends reserving or acquiring water for "public good" uses such as recreation, as well as making other changes to allow water to be marketed as freely in Colorado as it is in the neighboring states.

Colby (1998) suggests that the claims of Native Americans have the effect of imposing high transaction costs on water trading in many western rivers. She argues that even though markets do not work well with high transaction costs, when those costs are compared with the costs of litigated solutions, water markets look like a much better alternative.

Howitt (1998) reports that spot and options markets performed well during California's droughts in the 1990s. Even though these markets are a fairly recent phenomenon, he thinks they are a promising option for stabilizing available water supplies in California and other similar areas. Permanent shifts in demand, however, require a much more active formal market for water rights.

Conclusions

Contrary to the claims of many critics, water markets have worked and are likely to be a better mechanism for reallocating water than the alternative methods. There are both formal and informal water markets at work today. In addition, there are spot market sales, sales of permanent water rights, and leasing arrangements that are similar to those used for land, including crop sharing and cash rents (Saleth 1998).

Where water is scarce and large amounts of the available water supplies were committed to particular uses a long time ago, the economic benefits from water markets are likely to be large. In contrast, if the allocation was made fairly recently, based on the most highly valued uses of water, and new opportunities are not available, then the gains will be much more modest, as is shown in the Spanish example developed by Garrido (1998b).

For markets to be effective, transaction costs must be kept low. To keep these costs low, the appropriate institutional and organizational arrangements need to be in place, as well as flexible infrastructure and management. As pointed out earlier, the critical first step is to establish tradable water rights or water use rights separate from land, as well as the mechanisms to deal with third-party effects.

If it is difficult to establish legally enforceable, permanent water rights, a "thick" spot market may provide almost the same security as ownership of permanent water rights. In other words, the ability to buy the water needed at a reasonable price may provide enough security so that firms are willing to invest in enterprises that are dependent on this purchased water. A contingent water market can provide additional security so that firms can be assured of a given volume of water at a set price.
With only a spot market and no contingent markets, firms may be subject to wide fluctuations in prices.

For those users needing certain supplies, spot water markets are probably cheaper alternatives than having to buy enough senior water rights so that one is guaranteed adequate supplies even in the worst drought. (Owners of senior water rights have the right to whatever water is available, before the more junior water rights owners.) In Pakistan, for example, the markets for groundwater have greatly improved the security of water supply, particularly in government irrigation projects. This security has allowed increased investment and increased production. It will be important to see if spot and contingent markets have similar effects on the productivity of water.

Finally, the evidence indicates that appropriately designed water markets, supported by sound institutions, are an effective mechanism for reallocating scarce water among sectors. Carefully designed water markets make it possible to meet the growing urban and industrial water demands without derailing growth in crop production. Market transfers among sectors may make it possible to significantly scale back investments in new water supply projects. Government inaction, ineffective institutions for water management, and high transaction costs, however, are likely to prevent water markets from reaching their full potential for reallocating scarce water resources.

Notes

K. William Easter is a professor of applied economics at the University of Minnesota, St. Paul; Mark W. Rosegrant is a research fellow at the International Food Policy Research Institute, Washington, D.C.; and Ariel Dinar is a principal economist in the Rural Development Department of the World Bank.

1. "Public goods" are goods that consumers cannot be excluded from using and that are not consumed during use but continue to provide the same benefits to other consumers (World Bank 1993).

2. Scarcity value is the opportunity cost of water. It is the present value of the sacrifices imposed on the future by using the resource today. Buffer stock value is the value of groundwater in stabilizing water supplies when the supply of surface water is uncertain (Tsur 1990).

3. The 1995 exchange rate of 24.5:1 was used to convert Pakistani rupees to U.S. dollars.

4. An acre-foot equals 1,233 cubic meters.

References

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


Acknowledging the importance of sound judicial systems to good governance and economic growth, the World Bank and several other donor organizations have funded judicial reform projects in more than two dozen developing countries and transition economies during the past few years. Yet little is known about the actual effect of judicial reform on economic performance or even about what elements constitute a sound reform project. This article surveys a wide range of current studies on judicial reform and finds some surprising results.

The recognition that good governance is essential for economic growth has sparked renewed interest in projects to reform judicial systems. Since 1994 the World Bank, the Inter-American Development Bank (IDB), and the Asian Development Bank have either approved or initiated more than $500 million in loans for judicial reform projects in 26 countries (Armstrong 1998). The U.S. Agency for International Development (USAID) has spent close to $200 million on similar projects in the past decade (GAO 1993), and other government and private groups are also funding programs to modernize the judicial branch of government (ACCT 1995; Blair and Hansen 1994; Metzger 1997). Today, the majority of developing countries and former socialist states are receiving assistance of some kind to help reform courts, prosecutors' offices, and the other institutions that together constitute the judicial system.

Although few now question the importance of judicial reform for development, little is known about the impact of the judicial system on economic performance. Nor is there any agreement on what makes for a successful judicial reform project. Some argue that reform cannot be achieved without a societywide consensus, while others contend that the reform project can help create this consensus. Fears are also being expressed that judicial reform programs will repeat the mistakes of the law and development movement, an earlier, American-sponsored initiative that unsuccessfully sought to export the U.S. legal system wholesale to the developing world.
The design and implementation of judicial reform projects are complicated by a lack of knowledge about the relationship between formal enforcement of the laws through the courts and traditional, extralegal—or informal—means of enforcement. Coincident with the growing emphasis on judicial reform, a body of research has emerged showing that the formal legal system is just one way of ensuring compliance with society's laws. A variety of studies, in settings as diverse as medieval Europe and contemporary Asia, show that informal mechanisms based on incentives provided by repeat dealings can ensure the performance of contracts that no court has the power to enforce. One early, and surprising, finding of this research is that in some instances the sudden introduction of a formal mechanism to resolve legal disputes can disrupt informal mechanisms without providing offsetting gains.

**Rationales for Judicial Reform**

Judicial reform is part of a larger effort to make the legal systems in developing countries and transition economies more market friendly. This broader legal reform movement encompasses everything from writing or revising commercial codes, bankruptcy statutes, and company laws through overhauling regulatory agencies and teaching justice ministry officials how to draft legislation that fosters private investment. Although the line between judicial and legal reform blurs at the margin, the core of a judicial reform program typically consists of measures to strengthen the judicial branch of government and such related entities as the public prosecutor and public defender offices, bar associations, and law schools (Blair and Hansen 1994; Dakolias 1996; IDB 1994; Shihata 1995). These measures aim to:

- **Make the judicial branch independent.** Included here are changes in the ways in which judges are selected, evaluated, and disciplined to ensure that decisions are insulated from improper influences. In some cases, the budget for the judicial branch or the authority to administer the funds allocated for the judicial function is transferred from the ministry of justice or other executive branch agency to the judges themselves. Independence can also encompass giving judges the power to declare acts of the executive and legislative branches of government in violation of the country's constitution or some other higher law.

- **Speed the processing of cases.** Providing management training, computers, and other resources to judges and court personnel reduces case backlogs and accelerates the disposition of new disputes. Revising the procedures for filing and resolving lawsuits helps to weed out procedures that invite delay and raise costs.

- **Increase access to dispute resolution mechanisms.** The creation of mediation and conciliation services and other alternatives to resolving disputes in the courts...
reduces court costs, as does introduction of small claims courts or justices of the peace and the establishment of legal aid societies. Actions may also include transferring responsibility for noncontentious matters, such as name changes, the probate of uncontested wills, and the registration of property, to administrative agencies so that the courts have more time for disputed cases.

- **Professionalize the bench and bar.** In-service training for judges, lawyers, and other legal professionals entails programs to establish codes of ethics and disciplinary procedures. Increasing the number of law schools, ensuring that these schools have adequate resources, and modifying the curriculum to reflect the demands of a market economy are also a part of this element.

USAID funds judicial reform as part of its larger effort to strengthen newly emerging democracies around the globe (Blair and Hansen 1994; GAO 1993; Walker 1995). The agency’s projects originated in the early 1980s to assist the then fragile democratic government in El Salvador to bring individuals accused of human rights abuses to justice. A small program was initiated to train judges and law enforcement personnel in investigative techniques. The program was subsequently expanded and then extended, first to the rest of Latin America and later to the nations of Central and Eastern Europe and the newly independent states of the former Soviet Union.

The IDB finances judicial reform projects for a combination of political and economic reasons (IDB 1995). On the one hand it sees judicial reform as an indispensable element in consolidating democratic institutions in Latin America by protecting basic human rights and promoting harmonious social relations. At the same time, it recognizes that a well-functioning judicial system is important to the development of a successful market economy. Judicial reform is part of the IDB’s recent initiative to help borrower countries modernize the machinery of government. Since this initiative was launched, the IDB has either approved or taken under consideration 16 separate reform projects (Armstrong 1998).

Judicial reform projects sponsored by the World Bank aim solely at enhancing a nation’s economic performance. The Bank is enjoined by its Articles of Agreement from interfering in the political affairs of its members, a prohibition it interprets as preventing it from supporting judicial reform unless the project “is relevant to the country’s economic development and to the success of the Bank’s lending strategy for the country” (Shihata 1995:170). In practice this means that it does not provide assistance to reform criminal codes, train police or criminal court judges, or manage penal institutions (World Bank 1995). This focus on the economic consequences of judicial reform has led to complaints that such projects are ineffectual. Becker (1997) asserts that the emphasis on narrow technical issues comes at the expense of more important, but arguably political, questions. (For a recent critique of a World Bank-sponsored project in Venezuela by two human rights groups, see Lawyers Committee 1996.)
Judicial Reform and Economic Development

Whatever the rationale for judicial reform, it is widely believed that reform will significantly improve economic performance (Sherwood 1995). One hypothesis focuses broadly on the importance of the judicial system in enforcing property rights, checking abuses of government power, and otherwise upholding the rule of law. A second, narrower one casts the relationship solely in terms of the judiciary's effect in enabling exchanges between private parties. Although neither hypothesis has been subjected to a rigorous empirical test, there is some indirect evidence, albeit tentative and inconclusive, supporting both.

The narrow hypothesis originates with the 16th century English philosopher Thomas Hobbes, who argued that without a judicial system, traders would be reluctant to enter into wealth-enhancing exchanges for fear that the bargain would not be honored. In Hobbes's words, when two parties enter into a contract, "he that performeth first has no assurance the other will perform after because the bonds of words are too weak to bridle men's ambitions, avarice, anger, and other passions without the fear of some coercive power" ([1651] 1962:8).

