Urban Concentration: The Role of Increasing Returns and Transport Costs

Paul Krugman

Very large urban centers are a conspicuous feature of many developing economies, yet the subject of the size distribution of cities (as opposed to such issues as rural-urban migration) has been neglected by development economists. This article argues that some important insights into urban concentration, especially the tendency of some developing countries to have very large primate cities, can be derived from recent approaches to economic geography. Three approaches are compared: the well-established neoclassical urban systems theory, which emphasizes the tradeoff between agglomeration economies and diseconomies of city size; the new economic geography, which attempts to derive agglomeration effects from the interactions among market size, transportation costs, and increasing returns at the firm level; and a nihilistic view that cities emerge out of a random process in which there are roughly constant returns to city size. The article suggests that Washington consensus policies of reduced government intervention and trade opening may tend to reduce the size of primate cities or at least slow their relative growth.

Over the past several years there has been a broad revival of interest in issues of regional and urban development. This revival has taken two main directions. The first has focused on theoretical models of urbanization and uneven regional growth, many of them grounded in the approaches to imperfect competition and increasing returns originally developed in the "new trade" and "new growth" theories. The second, a new wave of empirical work, explores urban and regional growth patterns for clues to the nature of external economies, macroeconomic adjustment, and other aspects of the aggregate economy.

Most of this work has focused either on generic issues or on issues raised by the experience of advanced market economies like the United States. Yet arguably the issues raised by the recent work are most salient for smaller, less-wealthy countries like Mexico and Brazil.

Why might the "new economic geography" be more relevant for the developing world than for industrial countries? First, the matter is an urgent one for real-world

241

policy. Urbanization in developing countries, and particularly the very large agglomerations such as Mexico City and São Paulo, is widely regarded as a problem. Rural-urban migration has, of course, been the subject of a vast literature in development economics, with many papers suggesting that its pace is excessive from a social point of view. Moreover, the sheer size of some cities that such migration now feeds reinforces these concerns. Although nobody can claim to have made a thorough welfare-economic study of the consequences of the emergence of huge cities in developing countries, many observers believe that something has gone wrong, that such giant cities are in some sense parasitic entities that drain vitality from their host economies—Bairoch (1988) has called these metropolises "Romes without empires"—that the environmental and social problems posed by cities with populations in the tens of millions are even greater in poor nations than in the West.

Associated with concern about urbanization and metropolitan growth is related concern about regional inequality. In many developing countries the regions that contain the big cities are also much richer per capita than other regions. The problem of core-periphery patterns within countries is not only economic and social but political as well: it is no accident that a separatist movement has emerged in Brazil's relatively rich south or that armed opposition to the central government surfaced in the bypassed southern regions of Mexico.

On the bright side, urbanization and unequal regional development may be analytically more tractable in developing than in industrial countries. The models developed in recent years, which stress the role of relatively measurable factors like economies of scale and transportation costs in determining urban growth, often seem to miss much of the story in advanced economies. For one thing, in huge economies like the United States or the European Union static economies of scale tend to seem relatively unimportant. For another, in advanced nations that are increasingly in the business of producing information rather than tangible goods, the nature of both the external economies that induce agglomeration and the transaction costs that make distance matter becomes more and more subtle. By contrast, developing countries have much smaller internal markets. For example, although Mexico's population is one-third that of the United States, its dollar purchasing power is about the same as that of metropolitan Los Angeles. Thus conventional scale economies remain relevant. And these countries still devote much more of their labor force and expenditure to tangible products that must be transported by road or rail.

Finally, the radical policy changes that have taken place, or may be about to take place, in some developing countries are likely to have major impacts on urban and regional development, impacts that we want to be able to predict. One need only consider the case of Mexico: the federal district in that country became dominant during a prolonged period of both import-substituting development strategy and extensive government involvement in the economy. As the country has shifted to an export-oriented trade policy, the manufacturing center of gravity has visibly shifted toward the country's northern states. Will the combining of that shift with privatization and deregulation undermine Mexico City's special role, or will other activities maintain its position?

For these reasons, then, it is natural to ask whether, and if so to what extent, the new tools of urban and regional analysis apply to developing countries. The literature on urban and regional issues in development is immense. This article explores a narrow, indeed largely technical issue: what can we learn from looking at urbanization and regional inequality in developing countries through the lens of the specific approach to economic geography that has emerged out of the new trade and growth theories? The article sketches out a minimalist new economic geography model designed to highlight the way a tension between forces of agglomeration and forces of dispersal determines city sizes. The implications of that tension are illustrated by examining a particular issue: how trade policy may affect the tendency of developing countries to have very large, primate cities. Two other factors also are explored that probably have even more important roles in determining urban structure: the centralization of government and the quality and form of transportation infrastructure.

Approaches to Urban Development

Urbanization—and uneven regional development, which is a closely related process—clearly involves a tension between the "centripetal" forces that tend to pull population and production into agglomerations and the "centrifugal" forces that tend to break such agglomerations up. The following tabulation lists the major types of centripetal and centrifugal forces that appear in various models of urban growth:

Centripetal forces
- Natural advantages of particular sites
    Harbors, rivers, and the like
    Central locations
- Market-size external economies
    Access to markets (backward linkages)
    Access to products (forward linkages)
    Thick labor markets
- Pure external economies
    Knowledge spillovers

Centrifugal forces
- Market-mediated forces
    Commuting costs, urban land rent
    Pull of dispersed resources, such as farmland
- Nonmarket forces
    Congestion
    Pollution

Several key distinctions among these forces are worth pointing out. Among centripetal forces there is a basic distinction between natural factors that favor a site—

such as a good harbor or a central position—and external economies that are acquired and self-reinforcing advantages of a site. Among external economies there is a further key distinction between "pure" external economies, such as spillover of knowledge between nearby firms, and market-size effects, whether in the labor market or in the linkages between upstream and downstream industries.

On the side of centrifugal forces there is a similar distinction between nonmarket diseconomies (such as congestion) and factors such as land prices that are fully mediated through the market. A narrower but sometimes important distinction appears between forces that push business out of a large city, such as urban land prices, and those that pull business away, such as the existence of a dispersed rural market.

Which forces actually explain the pattern of urbanization in developing countries? The answer is, of course, all of them. Nonetheless, to say anything useful we must always rely on simplified models. The typical analytical approach therefore takes "one from column A and one from column B" and thus gets a particular story about the tension between the agglomeration and dispersion that creates an urban system. Several such approaches have achieved wide influence.

Neoclassical Urban Systems Theory

At least within the economics profession the most influential approach to urban development is probably what we might call neoclassical urban systems theory. This approach models the centripetal forces for agglomeration as pure external economies (therefore allowing the modeler to assume perfect competition)[1] and the centrifugal forces as arising from the need to commute to a central business district within each city, a need that leads to a gradient of land rents within each city. In the simplest case the tension between these forces leads to an optimal city size, though there is no guarantee that market forces will actually produce this optimal city.

This neoclassical approach has been extensively developed by Henderson (1974, 1977, 1988) and his followers, who added two important elaborations. First, Henderson pointed out that if cities are the "wrong" size, there are potential profit opportunities for a class of "city developers"; and as an empirical matter, large forward-looking private agents who seem to try to internalize external economies do play a large role in urban development in the United States. Thus Henderson-type models adopt as a working hypothesis the assumption that competition among developers produces cities of optimal size.

Second, according to Henderson, external economies may well be industry-specific (textile plants may convey external benefits to neighboring textile plants; metalworking plants may do the same, but it is hard to see why metalworkers want textile workers nearby). On the other hand, diseconomies of commuting and land rent depend on the overall size of a city, not the size of an individual industry within that city. Thus Henderson-type models predict the emergence of specialized cities, with each city's "export" sector producing a range of industries with mutual spillovers, and with industries that do not benefit from these spillovers seeking other locations. Since cities are specialized, this approach explains the existence of an

urban system with many different types of cities; inasmuch as the optimal size of a city depends on the relative strength of external economies and city-size diseconomies, and external economies are presumably stronger in some industries than in others, cities of different types will be of different sizes. Neoclassical urban systems theory therefore offers a framework that explains the existence not only of cities but also of a system of cities of differing sizes.

While the insights gained from this approach are impressive, it has important limitations. First, the external economies that drive agglomeration are treated for the most part as a kind of black box, making it difficult to think about what might influence their strength and thus making it hard even to start to predict how policy or other changes might affect the urban system. Second, the reliance of much of this literature on the assumption of competition between city developers, while a useful clarifying device, strains credibility when applied to huge urban areas: the Irvine Corporation may arguably have played a major role in developing a particular "edge city" within metropolitan Los Angeles, but could any private agent internalize the externalities of São Paulo? Finally, neoclassical urban systems theory is entirely nonspatial: it describes the number and types of cities, but says nothing about their locations. In the past few years an alternative approach has emerged that shares much of the framework of urban systems theory but attempts to deal with these issues.

Monopolistic Competition Theory

In this new literature agglomeration economies are not assumed but are instead derived from the interaction among economies of scale at the plant level, transportation costs, and factor mobility. Economies of scale at the plant level inevitably imply imperfect competition; this imperfection is modeled using the same (unsatisfactory) monopolistic competition approach that has played such a large role in trade and growth theory over the past fifteen years. The "new economic geography" literature, begun in Krugman (1991a,b), bears considerable resemblance to the urban systems approach, but the black-box nature of external economies is gone, there is a spatial dimension, and the models no longer rely on the assumption of city developers who enforce optimal outcomes. In some respects, in fact, the new approach seems closer in spirit to the "cumulative process" description of urban and regional development associated with geographers such as Pred (1966).

The model described below is in this tradition, so it is worth noting the considerable limitations of this approach. Two points stand out. First, multiple-city systems are difficult to model using this approach. Where the urban systems approach easily tells a story of multiple cities of a number of different types, in monopolistically competitive spatial settings (see, for example, Krugman 1993b) multiple-city systems can at this point be modeled only with considerable difficulty, and initial efforts to get some kind of urban hierarchy have encountered surprisingly nasty problems (Fujita and Krugman 1993). Second, going from the black-box external economies of the urban systems model to the derived agglomeration effects of the monopolistic competition model may involve a degree of misplaced concreteness. We will have

a seemingly clear story about linkage externalities in the manufacturing sector, but it may be that, say, informational externalities in the service sector are equally important even in developing countries. Attempts to get specific, to open up the black box, always run this risk; nonetheless, it seems greater than usual in this case.

