Market Failures as Bases
of Urban Policies

George S. Tolley

Urbanization in developing countries remains controversial. At the heart of many economic development efforts has been promotion of industrialization, which fosters the growth of urban centers. But urbanization has also been viewed as undesirable because it takes people away from a rural way of life that is thought to be intrinsically better and because of the economic and social problems found in cities.

Urban poverty has been viewed as one of the ills of growth. Recurrent assertions that overurbanization has taken place refer to environmental problems in cities, wage policies that induce excess rural-to-urban migration, and government and tax policies that favor cities. The arguments have usually been made in qualitative terms and have seldom been related seriously to the functioning of markets.

This chapter offers an analysis of these and other issues that bear on urbanization performance. The causes of urban poverty are first brought out, and the relation of poverty to overall economic development considerations is discussed. Attention then turns to accounting problems that distort measured changes in national income when urbanization occurs.

Overurbanization is then analyzed. Models with measurement are developed that deal with pollution, congestion, urban wages, and government fiscal policies. The models give indications of the city size effects and income losses attributable to urbanization externalities. The reasons why large cities in developing countries are so large and are getting larger are analyzed, with special attention to transport.

The final section draws policy implications. Possibilities are considered for internalizing externalities as a first-best policy that is superior to trying to influence urbanization as such. Options and pitfalls in undertaking direct spatial policies to influence urbanization are discussed, and the role of externalities in nonspatial policies is considered. The need to improve methods for evaluating the spatial effects of infrastructure decisions is emphasized, and evaluation needs in the management of urban services in general are brought out.

Urbanization and Poverty

The "rings of thorns"—or shantytowns on the outskirts of many of the large cities of the developing world—have been the subject of much concern, and they are, at least subconsciously, partly responsible for the view expressed by some that excessive urbanization has occurred. The growth of the numbers of poor people is an especially noted symptom of urbanization, and there may be a lurking hope that curbing urbanization would reduce poverty.

Interest persists in estimating the numbers of people in poverty according to some definition. Aside from the problem of choosing the level of income below which poverty is deemed to exist (a workable procedure if the choice is recognized as arbitrary), the problem arises of measuring the extent of poverty among different groups within the population. That task is made difficult by the varying circumstances of people, which make simple comparisons of money income invalid.

Explaining the extent of urban poverty requires an understanding of what people earn and is more difficult than explaining the overall degree of urbanization (discussed in chapter 2). Yet a foremost reason why poor people are found in large numbers in cities of developing countries is the sheer overall magnitude of poverty in those countries and its distribution among urban and

rural areas because of migration. Through migration, rough equalization of real income levels between urban and nonurban areas can be expected to be achieved within each income stratum, from rich to poor. Lack of knowledge of opportunities in different areas is not an apparent barrier to migration, nor is lack of such contacts as friends and relatives who can help the migrant get settled in a new environment.

Equalization of real earnings between rural and urban areas may not be exact. Whether equalization takes place cannot be proved or disproved by looking at urban and rural money incomes at any one time. Observed money incomes are not directly commensurate in rural and in urban areas. People place different values on differences in the availability of goods, prices, and general environments. Thus the equalization is not between money incomes but rather between real incomes, including the values people place on differences between rural and urban areas.

The hypothesis that a rough equalization of real earnings takes place is supported by two arguments. The first is introspective: it does not stand to reason that people, in making choices about where to live, will ignore opportunities in other places. If they pay some attention to alternatives, differences in real earnings between places will not get far out of line. The second argument is empirical: if people did in fact ignore real income differences, the erratic pace of demand for products and supplies by people in different places, as determined by demographic considerations, would produce wide swings in observed relative differences in money incomes. Although some annual fluctuations in the ratio of money incomes between rural and urban areas take place, over the years the fluctuations are limited. Thus, explaining poverty in cities becomes a matter of explaining how the demand for the services of poor people in urban areas interacts with their conditions of supply. The same type of explanation applies to people in rural areas. The supplies in each case are determined by the total number of persons of a given type in the economy.

Suppose we let N_u and N_a refer not to the total number of persons in urban and rural areas, as in chapter 2, but to numbers of persons who supply unskilled or common labor in the two types of area. Then the same kind of explanation of the rate of urbanization of the total population that was given in chapter 2 can be used to explain the rate of urbanization of unskilled labor. The income elasticities will reflect the income elasticities of the final products in which the unskilled labor is used, and they will also reflect the effects of the accumulation of physical capital and the growth in numbers of skilled laborers on the demand for unskilled labor. Whereas the demand for unskilled labor in rural areas consists primarily of demands for field work and lower-income

farming, the demand in urban areas consists of demands for common labor in a variety of activities, including factory work and services. Services in urban areas are supplied through work in formal firms and as domestic help as well as in casual and informal employment. Common labor is itself an aggregation and simplification that masks many gradations of productivity within the unskilled category. In speaking of the poor, one is speaking primarily of families whose breadwinners are at the lower ends of the unskilled distribution.

Although, as noted, the demand for unskilled labor is affected by accumulations of physical capital and growth in the number of skilled laborers, these two types of productive factors will also tend to grow as the output of urban commodities rises. The number of unskilled laborers, the number of skilled laborers, and the amount of physical capital will all tend to grow with urbanization. Performance levels of skilled laborers may rise, owing to education and to experience with more sophisticated techniques, and physical capital may accumulate more rapidly than population. These developments may affect the demand for unskilled labor, but they operate to some extent in rural as well as in urban areas.