Twentieth century development economists have revived Hobbes's thesis. North (1990:54) asserts that the absence of low-cost means of enforcing contracts is "the most important source of both historical stagnation and contemporary underdevelopment in the Third World." In Williamson's (1995) view, a "high-performance economy" is one that is characterized by a significant number of long-term contracts—just the type of business relationship that is unlikely to thrive in the absence of a well-functioning judicial system. When the judiciary is unable to enforce contract obligations, a disproportionately large number of transactions takes place in the spot market, where there is less opportunity for breaching contracts. Or, alternatively, firms circumvent the judicial system altogether by vertical and conglomerate integration, turning arms-length transactions into intrafirm ones. In either case, argues Williamson, the results are higher transaction costs and a "low-performance economy."

Survey evidence from Ghana supports Williamson's argument that the absence of a judicial system raises transaction costs—but not in the way he posited. As reported in Fafchamps (1996), businesses in Ghana rely upon a network of traders to serve as go-betweens. Rather than solicit a supply of lumber, say, from an unknown company directly, a firm will enlist a trader that it knows and that knows the lumber company. The personal relationships provide the buyer and seller with some assurance that the lumber will be delivered and payment received, but at a price: the reliance on intermediaries raises the costs of doing business.

Constructing a direct empirical test of Hobbes's hypothesis is a formidable task. As Sherwood, Shepherd, and de Souza (1994) note, it would require determining what transactions are not taking place and then quantifying the resulting losses. Clugue
and others (1995) proposed an indirect test instead, based on the assumption that the greater the percentage of money held in bank accounts and other financial assets, the more confidence citizens have in the judiciary and other institutions required to enforce bargains. Conversely, they reasoned, when a large percentage of the money supply is held outside banks and other financial institutions, the greater the likelihood that a substantial number of exchanges are consummated in simultaneous, spot-market transactions.

Cross-country regressions for a large sample of industrial and developing countries using a measure of the stock of money held in the financial system yielded the predicted results. The greater the percentage of the economy’s money in the system, the higher the level of investment and, to some extent, growth. But as Castelar Pinheiro (1998) observes, these results are problematic. A large portion of the money held in financial institutions consists of currency and other liquid assets available for immediate withdrawal. Hence, it cannot readily be assumed that a high ratio of funds held inside the system necessarily means that fewer spot or simultaneous transactions are taking place.

Surveys of Latin American entrepreneurs provide somewhat more, if also indirect, support for Hobbes’s thesis. In Peru almost a third of those responding to a World Bank poll said they would not switch from a trusted supplier to a new one—even if a lower price were offered—for fear the new supplier could not be held to the bargain (Dakolias 1996). A similar survey in Ecuador found that businesses were hesitant to invest because of the uncertainty of and potential lack of timeliness in enforcing contract rights. In-depth interviews of Brazilian entrepreneurs suggest that domestic investment would increase 10 percent if the Brazilian judiciary were on a par with those in the advanced market economies (Castelar Pinheiro 1998).

A second—far broader—hypothesis posits a more complex relationship between judicial reform and economic development. This view holds that economic development depends on a legal system in which not only are contracts between private parties enforced, but the property rights of foreign and domestic investors are respected and the executive and legislative branches of government operate within a known framework of rules (Dakolias 1996; Shihata 1995; World Bank 1992, 1994, 1997). This way of defining the rule of law assigns a prominent place to the judicial system: “[T]he judiciary [is] in a unique position to support sustainable development by holding the other two branches accountable for their decisions and underpinning the credibility of the overall business and political environment” (World Bank 1997:100).

The argument that the rule of law fosters economic development has been made many times. The 15th century jurist John Fortescue ([1471?] 1979) asserted that medieval England’s prosperity was traceable to the quality of English legal institutions. Almost 300 years later Adam Smith ([1755] 1980:322) observed that “a tolerable administration of justice,” along with peace and low taxes, was all that was
necessary to “carry a state to the highest degree of opulence.” Max Weber, the 19th century German sociologist, was the first to look carefully at the relationship among the rule of law, a well-functioning judiciary, and economic development (Trubek 1972), but according to Hayek (1960), credit for recognizing the judiciary’s importance in enforcing the rule of law belongs to the writers of the American Constitution and the German philosophers who elaborated the concept of the Rechtsstaat. The former showed why judicial review of legislative actions was crucial, while the latter demonstrated the importance of subjecting the actions of the executive and its administrative agencies to judicial scrutiny.

Weber’s comparative analysis of the role of law in China and the West was perhaps the first systematic effort to develop empirical support for the claim of a relationship (Bendix 1960). Most recently, in a survey of 3,600 firms in 69 countries, more than 70 percent of the respondents said that an unpredictable judiciary was a major problem “in their business operations” (World Bank 1997:36). The report also found that the overall level of confidence in the institutions of the government, including the judicial system, correlated with the level of investment and measures of economic performance.

But rigorous econometric methods for verifying the rule-of-law hypothesis and the role played by the judicial system are still in their infancy. Castelar Pinheiro (1998) reviews three recent efforts using cross-country regression analysis: Brunetti and Weder (1995), Knack and Keefer (1995), and Mauro (1995). Each uses a proxy for judicial system performance, such as entrepreneurs’ perception of the political risk involved in conducting business in a given country, to explore the correlation between a better judicial system and higher rates of investment, growth, and other indicators of economic performance. All three studies report a relationship between the proxies selected and different indicators of economic development, but as Castelar Pinheiro notes, each suffers from several methodological problems that make the results suggestive at best. The proxies for judicial system performance are often questionable, and there are problems with the endogeneity of the independent variables. These studies also do not rule out competing explanations such as increases in trade and investment or even the effects of other institutional reforms such as the introduction of an independent central bank.

Nor do cross-country regressions settle the question of the direction of causality. It may be that higher levels of development permit the state to spend more on the judicial system (Posner 1998). Or as Pistor (1995) observed in a review of judicial and economic reform in the transition economies, the same factors that contribute to economic reform and development may also be responsible for improvements in the judiciary. Both may be a result of preexisting attitudes and beliefs in society at large, or what has recently been termed “social capital” (Ellickson 1997; Solow 1995; World Bank 1997).
Hirschman's (1994) suggestion about the relationship between political and economic progress may apply equally to the relationship between judicial reform and development. He argues that political and economic progress are not tied together in any straightforward functional way. Rather, given the historical record, the relationship is probably better modeled as a series of on-and-off connections, or of couplings and decouplings. At some stages in the development process, the two may be interdependent, while at other stages they may be autonomous. There is no reason not to believe that a similar dynamic may be at work in the interplay between the evolution of the judiciary and economic growth, and the legal transplant school of comparative law has marshaled an enormous body of evidence showing that substantive law develops independent of economic and social variables (Ewald 1995).

In sum, while history and comparative analysis support the view that a better judicial system fosters economic growth, there is, as Weder (1995) observes, no clear, empirical evidence showing the economic impact of a weak judicial system. The most that can be said at the moment is that the weight of opinion and evidence suggests the existence of some type of relationship.

The Prerequisites of Successful Judicial Reform

Judicial reform can threaten those with a stake in the status quo. As both Eyzaguirre (1996) and Blair and Hansen (1994) note, inefficiencies in court procedures and management often provide opportunities for rent-seeking by attorneys, judges, and judicial support personnel. In Argentina, for example, the judicial clerks have protested a proposal by Fundación de Investigaciones Económicas Latinoamericanas (FIEL 1996) that they work more than the current 132 days a year. (The increase would raise the work year to at least 163 days, the average for executive branch personnel, if not to the 231 average for Argentine private sector employees.) The support staff is also challenging a recommendation to curb their power over case management and courtroom scheduling.

Reform may also engender opposition from the nation's organized bar. In Uruguay lawyers objected to the introduction of procedures that would speed up civil and criminal trials, fearing that speedier trials would mean less work for them (Vargas 1996). Reform can threaten lawyers' incomes in other ways as well. The practice of law is almost invariably a state-sanctioned guild or cartel, but as Posner (1995) explains, unlike an oil or steel cartel, "legal services" are difficult to define. The state must therefore specify what tasks are for lawyers and what tasks can be performed just as well by nonlawyers. In Peru, for example, attorneys and public notaries vigorously opposed measures to cut the costs of registering land belonging to the urban poor because the measures would allow engineers, architects, and other professionals...
to provide services that had once been the exclusive preserve of the legal profession (World Bank 1997).

Given the opposition that judicial reform is certain to generate, one view holds that no program should be undertaken absent a broad consensus in the country on the need for significant change. Dakolias (1996) recommends extensive consultation with committees representing judges, members of the bar, and other affected groups during the preparation and implementation of the project. Shihata (1995) adds that this consensus must include a long-term commitment on the part of the government to provide the resources required for an effective judiciary.

Blair and Hansen (1994) reached a similar conclusion in an evaluation for USAID of judicial reform projects in Argentina, Colombia, Honduras, the Philippines, Sri Lanka, and Uruguay. Absent a high level of support from the ministry of justice, senior executive branch officials, legislators, and judges, the authors argue that judicial reform is unlikely to succeed. When such support is lacking, they recommend that both public and private donors forgo judicial reform altogether. Instead, they suggest that donors concentrate on building a consensus for reform by opening a dialogue with the government and by encouraging bar associations, business groups, and other nongovernmental organizations to campaign publicly for reform.

In the six cases examined, Blair and Hansen found that training judges, improving management systems, and supplying computers and other resources to the judiciary had little impact in countries where a consensus for judicial reform was lacking. The lesson they draw is that these traditional components of judicial reform, often termed “institutional strengthening,” should not be initiated until more basic reforms have been achieved. Legal changes permitting the use of alternative dispute resolution mechanisms, creating or broadening legal aid programs, and ensuring that judges are appointed on the basis of merit should come first. Only after such structural reforms and access-enhancing measures are in place do they support institutional strengthening.

Several other evaluations of judicial reform in Latin America appeared to confirm Blair and Hansen's findings. Even before their results were published, the U.S. General Accounting Office (GAO 1993), in an analysis of USAID-sponsored programs in Central American and Colombia, concluded that providing computers, training, and other technical assistance to the judiciary in countries where a strong commitment to reform was lacking had not been productive. Buscaglia, Dakolias, and Ratliff (1995) report that Latin American judges often questioned the value of training in the absence of more fundamental reforms in the judicial system. In many cases, once a judge had been trained, he or she quickly left the bench for a more lucrative position in the private bar.

In a study of judicial reform projects in 15 Latin American countries, Martínez Neira (1996) found there had been too much emphasis on increasing the number of judges, courts, buildings, and computers at the expense of more fundamental changes.
in the legal system. He contends that this imbalance resulted from a lack of consensus on the scope of reform. Without such a consensus, judges, clerks, attorneys, and other actors in the legal system are free to pursue their own agendas. Pérez Perdomo (1993) makes a similar point, arguing that too many Latin American reform programs reflect only the needs and perspectives of judges and others with an institutional role in the judicial system.