Finally, we should point out one additional risk in both the urban systems and the monopolistic competition approaches to urban modeling: we may be trying to explain too much, engaging in a kind of Rorschach test in which we are trying to find deterministic explanations of essentially random outcomes. While this notion does not exactly constitute a rival theory of urban systems, the idea that they are largely random creations requires at least some discussion.

Random Urban Systems

The general idea suggested by the tabulation above—that city sizes are determined by a tension between centripetal and centrifugal forces—seems to imply the conclusion that there will in any economy be a typical, equilibrium city size. In fact, one sees a whole range of city sizes. The urban systems theory explains that there are different types of cities, each with a characteristic size, and that the size distribution is actually a type distribution. While this argument surely has some validity, it may not be a full explanation. For one thing, urban specialization is increasingly difficult to detect in advanced countries. It is a familiar point that the mix of activities within U.S. metropolitan areas has become increasingly similar since 1950, and the influential study by Glaeser and others (1992) finds, as well, that individual industries seem to grow fastest in more diverse metropolitan areas.

Moreover, the size distribution of cities is suspiciously smooth and regular. City sizes in many countries are startlingly well described by a power law of the form

$$(1) \qquad\qquad N(S) = AS^{-\alpha}$$

where $N(S)$ is the number of cities that are the same size as or larger than S. Furthermore, the exponent α is generally quite close to 1. In fact, when equation 1 is estimated for U.S. metropolitan areas, α is almost exactly 1, and it has remained close to 1 for at least a century. International evidence is not quite so strong, perhaps because of definitions of city boundaries: Rosen and Resnick (1980) show that when data for metropolitan areas rather than cities proper are used for a number of countries, α almost always moves substantially closer to 1.

Why should this matter? Because while a relationship like equation 1 is difficult to explain with an equilibrium story about determination of city size, it is quite easy to justify with a nihilistic story of the kind analyzed by Herbert Simon (Ijiri and Simon 1977). Suppose that for all practical purposes there is no equilibrium city size—that approximately constant returns to scale appear over some wide range of sizes. And suppose that cities grow through some random process, in which the expected rate of growth is independent of city size. Then as long as the random process generates a widely dispersed distribution of city sizes, that distribution will

be well described by a power law like equation 1. (A suggestive explanation of this result is given in the appendix.)

Worse yet, such a nihilistic approach can even explain the tendency of the exponent of the power law to be close to 1. Suppose that there is some minimum viable city size, say S_0, and that the distribution of city sizes above that minimum is well described by equation 1. Then the *average* city size is

$$(2) \qquad \bar{S} = S_0 \, (\alpha/\alpha - 1).$$

In other words α close to 1 is equivalent to the statement that the average city size is large relative to the minimum. And it is easy to imagine why this might be the case. Suppose that urban population has grown substantially over a period during which, for whatever reason, few new cities have been founded. Then the existing cities must on average grow much larger than the minimum viable size, and the estimated α will be close to 1.

This nihilistic approach raises real questions about any kind of equilibrium model of an urban system; indeed, if this interpretation is correct, there may be no optimal or equilibrium city size, simply a random process that generates population clusters of many sizes. At some level this interpretation cannot be completely right: surely city size must matter. (This is the same issue that arises in studies of the size distribution of firms, which also seems to obey power laws.) Yet the data may contain less information than we think.

On the other hand, this approach suggests that estimates of relationships like equation 1, together with related measures like "primacy," may be a useful summary indicator of the structure of a country's urban system. Primacy describes the size of the largest city relative either to total population or to some other measure, such as the population of the *n* largest cities. Many have studied city size distributions: Carroll (1982) provides a survey; Rosen and Resnick (1980) is a particularly clear example; Ades and Glaeser (1993) is a recent study inspired by the new economic geography literature. This literature suggests several stylized facts that may help us to think about urbanization in developing countries.

Stylized Facts

While urban experience varies widely across nations, there seem to be four interesting empirical regularities about urban size distributions.

First, per capita income is negatively related to measures of urban concentration, whether one uses α from equation 1 or measures of primacy such as the share of the largest city in the population of the top ten. This observation confirms an impression of giant metropolitan areas in developing countries: to a large extent, of course, the developing world has big cities simply because it has so many people, but even in this light the biggest cities in these countries are disproportionately big.

Second, the concentration of urban population is closely related to the concentration of political power. Countries with federal systems, and thus geographically

diffused power, have flatter distributions of city size and, in particular, smaller biggest cities than countries that do not have federal systems. Thus Tokyo, the largest city in centralized Japan, is considerably larger than New York, the biggest city of federal America, even though the United States has twice Japan's population. Australia and Canada, though developed at about the same time, have much less urban concentration than do Argentina or Chile. Dictatorships have more concentrated urban centers than do more pluralistic systems, according to Ades and Glaeser (1993).

Third, the nature of transportation infrastructure has an important effect on urban concentration. Countries in which the capital city has a uniquely central position—something that Rosen and Resnick (1980) proxy by a measure of rail density—tend, not too surprisingly, to have more populous capitals. Obviously, this effect often works in tandem with centralization of political power.

Finally, a less dramatic but still visible relationship is apparent between trade openness and urban structure. More open economies, as measured by the share of exports in gross domestic product, tend to have smaller biggest cities. (This is an other-things-equal proposition. Countries with small populations tend to be open, and also to have a large share of their population in the biggest city—consider Singapore. But countries that are more open than you would expect given their population tend to have smaller biggest cities than you would expect given their population.)[2]

At this point, then, we have described a menu of ways (far from inclusive) to think about urban systems in developing countries and have very briefly set out some stylized facts. The next step is to sketch a particular model as a basis for trying to understand those facts.

A Model of Urban Concentration

This section presents a formal model of urban concentration; the full model is presented in the appendix. As pointed out above, numerous centrifugal and centripetal forces may affect urban concentration. All of them probably play some role in practice, yet the modeler normally chooses only a few to include in any given analysis. In my own work I have generally chosen to include only the centripetal forces that arise from the interaction among economies of scale, market size, and transportation costs, that is, backward and forward linkages. Other external economies are undoubtedly at work in real urban areas, but they are omitted in the interest of keeping the models as simple as possible and of keeping a reasonable distance between assumptions and conclusions.

For similar reasons we can handle only one centrifugal force at a time. It turns out to be useful to move back and forth between two different approaches. One, which is close in spirit to the neoclassical urban systems literature, involves commuting costs and land rent. The other involves the pull of a dispersed rural market. This second approach has already been described in a number of published articles, for example, Krugman (1991a,b, 1993b); thus the formal model described here does not include this effect.

As we will see, attempting to make sense of the stylized facts described above is easiest when keeping both approaches in mind. The role of trade openness in urban concentration is most easily understood by focusing on urban land rent, while one cannot model the effects of political centralization and infrastructure without some kind of backdrop of immobile population and purchasing power.

Imagine, then, a stylized economy consisting of three locations, 0, 1, and 2. Location 0 is the "rest of the world," while 1 and 2 are two domestic locations (say, Mexico City and Monterrey). There is only one factor of production, labor. A fixed domestic supply of labor L is mobile between locations 1 and 2, but there is no international labor mobility.

In this radically oversimplified model the issue of urban concentration reduces to just one question: how equally or unequally will the labor force be distributed between the two locations? It is, of course, a considerable stretch to relate results of this kind to the realities of multicity urban systems, but as always the hope is that despite their oversimplifications simple models yield useful insights.

To generate diseconomies of urban concentration, we assume that in each location production must take place at a single central point. Workers, however, require land to live on. To make matters simple, we make several special assumptions. First, each worker needs a fixed living space, say, one unit of land. Second, the cities are long and narrow, so that workers are effectively spread along a line. This assumption implies that the commuting distance of the last worker in any given location is simply proportional to that location's population (as opposed to depending on the square root of population, as it would in a disk-shaped city).[3]

The diseconomies arising from the need to commute will be reflected both in land rents and in commuting costs. Workers who live in the outskirts of the town will pay no land rent but will have high commuting costs. Workers who live closer to the city center will avoid these costs, but competition will ensure that they pay an offsetting land rent. The wage net of commuting costs will decline as one moves away from the city center, but land rents will always exactly offset the differential. Thus given any wage rate at the center, the wage net of both commuting and land rents will be a decreasing function of city size for all workers.

To explain agglomeration in the face of these diseconomies, we must introduce compensating advantages of concentration. These must arise from economies of scale. Unless economies of scale are purely external to firms, however, they must lead to imperfect competition. So we must introduce scale economies in a way that allows a tractable model of imperfect competition.

Not surprisingly, the easiest way to do this is with the familiar tricks of monopolistic competition modeling. We suppose a large number of symmetric potential products, not all actually produced. Each producer acts as a profit-maximizing monopolist, but free entry drives profits to zero. The result will be that a large concentration of population produces a large variety of differentiated products. (One might think that the average scale of production will also be larger. Unfortunately, in the Dixit-Stiglitz–type model used in the appendix, this plausible effect does not materialize: all scale gains appear in the form of variety rather than production).

Will this advantage make such a location attractive despite high land rent and commuting costs? Only if there are costs of transacting between locations, so that a location with a large population is a good place to have access to products (a forward linkage) and to markets (a backward linkage). Thus we next introduce transportation costs, both between domestic regions and between these regions and the rest of the world. For technical reasons involving the way that monopolistic competition must be modeled, it turns out to be extremely convenient, if silly, to assume that transport costs are incurred in the goods shipped, an assumption sometimes referred to as the iceberg assumption: if one unit of a good shipped between regions is to arrive, $\tau > 1$ units must begin the journey. The same applies to international shipments, except that the transport costs may be different.

We may think of interregional transport costs as "natural" consequences of distance (albeit affected by investments in infrastructure). The costs of transacting with the rest of the world, however, involve not only natural costs but artificial trade barriers. Thus the level of transport costs to and from the outside world can be seen as a policy variable.

And that's it (except for the details laid out in the appendix). The interaction among economies of scale, transport costs, and labor mobility is enough to generate economies of agglomeration; the need to commute generates diseconomies of city size; the tension between centrifugal and centripetal forces provides a framework for thinking about urban structure.

To understand how this model works, consider what would happen in the absence of foreign trade, and within that special case ask only a limited question: Under what conditions is concentration of all population in either location 1 or 2 an equilibrium? Once we have seen this case, it will be easier to understand the results when the model is opened up.

Suppose, then, that the cost of transacting with the outside world is very high, so that we can ignore the role of the rest of the world. Furthermore, consider the determination of relative real wages when almost all domestic labor is in region 1. If the real wage rate of a worker in location 2 is less than that of a worker in region 1 in this case, then concentration of all labor in region 1 is an equilibrium; otherwise it is not.