Changes in the numbers of unskilled laborers in urban and rural areas often go along, more or less, with changes in urban and rural populations. Without claiming an exact proportionality, it seems that changes in unskilled labor and population in an urban area tend to move similarly. If so, the explanation of quantitative changes in percentage terms for the urbanization of unskilled labor is similar to that already considered for the population as a whole. In some countries lower-income and subsistence agriculture become outmoded, leading to a low or even negative apparent income elasticity for unskilled labor in agriculture. In this case the general approach to the explanation of urbanization of the unskilled population remains the same as for the total population, but the quantitative outcome is different. The implication in this case is that the rate of urbanization of the unskilled population will be more rapid than that of the population as a whole.

There are three reasons why the absolute level of poverty diminishes during development. First, the application of new techniques and capital raises the marginal productivity of unskilled labor. This phenomenon is in many ways the essence of development. If development does not succeed in raising the well-being of all strata, it does not succeed in its purpose. Moreover, looking at the various gradations of development around the world, one sees that a primary distinguishing feature among countries is the extent to which the level of well-being of all strata appears to be raised.

It may be hypothesized that absolute poverty tends to decline at first slowly and then at an accelerating rate. If

labor is paid its marginal product, the percentage rate of increase in wages of unskilled labor—which determines the rate of decline of poverty—will equal the percentage rate of increase in marginal productivity. The percentage rate of increase in marginal productivity depends in turn on changes in techniques and in quantities of associated factors of production that directly raise marginal productivity and on changes in the proportions of labor and other factors of production used in a given activity. If more unskilled labor is used with the other factors of production, its wage will go down, and if less is used its wage will go up. For the economy as a whole, any increase in unskilled labor in one activity must be accompanied by decreases elsewhere, assuming that there is a given amount of unskilled labor in the economy at any one time. Moreover, because real incomes for unskilled labor tend to be equalized among activities, the percentage rate of change in wages of unskilled labor will tend to be the same in different activities.

Combination of these conditions gives insights on how wages of unskilled labor in an economy change over time. The percentage change in the wage, $\dot{w}$, will equal the increase in marginal productivity $\dot{A}$ (attributable to improvements in techniques and changes in amounts of associated factors) plus the product of the elasticity of demand for labor in an activity, β, and the percentage change in employment of labor in that activity, $\dot{n}$. The economy is then divided into urban and nonurban activities, and the conditions for the percentage change in wage are $\dot{w} = \dot{A}_u + \beta_u \dot{n}_u$ and $\dot{w} = \dot{A}_a + \beta_a \dot{n}_a$. The first condition is multiplied by the proportion of the labor which is urban, n_u, and the second condition by the proportion which is nonurban, n_a, and the results are added. Since the total of the proportions is unity ($n_u + n_a = 1$), the left-hand side of each equation, after adding, equals $\dot{w}$. On the right-hand side, the elasticities of demand for labor, β_u and β_a, will be equal if the factor shares of the labor are the same in urban and nonurban activities, if their elasticities of substitution with other factors are the same, and if the price elasticities of demand for the products produced are the same. These assumptions are not likely to be fulfilled exactly, but they serve as a useful benchmark approximation. The assumptions will in fact be fulfilled if the product prices are given, as may happen if the economy is importantly involved in international trade and if the aggregate production functions for the two types of activities are similar as regards shares and elasticities of substitution—which again may not be unreasonable. If the elasticities are equal, one part of the right-hand side will consist of two terms which combine to form the expression $n_u \dot{n}_u + n_a \dot{n}_a$. But since $\dot{n} = dn/n$ and $dn_1 + dn_2 = 0$, in view of the given amount of the labor in the economy, $n_u \dot{n}_u + n_a \dot{n}_a$ equals zero and the expression

disappears. All that is left on the right-hand side is the weighted sum of the increases in marginal productivity that would occur in the absence of changes in amounts of labor in the different activities. The percentage change in wage is

$$(4\text{-}1) \qquad \dot{w} = n_u \dot{A}_u + n_a \dot{A}_a.$$

Refinements to allow for differences in elasticities of demand for labor could be introduced, but the terms shown would probably still dominate the results. The result shown generalizes to many activities and can be written $\dot{w} = \Sigma n_i \dot{A}_i$.

The percentage change in the marginal product of labor in agriculture, $\dot{A}_a$, may be small, either because progress in agriculture is slow or because a low-income or subsistence sector makes up a large part of agriculture and is not progressing. At the early stages of development, when most labor is in agriculture, n_a will be close to one and n_u will be close to zero. Even though there is rapid progress in urban productivity, and hence a large value of $\dot{A}_u$, the percentage rise in the wage will be small because of the low weight of $\dot{A}_u$. As urbanization proceeds during development, the fall in n_a and the rise in n_u will increase the weight of $\dot{A}_u$ and decrease the weight of $\dot{A}_a$. As a result, the rise in the wage rate will accelerate until, if the extreme of complete urbanization occurred, the rise in the wage would be fully equal to the upward shift in urban productivity.

Besides the rise in the return to unskilled labor, owing to improved techniques and the use of associated factors of production, two additional reasons for a decline in absolute poverty may be noted. First, people acquire skills and move out of the unskilled labor category. They do this because economic development, by raising the marginal productivity of skilled labor, increases the returns to education and other human capital investments by which people transform themselves from unskilled to skilled laborers. All persons who do so lift themselves from poverty categories to a higher-income status.

The second reason is closely related to the first. After some people have transformed themselves into skilled laborers, there are fewer unskilled laborers in the economy. Unskilled labor becomes scarcer than if the same proportion of the population had remained unskilled. In view of the elasticity of demand for labor, the greater scarcity of unskilled labor will lead to a rise in its wage.