Not everyone agrees with these criticisms, however. In a response to the General Accounting Office’s critique, USAID (1993) contended that such programs can be seen as the vehicles for developing a consensus. The collaboration between outside experts and judges and others within the country and the ensuing public discussion can help to generate the necessary political commitment, the agency maintained.

Hammergren (1998), who subscribes to this view as well, also takes issue with Blair and Hansen’s recommendation that institutional strengthening should always follow structural reforms and measures to increase access to the judicial system. She observes that institutional strengthening can pave the way for broader reforms that, if attempted first, may engender such strong opposition that reform will be stalled. She argues that institutional-strengthening measures do not have to end up simply serving the needs of judges, lawyers, and others with a stake in the status quo. Such measures can include ways of making the judicial system more accountable to the public and, as MacLean (1996) has stressed, give it a public service orientation. Hammergren also cautions that nongovernmental organizations have their own interests that may be at odds with a broader reform agenda. Business groups may, for example, be interested solely in the creation of commercial courts or steps that reduce legal fees.

The Law and Development Movement

Are the legal technical assistance programs sponsored by the World Bank ignoring the lessons learned in earlier attempts to foster development through law (McAuslan 1997; Thome 1997)? In the 1960s USAID, the Ford Foundation, and other private American donors underwrote an ambitious effort to reform the judicial systems and substantive laws of countries in Asia, Africa, and Latin America. This “law and development” movement engaged professors from Harvard, Yale, Stanford, Wisconsin, and other leading law schools and within a few years had generated hundreds of reports on the contribution of law reform to economic development (Merryman 1977). Yet after little more than a decade, both key academic participants (Merryman 1977; Trubek and Galanter 1974) and a former Ford Foundation official (Gardner 1980) declared the program a failure, and support quickly evaporated.

The guiding assumption of the law and development movement was that law is central to the development process. A related belief was that law was an instrument
that could be used to reform society and that lawyers and judges could serve as social engineers. As Merryman (1977) notes, not everyone subscribed to this view. A few participants in the movement argued that only minor changes could be effected through legal reforms, and others contended that law reform should follow broader changes in society, that is, that the proper aim of reform was to adjust the legal system to social and economic changes that had already taken place. But the dominant view of law and development practitioners and theorists alike, although still unproven (Vorkink 1997), was that law reform could lead social change—that law itself was an engine of change.

A second important belief was that educating the bench and bar in developing countries would advance reform efforts. The gap between the law on the books and the law in action in developing countries was widely appreciated, and one of the solutions advanced was professional education (Burg 1977). It was thought that if lawyers and judges were properly educated about law’s role in development, they could be enlisted to close the gap. The idea was to turn members of both professions into legal activists through education. Yet as one sympathetic chronicler of the movement observed, this idea was supported by nothing more than “hopeful speculation” that education could overcome values instilled by family, class, religion, and other social forces (Lowenstein 1970:246-47).

The postmortems on law and development identify a number of pitfalls that advocates of judicial reform ought to bear in mind (Burg 1977; Gardner 1980; Merryman 1977; Trubek and Galanter, 1974). One is that the movement lacked any theory of the impact of law on development. Practitioners thus had no way to prioritize reforms or predict the effects of various measures. A second failing was too little participation by the lawyers and others in the target country who either would have to carry out the reforms or would be affected by them. Foreign legal consultants, through a combination of expertise and access to funding, were often able to dictate the content and pace of reform. A third problem was that the movement focused on the formal legal system to the exclusion of customary law and the other informal ways in which many people in developing nations order their lives (Trubek and Galanter 1974).

But perhaps the most significant reason for its failure was the naive belief that the American legal system (and the legal culture generally), which Trubek and Galanter (1974:1062) refer to as “liberal legalism,” could be easily transplanted to developing countries. In the United States judges play a significant role in policymaking, and as a result, lawyers are often able to engineer significant changes in policy through litigation. This is not true in civil law systems or indeed even in the United Kingdom and other nations that share the same common law background as the United States. As Merryman (1977:479) put it, the law and development movement reflected the American “legal style,” and this was a style that those in other cultures did not find particularly attractive.
At a 1995 conference hosted by the British Council, participants debated whether the mistakes of the law and development movement are likely to be repeated. Faundez (1997:13) argued that although the old programs and the Bank’s new initiatives appear to be quite similar on the surface, the context in which the Bank’s current programs are being carried out is significantly different. Behind the law and development movement was the premise that the state “would initiate and promote the process of economic development.” By contrast, today the Bank sees law as facilitating market transactions by defining property rights, guaranteeing the enforcement of contracts, and maintaining law and order. Because the state is no longer the protagonist of social change, as in the law and development model, there is less room for error.

Yet as his analysis proceeded, Faundez seemed to be less sure that the mistakes of the law and development movement would be avoided. He recognized that current development theories, inspired by the work of Douglass North and other neo-institutional economists, still contemplate a role for the state. It is, to be sure, a different one from the activist theory implicit in the law and development movement, and it is one that is informed by economic analysis. “But it is unlikely that by shifting the focus of attention from legal institutions to economic analysis this new approach will manage to avoid the problems which so frustrated and disappointed members of the law and development movement” (Faundez 1997:14). His concern is that all the unanswered questions that lurked behind the law and development movement—the role of law and the formal legal system in development, the relationship between law and politics, and among democracy, authoritarianism, and development—still remain.

If Faundez is ultimately uncertain that the Bank will not repeat the mistakes of the law and development movement, both McAuslan (1997) and Thome (1997) have no doubts that it will. McAuslan advances a series of reasons why this is likely to happen. Like Faundez, McAuslan underlines the absence of any empirical data connecting reform with development and the consequent disagreement even among reformers over priorities and strategy. In a commentary on McAuslan’s article, Thome goes a step further. He believes the problem is not so much a lack of empirical data as the failure to reflect the data that are available. He asserts that all law reform, and judicial reform in particular, rests on three premises: first, that development requires a modern legal framework resembling that in the United States; second, that this model establishes clear and predictable rules; and third, that the model can be easily transferred. Yet, he says, empirical research has refuted all three assumptions.

McAuslan is also critical of what he argues is the Bank’s focus on law reform to facilitate market transactions. The emphasis should be on promoting good governance and alleviating poverty. An efficient and equitable market economy requires well-functioning state-run institutions that can curb the abuses likely to arise as the market economy develops. He fears that in emphasizing the role of the judiciary in fostering economic growth, these considerations will be pushed aside. He cites as an example land-grabbing by elites as property rights are defined and allocated by the

Richard E. Messick 127
state. Without judicial and administrative bodies capable of curbing such behavior, income inequalities will be exacerbated, and political instability may result.

McAuslan and Thome are not entirely negative. They do note that some Bank projects reflect the lessons of the law and development movement. The Financial and Legal Management Upgrading Project (FILUP) in Tanzania, for instance, has involved local lawyers from the beginning, both in studying the legal system and in developing proposals for change. Even the legislative drafting project for China, which they say is premised on an extreme view of the importance of law in the development process, is training local lawyers in the skills necessary for market reform. But even though the failings of the earlier law and development programs may be clear, both critics assert that pressures to produce results quickly will work against the gradual and incremental approach to law reform warranted by our current state of knowledge about the relationship between law and development.

**Integrating Judicial Reform with Informal Enforcement Mechanisms**

Judicial reform aims to buttress the rule of law and assure entrepreneurs that contracts will be enforced. Yet other institutions within society perform these same functions. Accountants and auditors issue standards and render judgments evaluating the performance of various economic actors. Credit bureaus provide an incentive to consumers and businesses alike to observe contracts by disseminating information about those who have failed to perform. And the media and nongovernmental organizations are often the first to detect and publicize arbitrary or illegal actions by government officials. Judicial reform projects that build upon or enhance the operation of such informal enforcement mechanisms will yield greater returns than those that do not. At the least, designers of reform projects must take the presence of these informal mechanisms into account. Otherwise, as Greif (1997) warns, the projects could backfire.

A forthcoming study by Kranton and Swamy of the effect that the introduction of formal courts had on rural credit markets in India illustrates just how a judicial reform project can go awry. Before there were formal courts, moneylenders relied on informal enforcement mechanisms to ensure that clients paid their debts. Resort to these mechanisms was costly and usually required the acquiescence of local leaders. Entry into the moneylending business was thus retarded and competition lessened. Although lack of competition meant that interest rates remained high, it also gave lenders a cushion that allowed them to extend payment terms and otherwise accommodate debtors experiencing difficulties in meeting their obligations.

The introduction of courts by the colonial authorities brought new entrants into the market. Interest rates declined, robbing lenders of the financial cushion that had
allowed them to carry borrowers in bad times. When drought hit, lenders went to court to foreclose on farmers’ land. Riots and widespread social unrest followed.

As Kranton and Swamy note, the lesson is not that judicial reform is never appropriate, but that where other markets, such as those for insurance and futures, are missing, care must be taken to ensure that judicial reform does not have unintended consequences. Fafchamps’s (1996) study of contracting in Africa supports this view. He found that rigid compliance with the terms of a written contract was difficult, if not impossible, in developing countries. Their economies are simply subject to too many exogenous shocks for contracts to be strictly enforced, which is why informal contract enforcement mechanisms build in such flexibility.

The problem in every case comes in determining how judicial reform will affect informal enforcement mechanisms, for a general theory of informal mechanisms and their interplay with formal mechanisms has yet to be advanced (Ellickson 1991). What is known is that when formal systems are deemed illegitimate, as they were in Moslem Central Asia after the Soviet takeover in the 1920s, disputes will be directed away from the formal system (Massell 1968). Ellickson (1991) submits that the division of labor between formal and informal mechanisms is affected by the technical complexity of the issues involved. He found, for example, that in northern California disputes between neighboring ranchers about the cost of a fence, which raised simple questions of fact and technology, tended to be resolved informally. By contrast, disputes involving the allocation of groundwater supply, where the facts were difficult to ascertain and resolution of the contested issues involved complex technical questions of return flow and allocation during shortages, were more likely to be presented to a court for adjudication.

The informal enforcement mechanisms that have drawn the most attention are reputation-based systems that permit merchants to carry on extensive trading relations over time and space in the absence of a court system that could ensure contract performance. Greif (1989) describes a system used by traders in North Africa and the Mediterranean in the 11th century, and in a later paper (1997), he shows similar reputation mechanisms at work in settings as diverse as the Wisconsin lumber industry, the New York diamond trading business, long-distance commerce in medieval Europe, and parts of contemporary Asia, Africa, and Latin America. The common denominator in all these examples is that the gains from repeat dealings provide the incentive necessary to ensure performance. That is, the discounted present value of the earnings stream that can be realized from future transactions exceeds the onetime wealth increase realizable from breaching the current agreement (Klein 1985).