We first note that the nominal wage paid at the center of city 2 (w_2) must be less than that at the center of city 1 (w_1). The reason is that almost all output from a firm in 2 must be sold in 1 and must therefore incur transport costs. At the same time the zero-profit output for firms is the same in each location. So goods produced at location 2 must have sufficiently lower f.o.b. prices to sell as much in 1's market as goods produced at location 1. It can then be shown that

$$(3) \qquad\qquad w_2/w_1 = \tau^{(1-\sigma)/\sigma} < 1$$

where σ is the elasticity of substitution among differentiated products.

This wage premium at location 1, which results from its dominant role as a market, essentially represents the backward linkages associated with the concentration of demand there.

Next we notice that if almost all labor is in location 1, almost all goods consumed in 2 must be imported, implying a higher price of these goods:

$$(4) \qquad\qquad T_2/T_1 = \tau$$

where T_i is the price index for goods (excluding land rent) at location i.

If the wage rate is higher in 1 and the price of consumer goods lower, must not real wages be higher in 1? No—because land rent or commuting costs (or both) are higher. With almost all of the labor force L concentrated in 1, the most remote workers in 1 must commute a distance $L/2$, and all workers who live closer to the center must pay a land rent that absorbs any saving in commuting costs. Meanwhile, the small number of workers in 2 pay almost no land rent and have essentially no commuting distance. So the real wage difference turns out to be

$$(5) \qquad\qquad w_1/w_2 = \tau^{(2\sigma - 1)/\sigma} (1 - \gamma L).$$

In this expression the first term represents the centripetal forces—the backward and forward linkages described in equations 3 and 4, which arise from the concentration of suppliers and purchasing power at location 1; the second term represents the centrifugal forces of commuting cost and land rent.[4]

Our next step is to examine the relation between trade openness and urban concentration.

Trade Openness and Urban Concentration

The previous section demonstrates how a concentration of labor in one location may be sustainable, despite the commuting and land rent diseconomies of urban size, through forward and backward linkages. Now suppose that the economy is open to international trade, albeit with some natural and perhaps artificial barriers. How does this change the story? It should be obvious that the effect is to weaken the centripetal forces while leaving the centrifugal forces as strong as before.

Consider a hypothetical primate city, a Mexico City or São Paulo, in a country with a strongly protectionist trade policy. Firms will be willing to pay a wage premium in order to locate at that center precisely because so many other firms, and thus the bulk of their market, are concentrated there. They also may be attracted by the presence of other firms producing intermediate inputs—something not explicitly represented in the model in the appendix, but similar in its effect. On the other side workers will face high land rents or commuting costs, but these will be at least partly offset by better access to the goods and services produced in the metropolis.

But now throw this economy open to international trade. The typical firm will now sell much of its output to the world market (and perhaps get many of its intermediate inputs from that market as well). To the extent that production is for world markets rather than for the domestic market, access to the main domestic market

becomes less crucial—and thus the wage premium that firms are willing to pay for a metropolitan location falls. At the same time, workers will consume more imported goods; they will therefore be less willing to accept high commuting and land costs in order to be close to the metropolitan suppliers. The result can be to make a previously sustainable metropolitan concentration unsustainable.

The easiest way to confirm this intuition is through numerical examples. Figures 1 and 2 show, for one set of parameters, how the qualitative behavior of our two-location model changes as the economy becomes more open (that is, as the cost of shipping goods to and from the world falls). Each figure shows how equilibrium real wage rates in the two locations vary as the share of the labor force in location 1 changes. If we assume that workers move toward whichever location offers the higher real wage rate, these figures show a picture of the economy's dynamic behavior. When the real wage differential is positive, labor moves toward location 1; when it is negative, labor moves toward location 2.

When the costs of transacting with the outside world are fairly high, so that the economy is not very open, there is an equilibrium, though unstable, in which labor is equally divided between the two locations (figure 1). If slightly more than half the labor is in location 1, that location will offer higher wages, inducing more labor to move there. This will strengthen the forward and backward linkages and induce still more labor to move there, and so on. Thus in this closed-economy case a cumulative process leads to a concentration of population in a single metropolis. (Obviously this result does not fully obtain in practice, but perhaps it suggests how a very large primate city is established.)

If the economy is more open, we get a result like that in figure 2.[5] Now the equilibrium in which the population is equally divided between the two locations

Figure 1. *Response of Labor Force to Relative Wages under High Costs of Transacting with Outside World*
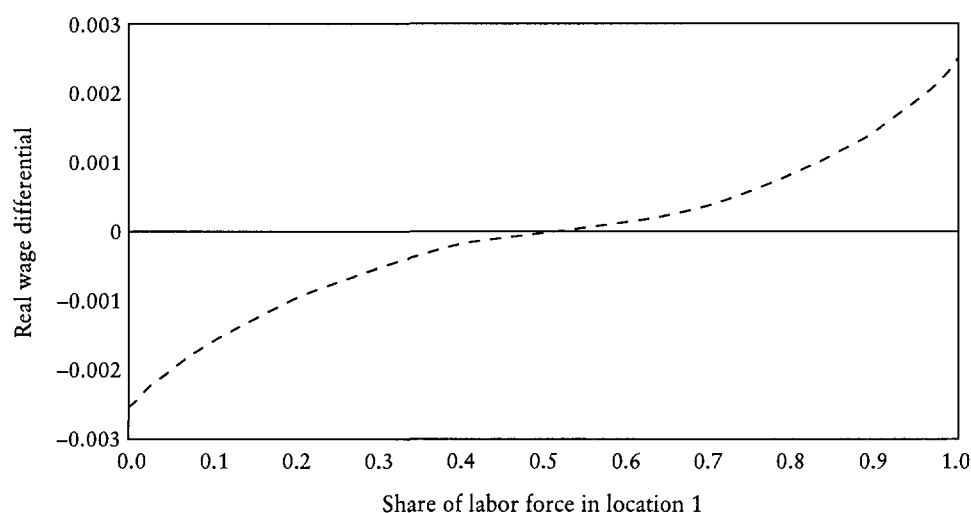
is stable, and a concentration of population in only one location is unsustainable. Thus in this situation we tend to have two equal-size cities rather than one very large metropolis.

It is, of course, obvious that Mexican industry has been shifting its center of gravity away from Mexico City as the country has shifted toward exports. In that case, however, the explanation lies at least partly in the role of access to the U.S. border, as well as in the role of the *maquiladora* program in fostering export industry in the country's north. Our analysis suggests, however, a more generic reason why inward-looking policies may encourage the growth of primate cities, and outward-looking policies may discourage that growth; the empirical evidence described above offers at least modest support for the belief that such a generic tendency exists.

## Political Centralization and Regional Inequality

While the theoretical and empirical relationship between trade policy and urban structure is a surprising, and thus gratifying, insight, it is surely not the most important reason why developing-country cities grow so large, or why regional inequality is so marked in developing countries. Almost surely the most important reason is the role of political centralization.

Political centralization has effects at several levels. The most obvious is that the business of government is itself a substantial source of employment: employment in Paris is larger than it is in Frankfurt in part simply because there are so many more people working for the government, or supplying nontraded services to those who work for the government.

Figure 2. *Response of Labor Force to Changes in Relative Wages under Relatively Low Costs of Transacting with Outside World*
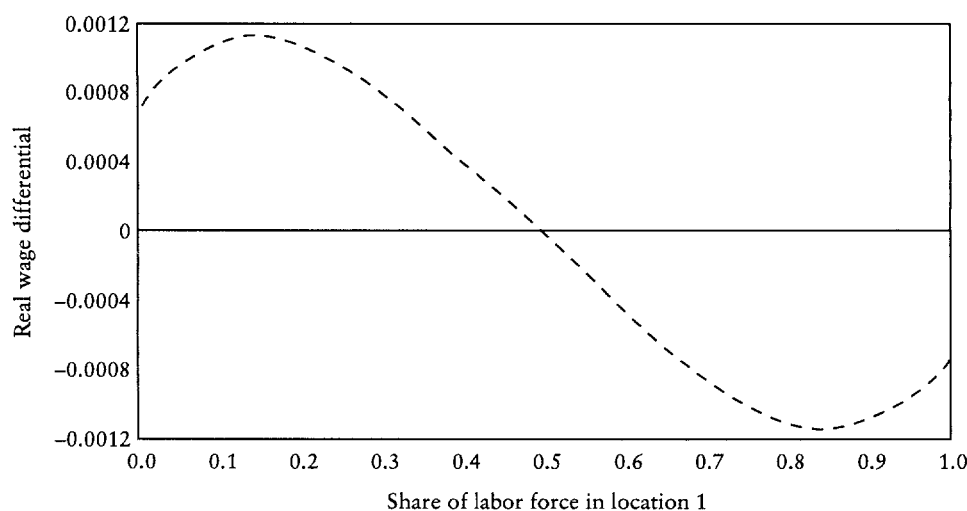
A more subtle source of urban concentration is the importance of access to the government, especially in highly interventionist states. In its simplest form this is simply a result of the concentration of lobbyists. More subtly, if government policies tend to be more responsive to those close at hand (if, say, subsidies or protection to prevent strikes are more forthcoming in the capital than in the provinces), this exerts a hard-to-measure but doubtless important attraction of the capital area for business.

Economic modeling per se cannot contribute much to our understanding of these political concerns. It can, however, help us understand a further consequence of political centralization: the multiplier effects on regional concentration that can result from asymmetric government spending.

Consider a variant on the approach described in the last two sections. Put the commuting and land-rent diseconomies to the side and suppose instead that there is an immobile rural population divided between two regions. Manufacturing will be drawn to concentrate in one region by the forward and backward linkages we have already seen in action, but against this force will be the pull of the market provided by the rural population. A model along exactly these lines is worked out in Krugman (1991b). I show there that the outcome depends on the parameters. For some parameters one gets the type of result shown by the dashed curve in figure 3: the stable equilibrium is one in which manufacturing is equally divided between the two regions.