Urbanization as a Source of Growth: An Illusion

The fact that per capita money incomes are generally higher in urban than in rural areas leads to an increase in measured per capita income under urbanization. In

growth source analysis, which relates changes in na-tional output to changes in inputs, the rural-urban shift shows up as a contribution to growth because differ-ences in factor earnings are assumed to reflect real differences. The result is sometimes taken as an indica-tion that urbanization is itself a source of growth, and thus that urbanization leads to additional development beyond that attributable to the demand and supply shifts that have been analyzed.

That the rural-urban shift is a source of growth may be partly or even wholly an illusion. A prerequisite of meaningful interpretation of the rural-urban shift is that education and other sources of differences in the average quality of labor in rural and urban areas be adequately measured. Two concomitant changes in economic de-velopment are, ordinarily, a shift of less-educated labor from rural to urban areas and an increase in education levels. If the increase in education levels is greater in urban than in rural employment, the shift from rural to urban employment will appear to give rise to growth unless the differences in education and other reasons for skill differences are carefully controlled. What is really a contribution to growth by increases in human capital may appear as a contribution by the rural-urban shift.

Even after differences in quality of inputs have been taken into account, measured contributions of the rural-urban shift may remain which are illusory. The fact that nonurban areas have more home production and other production which does not pass through the market-place and therefore cannot be measured is often pointed out.

A perhaps more important consideration is that the prices of local goods tend to be higher in urban than in nonurban areas. A prime example is residential rent, but the phenomenon also pervades services. Higher money wages that make labor as well off in urban as in rural areas—as will tend to be brought about by migration—will have a multiplier effect and will raise prices of local goods produced by the labor. The higher money wages have a feedback effect and raise money wages still more. Thus, in a market economy with labor mobility, the prices of local goods can be expected to vary, and differ-ences in money wages will appear even if real wages are equated in different locations. Differences in prices be-tween areas may be wrongly interpreted as causing changes in total output when there is a rural to urban population shift.

Whether real effects of the rural-urban shift remain after measurement is carried out correctly is moot. It is sometimes said that a differential in real earnings be-tween rural and urban areas is required to induce the shifts in resources that occur during growth. Expecta-tional considerations should be recognized, however. A differential may be required only if the shifts are not foreseen. If the shifts are foreseen—and the fact that the rural-urban shifts are large and of long duration during development gives reason to believe that they are—people can still make migration choices that equate real earnings. A differential might be required to compensate for the costs of moving, but empirically these costs appear to be small. A factor that could actually make urban wages lower in real terms than earnings in rural areas is the prospect of greater opportunity for increases in earnings in cities. If there is little opportunity for career advancement in rural areas, lifetime rural income profiles will be flatter. The steeper profiles in urban areas could lead migrants from rural areas to accept lower entry-level earnings in urban areas.

A final possibility, and perhaps the only actual source of real effects of a rural-urban shift, is the delayed de-cline of subsistence-type agriculture. As emphasized earlier, whole subsectors may become outmoded, in which case there is likely to be a generational process of replacement of human capital. Assuming that this large-scale change was not foreseen, older persons may be found in low-income career paths in subsistence-type agriculture even though their earnings are lower than earnings available elsewhere. Their offspring who re-place them in the labor force then migrate to cities and find more productive employment. This intergenera-tional shift from low-productivity to high-productivity employment leads to an increase in total national in-come. It should be noted that this is not an independent source of growth but rather a manifestation of a lag in adjustment; it reflects a failure to respond promptly to past sources of growth.

Has Overurbanization Occurred?

So far in this chapter no reasons have been found to question market performance in response to the demand and supply changes that determine urbanization. This section considers possible benefits and costs of the pro-cess of urbanization that people's responses do not take into account.

Environment

As noted above, differences in living conditions be-tween rural and urban areas cause money income differ-ences because of people's responses to these living con-ditions. The question now is whether people, in their migration responses that influence the degree of urban-ization, impose costs or gains on other people which those making the decisions have little or no incentive to consider. Two of the many costs which are not fully internalized are pollution and congestion.

Commuting costs, pollution, and congestion are among the many reasons that money wages are generally observed to be higher in urban than in nonurban areas. New entrants to a city pay the full costs of longer commuting distances in the form either of long travel to work or of high rents if they do not commute from the agricultural edges of the city. New entrants likewise bear higher costs of pollution and congestion than in rural areas, but only the average of such costs—not the extra costs that they impose on all others in the city because their presence leads to additional air pollution and slowing of traffic.

A way of approaching the question of market performance is to ask how much higher money wages must be in a larger city than in a smaller one to compensate for the increased costs in the former. As a simplified numerical example, take a developing-country city with a population of 600,000 and a work force of 150,000 persons employed full-time at an average annual wage of $2,000.

Consider first the commuting distance effect. When a laborer migrates from farm to city, the edges of the city are moved farther out into the agricultural periphery. Anyone living in from the periphery must pay higher rents because of the commuting advantage of interior residential sites over peripheral locations. As the city grows money wages tend to increase in relation to money wages in nongrowing places. Suppose that in such a city each worker commuting from the periphery of the city spent, on average, 15 minutes longer a trip traveling to work than he would from peripheries of typical smaller cities. If the extra time is multiplied by 250 work days a year and two trips a day, and each hour is valued at $0.60, the added cost of travel time in this city is $75 a year. To compensate, the wage in the large city would have to be high enough to yield at least this much additional income.