The incentive to maintain a good reputation operates in other settings besides merchant-to-merchant relations. Credit bureaus—business associations that exchange information about the payment history of their customers—count on consumers’ desire to buy on credit in the future to assure payment of current obligations (Klein 1992). A similar principle is behind consumer testing laboratories, better business
bureaus, and other groups that market seals of approval or provide quality guarantees (Klein 1997). When these groups certify that a product or business meets a certain standard, they are providing a visible sign of good reputation that can be used to generate future sales.

Development itself can affect the mix of formal and informal mechanisms in an economy. According to Besley (1995), one of the reasons informal financial institutions such as rotating savings and credit associations continue to operate in the developing world is that they spend far less than do banks and other formal financial institutions to ensure that borrowers repay their debts. Because their borrowers typically come from the same village, these institutions can rely on group pressure and other informal methods to see that the loans are repaid. Besley predicts that these institutions will lose their comparative advantage as the number of close-knit communities declines with the changes brought by economic development and that they will ultimately be supplanted by formal firms.

Milgrom, North, and Weingast (1990) also stress how increases in the costs of operating an informal enforcement mechanism lead users to turn to the formal legal system. They model a reputation-based system that resembles the one devised by long-distance traders in medieval Europe. Enforcement depends on each trader determining whether the other party to the contemplated exchange has failed to honor a contract, or cheated, in the past. If the other party is a cheater, the honest trader refuses to exchange with him. The threat of a boycott deters cheating.

But traders incur costs in ascertaining the past history of those with whom they contemplate exchanging—costs that increase as the economy grows. The number of potential trading partners on which information must be gathered expands, and the number of queries rises as the number of potential exchanges increases. Accordingly, Milgrom, North, and Weingast argue that eventually the costs to traders will exceed the costs of operating a formal judicial system. Rising transaction costs, they note, explain why national courts replaced the law merchant system of informal enforcement in Europe during the late Middle Ages.

Although the value of their analysis is its explicit focus on transaction costs and how changes in these costs dictate the choice between an informal and formal enforcement mechanism, to the extent that their work implies that development always makes informal mechanisms more costly than formal mechanisms, it is misleading. Development lowers at least some of the costs involved in operating an informal enforcement mechanism. The use of faxes, computers, and other technologies, for example, reduces the costs of compiling and disseminating information about the credit history of consumers and businesses (Ellickson 1991). How these increases and decreases net out, however, remains to be explored.

The line between formal and informal mechanisms may be fuzzy. In some cases a hybrid system appears. For example, the courts may enforce social customs or practices sanctioned by merchants (Benson 1989). Or bankers may hold titles to farmers’
land while their loans are outstanding, as in Thailand (Siamwalla and others 1993) and Honduras (Stanfield and others 1990). Although actual possession gives the banks no formal legal right to foreclose on the land in case of default, bankers consider the leverage from holding the title to be a sufficient guarantee of repayment.

Much remains to be learned about the working of informal enforcement mechanisms and their relationship to the formal legal system. But at least some of the ways in which judicial reform can build on or complement informal systems are already apparent. Some are obvious. Projects should capitalize on the power of the media to police reform efforts by providing as much information about the judicial system as possible. In Argentina, for example, FIEL (1996) has proposed releasing information regularly about judicial caseloads, case backlogs, and other indicators of judicial productivity.

At a minimum, reform measures should try to bolster or complement informal enforcement mechanisms. In the case of reputation mechanisms, this could mean disclosing the identity of the parties to lawsuits, the status of cases in litigation, and the disposition of closed cases, including the amount of any damages awarded. Going further, current laws need to be reviewed to be sure they pose no obstacles to the easy and inexpensive dissemination of truthful public information about firms and individuals.

Whenever reputation information circulates, there is the possibility for abuse. False and defamatory material may be disseminated, jeopardizing privacy interests and compromising opportunities to make a fresh start. There are, however, many ways to strike a balance between these interests and the interests served by a well-functioning reputation system. In the United States, the Federal Fair Credit Reporting Act provides one model for balancing debtor and creditor interests. Another is Brazil’s juridically sanctioned process for publicizing information about those who have failed to pay their debts (World Bank 1997).

Often, enforcing reputation mechanisms depends on an agreement among the participants to boycott anyone who has a history of breaching an agreement (Milgrom, North, and Weingast 1990). But some U.S. courts have ruled that a boycott is illegal when one or more of the firms participating in the boycott is a competitor of the party boycotted (ABA 1997). These rulings have been sharply criticized for misapprehending the nature of anticompetitive agreements (Heidt 1986), but their influence may still be reflected in competition law. Accordingly, a program to foster informal contract enforcement mechanisms should also encompass a review of the applicable competition law.

On the basis of research in Ghana, Fafchamps (1996) recommends helping local business communities develop systems to share information about the reliability of suppliers and customers. Klein’s (1992) analysis of credit bureaus shows that freerider problems and other market failures are endemic in the start-up phase and can prevent the formation of credit reporting entities. In the United States credit bu-

Richard E. Mezick
reaus began as small, nonprofit associations, often run as an adjunct to the local chamber of commerce. Members were drawn from a tight-knit group of local merchants, and social ties supplied the incentives to overcome the problems of market failure. Where such incentives are missing, alternative ways of fostering the growth of credit reporting agencies should be considered. In Taiwan (China), the government's check clearinghouse serves as a substitute for a private credit bureau, charging a small fee for information about individuals who have bounced checks (Winn 1994).

Several other measures can be included in judicial reform projects to complement informal enforcement. One possibility is to transfer the responsibility for matters such as the registration of property rights to administrative agencies. This transfer can foster hybrid enforcement mechanisms. For example, technology now permits the creation of essentially paperless registries for land and other types of property. But paper titles may serve a purpose by permitting the development of a type of informal mortgage based on physical possession of the certificate.

Conclusion

Although judicial reform projects are an accepted part of the development landscape, crafting an effective project poses several challenges. Many questions about how judicial reform affects the economy, and society generally, remain to be answered. Not surprisingly, one result is that no consensus on the prerequisites for a successful project has emerged. Accordingly, in the absence of a better theoretical understanding of the impact of judicial reform, care is required in designing and implementing projects (Greif 1997; Dakolias 1996; World Bank 1995; IDB 1995). Reform must be preceded by an in-depth analysis of country needs—an analysis that must be continually reviewed as implementation proceeds.

Notes


1. An exhaustive list of potential interventions appears in Dakolias (1996). Other useful discussions of the range of possible reforms include the articles collected in Rowat, Malik, and Dakolias (1995).

2. Although Rechtsstaat is frequently rendered into English as "rule of law," as Fletcher (1996) notes, this translation can be highly misleading. The word "law" in the English phrase "rule of law" can mean either positive law, that is, any law enacted by a duly constituted government, or natural law, that is, precepts that meet some test of morality and justice. The German Recht is far closer in meaning to the latter than to the former, and thus a more accurate translation of Rechtsstaat would be "state ordered or ruled by natural law or justice." Similar translation problems arise with other European languages that, like German, have different terms for positive and natural law.
References

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


The Costs and Benefits of Regulation: Implications for Developing Countries

J. Luis Guasch • Robert W. Hahn

The literature on the benefits and costs of regulation demonstrates that this issue can be explored systematically using standard economic analysis. It also shows that regulation can have a significant adverse impact on economic growth. Specifically, regulation aimed at controlling prices and entry into markets that would otherwise be workably competitive is likely to reduce growth and adversely affect the average standard of living. In addition, process regulation can impose a significant cost on the economy. Nonetheless, social regulations may have significant net benefits for the average consumer if designed judiciously.

There are several policies developing countries might consider adopting to improve their general approach to regulation. The appropriate regulatory tools and framework will depend on many factors, including bureaucratic expertise, resource availability, political constraints, and economic impacts. There is a general need to enhance the capability for evaluating regulation at the local and national levels.

The past two decades have witnessed two trends in regulation. First, there has been an unparalleled rise in new regulations related to health, safety, and the environment. Second, there has also been substantial economic deregulation of certain industries in some countries, including airlines, trucking, railroads, financial markets, energy, and telecommunications. At the same time, to complement the objectives of their far-reaching privatization programs, developing countries have begun to examine regulations that keep prices inefficiently high.

The increased interest in regulatory reform can be explained in part by a growing understanding of the impacts of regulation. The rationale that certain sectors had to be regulated because they were "natural monopolies" vital to national social or strategic interests is no longer considered valid. Moreover, the costs that regulations can impose on the economy are now better understood. Indeed, scholars now appreciate that regulation is subject to political influences and is rarely implemented with the sole purpose of improving economic efficiency; in many cases regulation has had
adverse effects on the economy. That argument forms the basis for the trend toward regulatory reform as globalization increases the pressure to reduce production costs and as officials react to the increased mobility of capital and labor by adjusting their policies to reflect the likely impact of regulations on price changes.

Not all regulation is on the decline, however. Citizens in many countries express a desire for more regulation in areas such as environmental protection, public health, and safety standards. Rising incomes partly explain the increased interest; as consumers become wealthier, they demand more amenities such as cleaner air and water and better sanitation. And as politicians seek to supply more of these goods and services, they will also explore more efficient ways of supplying them.

Current political concerns with limiting tax increases in many countries are also creating incentives to use certain kinds of regulation. When legislators rein in spending and tax levels, regulation can be a useful substitute for achieving political objectives, such as redistributing income to particular interest groups in exchange for political support. In this kind of political environment, legislators adopt regulatory requirements or mandates whose costs are not directly paid for by taxpayers; although less visible, these costs are nonetheless real. From the government's perspective, the effort appears to be relatively low cost. The federal budget is barely affected when a major change is mandated by regulation.

Why Regulate?

The most common economic arguments for regulatory intervention are market failure and considerations of equity. In the case of social regulation, a primary rationale is that without government intervention, individual companies may not take into account the full social cost of their actions. A firm may pollute excessively unless it incurs some implicit or explicit cost for polluting, for example, or workers may not have adequate information on health and safety hazards in the workplace to make fully informed choices. The argument for economic regulation has to do with the potential for improving production efficiency. If economies of scale exist, a single firm may, in theory, be able to produce more efficiently than several competing firms, but its monopolistic power may need to be restrained through regulation. There is some justification for pursuing these objectives, but experience suggests that such rationales are often not persuasive in practice.

Correcting market failures and ensuring equity are laudable goals, but achieving those goals through regulation is not always successful. Just as there is potential for many kinds of market failure, so too is there potential for government failure. Economic regulation involves an understanding of the cost and demand structure of an industry, but a regulator typically does not have access to such information. Simi-
larly, health, environmental, and other social regulations must frequently be based on very limited information.