But now suppose that a government collects taxes from the rural population in both regions but spends it all in one region. Obviously the latter region becomes the larger market, thus attracting more manufacturers. However, the forward and backward linkages that are generated attract still more manufacturing to that region, fos-

Figure 3. *Response of Manufacturing to Relative Wages*

Share of manufacturing in location 1
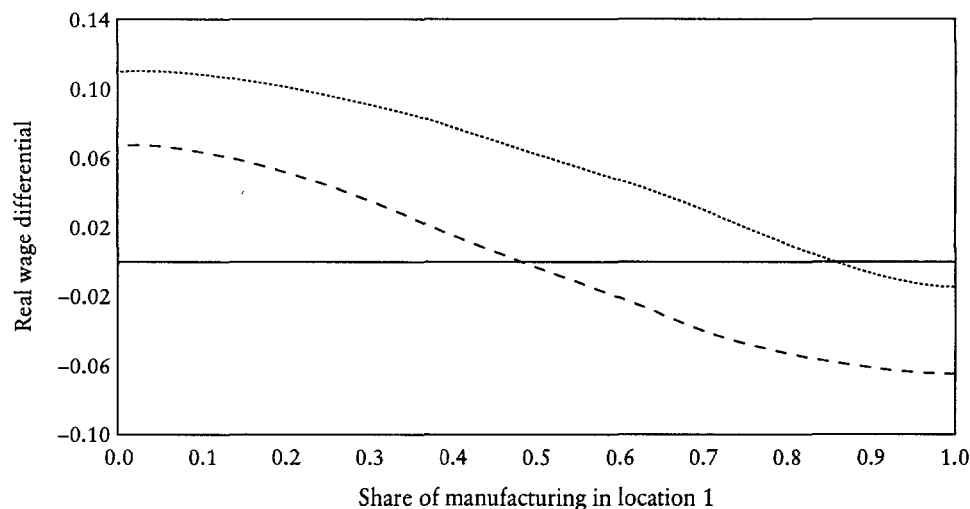
tering a cumulative process of concentration. In figure 3 we start with an economy in which the natural state of affairs has 50 percent of manufacturing in each region. In this example a tax equal to 20 percent of rural income was collected in both regions, but spent only in region 1. The result is shown by the upward shift in the schedule relating the real wage differential to the allocation of manufacturing between the regions (the dotted curve). In this case the multiplier effects cause a concentration of approximately 85 percent of manufacturing in the favored region. The direct transfer of resources from the periphery to the core is only 8 percent of GDP, but the end result is to raise the favored region's share of GDP (before taxes) from 50 to 74 percent.

Although it is not explicitly modeled here, there ought to be an interaction between the strength of multiplier effects producing regional concentration and the degree of openness of the economy. Locating manufacturing near the capital in order to take advantage of the market that the government and its employees provide will be much less attractive in a very open than in a very closed economy.

Transportation Infrastructure

The extent and form of a country's investments in transportation infrastructure can affect the tendency to form large urban centers in at least two ways.

First, the higher transport costs are within a country, the stronger the advantages in terms of backward and forward linkages of locating production near an established metropolitan concentration. This effect may be seen directly in equation 5, which asks whether the linkages are strong enough to sustain an established concentration in the face of the diseconomies of urban scale. In this expression the higher are the transport costs, the more likely is the condition for sustainability to be satisfied.

The implication is that the tendency to concentrate economic activity in a single large city may be reinforced if the government neglects the transportation network. This makes intuitive sense, and corresponds to workaday perceptions about the contrast between location decisions in advanced and developing economies. In advanced economies good transportation to markets (and good communications) is available virtually everywhere, whereas in developing countries roads and telecommunications often peter out quickly as one moves away from the capital.

A more subtle issue involves the form of the transport system. A system that is centered on the primate city is more likely to promote concentration than one that does not favor movement of goods and services in any particular direction.

This point also seems intuitively obvious, but it may be worth sketching out how it works in formal models. Imagine, as in Krugman (1993a), a country with not two but three regions. And suppose that instead of being equal in all directions, transport costs between location 1 and both other locations are lower than those between 2 and 3, so that 1 is in effect the hub of the transport system. Then it is straightforward to show that even if all three regions offer the same size market, region 1 will be a preferred location for goods produced subject to scale economies: it offers bet-

ter access to the national market than does either of the other locations. Of course, such an advantage will not usually stand alone. Typically, concentration of population and centralization of the transport system reinforce one another: transport links point toward the primate city because that is where the markets and suppliers are, and business concentration is all the greater because of the role of that city as a transport hub.

One might speculate that the apparent tendency of developing countries to have more concentrated distributions of urban size is due to an important extent to the way that their relative poverty leads to a limited transport system. In advanced countries the volume of traffic is sufficient to ensure that good roads link even minor centers; railway lines will often provide direct connections that bypass the biggest cities.[6] In developing countries traffic is sufficient to support good roads pointing only toward the capital, if any at all. Here, too, there is probably a political linkage—a system that centralizes political power in the capital is likely to concentrate investment in infrastructure either near it or on projects that serve it.

Policy Implications

One wants to be very careful about drawing policy implications from any discussion of urbanization and regional growth. By its nature this is a subject that deals extensively with external economies and diseconomies; while neoclassical urban systems theory may suggest that competition among city developers yields optimal results, the newer literature does not contain any such suggestion. Yet the extent and even the direction of the deviations from optimality may be sensitive to the particular form of the external effects. One could in principle argue that since the growth of cities necessarily involves positive external economies, the biggest cities tend to be too small. Or one could argue that the diseconomies of congestion and pollution— or the inability of markets to internalize the benefits of creating new cities—mean that primate cities are too big. Most people have an instinctive feeling that the biggest cities are too big. I share that prejudice, but it must be said that it is only a prejudice at this point.

That said, the general moral of the models described here seems to be that a desire for cities in developing countries to be not quite so big may be fulfilled indirectly by the kinds of liberal economic policies currently favored by most international institutions for other reasons. Liberal trade policy appears likely to discourage primate city growth; so does a reduction in state intervention and a decentralization of power. Investment in better transportation infrastructure—a traditional role of government—also seems to work in the same direction.

The tentative conclusion, then, is that neoliberal policies seem likely to have the unexpected side benefit of partly alleviating any problems created by the growth of very large cities. The definite conclusion is that whatever the changes made in economic policies, their implications for urban and regional development within countries are an important, neglected issue.

Appendix

In this appendix I present the formal structure of the model of the determinants of urban concentration sketched out in the second section of the article and illustrated in the third. For a full description of how the model is solved, and an exploration of its properties, see Krugman and Livas Elizondo (1992).

A Formal Model of Urban Concentration

We consider an economy with three locations: 0, 1, and 2. Labor is mobile between 1 and 2, but not from the rest of world.

Each location is a linear city, populated by workers who must work in a central business district but require one unit of land to live on. Thus if a location has a labor force L_j, the distance the last worker must commute is

$$(A.1) \qquad d_j = L_j/2.$$

We assume that commuting costs are incurred in labor: a worker is endowed with one unit of labor, but if he must commute a distance d, he arrives with a net amount of labor to sell of only

$$(A.2) \qquad S = 1 - 2\gamma d.$$

These assumptions immediately allow us to describe the determination of land rent given the labor force at a location. Let w_j be the wage rate paid at the city center per unit of labor. Workers who live at the outskirts of the town will pay no land rent, but will receive a net wage of only $(1 - \gamma L_j)\, w_j$ because of the time spent in commuting. Workers who live closer to the city center will receive more money, but must pay an offsetting land rent. The wage net of commuting costs declines as one moves away from the city center, but land rents always exactly offset the differential. Thus the wage net of both commuting and land rents is $(1 - \gamma L_j)\, w_j$ for all workers.

The total labor input of a location, net of commuting costs, is

$$(A.3) \qquad Z_j = L_j\,(1 - 0.5\gamma L_j)$$

and the location's total income—*including* the income of landowners—is

$$(A.4) \qquad Y_j = w_j Z_j.$$

Next, we assume that everyone in the economy shares the constant elasticity of substitution utility function

$$(A.5) \qquad U = \left(\sum_i C_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}.$$

To produce any good i at location j involves a fixed as well as a variable cost:

(A.6) $$Z_{ij} = \alpha + \beta Q_{ij}.$$

The properties of monopolistic competition models like this are by now very familiar. As long as many goods are produced, and as long as we make appropriate assumptions on transportation costs (see below), each producer faces an elasticity of demand equal to the elasticity of substitution, and will therefore charge a price that is a constant markup over marginal cost:

(A.7) $$P_j = (\sigma/\sigma - 1)\, \beta w_j.$$

Given this pricing rule and the assumption that free entry will drive profits to zero, there is a unique zero-profit output of each product:

(A.8) $$Q = (\alpha/\beta)\, (\sigma - 1).$$

And the constancy of output of each product implies that the number of goods produced at each location is simply proportional to its net labor input after commuting:

(A.9) $$n_j = (Z_j/\alpha\sigma).$$

It will save notation to make two useful choices of units. First, units are chosen to make the f.o.b. price of goods produced at any given location equal to the wage rate at the region's city center. Thus:

(A.10) $$P_j = w_j.$$

Second, there is no need to count goods one at a time. They can be equally well counted in batches, say, of a dozen each. To save notation, the batch size is such that

(A.11) $$n_j = Z_j.$$

To preserve the constant elasticity of demand facing firms, the costs of transacting between locations must take Samuelson's "iceberg" form, in which transport costs are incurred in the goods shipped. Thus we assume that when a unit of any good is shipped between location 1 and location 2, only $1/\tau$ units actually arrive; thus the c.i.f. price of a good shipped from either domestic location to the other is τ times its f.o.b. price. Only a fraction $1/\rho$ of a good imported from location 0 is assumed to arrive in either location 1 or 2. For simplicity, exports are assumed to take place with zero transport costs.[7]

We take τ to represent "natural" transport costs between locations. The parameter ρ, however, is meant to be interpreted as combining natural transport costs with artificial trade barriers. It would be straightforward (and would yield similar results)

in this model to introduce an explicit ad valorem tariff whose proceeds are redistributed, but here we simply imagine that any potential revenue is somehow dissipated in waste of real resources.