Next, consider the pollution effect. Particularly in the drier summertime, dust and other suspended particulates are an important source of pollution. Much of this is natural, but human activity also causes much of it, and smaller cities are therefore generally less dusty than larger ones. A typical level of suspended particulates would be 150 micrograms a cubic meter of air, whereas in smaller cities and towns a representative level would be perhaps 100 micrograms. Studies of the relation between property values and air quality indicate that individuals are willing to, and in fact do, pay higher rents to escape unfavorable air quality. There have been numerous more direct estimates of the physical damage caused by pollution. A rough estimate of the annual cost of pollution is $1 per microgram for a family of four persons, a figure that reflects damage to health, shortened life spans, deterioration of durable property, and

impairment of aesthetic values. This damage value would imply that the cost of pollution to a family in the larger city is $50 more (50 micrograms difference multiplied by $1) than the cost to a counterpart family in a smaller city with better air. If air pollution were the only unusually adverse characteristic of larger cities, and if the population were stable, the wage in the city of 600,000 would have to be $50 more than in smaller places.

If the pollution effect is added to the distance effect, the estimated amount by which wages in the larger city must exceed wages in smaller places to compensate for the disamenities of the larger city is $125. This is a conservative estimate because many of the numerical estimates used in the derivation are conservative and because a number of disamenities (such as other pollutants and noise, visual blight connected with slums, and crime) have been neglected. If these figures were correct and if no other factors influenced decisions to migrate, people would disperse themselves among cities in such a way that the wage rate was, say, $2,000 in smaller cities and $2,125 in the larger city to equalize the advantages of living in one or the other city.

The factors considered so far represent average costs per worker or per family, which are higher in larger than in smaller cities. The question becomes whether the addition of a worker in a city adds the same amount to costs, or a greater amount.

One reason the costs may be greater is that travel for everyone else in the city may be slowed by the extra congestion. Suppose that in a city of 600,000 half of the labor force, or 75,000 workers, commutes to the central employment location. Each travels an average of 5 miles per trip over congested streets at a speed of 20 miles per hour (the speed would be 35 miles per hour if there were no traffic congestion). If half the new entrants commute to the center of the city (the same assumption as for the existing labor force), the average new entrant into the city's labor force makes five rush-hour trips per week over this highway system. The effect on traffic speeds and on commuting times for the other motorists can be estimated through the use of the engineering relationship $v = \bar{v} - kq$, where v is the speed under congested conditions, $\bar{v}$ is the speed under noncongested conditions (35 miles per hour), q is the traffic volume (75,000 cars per hour), and k is a constant (here, $2 \cdot 10^{-4}$), giving v equal to 20 miles per hour (the existing speed noted above). Now recalculate v with $q = 75,001$ cars per hour to estimate a new speed. Subtract the old from the new speed to obtain the change in speed. Divide this change into the 5-mile trip on congested streets and multiply by the value of time, $0.60 per hour, to obtain the added cost that an extra trip imposes on a trip taken by someone already in the

city, $1.50 \cdot 10^{-6}$. Multiply this result by 0.5, since the commuting assumption implies that only one-half trip per new worker is added to the congested streets. Then multiply by the 75,000 affected workers and by the 500 work trips they each make per year (250 work days times two trips per day—home to work and work to home) to arrive at a yearly figure of $28 extra commuting cost that a new worker imposes on the population already in the city.

The impact of the additional family on damages from air pollution can be computed with the use of data already cited. An inference from observations on air quality and metropolitan population size is that the average particulate level is likely to rise by $4 \cdot 10^{-5}$ micrograms a cubic meter as a result of adding one family. If the damage value of $1 a microgram is applied and it is assumed that there are 150,000 families in the urban area, the overall damage value is 0.00004 times $1 times 150,000, or $6.

Taking account of both congestion and pollution, the additional costs imposed on all other families when a family moves into an urban area of 600,000 population amount to $28 plus $6, or $34 a year.

These results should be considered in relation to the previous conclusion, that wages in this city would be about $125 higher than in other places of smaller size. When a family moves into the area, if there is a free market equilibrium of city sizes generally, the personal gain is the $2,125 wage, for which $125 urban costs are incurred in this city and a $2,000 wage elsewhere is forgone. The other families in the large city suffer a combined loss of about $34. Thus there is a net loss of $34.

Although the example is only suggestive, it gives a perspective on the magnitude of the effects connected with the physical environment. The $34 annual cost neglected in migration decisions is 1.6 percent of the wage rate. The demand for labor in any one city may be quite elastic owing to possibilities for undertaking production elsewhere. For example, if the elasticity of demand for labor is unity, the number of people in the city is made 1.6 percent larger by the neglected costs. If the elasticity is as high as 3, the city is made about 5 percent larger.

Protected Employment

A phenomenon that has received much attention is the attraction to cities of persons who are willing to accept unemployment in return for being in a position to obtain protected employment should the opportunity arise. Examples of protected employment include employment effectively covered by legally decreed minimum wages, government employment in which wages

of employees are set through a political or negotiating process, and situations in which foreign corporations may decide for a number of reasons to pay more than the market-clearing wage. The wage in protected employment is higher than elsewhere in the economy.

A simple, well-known condition that has been used to analyze protected employment is based on the hypothesis that workers will distribute themselves so that the wage received in unprotected employment ends up being just as great as expected or prospective earnings in protected employment, taking account of the greater probability of unemployment that workers face if protected employment is sought. People are willing to face the unpleasant prospect of greater unemployment, which requires drawing down savings or asking kin for largesse, in return for the hope of higher earnings if they eventually succeed in finding a protected job. A queue of people seeking protected jobs forms, and it grows until expected well-being in the protected employment sector is driven down to that in unprotected employment.