Political problems also lead to inefficient economic results. Because regulation redistributes resources and rents, politicians often use it to secure political gains rather than to correct market failures. A large array of regulatory instruments, such as quotas, licenses, and subsidies, is used to channel significant amounts of wealth to influential groups in society. In the United States, for example, price supports on peanuts resulted in an average annual consumer-to-producer transfer of $225 million (in 1987 dollars) with an associated deadweight loss of $34 million (Rucker and Thurman 1990). Wealth transfers are also a consideration in social regulation. Environmental and energy mandates frequently carry a heavy price tag.

Of course, if regulation becomes very inefficient and visible, pressure for reform may build. Firms with new technologies may lobby for reduced regulation; consumers and businesses may find ways to buy products and services at lower prices and opt out of the regulated markets. These considerations are particularly apt when demand can be met by tradable goods. Then the pressure to deregulate will come from domestic producers who must compete with less regulated imports. In addition, producers of tradable goods that rely on heavily regulated suppliers will have an interest in facilitating deregulation of these sectors to lower their overall production costs.

Estimating the Impact of a Regulatory Change

Perhaps the most difficult task in estimating the impact of a regulatory change is specifying the counterfactual: What would have happened in the absence of that change? By comparing the effects of the counterfactual with the change induced by the regulation, it is possible to estimate the differences in costs and benefits between the two conditions and to calculate the impact on producers and consumers.

Once a counterfactual has been specified, there are five general approaches to estimating the cost of regulation—econometric analysis, expenditure evaluation studies, engineering cost analysis, productivity studies, and general equilibrium analysis.

- **Econometric studies** typically evaluate output markets directly or use production and cost functions to measure the impact of regulatory change. Although such studies do provide a formal statistical apparatus with which to test hypotheses, their formulation is typically quite general, glossing over the precise nature of actual production functions. Macroeconomic models are sometimes used in conjunction with econometric estimation to assess the economywide effects.
- **Expenditure evaluations** frequently rely on surveys of firms or businesses to determine costs of compliance. Direct surveys produce easily quantified (and often large) estimates of the cost of regulation, but such surveys face several
problems. The first involves potential respondent biases. For example, a firm or corporation may inflate its estimated costs in hopes that, if others follow suit, politicians will consider providing regulatory relief. More important, however, direct expenditure studies do not specify a counterfactual. For example, an automobile company may choose to install stronger bumpers on its cars even without a regulation forcing it to do so. Attributing the added cost of such bumpers to government regulation overstates the impact of regulation.

• **Engineering approaches** calculate the added cost of installing equipment directly, adjusting for quality changes. Again, the question is what kind of car would have been built in the absence of specific environmental regulations.

• **Productivity studies** chart the difference between observed productivity changes over time and those that would have occurred in the absence of one or more federal regulations. These studies suffer from several problems, such as their reliance on expenditure data and an inability to specify the determinants of macroeconomic performance over time.

• **General equilibrium models**, which have become more popular recently, examine how a perfectly competitive market responds to a new policy, such as a change in regulation. The effects of a regulation can be linked to changes in output, employment, and in some cases welfare. Although general equilibrium models are not without their problems, including substantial data requirements, their results provide a better picture of regulatory effects in some cases. Simply stated, the methodological issue boils down to defining the conditions under which it is reasonable to assume away all but the most important effects.

The two basic approaches to measuring benefits rely either on asking people what they are willing to pay for changes in regulatory standards (contingent valuation) or on inferring from observed behavior the amount individuals actually pay for such quality changes. Although contingent valuation is particularly useful when markets do not exist for the commodity to be valued (for example, wilderness areas), it suffers from a likely divergence between what people choose to tell the interviewer and how they would behave under actual, rather than hypothetical, conditions.

Researchers rely on studies of averting behavior and on hedonic (shadow) price or wage methods to infer a willingness to pay. For example, certain neighborhoods have constructed barriers to mitigate the effects of highway or airport noise. The benefits of the noise reduction are then assumed to be at least the cost of the expenditure. Shadow price or wage methods attempt to evaluate the marginal value of quality improvements in specific amenities. For instance, the value workers implicitly place on safety is assumed to be the wage premium received by those working in more hazardous, although otherwise identical, circumstances. As such, econometric analyses of the implicit wage (or price) premiums can reveal the amount workers are willing to pay for improved workplace safety and, in the aggregate, their willingness to
pay to prevent an expected fatality. The technique has been fruitfully applied in a number of settings, including the valuation of reduced crime, reduced highway or airport noise, cleaned-up hazardous waste sites, and other location-specific amenities.

Hedonic estimation procedures are useful but rely on very indirect methods that can, under certain circumstances, lead to identification problems. For instance, specifying all the relevant demand and supply characteristics that determine where people choose to live is a daunting econometric task, and one that may be severely biased if any of the determinants have been omitted. Moreover, in the case of estimating the value of improved visibility and health, statistical problems often arise. Further, people may not be completely informed about certain risks, such as those associated with particular jobs, hazardous waste sites, and polluted air. Despite these problems, rapid advances in this relatively new technique promise improved empirical estimates of commodities not explicitly traded in the marketplace.

Estimates of Benefits and Costs

In the first study to synthesize data on the costs and benefits of regulation, Hahn and Hird (1991) distinguished between transfer costs and efficiency costs. Transfers represent payments from one group to another (for example, producers to consumers); efficiency costs represent net losses in producer and consumer surpluses. Both measures are important, but for different reasons. Transfer payments provide a measure of the winners and losers from regulatory change, while changes in net surplus provide an indication of the overall impact of a regulation on the economy or industry under investigation.

Tables 1 and 2, which show estimates of the costs of economic regulation and the costs and benefits of social regulation in the United States, demonstrate that it is possible to explore systematically the costs and benefits of regulatory activity using standard economic analysis. According to the analyses summarized in the tables, the efficiency costs appear to be much smaller than the transfer costs—information that should be taken into account when considering the effects of regulatory intervention. For systematic economic studies of federal regulations in the United States, see Weidenbaum and DeFina (1978); Litan and Nordhaus (1983); Hahn and Hird (1991); Hopkins (1992); Winston (1993); and Office of Management and Budget (1997).

Hopkins (1992) argues that the costs of process regulation are also substantial. Table 3 shows that the total cost of federal regulation in 1991 was estimated at $542 billion, or about 9.5 percent of gross domestic product (GDP), including transfers. The largest component of that cost was process regulation—the $189 billion in annual expenditures related to government paperwork requirements, primarily for tax compliance. These costs do not necessarily represent efficiency costs, however; one must consider all aspects of a tax system in evaluating its impact on efficiency. None-
<table>
<thead>
<tr>
<th>Sector</th>
<th>Efficiency costs</th>
<th>Transfers</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>International trade</td>
<td>17.3</td>
<td>85.6–110.6</td>
<td>Hufbauer, Berliner, and Elliot (1986)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>&lt; 14.1</td>
<td>&lt; 42.3 (^{a})</td>
<td>Wenders (1987)</td>
</tr>
<tr>
<td>Agricultural price supports</td>
<td>6.7</td>
<td>18.4</td>
<td>Gardner (1987)</td>
</tr>
<tr>
<td>Air transport</td>
<td>3.8</td>
<td>7.7</td>
<td>Morrison and Winston (1986, 1989)</td>
</tr>
<tr>
<td>Rail transport</td>
<td>2.3</td>
<td>6.8 (^{a})</td>
<td>Winston (1985)</td>
</tr>
<tr>
<td>Postal rates</td>
<td>—</td>
<td>4–12</td>
<td>President's Commission on Privatization (1988)</td>
</tr>
<tr>
<td>Milk marketing orders/price supports</td>
<td>0.4–0.9</td>
<td>0.9–3.5</td>
<td>Ippolito and Masson; Buxton and Hammond (both cited in MacAvoy 1977)</td>
</tr>
<tr>
<td>Natural gas (^{b})</td>
<td>0.2–0.4</td>
<td>5.0</td>
<td>Loury (1983)</td>
</tr>
<tr>
<td>Barge freight</td>
<td>0.2–0.3</td>
<td>0.6–0.9 (^{a})</td>
<td>Litan and Nordhaus (1983)</td>
</tr>
<tr>
<td>Davis-Bacon Act</td>
<td>0.2 (^{a})</td>
<td>0.5</td>
<td>Thiebolt (1975) (updated)</td>
</tr>
<tr>
<td>Credit</td>
<td>0.05–0.5</td>
<td>0.15–1.6 (^{a})</td>
<td>Litan and Nordhaus (1983)</td>
</tr>
<tr>
<td>Ocean freight</td>
<td>0.05–0.08</td>
<td>0.15–0.22 (^{a})</td>
<td>Jantscher (1975)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>45.3–46.5</strong></td>
<td><strong>172.1–209.5</strong></td>
<td>—</td>
</tr>
</tbody>
</table>

---

— Not available.

\(^{a}\) Figure estimated using 3:1 ratio of transfers to efficiency costs.

\(^{b}\) Cost of natural gas regulation is expected to approach zero as all price controls are lifted.


---

<table>
<thead>
<tr>
<th>Sector</th>
<th>Costs</th>
<th>Benefits</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>55.4–77.6</td>
<td>16.5–135.8</td>
<td>Hazilla and Kopp (1990); Freeman (1990); Portney (1990)</td>
</tr>
<tr>
<td>Highway safety</td>
<td>6.4–9.0</td>
<td>25.4–45.7</td>
<td>Crandall (1988)</td>
</tr>
<tr>
<td>Occupational safety and health (OSHA) (^{a})</td>
<td>8.5–9.0</td>
<td>Negligible</td>
<td>Crandall (1988); Denison (1979); Viscusi (1983)</td>
</tr>
<tr>
<td>Nuclear power</td>
<td>5.3–7.6</td>
<td>—</td>
<td>DOE policy study (cited in Litan and Nordhaus 1983)</td>
</tr>
<tr>
<td>Drugs</td>
<td>&lt; 1.5–3.0</td>
<td>—</td>
<td>Peltzman (1973)</td>
</tr>
<tr>
<td>Equal employment opportunity</td>
<td>0.9</td>
<td>—</td>
<td>Weidenbaum and DeFina (1978); Litan and Nordhaus (1983)</td>
</tr>
<tr>
<td>Consumer product safety</td>
<td>&gt; 0.034</td>
<td>—</td>
<td>Executive Office of the President (1990) (administrative costs only)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>78–107.1</strong></td>
<td><strong>41.9–181.5</strong></td>
<td>—</td>
</tr>
</tbody>
</table>

---

— Not available.

\(^{a}\) OSHA, Occupational Safety and Health Administration; DOE, Department of Energy.