Given these transport costs and the utility function, we may define true consumer price indexes for manufactured goods in each location. First, let us define the shares of the three locations in the total number of products produced, which are equal to their shares of net labor input:

$$(A.12) \qquad \lambda_j = \frac{n_j}{\sum_k n_k} = \frac{Z_j}{\sum_k Z_k}.$$

Let the wage rate in location 0 be the numeraire; then the true price indices are

$$(A.13) \qquad T_0 = K\left(\lambda_0 + \lambda_1 w_1^{1-\sigma} + \lambda_2 w_2^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$$

$$(A.14) \qquad T_1 = K\left[\lambda_0 \rho^{1-\sigma} + \lambda_1 w_1^{1-\sigma} + \lambda_2 \left(w_2 \tau\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

$$(A.15) \qquad T_2 = K\left[\lambda_0 \rho^{1-\sigma} + \lambda_1 \left(w_1 \tau\right)^{1-\sigma} + \lambda_2 w_2^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

where

$$(A.16) \qquad K = \left(n_0 + n_1 + n_2\right)^{\frac{1}{1-\sigma}}.$$

We will take Z_0 as given. Suppose we know the allocation of labor between locations 1 and 2. Then we can determine Z_1 and Z_2. As we will see, we can then solve the model for equilibrium wage rates w_j. Labor is, however, mobile, and we will have a full equilibrium only if all domestic workers receive the same net real wage. This net real wage in location j can be defined as

$$(A.17) \qquad \omega_j = w_j (1 - \gamma L_j)/T_j.$$

A situation in which real wages are equal in the two domestic locations is an equilibrium. Such an equilibrium may, however, be unstable under any plausible adjustment story. To get some rudimentary dynamics, we impose a simple Marshallian adjustment mechanism,

$$(A.18) \qquad dL_1/dt = -dL_2/dt = \delta (\omega_1 - \omega_2).$$

We have now laid out a complete formal model. It is not a model with a closed-form analytical solution. However, if one is willing to rely on numerical examples, it is straightforward to solve the equations on the computer for any given parameters and see how the wage differential depends on the allocation of labor between

the two locations, thereby deriving diagrams like figures 1 to 3. As explained in the text, such pictures allow us to see how the patterns of urban or regional concentration change as the parameters change.

City Growth and Power Laws

As mentioned in the text, the size distribution of cities is startlingly well described by a power law of the form

(A.19) $$N(S) = AS^{-\alpha}$$

where $N(S)$ is the number of cities with populations larger than S, and the exponent is very close to −1. (As an illustration, figure A.1 plots the log of metropolitan area rank against the log of city population for the United States in 1991.)

If the distribution of cities were continuous and there were no maximum city size, equation A.19 would be equivalent to saying that the density of cities of size S is
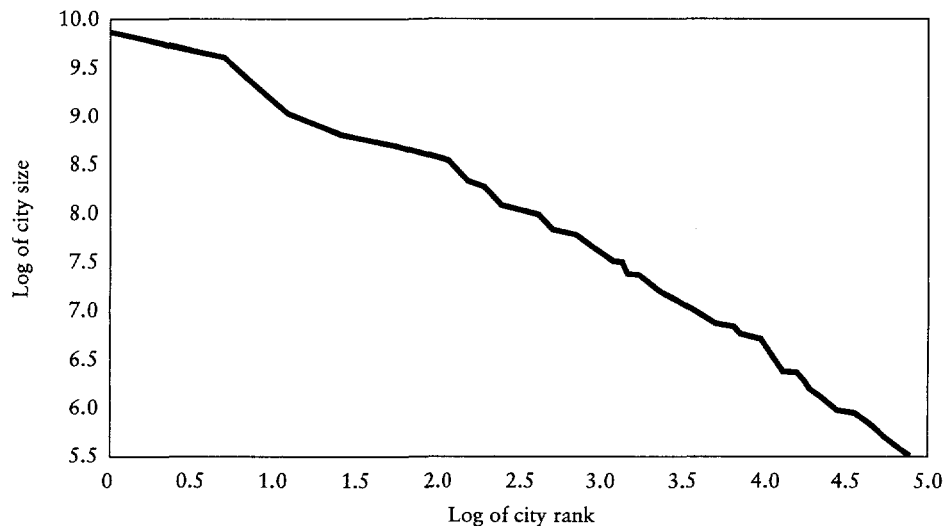
(A.20) $$n(S) = \alpha AS^{-\alpha - 1}.$$

Now imagine that cities come only in discrete sizes, with units of, say, 10,000 people. Let k be the number of units in the population, and $n(i,k)$ be the number of cities with i units; then equation A.19 with an exponent of −1 becomes the statement that

(A.21) $$n(i,k) \approx B(k) i^{-2}.$$

Figure A.1 *Relation of U.S. City Rank and City Size*

Why should something like equation A.21 be true? In 1955 Herbert Simon offered an ingenious explanation, which is a bit short of a formal proof. I offer here a heuristic version of Simon's argument, which is in turn less than rigorous, so it should be viewed only as a suggestive justification.

Imagine that urban growth proceeds according to the following process: new units arrive in the economy successively over time; each new unit is attached to an existing city with a probability that is proportional to the number of units already there. (In Simon's original formulation, some units form the nuclei of new cities; I return to that issue below.) Thus a city of size i has a probability i/k of getting the next unit.

What is the expected change in the number of cities of size i when a new unit is added? That number can change in two ways. First, a city of size $i - 1$ can acquire the new unit, in which case it becomes a city of size i, adding 1 to the total. Second, a city of size i can acquire the unit, in which case it becomes a city of size $i + 1$, reducing the number of i cities. It therefore follows that

$$(A.22) \qquad E[\Delta n(i,k)] = \frac{(i-1)n(i-1,k)}{k} - \frac{in(i,k)}{k}.$$

Now comes the crucial ad hoc step. Simon asks us to imagine that the frequency distribution of city sizes approaches a steady state. This cannot be quite right, since the largest city keeps on getting bigger. But suppose that it is approximately true. Then the number of cities of size i must grow at the same rate as the population, implying

$$(A.23) \qquad E[\Delta n(i,k)] = \frac{n(i,k)}{k}.$$

From equations A.22 and A.23 it follows that

$$(A.24) \qquad \frac{n(i,k)}{n(i-1,k)} = \frac{i-1}{i+1}$$

and thus that

$$(A.25) \qquad n(i,k) = \frac{i-1}{i+1}\frac{i-2}{i}...\frac{1}{3}n(1,k)$$

or

$$(A.26) \qquad n(i,k) = \frac{2}{i(i+1)}n(1,k) \approx 2n(1,k)i^{-2}$$

for large i.

That is, in the upper tail of the size distribution, equation A.21 should be approximately true!

This derivation is a bit slippery. It can be bolstered, however, by simulation results; these show that a wide variety of stochastic growth models will produce upper tails for which equation A.21 is very close to true. For example, in Krugman

Figure A.2 *Relation of Simulated City Rank and Size, Top 100 Cities*
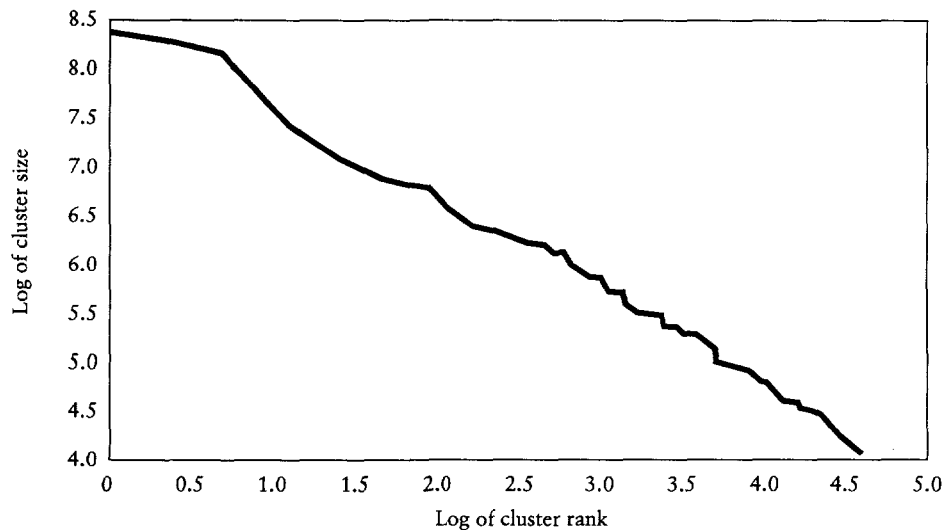
(forthcoming) I consider a model of the following form. A city is begun by an entrepreneur who starts a business. She has two foremen, each of whom with probability p leaves to set up a new factory in the same town. Each foreman has two foremen. Suppose that the probability of defection is close to 0.5, as it must be if towns are to grow very large. Then the results are startling. In figure A.2, I started with 1,000 original businesses, and set $p = 0.49$; the figure shows the relationship between rank and size for the top 100 "cities."

There is a close affinity between Simon's work and the trendy current work on "self-organized criticality," which attempts to explain such observed power-law relationships as the Gutenberg-Richter law relating the sizes and frequencies of earthquakes (Bak 1991).

Notes

1. It is possible, without any real change in the structure, to derive external economies from a monopolistically competitive sector that produces nontraded inputs. See Abdel-Rahman (1988) and Rivera-Batiz (1988).

2. Before the Rosen and Resnick study (1980) most writing on primacy assumed that export orientation would tend to *increase* primacy. The ruling image was of a primary exporting country in which the primate city would be the country's main port; the implicit argument was that the economies of scale involved in building infrastructure for exporting were larger than those involved in selling to the domestic market. One can hardly deny that this effect has existed in some times and places; the evidence that the effect runs the opposite way is not overwhelming. This kind of ambiguity arises in any attempt to summarize the richness of cross-national variation with a short list of explanatory variables.

3. In what are commuting costs incurred? It is easiest to assume that they are incurred only in workers' time, and that time spent commuting is time not spent working. In this case, as shown in the appendix, the net wage rate of the most remote worker in a city of population L takes the form $w (1 - \gamma L)$, where w is the wage at the center.

4. In this story all workers end up concentrated in one location, earning the same wage. It thus appears to be a model of urban concentration but not of regional inequality. Suppose, however, that not all workers are mobile. Then it is apparent that a core-periphery pattern could emerge in which mobile workers agglomerate in one region, leaving behind an impoverished rump of those workers who for whatever reason cannot or will not move. Stories along the same lines can surely also be relevant to the extreme regional inequality one sees in some developing countries.

5. Krugman and Livas Elizondo (1992) show that there may be a range of parameters in which both dispersed and concentrated stable equilibriums exist. Numerical examples suggest, however, that this range is quite narrow.

6. Cronon (1990) shows that the rapid growth of Chicago in the nineteenth century came to an end largely because the growing density of the U.S. rail network made its traditional position as a rail hub increasingly unimportant.

7. Even though we make exports costless, an increase in p, which reduces imports, must necessarily decrease exports as well. The mechanism through which this happens is a rise in the prices of domestic relative to foreign output—in effect, a real overvaluation that prices domestic goods out of world markets.

References

Abdel-Rahman, H. 1988. "Product Differentiation, Monopolistic Competition, and City Size." *Regional Science and Urban Economics* 18: 69–86.

Ades, A., and E. Glaeser. 1993. "Trade and Circuses: Explaining Urban Giants." Department of Economics, Harvard University, Cambridge, Mass.

Bairoch, P. 1988. *Cities in Economic Development.* Chicago: University of Chicago Press.