In urban protected employment, which is not all urban employment and may be only a small part of it, the probability of employment is sometimes approximated as the ratio of total protected employment, n_{ue}, to the total number of people seeking protected employment, n_u. The rationale is that the total supply provides a pool from which those seeking employment have a chance of being drawn. The wage condition is that the wage in unprotected employment equals the expected wage in protected employment, which in turn is the probability of employment times the wage received if employed in protected employment, giving $w_a = (n_{ue} / n_u)w_u$. In terms of percentage changes the wage condition is $\dot{w}_a = \dot{n}_{ue} - \dot{n}_u + \dot{w}_u$, which is the first of five conditions for this model.

Meanwhile, the change in total employment in the economy as a whole as a consequence of a rise in the protected employment wage is the weighted sum of percentage changes in protected and unprotected employment. This implies as a second condition $n_e\dot{n}_e = n_a\dot{n}_a + n_{ue}\dot{n}_{ue}$. As in the earlier analysis, the weighted sum of percentage changes in people seeking employment is zero, given the total population in the country. This implies as a third condition $n_a\dot{n}_a + n_u\dot{n}_u = 0$. Other conditions from the earlier analysis are that the percentage changes in employment equal the percentage change in wage multiplied by the elasticity of demand for labor. The fourth and fifth conditions are thus $\dot{n}_a = \beta_a\dot{w}_a$ and $\dot{n}_{ue} = \beta_u\dot{w}_u$.

The five conditions determine the percentage effects of $\dot{w}_u$ (the percentage excess of the protected employment wage over the unprotected wage) on the unprotected wage, total employment, unprotected employment, protected employment, and numbers of people

seeking protected employment ($\dot{w}_a$, $\dot{n}_e$, $\dot{n}_a$, $\dot{n}_{ue}$, and $\dot{n}_u$). The effect on the total number of people in urban areas is given by the number of persons seeking protected employment. If that number increases, migration from rural to urban areas may be induced as people come to cities in hopes of finding a job at higher real wages. Solution of the five conditions for the percentage change in numbers of people seeking protected employment gives

$$(4\text{-}2) \qquad \dot{n}_u = \frac{1 + \beta_u}{1 - n_u/n_a\beta_a}\dot{w}_u.$$

A conclusion apparent from this result is that the existence of the higher wage may either raise or lower the total number of people who offer themselves for protected employment. The elasticity of demand for labor in protected employment, β_u, is negative, and if it is greater than one in absolute value, the sign of the effect of the rise in protected wage as given by the numerator of equation 4-2 is negative. With constant factor shares and with other product and factor prices unchanged, the absolute value of the elasticity of demand for labor will be the reciprocal of labor's share and will thus be greater than one, which ensures a negative effect. If product prices are raised, the absolute value of the elasticity will be even greater. With an absolute value of elasticity in the vicinity of one, which is not too unlikely, the effect on the number of people seeking protected employment is about zero. The maximum increase in the number of people seeking protected employment is obtained in the unlikely event that the elasticity of demand for protected labor is zero and the elasticity of demand for unprotected labor is infinite. In that case the coefficient in equation 4-2 reduces to one, an indication that the increase in the number of people seeking protected employment is proportional to the excess of the protected wage over the unprotected wage. Wages are likely to be raised substantially only in cases in which the amount of protected employment is rather small. Only in very unusual cases would there be a large effect on wages in protected employment if the protected employment were a sizable fraction of urban employment. Thus, even in the extreme case of zero elasticity of demand for labor, the effect of wage protection on the number of people in urban areas would be limited. It has been shown that because of labor turnover, the elasticity required to produce an increase in people seeking protected employment is even smaller than unity. The effect on the number of people in urban areas is further limited by the fact that in the usual case, in which the protected wage applies to only a fraction of the workers in urban areas, the people seeking protected employment come not only from rural areas but also from other employment in urban areas, and to the extent that they come from other urban employment, there is no net effect on urbanization.

The unemployment induced by wage protection, as a percentage of total protected employment, is derived by subtracting the percentage change in protected employment from the percentage change in the number of people seeking protected employment. It is n_u/n_{ue} times the equation 4-2 coefficient plus the absolute value of the elasticity of demand for labor. Whether the unemployment is substantial depends on the magnitude of the wage increase and on the amount of protected employment. The unemployment, again, may be limited. To the extent that workers rotate or are in the same families, the unemployment does not affect the distribution of income.

Still more refined models might allow for workers to be doing something else while they are looking for jobs. It seems possible that the refinements would lead to even smaller estimated effects of wage protection on urbanization and unemployment. Reliable empirical evidence is difficult to obtain because of the difficulty of obtaining and interpreting unemployment statistics.

Benefits from Government Expenditure and Tax Policies

The benefits of such urban services as water supply, publicly provided health facilities, unemployment compensation, social security, and maintenance of low food prices may make cities attractive if, on net, the charges and taxes resulting from residence in urban areas are less than the benefits. The net benefits of urban residence that result from government actions can be expected to be reflected in a compensating wage differential, along with the other differences between urban and rural areas that go into determining urban-rural wage differentials.