Table 3. Costs of Federal Regulation in the United States, Selected Years
(billions of 1991 dollars)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental regulation</td>
<td>42</td>
<td>87</td>
<td>115</td>
<td>178</td>
</tr>
<tr>
<td>Other social regulation</td>
<td>29</td>
<td>30</td>
<td>36</td>
<td>61</td>
</tr>
<tr>
<td>Economic regulation-efficiency</td>
<td>120</td>
<td>73</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Process regulation</td>
<td>122</td>
<td>153</td>
<td>189</td>
<td>221</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>313</td>
<td>343</td>
<td>413</td>
<td>533</td>
</tr>
<tr>
<td>Economic regulation-transfers</td>
<td>228</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td><strong>Total costs</strong></td>
<td>540</td>
<td>473</td>
<td>542</td>
<td>662</td>
</tr>
</tbody>
</table>


Nevertheless, their sheer magnitude suggests that reducing paperwork would dramatically improve efficiency. ¹

Outside the United States, few studies have estimated the costs of regulation. In Australia the total cost of regulation was estimated to be 9–19 percent of GDP in 1986 (OECD 1996). Mihlar (1996) estimates that the costs of regulation in Canada amounted to 12 percent of GDP. Based on an assumed ratio between private compliance costs and spending on regulatory programs, he extrapolated national regulatory costs from federal and provincial administrative budgets. Although crude, the calculation gives a rough estimate of the size of the regulatory burden.

Because these cost estimates are often cited without careful analysis, several points about them are worth noting. First, the figures are highly uncertain and often incomplete. Where there are uncertainties in the data, these should be conveyed as clearly as possible to policymakers. Second, the figures developed using this approach to cost estimation are likely to understate the total impact of regulatory costs because they do not include the adverse effect that regulation typically has on innovation. Third, as shown in table 4, the cost of regulation as a fraction of GDP is significant for countries where such estimates are readily available, ranging from 7 to 19 percent. The Organisation for Economic Co-operation and Development, using a country-based macroeconomic model, has estimated that regulatory reform could increase GDP in the long run by as much as 3.5 percent in the United Kingdom and by as much as 6 percent in France, Germany, and Japan (OECD 1997a).

The Adverse Impacts of Regulatory Intervention

Many studies have attempted to estimate the adverse impacts of regulatory intervention:

- Christainsen and Haveman (1981) examined the effect of regulation on labor productivity and concluded that more than 10 percent of the contraction in
Table 4. Costs of Regulation and Gains from Deregulation
(percentage of GDP)

<table>
<thead>
<tr>
<th>Economy</th>
<th>Costs of regulation</th>
<th>Projected benefits of further economic deregulation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>11.8</td>
<td>—</td>
<td>Mihlar (1996)</td>
</tr>
<tr>
<td>European Union</td>
<td>—</td>
<td>4.5–7.0</td>
<td>Emerson and others (1988)</td>
</tr>
<tr>
<td>Germany</td>
<td>—</td>
<td>0.3</td>
<td>Lipschitz and others (1989)</td>
</tr>
<tr>
<td>Japan</td>
<td>—</td>
<td>2.3–18.7</td>
<td>OECD (1997a)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>—</td>
<td>0.5–1.1</td>
<td>Sinderen and others (1994) (cited in OECD 1997a); van Bergeijk and Haffner (1996)</td>
</tr>
<tr>
<td>United States</td>
<td>7.2–9.5</td>
<td>0.3</td>
<td>Hopkins (1992); Winston (1993)</td>
</tr>
</tbody>
</table>

— Not available.

Note: These numbers are underestimates of the effects of deregulation because the studies do not include all sectors in which deregulation can be beneficial. Further qualifications and elaborations of these estimates are available from the authors.

the growth of labor productivity in the mid-1970s was attributable to the expansion of federal regulations. Guasch (forthcoming), who also looked at labor costs, found that job growth was more robust in countries with more flexible labor markets (those at the top of table 5) than in those with strict controls. Although many other factors affect employment, there are strong reasons to believe that flexible labor market policies are likely to increase employment.

- In examining the long-term growth effects of regulation on eight industries from 1973 to 1987, MacAvoy (1992) found economywide losses of 1.5–2 percent of U.S. gross national product.
- Studies examining environmental, health, and safety regulations have yielded qualitatively similar impacts. For example, Jorgenson and Wilcoxen (1990) found that the costs of pollution control were associated with a reduction of more than 2.5 percent in the U.S. gross national product from 1974 to 1985. Robinson (1995) concluded that environmental and occupational health and safety regulations cumulatively reduced multifactor productivity in the manufacturing sector by more than 10 percent from 1974–75 to 1985–86.
- Research on the relationship between regulation and output growth in OECD countries by Koedijk and Kremers (1996) concluded from an index of regulatory intensity that countries with the least regulation enjoyed the highest growth in output per person. The measures the authors construct are admittedly rough, but they may serve as a proxy for the degree to which markets are regulated in different countries.
### Table 5. Labor Regulations

<table>
<thead>
<tr>
<th>Country</th>
<th>Payroll taxes (as a percent of the wage bill)</th>
<th>Severance payments</th>
<th>Collective bargaining</th>
<th>Employment growth, 1992–95</th>
<th>Unemployment rate, 1996</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Australia</td>
<td>27.8</td>
<td>Low</td>
<td>Centralized</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Chile</td>
<td>20.9</td>
<td>Low</td>
<td>Firm level</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>22.9</td>
<td>None</td>
<td>Firm level</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Malaysia</td>
<td>24.3</td>
<td>Low</td>
<td>Firm level</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>New Zealand</td>
<td>11.5</td>
<td>None</td>
<td>Firm level</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>20.1</td>
<td>None</td>
<td>Firm level</td>
<td>1.8</td>
</tr>
</tbody>
</table>

#### More flexible markets

#### Less flexible markets

<table>
<thead>
<tr>
<th>Country</th>
<th>Payroll taxes (as a percent of the wage bill)</th>
<th>Severance payments</th>
<th>Collective bargaining</th>
<th>Employment growth, 1992–95</th>
<th>Unemployment rate, 1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>50.0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>High</td>
<td>Centralized</td>
<td>−0.7</td>
<td>17.2</td>
</tr>
<tr>
<td>France</td>
<td>54.7</td>
<td>High</td>
<td>Centralized</td>
<td>−0.4</td>
<td>11.6</td>
</tr>
<tr>
<td>Italy</td>
<td>52.8</td>
<td>High</td>
<td>Centralized</td>
<td>−1.7</td>
<td>10.2</td>
</tr>
<tr>
<td>Spain</td>
<td>38.2</td>
<td>High</td>
<td>Centralized</td>
<td>−1.6</td>
<td>22.4</td>
</tr>
</tbody>
</table>

*Note:* The data for France, Spain, Italy, and Japan correspond to 1994, those for Malaysia to 1995, and those for Argentina and Chile to 1996. Severance payments are based on OECD indexes.

* a. Employment growth is measured as annual average percentage growth.

* b. Argentina amended its labor laws in 1996, and payroll taxes now average 41.0.

*Source:* Guasch (forthcoming).

### Assessing the Gains from Regulatory Reform

In the area of social regulation, Hahn (1996), who reviewed more than five regulatory impact analyses (RIAs) covering health, safety, and environmental regulations from 1990 to mid-1995, reported considerable variation in the type and quality of analysis performed by government agencies. Cost-benefit analyses were often incomplete, and the costs of imposing the regulations were reported to be greater than the monetary benefits in more than 80 percent of the regulations. Based on this analysis, the net benefits from these regulations amounted to about $280 billion (1994 dollars) since 1990. Figure 1 shows the distribution of net benefits for 54 rules. The left side of the figure shows the number of rules with net costs that fall in various categories. The right side of the figure shows the number of rules with net benefits that fall in various categories. The aggregate net benefits are positive because many of the rules have substantial benefits. Eliminating those that would not pass a benefit-cost test could increase the present value of net benefits by more than $115 billion.

For various reasons, however, the RIA numbers cannot be taken at face value. Both theory and empirical evidence suggest that government agencies are likely to overstate substantially the aggregate net benefits of their programs (for instance, protecting the environment or improving safety in the workplace) to show that they are meeting the demands of interest groups.
A useful measure of the impact of regulations is how many lives are saved. Morrall (1986), in a review of several final and proposed regulations calculates that the amount needed to avoid each premature death varies over eight orders of magnitude—from roughly $100,000 to more than $5 trillion (1990 dollars)! This suggests that regulations could be developed that would prevent many more premature deaths while still saving consumers money. Tengs and Graham (1996) found that reallocating regulatory expenditures within the United States to those investments that are most effective could avert an additional 60,000 deaths, or twice the current number. In addition, reallocating $8 billion (1994 dollars) in regulatory expenditures from the United States to developing countries could save more than 100 million additional life-years (Hahn 1996).

The Benefits of Deregulation in Industrial Countries

The overall welfare gains from deregulation in the United States have been substantial. Aggregate welfare gains from eliminating entry and exit restrictions and freeing prices to market levels ranged from $35 billion to $46 billion (1990 dollars) a year, of which consumers gained $32 billion to $43 billion from lower prices and better services, and producers gained about $3 billion a year from increased efficiency and
lower costs (table 6). Winston (1993) estimates that additional gains from eliminating remaining distortions could be more than $20 billion a year. Even so, there is evidence that the gains are likely to be significantly understated. In a recent paper, Winston (1998) notes that although industry may adjust prices to reflect marginal costs quickly after deregulation, it takes time to optimize production. He argues that policymakers and the public tend to notice only the short-term effects and therefore undervalue the benefits of deregulation; the positive impact that deregulation has on innovation is frequently overlooked. Such innovations increased productivity and reduced operating costs by one-fourth to more than one-half in different industries.

Sectoral studies yield similar results on the adverse consequences of economic regulation. Caves, Christensen, and Swanson (1981) compared the productivity growth of U.S. railroads from 1956 to 1974 with the growth of Canadian railroads during the same period. Both industries had access to the same technology, but Canada’s railroads were not as heavily regulated as those in the United States. The authors argue that regulation substantially reduced productivity growth, estimating that if the U.S. railroads (with a growth rate of 0.5 percent) had experienced the same growth as Canada’s (3.3 percent), the cost of providing rail services in 1974 would have been $13.8 billion (1985 dollars) lower. Willig and Baumol (1987) estimated that, after rail deregulation in the United States, annual operating expenses dropped 26 percent from 1980 to 1985, while traffic volume remained virtually unchanged and investment increased.

The empirical evidence on the trucking, airline, telecommunications, and financial industries is impressive.

- Average unit costs in the U.S. trucking industry declined from 30 cents a ton-mile in 1977 before deregulation to 10 cents a ton-mile in 1983 after
deregulation (in 1977 dollars). The annual welfare loss from regulation of rail and motor carrier rates was estimated at $1 billion to $4 billion (1977 dollars) (Braeutigam and Noll 1984; Winston and others 1990).