Bak, Per. 1991. "Self-Organizing Criticality." *Scientific American* 264 (January): 46–53.

Carroll, G. 1982. "City-Size Distributions: What Do We Know After 67 Years of Research?" *Progress in Human Geography* 6: 1–43.

Cronon, W. 1990. *Nature's Metropolis: Chicago and the Great West.* New York: Norton.

Fujita, M., and P. Krugman. 1993. "Monopolistic Competition and Systems of Cities." Department of Regional Sciences, University of Pennsylvania.

Glaeser, E., H. D. Kallal, J. Scheinkman, and A. Shleifer. 1992. "Growth in Cities." *Journal of Political Economy* 100: 1126–52.

Henderson, J. V. 1974. "The Sizes and Types of Cities." *American Economic Review* 64: 640–56.

———. 1977. *Economic Theory and the Cities.* Oxford: Oxford University Press.

———. 1988. *Urban Development.* Oxford: Oxford University Press.

Ijiri, Y., and Simon, H. 1977. *Skew Distributions and the Sizes of Business Firms.* Amsterdam: North-Holland.

Krugman, P. 1991a. *Geography and Trade.* Cambridge, Mass.: MIT Press.

———. 1991b. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99: 483–99.

———. 1993a. "The Hub Effect: Or, Threeness in Interregional Trade." In W. Ethier, E. Helpman, and J.P. Neary, eds., *Theory, Policy, and Dynamics in International Trade.* Cambridge: Cambridge University Press.

———. 1993b. "On the Number and Location of Cities." *European Economic Review* 37 (March): 293–98.

———. Forthcoming. *The Self-Organizing Economy.* London: Basil Blackwell.

Krugman, P. and R. Livas Elizondo. 1992. "Trade Policy and The Third World Metropolis." Working paper 4238. National Bureau of Economic Research, Cambridge, Mass.

Pred, A. 1966. *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800–1914.* Cambridge, Mass.: MIT Press.

Rivera-Batiz, F. 1988. "Increasing Returns, Monopolistic Competition, and Agglomeration Economies in Consumption and Production." *Regional Science and Urban Economics* 18: 125–53.

Rosen, K., and M. Resnick. 1980. "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy." *Journal of Urban Economics* 8: 165–86.

Simon, Herbert. 1955. "On a Class of Skew Distribution Functions." *Biometrika* 52: 425–40.

COMMENT ON "URBAN CONCENTRATION: THE ROLE OF INCREASING RETURNS AND TRANSPORT COSTS," BY KRUGMAN

Andrew M. Isserman

P aul Krugman does many important things extremely well. Whether his research has created the basis for a new regional economics or economic geography remains to be seen, but his research and proclamations already have infused the field with new enthusiasm, energy, and hope. He demonstrates that its messy problems can be studied with today's formal mathematical models and argues that the field will grow in prestige and importance. Yet some charge Krugman with ignoring prior work and, worse yet, with presenting no new insights. To use his own words, "I am having a terrible time with my current work on economic geography; referees tell me that it's obvious, it's wrong, and anyway they said it years ago" (quoted in Gans and Shepherd 1994, p. 178). How valid are these criticisms with respect to this article?

Is the Model So Simple That It Yields Misleading Conclusions?

Krugman argues that people concentrate in cities because of the greater variety of goods available. Firms concentrate there because cities offer larger markets for their goods. In perhaps a spatial wrinkle on Say's law, firms create their own markets by concentrating their locations. Wages are lower in the countryside because firms must absorb the costs of reaching the city market; prices are higher because of the cost of transporting goods there from the city. The lower wages and higher goods prices in the countryside do not stimulate everyone to move to the city because higher rent and commuting costs in the city balance its higher wages and lower prices.

Krugman predicts that free trade can reduce the pressures on large cities by opening up alternative markets to the firms concentrated in the cities. Firms no longer need to seek out locations that maximize their access to the national market. Likewise, having access to goods from around the world, consumers need not locate in the city and pay high commuting and land costs.

Central to evaluating the model and its predictions is how the model defines its key components—primate cities, countryside (or smaller cities), traded goods, transportation infrastructure, and migration. The main advantage of modeling is simpli-

Andrew M. Isserman is director, Regional Research Institute, and professor of economics and geography at West Virginia University.

fication. Its focus on selected attributes can yield new insights. The main danger of modeling is that those insights might be wrong because of ill-chosen simplification or gross mischaracterization.

Primate cities in the model are production points that can offer wage premiums because they are the market centers and incur low transport costs to reach that market. Commuting costs increase and rents decrease with distance from the center. Primate cities have more people than other places. That's about it for the primate city. The countryside must offer lower wages to absorb the costs of transporting goods to the city market. That's about it for the countryside. As Krugman says, the model is simple.

Now, is there mischaracterization and oversimplification? First, the primate city is not only a manufacturing center. Typically, it is the center of just about everything. Thus, even if a greater proportion of its production is sold to foreign markets as a result of free trade, only one part of its economic base is freed from dependence on the primate population concentration. Its other functions, particularly finance, government, trade, and communication, might expand and reinforce its dominant position. Krugman acknowledges some of these roles; his formal model does not. Consequently, the model may yield erroneous conclusions regarding the overall effects of free trade on population concentration.

The posited manufacturing trade effect might not occur at all. Its basis is the substitution of foreign markets and goods for primate city markets and goods. For that substitution to occur, primate city manufacturers must be able to produce for the international market, and city residents must be able to afford the goods produced by the international market. To the extent that these assumptions do not hold, the posited trade effect will be weaker. To the extent that local producers will be eliminated by foreign competitors or by competitors newly located at now optimal points that serve both the world and primate markets, the trade effect will be stronger. In the model's theoretical extreme the primate city will cease to exist if all manufacturing moves to the third point, that is, out of the country to the rest of the world. We cannot know the nature or size of the trade effect on population concentration until we know more about the trade, goods, income levels, and distribution systems of the particular country.

Meanwhile, what is going on in the countryside? Presumably, some small cities and rural places will grow because of their production for world markets. Rural areas, however, produce food, natural resources, and other goods and services that are tied to place-specific attributes, many of which cities cannot produce. That these goods are missing from the model becomes a problem when drawing conclusions about the effects of trade on urbanization. Trade liberalization may have very big effects on agriculture, for example. Unlike manufactured goods, agricultural goods might be perfect substitutes for goods now consumed in the world market. Agricultural trade may evolve precisely as Krugman postulates for the primate city's manufactured goods.

If so, trade liberalization might have serious repercussions for primate cities, but opposite to those Krugman posits. Imports might destroy domestic agriculture, as is

feared in France, Japan, the Republic of Korea, and many other countries. The plausible outcome in developing-country settings is more, not less, migration from the countryside to the primate city. Alternately, trade liberalization might enable the country to expand its exports, quite plausibly, by making agriculture more capital intensive in order to increase output for world markets. This change also might destroy the current agricultural system and cause additional rural unemployment and migration to the primate city.

The nature of transportation infrastructure in the model is important, too. What is assumed in the model and what is implicit in the trade liberalization predictions appear to be inconsistent. Crucial to Krugman's predicted trade effect is the ability to produce outside the primate city for world markets. Yet, according to the model, production is concentrated in the city largely because high transportation costs to the city from elsewhere in the country push people to the city as consumers and push firms there as producers.

Enter trade liberalization, and the transportation system seems suddenly different. Now it suffices to serve more points. For Krugman's Mexican border case that argument is easy: the transportation system that matters is outside the country once the border or port is reached. For other cases the story becomes a bit mysterious. We must now assume an adequate transportation infrastructure from a place to the world market, but before we assumed no system existed from there to the primate city that could take away the latter's locational advantage. Somehow, for the predicted trade effect to occur, we must change the world of the model or at least the country's transportation infrastructure. Creating additional markets through trade liberalization will weaken the domination of the primate city only if those markets can be reached at reasonable cost from several points within the country. Again, the actual outcome will depend on the particulars of each national case.

Finally, migration is implicit in the model. People move to the primate city until increases in land rents and commuting costs bring the urban system into equilibrium. But in reality people also move abroad. Trade liberalization that fosters greater production at home might reduce emigration—as NAFTA supporters argued—and thus might spur migration within the country, particularly to places where output and employment demand increase. Those places might well include the primate city, again confounding Krugman's posited trade effect.

I would be delighted to learn that the messy issues raised here can be readily resolved within the model. I fear, however, that the model is too simple and the predicted effects of trade on population concentration might not hold.

Although a look through Krugman's "new lens" may not give us an accurate view of the world, it does help reveal obstacles that block our view. Those obstacles deal with the nature of cities, countryside, production, infrastructure, and migration in individual countries—all standard elements of the old economic geography. By concluding that the knowledge of the old economic geographers, the area specialists, can teach us a lot about the effects of trade liberalization on urban concentration, I may be committing heresy for one trained in economics. My response to Krugman's model and his "what can we learn" question is, however, a call not for a return to

the old economic geography but for recognition of the virtues of both old and new. Perhaps the main implication of Krugman's article is that understanding the relationship between trade and urbanization in developing countries calls for still richer models that are rooted in critical "ground truth" on key parameters and variables.

Does Krugman's Model Teach Us New Things?

Are Krugman's policy conclusions regarding trade, centralization, taxes, and infrastructure surprising insights into our world? Do they flow directly from the model, or are they actually the product of "older," less-formal reasoning? What have we learned through that new lens? Krugman concludes that the desire that cities in developing countries not be quite so big may be served indirectly by liberal trade policy, a reduction in state intervention and a decentralization of power, and investment in better transportation infrastructure. He sprinkles these conclusions with qualifiers—"may be served," "appears likely," and "seems to work"— perhaps to express a wise author's recognition that this is the best his models can do at this point.

That changes in trade policy can create production advantages for certain regions is conventional knowledge, not just of economists (Ruane 1983; Williamson 1986), geographers (Sheppard 1982; Warf and Cox 1993), and sociologists (Portes 1989), but also of French farmers and other folk. That trade liberalization will discourage the growth of primate cities can be predicted simply by pointing to opportunities being created elsewhere in the country, but neither such common sense reasoning nor Krugman's formal model means that the prediction will be correct. Too many factors, absent from both, will determine what will happen in a particular country.

Krugman argues that political centralization is "almost surely the most important reason" that cities in developing countries grow so large. Yet it does not appear explicitly in the model. Centralization can be linked to the model, as Krugman does by pointing out that government is itself a major employer. He also asserts that if the government is more interventionist, access to government is more important, so more lobbyists will be contributing to the economic base and size of the city. Thus decentralized government and less-interventionist government will mean fewer people in the primate city. That argument may well be true, but it does not qualify as a new insight *provided by the model.* Furthermore, trade liberalization need not be accompanied by decentralized and less interventionist government despite being part of the same Washington consensus.