The role of the tax structure in determining net benefits of urban residence deserves special analysis. Taxes imposed solely on the site value of land can be expected to be neutral with respect to location decisions, but taxes on land improvements or on capital structures, such as property taxes, may affect incentives to add to capital. If agricultural and urban property are taxed differently, the location of capital may be affected. Since the owner of nonhuman capital does not have to reside where the capital is employed to collect the income, a condition for location neutrality is that the marginal tax rate on nonhuman capital be the same in different locations in money terms. The owner of human capital (the person himself) must reside where the services are rendered. Thus, a condition for location neutrality for human capital is that marginal tax rates on human capital in different places be the same in real terms.

The urban wage tends to be lower than it would be without net government urban benefits by the percentage at which the net benefits are valued. This gives a first condition that may be expressed in percentage terms as $\dot{w}_a = \dot{w}_u + \dot{g}$, where $\dot{g}$ is the percentage of the wage at which the government benefits are valued. That condition may be combined with three earlier conditions—that the total population is distributed either to rural or to urban areas ($n_a\dot{n}_a + n_u\dot{n}_u = 0$) and that the changes in employment are governed by the elasticities of demand for nonurban labor ($\dot{n}_a = \beta_a\dot{w}_a$) and for urban labor ($\dot{n}_u = \beta_u\dot{w}_u$). Four conditions thus determine percentage changes in urban and nonurban wages and employment ($\dot{w}_u$, $\dot{w}_a$, $\dot{n}_u$, and $\dot{n}_a$) as a function of the percentage of the wage at which government benefits are valued, $\dot{g}$.

Solution for the percentage increase in urban employment yields

$$(4\text{-}3) \qquad \dot{n}_u = \frac{-\beta_u}{1 + n_u\beta_u/n_a\beta_a}\dot{g}.$$

Since the elasticity of demand for labor, β_u, is negative, the effect of government-supplied urban benefits is unambiguously positive. Furthermore, in contrast to the effects on urbanization (considered above) of environmental externalities and protected wages, it appears that government benefits could substantially affect the number of people in urban areas, at least on a one-time basis. For example, if the elasticity of demand in absolute terms is about 1 and if the proportions of people in rural and in urban areas are about the same (that is, the denominator is 2), the value of the coefficient in equation 4-3 will be about one-half. One reason the effect may be substantial is that it applies to all residents of urban areas, on the assumption that none of them can be effectively excluded from government-supplied urban benefits. If these benefits are valued at 10 percent of the wage, the number of people in urban areas will increase by 5 percent.

Primal Cities

In the developed countries there are signs of decentralization not only to suburbs but to smaller towns and even rural areas. But developing countries are moving toward even greater urban concentrations, evidence not only of rapid overall urbanization but also of more pronounced tendencies than in developed countries to have a preponderance of population in one or a few cities.

The tendency to concentrate population in one or a few cities is as much a policy concern as urbanization as

such, and perhaps it is a greater concern. The reasons for tendencies toward concentration are in part the same as those for urbanization in general; natural population growth is a prime reason. Still, the question remains why population should be so heavily concentrated in a few places. Perhaps the single most important reason for tendencies toward large cities is that it is economical to centralize production to avoid the costs of transporting goods back and forth. Only when these advantages are overcome by rising costs of various kinds within a city does it become economical to decentralize. This is a pervasive factor the world over, in developed and in developing countries.

A possible reason for the greater concentration in developing than in developed countries is that the effect of the urbanization externalities considered above may be greater. In particular, net benefits from government actions could act to promote concentration in the largest cities. It would appear that the question of urban bias (if it exists) should be rephrased to deal not with exploitation of rural by urban people but rather with the effects of urban bias on city size. The existence of net benefits in urban areas encourages people to move toward those benefits. That movement dissipates temporarily induced rural-urban differences in well-being but may have significant effects on city size. But although uninternalized externalities may contribute somewhat to larger city size, it does not appear possible to make a really strong case that they are the prime cause of pronounced large urban concentrations in developing countries. It seems necessary to look elsewhere for a full explanation.

Another hypothesis, which appears more likely, is that intercity transport, which would permit decentralization to smaller cities, is a relatively expensive commodity in developing countries. Provision of basic road facilities is highly capital-intensive, as is the operation of trucks and other transport equipment. If this hypothesis is valid, a gradual tendency toward less centralization may be observed as development becomes more complete and the price of human inputs rises in relation to the price of physical capital. This tendency toward a lowering of the relative price of intercity transport, however, may set in only at a late stage of development.

A final, related, possibility is put forth more tentatively. Infrastructure decisions may not provide as adequately as would be beneficial for development outside large urban centers. Development of more adequate evaluation tools that would allow more accurate assessments of infrastructure investments might lead to more investment outside large cities.

These hypotheses need to be tested as part of an investigation of the fundamental determinants of city

size, as represented by the study by Henderson (chapter 7).

Economic Policy and Urbanization

Most of the reasons for rapid urbanization in developing countries are closely related to the economic development process itself. Wholesale reversal of urbanization is surely unthinkable and would amount effectively to a great interference with development, if not its virtual denial. The more relevant question becomes whether something marginal should be done to try to affect urbanization and in particular the concentration of urbanization in a few very large places.

Externalities

Three types of policies toward urbanization may be identified. The first is the internalization of externalities. If pollution, congestion, and related environmental externalities increase with city size, as suggested earlier, their internalization will lead to consideration of environmental costs in private decisions and will encourage decentralization. To affect decentralization, the policies must distinguish between the severity of environmental costs in different places. For example, if pollution abatement measures are tailored to local circumstances, greater controls will be imposed where control costs are smaller and damages are greater. In environmental policy as now practiced, however, uniform environmental standards are often imposed across all or a large part of a nation, so that control requirements are the same irrespective of city size, and thus incentives to decentralize are thwarted. If environmental controls are to play a role in decentralization policy, a less procrustean approach (which would also contribute to the efficiency of environmental policy) is needed.