- The airline industry reduced total costs per unit of service by approximately 25 percent; labor costs were cut as well (by 17 percent at American Airlines and 24 percent at United Airlines) with little effect on output in the first few years following deregulation (Caves and others 1987). In addition, excess capacity declined and productivity rose. Morrison and Winston (1995) estimate the net annual gains to passengers at $18.4 billion (1993 dollars).

- By 1996 long-distance telephone rates in the United States had fallen by more than 70 percent as a result of the divestiture of AT&T in 1984 (Taylor and Taylor 1993; Wall Street Journal 1991). The emergence of profitable services such as cellular telephony and voice messaging after divestiture shows how regulation can slow the introduction of new products and discourage innovation. Although the concept of cellular phones was discussed in the late 1940s and the technology was available in 1973, the Federal Communications Commission did not begin to issue licenses until ten years later—a delay that, by one estimate, cost the U.S. economy more than $25 billion a year in 1983 (Rohls, Jackson, and Kelly 1991), or about 2 percent of gross domestic product. Similarly, the delay in introducing voice messaging services cost more than $1.3 billion (1994 dollars) a year (Hausman and Tardiff 1996).

- Postderegulation effects have been observed in the securities, investment, and banking sectors. For example, when brokerage fees were deregulated, rates dropped by 25 percent and savings from overall consolidation and cost reduction amounted to 30 percent (Jarrell 1984). Studies have shown that even after accounting for changes in the services offered, the cost reductions were significant. In the United States partial deregulation of the banking and savings and loan industries resulted in employment cuts of more than 20 percent during 1984–93 and an increase in productivity (as measured by revenue per employee) of more than 300 percent (Guasch and Spiller forthcoming).

Although the database outside the United States is less extensive, there is reason to believe that the gains from deregulation of many industries elsewhere could be substantial (see table 4). For example, airline fares in Europe are roughly twice as expensive as in the United States (Airfare Management Unit 1995, 1996; Consulting Services Group 1995, 1996), but profitability is well below that of U.S. carriers. Lifting price and entry restrictions could reduce fares and benefit consumers. Indeed, the high-cost carriers, such as Iberia and Air France (both state-owned), have survived only with government aid. Good, Röller, and Sickles (1993), who argue that liberalization would lead to competition between international carriers and a convergence of cost structures, estimate that if the European airline industry were as efficient as
the U.S. airline industry, it would have saved approximately $4 billion a year in 1986 dollars.

Deregulation of electricity markets in Europe also offers significant opportunities for gains (Electricity Association Services Ltd. 1996). In Germany, for example, strict regulations require domestic companies to purchase electricity from regional producers even though cheaper power is often available nearby. The extent of the potential gains for German consumers is difficult to estimate, but in the United Kingdom, energy deregulation resulted in a 70 percent increase in productivity and an 18 to 21 percent reduction in franchise contract prices (OECD 1997a). Elsewhere in the European Union, firms pay over 50 percent more for their electricity than do their American counterparts. Moreover, the impact of higher energy prices on the overall economy can be quite significant (Navarro 1996). For example, a 30 percent increase in electricity prices tends to raise the prices of goods such as paper and pulp, metals, chemicals, and glass by roughly 2.5 percent.

Benefits of Deregulation in Developing Countries

In countries that have deregulated, the efficiency gains have been quite significant. For example, deregulation of entry into the long-distance telephone market in Chile has cut rates by 50 percent, making them close to U.S. rates (Guasch and Spiller forthcoming). In some Latin American countries, private sector participation in the telecommunications sector has cut waiting time for installation of new lines from a minimum of two years to a matter of weeks. At the port terminals in Buenos Aires, competition in operations has led to an 80 percent reduction in the fees charged. And opening port operations to multiple parties in the port of Montevideo has increased productivity by 300 percent. All those results were achieved within a year of deregulation (Guasch 1996).

A study of Argentina (Fundación de Investigaciones Económicas Latinoamericanas 1991) assesses the welfare cost of regulations and other government interventions in the 1980s (table 7). The total costs of regulation and state intervention amount to more than $4 billion a year (1990 dollars), and this is only for the selective listed interventions. Using a general equilibrium model, Chisari, Estache, and Romero (1997) estimate the gains from privatization and regulation in Argentina at about 1.3 percent of gross domestic product, or $3.3 billion. They also find that all income classes benefit.

It would be useful to assemble data on regulatory costs in other developing countries comparable to those assembled for Argentina. Yet there is no shortage of specific cases where economic regulation has had adverse consequences. For example, Uruguayan firms and consumers are paying an implicit tax of at least 30 percent for water, phone, and electricity, thus hindering the competitiveness of Uruguay’s products compared with those of Argentina, Brazil, and Paraguay, its fellow members in

**J. Luis Guasch and Robert W. Hahn**
### Table 7. The Costs of Regulation in Argentina

(millions of 1991 dollars)

<table>
<thead>
<tr>
<th>Sector</th>
<th>Period</th>
<th>Average annual cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High reserve requirements and subsidized credit by the central bank</td>
<td>1987</td>
<td>1,000</td>
</tr>
<tr>
<td>Inflation taxes on checking accounts</td>
<td>1983–87</td>
<td>670</td>
</tr>
<tr>
<td>Fuel price controls</td>
<td>1977–87</td>
<td>350</td>
</tr>
<tr>
<td>Health services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra costs from double affiliation</td>
<td>1986</td>
<td>150</td>
</tr>
<tr>
<td>Idle capacity in public hospitals</td>
<td>1987</td>
<td>172</td>
</tr>
<tr>
<td>Fish export subsidies</td>
<td>1986–87</td>
<td>12</td>
</tr>
<tr>
<td>Efficiency costs from domestic consumption restrictions in cattle markets</td>
<td>1984</td>
<td>104</td>
</tr>
<tr>
<td>Efficiency costs of the special fund for tobacco</td>
<td>1987</td>
<td>30</td>
</tr>
<tr>
<td>Air transport regulations</td>
<td>1988</td>
<td>75</td>
</tr>
<tr>
<td>Restrictions on rail transport of cement, wine, and grain</td>
<td>1987</td>
<td>95</td>
</tr>
<tr>
<td>Truck transport</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costs of road deterioration</td>
<td>1987</td>
<td>100</td>
</tr>
<tr>
<td>Costs of provincial regulations on the transport of grains</td>
<td>1987</td>
<td>30</td>
</tr>
<tr>
<td>Port restrictions on price and entry</td>
<td>1987</td>
<td>90</td>
</tr>
<tr>
<td>Regulations imposed on business</td>
<td>1965–87</td>
<td>1,200</td>
</tr>
<tr>
<td>Regulations on employment in the public sector</td>
<td>1987</td>
<td>120</td>
</tr>
</tbody>
</table>

**Note:** The costs of regulation measure different concepts, such as efficiency losses in the economy, cost premiums to consumers, tax reductions, and subsidies. Thus, it might not be technically correct to total the costs.

**Source:** Fundación de Investigaciones Económicas Latinoamericanas (1991).

The costs of various kinds of process regulation caused by inefficient bureaucracies and high levels of corruption can add substantially to consumer burdens in developing countries. For example, customs administrations tend to be plagued by inefficiency and corruption, imposing a high cost on traded goods. The Nigeria Manufacturers Association (1996) says that permission to clear goods in that country requires 27 stages and takes five to eight weeks. Inefficient regulation of port operations has contributed to implicit tariffs of 5 to 15 percent on exports in Latin America (Guasch and Spiller forthcoming). Surveys indicate that managers spend between 10 and 30 percent of their time managing process regulation, incurring costs on produced goods or services in the range of 5 to 15 percent (World Bank 1997).

---

*The World Bank Research Observer, vol. 14, no. 1 (February 1999)*
Box 1. Montevideo Taxicab Market

Entry restrictions for taxicabs in Montevideo, Uruguay made the 1990 market price of a taxicab license $60,000 (in 1990 dollars). Although the license is nominally lower than the $125,000 price in New York, lower Uruguayan per capita income means that the market value of the license as a proportion of per capita income is more than four times higher. The regulation of the taxicab market has led to a scarcity of taxicabs, as reflected in difficulty in hailing taxicabs, in high costs borne by consumers, and in capture and wasteful rent-seeking activity by the taxi owners association.

Source: Guasch and Spiller (forthcoming).

As noted earlier, Mexico is reviewing regulations for major federal agencies. The purpose of the review is to eliminate unnecessary regulations, simplify regulations that are unnecessarily burdensome, and make the process more transparent (box 2). By the end of 1997, approximately half of all regulations (called formalities) had been reviewed in seven of twelve ministries. Of those reviewed, 38 percent were scheduled to be eliminated and an additional 54 percent were scheduled to be simplified in 1998.

Conclusions

Regulations receive relatively little scrutiny, both because politicians wish to hide the cost of regulation from citizens and because estimating the costs and benefits of regulation is difficult. A better information base on the economic impact that different regulations have would enhance public decisionmaking. There are several poli-

Box 2. Regulatory Reform in Mexico

The government of Mexico is undertaking an examination of regulatory structure at the federal, state, and local levels. The aims of the Agreement for the Deregulation of Business Activity include streamlining federal regulations, reducing corruption by codifying regulation, and helping to promote more efficient and effective regulation. The program has enjoyed some early successes. Recent legislation simplifies administrative procedures, requires a quicker administrative response time, and reduces paperwork for foreign investors. In addition, a series of legal reforms aims to simplify court proceedings and reduce the costs of commercial lending. As a result of these reforms, Mexico City's Superior Court reports that the number of civil suits filed fell by 24 percent from 1995 to 1996. Agency-by-agency rule simplification and elimination are also proceeding swiftly. For example, the approval time for a business to begin operation has been reduced from an average of more than 200 working days to a maximum of 21 working days. Finally, a complete inventory of federal rules in effect is available on the Internet; easy access should help to reduce corruption and compliance costs.

cies that developing countries might consider; the recommendations here are purposefully general. In that spirit, the first important point to make is that effective policies will differ across countries. The appropriate regulatory tools and framework will depend on several factors, including bureaucratic expertise, resource availability, political constraints, and economic impacts.

There is a general need, moreover, to enhance the capability for evaluating regulation at the local and national levels (Hahn forthcoming), as illustrated by the absence of even rudimentary data in many countries on the effects of regulation. Countries should attempt to develop a "regulatory budget" that would show the economic effects of regulations and that would be published along with the government's fiscal budget. Such a capability will take time to develop.

Several jurisdictions, including some in developing countries, are putting procedures in place that would require a benefit-cost analysis before significant regulations could be implemented (OECD 1997b). This is likely to have a constructive influence on public policy by providing better information and holding government officials and political leaders more accountable (Hahn and Litan 1997). In the short term, it is important for agencies charged with administering regulations to begin by assembling crude cost and benefit data. For example, an agency could specify the rationale for a proposed regulation, the likely direct and indirect costs, a qualitative description of benefits, an assessment of other alternatives, including the status quo, and an explanation of why other alternatives were not selected if they are likely to be better for the average citizen.