Krugman's political section contains an interesting formal modeling exercise. It shows that if the government taxes rural income and spends the revenues in the city, further concentration in the city will result. That illustration, a nifty by-product of the model, is also not a new insight. Agricultural economists and others have argued for a long time that overtaxation of the countryside through low food prices has subsidized cities and stimulated urban and industrial growth. It was a key concept of Soviet economic planning, it underlies the Chinese economic system, not to mention feudal economies, and it is a shibboleth of the urban bias literature.

Finally, there is Krugman's argument for investment in infrastructure to discourage primate city growth. This recommendation comes from the model and is a direct reflection of the role played by transportation costs. Yet it is also an old point in regional and development economics (Hansen 1965; Hirshman 1958; Williamson 1965). Poor transportation linkages within the country and transportation systems focused on the primate city reinforce the city's dominant role. As Krugman notes, these points seem intuitively obvious, as does the related policy recommendation.

Conclusion

Perhaps the most important argument of the article, even if it is not based on the model, is that a whole constellation of popular policies can contribute to urban deconcentration. That the Washington consensus policies can lead to population decentralization is a striking, intriguing observation. Does it matter that it does not follow directly from the model? Does it matter that each piece of that conclusion can be found in earlier work? Does it matter, in short, whether Krugman is guilty of his referees' charges: "it's obvious" and "someone else already said it?" No! To require that a model teach us new things is to set a very demanding standard. Most models do not meet it. Usually, we are content if our modeling exercises reproduce known things or focus our attention on important parameters. Even if some of Krugman's conclusions are wrong, others obvious, and none virginal, the article focuses our inquiry, raises stimulating questions, and makes us appreciate the limits of our modeling capabilities and our knowledge.

Yes, Krugman is right. His style—"maybe by claiming more originality than I really have"—not to mention his many achievements, invites referees and commentators to apply unusual standards to evaluate his work. I propose another unusual challenge. At the 1992 conference, Bank President Lewis Preston stressed that research into why development takes place in some settings, and why poverty persists in others, is central to the Bank's mission (Preston 1993). Krugman's article has refocused our attention on markets, transportation costs, and government centralization to explain why development occurs in some places and not in others. He has not yet focused his remarkable skills and talents on why poverty persists. I encourage the World Bank to invite him to speak at this conference again, but this time on spatial aspects of poverty. Poverty economics, too, will benefit immensely from being stirred up by his ideas.

Paul Krugman's work and his models are worthy and exciting additions to the classic location theory of von Thunen, Weber, Losch, and Christaller. I agree wholeheartedly with his "definite conclusion...that whatever the changes made in economic policies, their implications for urban and regional development within countries are an important, neglected issue."

Note that this time Krugman offers no qualifiers. This conclusion is "definite." Although we in regional economics and economic geography have not neglected these issues, to our chagrin others in academia and public agencies have ignored them. The stimulating work of Paul Krugman will make it more difficult for them to continue to do so.

References

Gans, Joshua S., and George B. Shepherd. 1994. "How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists." *Journal of Economic Perspectives* 8 (1): 165–79.

Hansen, Niles M. 1965. "Unbalanced Growth and Regional Development." *Western Economic Journal* 9 (1): 3–14.

Hirshman, Albert O. 1958. *The Strategy of Economic Development.* New Haven, Conn.: Yale University Press.

Portes, Alejandro. 1989. "Latin American Urbanization During the Years of Crisis." *Latin American Research Review* 24 (3): 7–44.

Preston, Lewis T. 1993. "Opening Remarks." *Proceedings of the World Bank Annual Conference on Development Economics* 1992: 7–8.

Ruane, Frances R. 1983. "Trade Policies and the Spatial Distribution of Development: A Two-Sector Analysis." *International Regional Science Review* 8 (1): 47–58.

Sheppard, Eric. 1982. "City Size Distributions and Spatial Economic Changes." *International Regional Science Review* 7 (2): 127–51.

Warf, Barney, and Joseph Cox. 1993. "The U.S.-Canada Free Trade Agreement and Commodity Transportation Services among U.S. States." *Growth and Change: Journal of Urban and Regional Policy* 24 (3): 341–64.

Williamson, Jeffrey G. 1965. "Regional Inequality and the Process of National Development: A Description of the Patterns." *Economic Development and Cultural Change* 13 (4/2): 3–45.

———. 1986. "Regional Economic-Demographic Modeling: Progress and Prospects." In Andrew M. Isserman, ed., *Population Change and the Economy: Social Science Theories and Models.* Boston, Mass.: Kluwer-Nijhoff Publishing.

COMMENT ON "URBAN CONCENTRATION: THE ROLE OF INCREASING RETURNS AND TRANSPORT COSTS," BY KRUGMAN

J. Vernon Henderson

Paul Krugman provides an interesting and useful overview of what the new economic geography has to say about urban development. Given the breadth of the article, I limit my comments to four items.

Key Modeling Differences between the Neoclassical Urban Systems and the New Economic Geography

Krugman suggests that a critical difference between the urban systems approach and the new economic geography is in the specification of imperfect versus perfect competition. In fact, the urban systems models consider imperfect competition among producers within cities (Abdel-Rahman 1988), across cities (Henderson and Abdel-Rahman 1992), and as among land developers (Henderson 1977, 1988). Rather, the differences are in the specification of the nature of agglomeration benefits, in the treatment of space, and in the potential role of large agents in land markets.

The urban systems models assume that agglomeration benefits arise from localized, external economies of scale. Microeconomic foundations for these scale economies include Dixit-Stiglitz diversity of nontraded intermediate inputs going into production of a city's export good (Abdel-Rahman 1988); search and matching in local labor markets (Helsley and Strange 1990); information spillovers (Fujita and Ogawa 1980, 1982); and Becker-Murphy intraindustry specialization. Information spillovers can involve a market where information products are bought and sold, but information is subject to spatial decay. In contrast, the new economic geography assumes that agglomeration benefits arise solely from savings on transport costs from concentrating producers and trade at a point. There is a reality issue here of what is important in modern metropolitan areas, as well as the welfare implications of choices of assumption.

The second difference is that the new economic geography tries explicitly to model space between cities and intercity transport costs. The urban systems models do not introduce national space explicitly, which is clearly a loss. However, partic-

ularly in empirical applications, they do consider the impact on prices received or paid for products in cities with better access to national and international markets.

Finally, while the urban systems models are solved with both atomistic and large agents in land markets, that approach tends to focus on large agents, whereas the new economic geography ignores them. By institutional facts, local governments are large agents in land markets since half of urban land is in the public domain and the other half is controlled by planning and zoning. In addition, new cities, such as edge cities in the United States, tend to be the product of individual private developers, who engineer massive reagglomerations of employment and population.

Although the explicit introduction of national space is highly desirable, it is hard to do so and still retain the essential features of systems of cities. By introducing space the new economic geography has been unable to model a system or hierarchy of cities in which cities are of different sizes and specialize in the production of different bundles of goods. This inability limits sharply the application of the model to questions of urban concentration, as we will see. And so far the introduction of national space has been at a very elementary level.

Urban Concentration Issues in the New Economic Geography

Krugman considers an example of how to analyze the impact of the opening of trade on national space and urban concentration. He starts in autarky with two sites, call them A and B, where because of agglomeration benefits production is concentrated at one site. Let's say the two sites are on the coast, equidistant from international markets, as in Krugman. With the opening of trade, if the pull of international markets is strong enough, we end up with two cities of equal size focused on international exports.

I have three comments on this treatment. First, in Krugman's model or in variants of it that he has developed, we could have started in a different region of parameter space, where in autarky there are two (stable) equal-size cities, in which case trade would have no impact on urban concentration. Second, if we start with population concentrated all at one site, trade need not induce a shift of some population to the other. Population can remain concentrated. Third, the results depend on the two sites being equidistant from international markets.

Suppose, instead, that we start with a coastal periphery site and an interior central site. If in autarky all production is at the coast, the introduction of trade has no impact on urban concentration. If, however, in autarky all production is in the interior, trade also will not affect urban concentration: either all production will remain in the interior or it will all shift to the coast.

These examples show that the impact of trade on national space is situation-specific, depending on the precise geography of the country. The analysis of urban concentration here also is limited to a country of just two cities. Countries, even small ones, with a few city-state exceptions, have dozens if not hundreds of cities. There, urban concentration concerns only the portion of the population living in

relatively big cities. In thinking about urban concentration, we may want a more generic or general framework.

Urban Concentration in Urban Systems Models

Urban systems models are full general-equilibrium models in which there are agglomerations of population and investment in cities (Henderson 1988). The models are designed to capture the essentials of what we see in large, urbanized economies. Cities endogenously specialize in production of different export goods or groups of goods. Because they produce different sets of export goods, there are different types of cities. The size of any city type depends on the extent of localized external economies of scale in its specific production. The greater are the scale economies, the larger is the city size for that type. The grouping of products into a city type involves a tradeoff between the benefits of greater specialization at the detailed industry digit level and the benefits of diversity, such as the exchange of information and products with related industries. The numbers of each type of city depend on its equilibrium size and the national demand for its product. The greater the demand, the more cities of that type.

Now to the issue of urban concentration. Consider a simple hierarchy of four types of cities differentiated by the products in which they specialize:

City size	*Goods produced*
Villages and small towns	Agriculture
	Traditional manufacturing
	Traditional textiles
	Food processing
	Simple metal working
	Nonmetallic minerals (clay, glass)
Medium-size cities	Modern manufacturing
	Primary metals
	Machinery
	Transport equipment
Metropolitan areas	High-tech development and modern services
	Instruments
	Electronics
	Finance
	Publishing
	Arts
	Instruments

In autarky we start with a composition of city types and numbers, based on the economy's stage of development, technology, and national demand. That could involve low urban concentration, with many villages and small towns and just a few

medium-size cities, or metropolitan areas producing modern manufacturing and services, or the other way around.

The introduction of trade will alter the composition of national output, which will alter the number of required cities of each type. A shift to more modern manufacturing or services involves a reduction in the relative number of villages and small towns and an increase over time in the absolute and relative number of medium-size cities and metropolitan areas, thereby increasing urban concentration. That's the standard urbanization of a country. In a growing economy, this is accomplished through stagnation or declines in the population of most villages and small towns, with some former small towns growing into new medium-size cities and some former medium-size cities growing into new metropolitan areas. Typically, the original one or two metropolitan areas will also grow as the composition of their internal production changes.