The earlier analysis suggests that the externalities connected with high wages in protected employment may either increase or decrease urbanization. The most important consequences may be unemployment and overinvestment in search costs. It is not clear that wage policy and urbanization policy as such are as closely connected as might appear at first sight. At any rate, public choice considerations make it difficult to eliminate wage protection once it is instituted. A promising approach may be to reduce the effects of protection by keeping annual nominal rises in protected wages low in relation to the inflation rate.

The analysis suggests that the largest opportunities for internalizing externalities may be in the provision of public services, the tax structure, and the operation of government transfer and price programs. With regard to public services, there would be no externalities if recipients of the services were charged marginal costs and if strong attempts were made to carry provision of services to the point at which marginal cost equaled marginal benefits. Quite possibly, neither of these two conditions can be fulfilled. In countries where public services are markedly more available in large urban areas than elsewhere and where the services are provided free or below cost, raising the charges or extending the services more widely throughout the country are possible actions. Extension of services may be the more realistic policy direction and may have distributional benefits.

Infrastructure Decisions

One method of providing infrastructure is to concentrate facilities at places where demand for infrastructure services is growing, by mechanically extrapolating recently observed trends in demand. The method is safe because it has a self-fulfilling element: provision of infrastructure services in an area encourages people to reside in that area. Demand then increases, and more infrastructure is built. Provision of infrastructure thus feeds on itself. For example, road networks that make the larger cities ever more powerful hubs of transport may be progressively reinforced in response to the signal of growing traffic brought on by previous investments and further contribute to the large size of the cities.

At the opposite extreme, and less frequently, infrastructure may be provided to stimulate areas of lagging development or low population. The degree of success may be disappointing because natural conditions and other locational characteristics are not overcome by the new infrastructure, with the result that the infrastructure is not used fully and the areas do not grow as much as was hoped.

Neither extreme appears optimal. Development of improved methods of evaluation could have important consequences for the spatial distribution of population. More attention could be given to the effects of infrastructure on the location of activity and to the quantitative estimation of benefits and costs that result from alternative patterns of development. In particular, evaluation techniques for transport facilities are sophisticated in some ways but need further development. For example, the travel savings method, which is widely used to evaluate the benefits of road investments, is to be recommended if the origin and destination of trips is not affected. More reliable methods, however, are needed to estimate benefits if changes in trips are brought about when road investments induce changes in the location of activities. Development of improved methods could

lead to identification of alternatives that are preferable to expansion in the largest cities; investments might be made, for example, in favorably situated medium-size cities.

Direct Interventions

Nearly every country has policies that try to alter incentives to choose particular locations. Often such policies are explicitly concerned with preventing the concentration of the population in large cities. The policies take many forms, including land use controls and subsidies. With notable exceptions, such as the studies by Lee (chapter 8) and Reif (chapter 9), few serious attempts have been made to quantify the effects of the policies. Many policies appear to be largely nominal, without any real teeth, but in a few cases they may be overeffective.

Direct interventions may be decided on either because the best economic policies, connected with externalities or infrastructure, are not politically feasible or because deliberate actions to influence location, going beyond externalities and infrastructure, are mandated—desirably or undesirably—by political considerations. In either case, there are efficient ways to design the interventions. A Pigovian tax subsidy approach may be mentioned as a way to achieve efficiency in interventions. It should be noted, however, that Pigovian approaches—despite a century of advocacy—have seldom, if ever, been found to be amenable to practical application. Under a tax subsidy approach that is designed to encourage a coherent amount of decentralization, a shadow price for additions to population in different places, such as in cities of different sizes, would be established. Firms would face subsidies or taxes according to the amount of their employment in different places, and the tax or subsidy would be proportional to the difference in shadow price of having people in different places. Alternative forms of inducements (such as interest-rate and energy subsidies) and quantitative controls (such as land use restrictions) would be judged according to how well they approximated the outcome that would be attained under the shadow price approach. The same standard could be used to design and evaluate more general overall strategies for spatial development.

Deliberate spatial interventions as described above are controversial. Realistically, in the absence of badly needed objective quantifications of the effects of spatial intervention policies, it is difficult to be optimistic that effective or desirable interventions will occur soon. Efforts to design interventions rationally and even bring in cultural diversity and national defense as spatial considerations, along with more usual externalities, are to be commended, but this achievement so far remains elusive.

Another kind of intervention is inevitable: the public management decisions required to provide urban services, regulate and facilitate land use, and provide for financing. Various facets of these problems are dealt with in the studies by Bahl and Linn (chapter 10), Ingram (chapter 11), Pachón (chapter 12), Thomas (chapter 13), and Keare (chapter 14).

Because public management decisions for cities go far beyond spatial considerations, they are at once easier and harder to deal with. A large body of ready-made knowledge and tools can be brought to bear on these more traditional subjects, but there is a worldwide lag in coordinating expertise and combining economics with engineering, planning, law, and other disciplines required to make public management decisions. Multiple goals also make public management decisions difficult. Income distribution considerations are difficult to deal with at best and may run counter to spatial objectives, to say nothing of more general economic development objectives. One of the hardest questions is how much effort to put into city management decisions. The mainsprings of development are not found in many of these decisions, but failure to deal with them adequately may put drags on development. Which ones will take care of themselves somehow and which ones call for special efforts are concerns for economic development.