Such analyses should not be too burdensome. For regulations that have a limited impact, no analysis may be necessary. For regulations with potentially large economic consequences, more resources should be devoted to evaluation. Ideally, such analyses should be both prospective and retrospective, so that analysts can learn how to improve their impact assessments by comparing their predictions with actual political outcomes. To start with, we recommend developing a low-cost information management system that highlights some of the more important economic impacts of regulation. Front-line agencies need to be involved in the process so that they become more sensitive to the economywide impacts of their proposals.

As administrative capabilities evolve, a more thorough cost-benefit analysis will be required to support regulatory reform. Because economic regulation often results in economic inefficiency, the burden of proof should be on those who wish to maintain such regulation. In the case of social regulation, flexibility should be encouraged so that consumers and producers are able to innovate in response to regulations. Thus, for example, performance standards for meeting a pollution goal are generally preferred to standards that dictate the use of a particular technology. Of course, the amount of flexibility in a regulatory policy should be based, in part, on the ability of the administrative agency to implement it effectively (Hartman and Wheeler 1995).
Although an economic analysis of regulatory policies can be helpful, regulations often have unexpected and perverse consequences (Ackerman and Hassler 1981). Thus it is better to proceed with extreme care and err on the side of less regulation, particularly when dealing with economic interventions. Where there is no clear economic rationale for a regulation, it should be removed (Hahn 1998). Licensing and price or quota interventions, for example, do not serve the public interest but instead transfer political favors to preferred constituencies (Huber and Thorne forthcoming; Guasch and Spiller forthcoming). Removing such distortionary favors may not be easy in many cases and may involve resource transfers to politically powerful constituencies.

As they consider reforms, policymakers need to give a great deal more thought to the design of regulatory frameworks. In some instances, partial deregulation may not lead to an improvement over the status quo. For example, removing price restrictions but retaining entry barriers could lead to inefficient pricing. Full deregulation can lead to problems with monopoly unless great care is taken in managing the transition to a deregulated environment. The point is that the strategy for regulatory reform is critical to its effectiveness. Another set of problems stems from a tendency for a single-mission agency (health or education, say) to consider its mandate exclusively and to overstate the benefits of its program and understate the costs. As noted above, one way to address this problem is to require the agency to develop more data on the costs of specific regulatory proposals. A second is to limit the agency’s mandate. Other options include sunset requirements that would limit an agency’s authority to a fixed period unless renewed by legislative mandate and requiring the approval of a central—independent—agency that is primarily concerned with the economywide impacts of regulations (Hahn 1997). Because officials are concerned about issues of equity and efficiency, the regulations they write tend to be unduly complicated. This complexity not only gives bureaucrats and lawyers control over decisions but also makes it difficult for average people to understand the economic implications involved. The more transparent the regulations, the more they are likely to reduce the potential for corruption and increase the perceived legitimacy of the system. Straightforward language makes careful scrutiny possible and limits the likelihood that political interest groups will capture the benefits. A move toward greater transparency will occur as people begin to understand some of the hidden costs of regulation.

Developing countries have begun to realize the benefits of reforming economic regulation, but much remains to be done in the area of social regulation. Yet, it is beginning to appear on the policy agenda, if not from domestic pressure, then from interest groups in industrial countries.

The overall lesson is not that regulation is generally undesirable but that it often has undesirable economic consequences. Moreover, these effects result partly from
political forces that lead to inequitable redistribution of wealth (Stigler 1971). We believe such forces can be mitigated by more sharply evaluating the consequences and tradeoffs involved before a regulatory policy is set in stone.

Notes

J. Luis Guasch is lead economist for the Latin America and the Caribbean Regional Office and for the Development Economics Group at the World Bank and is professor of economics at the University of California, San Diego. Robert W. Hahn is director of the American Enterprise Institute (AEI)-Brookings Joint Center for Regulatory Studies, a resident scholar at AEI, and a research associate at the Kennedy School of Government, Harvard University. The authors thank Ed Campos, Michael Klein, Brian Levy, Joseph Stiglitz, and the referees for their helpful comments and Jonathan Siskin, Lisa Taber, and Fumie Yokota for their excellent research assistance.

1. Hopkins’ estimate for the total cost of regulation includes transfer costs and process costs. Subtracting transfer costs yields an estimate of $413 billion, or more than $1,500 per person for 1991. Process costs account for about half of the $1,500. For a critique of Hopkins’ analysis, see Office of Management and Budget (1997).

2. Employment declined from 260,000 in 1987 to 190,000 in 1990.

References

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


Comparative Perspective." World Bank, Private Sector Development Department, Washington, D.C. Processed.


The Twelfth World Congress of the IEA will be organised by the Asociación Argentina de Economía Política. The five day programme includes plenary sessions, invited sessions and presentation of contributed papers. Invited lectures will cover two topics: "Inequality in the World Today", with a programme arranged by Professor Richard Freeman (Harvard, USA), and "The Current State of Macroeconomics", arranged by Professor Jacques Drèze (CORE, Belgium).

Call for Papers: Submissions of papers related to any economic area are welcomed. The Programme Committee will screen the contributed papers. Authors will be notified by April 15. Three volumes of proceedings on macroeconomics, inequality and Latin American issues will be published. Those wishing to participate with a paper should send three copies of the paper by February 14, 1999, to Prof. David de la Croix, IRES - Université Catholique de Louvain, Place Montesquieu 3, B-1348 Louvain - La - Neuve, Belgium, Tel: +3210473453 Fax: +3210473945 e-mail: delacroix@ires.ucl.ac.be

Programme Committee Chairman: David de la Croix (Université Catholique de Louvain), Bina Agarwali (University of Delhi), Kenneth Arrow (Stanford University), Anthony Atkinson (Oxford University), Costas Azariadis (University of California), Partha Chander (Indian Statistical Institute), Vittorio Corbo (Catholic University of Chile), Karel Dyba (Ministry of Economics, Czech Republic), Victor El’as (University of Tucumán), Jean-Paul Fitoussi (OFCE - Paris), Jean-Michel Grandmont (CNRS), Daniel Hamermesh (University of Texas), Yujiro Hayami (Aoyama Gakuin University), Seppo Honkapohja (University of Helsinki), Valeri Makarov (Russian Academy of Sciences), Mustapha Nabi (World Bank), Fernando Navajas (University of La Plata), Dan O’Flaherty (Columbia University), Luigi Pasinetti (Università Cattolica del Sacro Cuore), Suzanne Scotchmer (University of California), Amartya K. Sen (Harvard University), Hans Werner Sinn (University of Munich), Rehman Sohban (Centre for Policy Dialogue-Dhaka), Erich Streissler (Universität Wien), Kotaro Suzumura (Hitotsubashi University), Jean-Pierre Urbain (University of Maastricht), Alan Woodland (University of Sidney) and Stefano Zamagni (Università degli Studi di Bologna).

Organising Committee Chairman: Enrique A. Bour (University of Buenos Aires, FIEL), e-mail: enrique@fiel.org.ar; Vice Chairman Rolf R. Mantel (University of San Andrés) e-mail: rrmantel@info4rek.com.ar; Executive Secretary Victor A. Beker (University of Belgrano, Univ Buenos Aires) e-mail: beker@ub.edu.ar

For Further Information: Congress Secretariat: MARIA GAZIANTI & ANOCIADOS, Carlos Pellegrini 9°y, 6th Floor (1089) Buenos Aires - Argentina e-mail: grazianti@inf.at.com.ar, Tel: +54113941726, +54113953435, Telephone numbers as from 1/1/99: +54113917726 +54111459555357 Fax: +5411393436. Fax number as from 1/1/99: +54113953436 Internet: http://www.aecp.org.ar/12worldcongress.
Introducing a publication with the most vital commodity of all:
thorough, reliable information.

Subscribe now to Global Commodity Markets, the only publication that gives you the complete picture on world commodities. With the accuracy and objectivity of the World Bank.

This new World Bank quarterly replaces—and significantly expands upon—its predecessor, Commodity Markets and the Developing Countries. You'll find price forecasts for 46 primary commodities and detailed market reviews for 27 major ones.

A full-service subscription includes not only the quarterly publication but also electronic Monthly Updates and electronic access to the quarterly. The Updates offer the most recent commodity prices and a brief review of market developments.

The detailed reviews, featured on generous two-page spreads, are packed with useful information, from end-of-year stocks to major producers to futures prices. The World Bank also provides objective market analysis.

A full-service subscription is just $645 per calendar year. You may also subscribe to the quarterly alone for $495. Subscriptions in 1999 are for a calendar year and will be prorated on a quarterly basis.

If you need dependable information on world commodities, subscribe now to Global Commodity Markets. It's your best vantage point for a global view.

ORDER FORM

Yes, sign me up for Global Commodity Markets.

☐ Full-service subscription, including the quarterly publication, electronic Monthly Updates, and electronic access to the quarterly, at $645 a year.*
☐ Quarterly subscription (print only) at $495 a year.*

* Subscriptions are for a calendar year and will be prorated on a quarterly basis.

Name _____________________________
Title _______________________________
Organization _________________________
Address ______________________________
City __________________ State ______ Zip ______
Phone __________ Fax __________
E-mail ______________________________

Payment: Orders from individuals must be accompanied by payment or credit card information. Credit cards accepted only for orders addressed to the World Bank. Check with your local distributor about acceptable credit cards. Please do not send cash.

☐ Check no. ________________________ in the amount of $ __________ enclosed. When ordering directly from the World Bank, make check payable in U.S. funds only drawn on a U.S. bank to: The World Bank. Please send your check with your order.

☐ Charge $ __________ to my:
☐ American Express ☐ Mastercard ☐ Visa
Card no. _____________________________
Expiration date __________ / ______

Name as it appears on the card (required for all credit card charges)
Signature ____________________________

Note: We must have a fax number or e-mail address to notify you of Monthly Updates. Please provide both if possible.

Mail order to: The World Bank, P.O. Box 7247-7956, Philadelphia, PA 19170-7956, USA
Fax order to: 202-476-2107  Order online: http://www.worldbank.org/html/extjob/publications.html  Questions? E-mail us at books@worldbank.org.
<table>
<thead>
<tr>
<th>Country</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>Tel: (32 2) 538-0841, Fax: (30 1) 364-8254, E-mail: <a href="mailto:order.dept@renoufbooks.com">order.dept@renoufbooks.com</a></td>
</tr>
</tbody>
</table>
Would you like to have more information about

THE WORLD BANK

Research Observer?

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


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

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