Of course, the shift in urban concentration can go the other way, as a country specializes in more traditional manufacturing, such as production of labor-intensive parts and components. With liberalization of international trade some countries, such as Mexico, seem to be following that path. Regardless, this type of analysis is a natural way to get at a change in urban concentration, even before we introduce specific effects due to the spatial dimensions of the geography of the country in question.

With this same model we can address the impact on urban concentration of government policies, such as protectionist trade policies or minimum wage laws affecting the national composition of output. But perhaps more critically in the long term, we can examine the impact of institutions on urban concentration. For example, empirical evidence suggests that central, as opposed to federal, systems of government bias investments in infrastructure toward large metropolitan areas such as the national capital. Decentralization of decisionmaking about investments in local public infrastructure may mean decreased urban concentration as smaller and medium-size cities compete more effectively in the national allocation of public investment.

But the effects may be more subtle. State-owned heavy industry may choose to locate in basically silly places, such as the national capital (far from raw material supplies), drawing other private producers there. Or the central government may restrict the location of foreign producers to a few key metropolitan areas. Consider the impact on urban centralization in Brazil, where protectionist policies favoring heavy industry in the postwar period were combined with the location of the government portion of the iron and steel industry in São Paulo or on the São Paulo–Rio axis and the restrictions on the locations of foreign automobile plants. That's like forcing iron and steel and auto plants to locate in New York City, rather than in the midwest. We don't even need to discuss São Paulo's air quality as a function of locating the heaviest polluters next to the greatest number of victims. Note that the suggestion here is that greater São Paulo has the wrong industrial composition, not that it is necessarily too large.

Even more subtle institutional examples exist. Take the lending policies of banks in a system dominated by state-owned banks, such as Indonesia. Often there is an effective spatial hierarchy, in which local banks in smaller towns are limited in the

sizes and types of business loans that they can make. Such restrictions create a spatial bias in capital markets, affecting firms' location decisions and encouraging urban centralization and concentration. More generally, access to a centralized bureaucracy in countries where the central government directly or indirectly is a major player in industrial markets forces centralization of firms.

The Rank Size Rule

My final comment is a more technical one. Krugman cites the rank size rule for an economy under which the power relationship Rank $= C_0$ (Urban Population)$^{-\alpha}$, has $\alpha = 1$. Rank is the rank of a city in the economy by its population size. We know from standard urban economics texts (for example, O'Sullivan 1992) that when the power law in Krugman's equation 1 is estimated for *all* U.S. metropolitan areas, α in fact is not 1.0, as Krugman suggests; it is 0.9. In a power relationship that is a big, rather than trivial, difference. Now rather than Rank times Population being a constant, Rank times Population declines exponentially against Rank, or Rank times Population $= C_1$ (Rank)$^{-.11}$. That's an 11 percent rate of decline!

Nonetheless, a general power law provides an excellent fit for the data. This power law may be derived from a variety of models, as examples, including the urban systems models. But most people feel it is a common statistical artifact, with no special meaning, suggesting neither random processes nor some fundamental law of nature.

References

Abdel-Rahman, H. 1988. "Product Differentiation, Monopolistic Competition, and City Size." *Regional Science and Urban Economics* 18: 69–86.

Fujita, M., and H. Ogawa. 1980. "Equilibrium Land Use Patterns in a Non-Monocentric City." *Journal of Regional Science* 20: 455–75.

————. 1982. "Multiple Equilibrium and Structural Transition of Non-Monocentric Urban Configurations." *Regional Science and Urban Economics* 12: 161–96.

Helsley, R., and W. Strange. 1990. "Matching and Agglomeration in a System of Cities." *Regional Science and Urban Economics* 20: 189–212.

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

————. 1988. *Urban Development: Theory, Fact, and Illusion.* New York: Oxford University Press.

Henderson, J. V., and H. Abdel-Rahman. 1992. "Urban Diversity and Fiscal Decentralization." *Regional Science and Urban Economics* 22: 103–22.

O'Sullivan, A. 1993. *Urban Economics.* Homewood, Ill: Irwin.

FLOOR DISCUSSION OF "URBAN CONCENTRATION: THE ROLE OF INCREASING RETURNS AND TRANSPORT COSTS," BY KRUGMAN

A participant from the World Bank emphasized the importance of government structure to the size and distribution of cities. One side effect of U.S. federalism is cities proportionately much smaller than in Latin America (where governments tend to be centralized). A third of Chile's population lives in Santiago, and a third of Argentina's lives in Buenos Aires. If New York City held a comparable proportion of the U.S. population, its population would be about 85 million.

Without question, said Krugman, in the cross-sectional work federalism is the single most clear-cut factor affecting the degree of urban concentration. Politics is more important—or at least more transparently important—than any strictly economic issues. To the question of whether it is the inherent concentration or the urban hierarchy that drives the political system, rather than the opposite, he responded that at some level it is all endogenous. Repeating a quotation from the Penguin Atlas of Ancient History—"As history is a branch of the biological sciences, its ultimate expression must be mathematical"—he said that at some level it is all quantum mechanics. And yet, while it is true that the political systems of Canada and the United States are endogenous, it is also true that they were set at a time when the countries were agricultural. So yes, the existence of multiple political centers in the United States led to the federal, decentralized U.S. system, but it was also predetermined. Those multiple centers were New England shipping and fishing versus Virginia tobacco exports, not Atlanta versus Boston.

A participant from the Bank asked Krugman how the opening of trade between cities would affect the size distribution of cities. It is a question not only of space between cities but one of space within cities as well. In South Africa, for example, there are true cities with economic nodes (such as Johannesburg), and there are dormitory towns further away (such as Soweto). As trade opens between those cities, does Krugman predict that Johannesburg will get a shock? Will Soweto emerge as a new metropolitan setting? Krugman responded with a "factoid," which, he said, was as far as one could go with such limited data: modern metropolitan areas—which are based largely upon automobile rather than rail systems, and in which fax machines and similar technologies are available—do seem to want to be polycentric.

Los Angeles, of course, has sixteen or so of what pass for downtowns, or edge cities, and developing countries' cities seem to be evolving in a similar direction. Mexico City has no single center; it looks more like Los Angeles than like Chicago in 1950. In that sense, one might expect greater interpenetration in the big picture—a big Johannesburg-Soweto sprawl in which races and types of activities are intermingled—together with intense segregation at the micro-level.

Krugman had discussed how political structures and issues influence the development of urban systems. A participant from the Bank wondered if a stronger argument might be made for the federal structure being the result of a preexisting urban system. Economic activity normally precedes political organization, said the participant; the United States has a federal structure because there were several economic centers, none willing to seek political dominance over the other. You have the same thing in Canada and elsewhere, including India. In a way the movement toward opening up in Mexico may have come from the realization that this was a way to solve problems of the urban system, rather than the reverse. Or look at the relations between European countries: the United Kingdom, which has a highly centralized political structure, is unwilling to cede power to another power center.

In empirical work, responded Krugman, researchers have tried to control for other factors in fairly large cross-sections of countries by dividing the sample into federal, decentralized systems and those that are more centralized. At the margin this is to some extent an arbitrary division, but the nature of the system itself does seem to contribute to deconcentration, over and above other factors that are controlled for. The capital market has a strong spatial bias toward concentration, said Krugman, offering as an example the built-in hierarchy in lending policies in Indonesia's banking system, which is dominated by state-owned banks: there are sharp limits on the size and class of loans to businesses in smaller cities, fewer limits on loans to businesses in medium-size cities, and very few limits on loans to businesses in larger cities. If you look at the data—especially before and after liberalization and similar changes—you will see that institutional behavior influences firms' decisions about location in a way that favors concentration and centralization.

Krugman said that common sense dictated that a truly unified European economy ought to have a "top dog" city larger than any of the national top dog cities in Europe. Following Zipf's law, Europe's primate city ought to contain 25 million or 30 million people. The interesting question is, what would happen if that city turned out not to be Paris? The geographic consequences of European integration involve important issues of city size, city roles, and regional roles. Europe is a vastly more polycentric economy than the United States. If post-1992 Europe is eventually going to look like the United States, some fairly interesting problems of political economy are going to arise in the course of adjustment.

A participant from the Stockholm School of Economics asked Krugman if the nihilistic and non-nihilistic models were not perhaps complementary in the sense that they address different things: one deals with binary choices and the other, in a sense, with the law of large numbers. Other models don't have an explicit growth process or dynamics attached to them, said the participant. If you get growth rates

distributed close to normal, you always get something that looks like a law of normal distribution, so these would be complementary rather than conflicting. Krugman wished it were as simple as the nihilistic and non-nihilistic views being different magnifications of the same phenomenon. The nihilistic story occurs wherever you have random city growth, where the expected rate of growth does not really depend on city size, provided growth goes far enough. The problem, he said, is that everything we have to say about economics suggests that there ought to be a maximum city size and that therefore expected city growth ought to depend on city size. In Henderson-like models, there ought to be a set of typical sizes for cities. It was unclear to Krugman whether there is some fix-up, whether lurking in the hidden logic of such models there is a way to create the optical illusion that markets don't care how big cities are. At the moment it appeared to him to be a paradox.

A participant from Oxford University was curious about Krugman's thinking on increasing returns and the relative shift that appears to occur between internal and external economies of scale. Krugman responded that plant-size economies are a decreasingly important factor in advanced countries, partly because manufacturing is now less important there and partly because the markets are now so much bigger relative to plant size than they used to be. The forces of agglomeration, he surmised, are increasingly external in form. The disturbing fact that the average size of enterprises is smallest in the largest cities would strongly suggest an external-economy explanation for what goes on.

That raises a substantial question about the kind of model described here, and, said Krugman, evokes the inevitable image of reinventing the wheel. But even if you don't like a particular model, there is something to be said for simply putting the spotlight on how important spatial issues are in daily life and economics. It seemed to him an obvious empirical fact that there are agglomeration economies in regional and spatial economics. A proliferation of different research approaches is beneficial because we are not competing for resources but collectively drawing the attention of economists and others to a field that has been stunningly neglected. Before the conference he had looked at two of the principle economics textbooks; he had found no reference to cities in one text and only one reference to cities in the other (a brief allusion to the Harris-Todaro model). He is bewildered about why regional science has not had a greater impact. Perhaps what is needed, he said, is buttoned-up, tied-down, neat demonstration models that offer a bridge to mainstream economics. A great deal of intelligent work is being done in regional science but it does not deliver the kind of neatly tied models that economists prefer.