Yet there is a way to improve urban management at present from a spatial point of view. It is quite practical to introduce quantification of the value of externality effects into the evaluation of projects. Orders of magnitude of the value of externalities at the margin can be developed along the lines illustrated in this chapter. These can be used as add-ons in social benefit-cost decisions concerning roads, water projects, and other expenditures that affect the location of activity. In this way, spatial externalities could be taken into account in the workings of government, without adopting explicit spatial policies. The required quantification would probably reveal the externalities to be small in most cases. Quantification would help keep externalities from being blown out of proportion by the emotional beliefs and political distribution motives that too often figure in discussions of externalities.

Bibliography

Brueckner, Jan K. 1981. "Labor Mobility and the Incidence of the Residential Property Tax." *Journal of Urban Economics*, vol. 10, no. 2 (September), pp. 173–82.

Chappie, Mike, and Lester Lave. 1982. "The Health Effects of Air Pollution: A Reanalysis." *Journal of Urban Economics*, vol. 12, no. 3 (November), pp. 346–76.

Cooley, Thomas F., and C. J. La Civita. 1982. "A Theory of Growth Controls." *Journal of Urban Economics*, vol. 12, no. 2 (September), pp. 129–45.

Cropper, M. L. 1981. "The Value of Urban Amenities." *Journal of Regional Science*, vol. 21, pp. 359–74.

Cropper, M. L., and A. S. Arriaga-Salinas. 1980. "Intercity Wage Differentials and the Value of Air Quality." *Journal of Urban Economics*, vol. 8, no. 2 (September), pp. 236–54.

Diamond, Douglas B., Jr., and George S. Tolley, eds. 1982. *The Economics of Urban Amenities*. New York: Academic Press.

Fare, Rolf, Shawna Groskopf, and Bong Joon Yoon. 1982. "A Theoretical and Empirical Analysis of the Highway Speed-Volume Relationship." *Journal of Urban Economics*, vol. 12, no. 1 (July), pp. 115–21.

Fogarty, Michael S., and Gaspar Garofalo. 1980. "Urban Size and the Amenity Structure of Cities." *Journal of Urban Economics*, vol. 8, pp. 350–61.

Graves, Philip E. 1980. "Migration and Climate." *Journal of Regional Science*, vol. 20, no. 3 (November), pp. 227–37.

————. 1983. "Migration with a Composite Amenity: The Role of Rents." *Journal of Regional Science*, vol. 23, pp. 541–46.

Grieson, Ronald E., and Richard Arnott. 1981. "Optimal Fiscal Policy for a State or Local Government." *Journal of Urban Economics*, vol. 9, no. 1 (January), pp. 23–48.

Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment, and Development: A Two-Sector Analysis." *American Economic Review*, vol. 60 (March), pp. 126–42.

Haurin, Donald R. 1981. "Local Income Taxation in an Urban Area." *Journal of Urban Economics*, vol. 10, no. 3 (November), pp. 323–37.

————. 1983. "The Effect of Property Taxes on Urban Areas." *Journal of Urban Economics*, vol. 7, no. 3. (May), pp. 384–96.

————. 1984. "The Effects of Output and Factor Subsidies or Taxes on an Urban Area." *Regional Science and Urban Economics*, vol. 14, pp. 533–46.

Henderson, J. Vernon. 1977. *Economic Theory and the Cities*. New York: Academic Press.

————. 1982. "Evaluating Consumer Amenities and Interregional Welfare Differences." *Journal of Urban Economics*, vol. 11, no. 1 (January), pp. 32–59.

————. 1982. "The Impact of Government Policies on Urban Concentration." *Journal of Urban Economics*, vol. 12, no. 3 (November), pp. 280–303.

————. 1985. "Property Tax Incidence with a Public Sector." *Journal of Political Economy*, vol. 93 (August), pp. 648–65.

————. 1986. "Efficiency of Resource Usage and City Size." *Journal of Urban Economics*, vol. 19, no. 1 (January), pp. 47–70.

Izraeli, Oded. 1977. "Differentials in Nominal Wages and Prices between Cities." *Urban Studies*, vol. 14 (October), pp. 275–90.

————. 1983. "Federal Individual Income Tax and Allocation of Labor between SMSAS." *Journal of Regional Science*, vol. 23, pp. 105–13.

McConnell, Virginia D., and Mahlon Straszheim. 1982. "Auto Pollution and Congestion in an Urban Model: An Analysis of Alternative Strategies." *Journal of Urban Economics*, vol. 11, no. 1 (January), pp. 11–31.

McDonald, John F. 1983. "An Economic Analysis of Local Inducements for Business." *Journal of Urban Economics*, vol. 13, no. 3 (May), pp. 322–36.

Porell, Frank W. 1982. "Intermetropolitan Migration and Quality of Life." *Journal of Regional Science*, vol. 22, pp. 137–58.

Power, Thomas M. 1981. "Urban Size (Dis)amenities Revisited." *Journal of Urban Economics*, vol. 9, no. 1 (January), pp. 85–89.

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In P. Mieszkowski and M. Straszheim, eds., *Current Issues in Urban Economics*. Baltimore, Md.: Johns Hopkins University Press.

Tolley, George S. 1974. "The Welfare Economics of City Bigness." *Journal of Urban Economics*, vol. 1, no. 3 (May), pp. 324–45.

Tolley, George S., Philip E. Graves, and John L. Gardner, eds. 1979. *Urban Growth Policy in a Market Economy*. New York: Academic Press.