NBER WORKING PAPERS SERIES

CITIES IN SPACE: THREE SIMPLE MODELS

Paul Krugman

Working Paper No. 3607

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1991

This paper is part of NBER's research program in International Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #3607
January 1991

CITIES IN SPACE: THREE SIMPLE MODELS

ABSTRACT

Urban agglomerations arise at least in part out of the interaction between economies of scale in production and market size effects. This paper develops a simple spatial framework to develop illustrative models of the determinants of urban location, of the number and size of cities, and of the degree of urbanization. A central theme is the probable existence of multiple equilibria, and the dependence of the range of potential outcomes on a few key parameters.

Paul Krugman
Department of Economics
MIT
Room E52-383
Cambridge, MA 02139

This paper offers three variations on a simple theme in location theory. The theme is a well-known one, but the specific framework within which the ideas are presented is somewhat new, and the variations may be of some interest.

The basic theme is that urban agglomerations arise at least in part out of the interaction between economies of scale in production and market size effects. Producers subject to scale economies have an incentive to concentrate their production at a limited number of sites; in order to economize on transportation costs, they prefer production sites that are close to large markets. But markets are large precisely where large numbers of producers have chosen to site their facilities. From this circularity, two general propositions follow. First, production tends to clump together in agglomerations considerably larger than the scale of any individual producer. Second, there are typically multiple equilibria: the location of urban centers is not uniquely determined by tastes, technology, and resources.

In two recent papers (Krugman 1991a, 1991b) I have considered location models along these lines. In those models, however, space was generally treated as consisting of two or three discrete regions, with transportation costs between regions but no transport costs within regions. In this paper space is instead treated as continuous (albeit one-dimensional), putting the models more squarely in the grand tradition of location theory.

All of the models share a common framework, in which the population can be divided into two parts: an immobile population of farmers, spread evenly along a line, and a mobile manufacturing labor force, which typically ends up concentrated at one or a few points along the line.

The first model demonstrates how the interaction of cost-minimizing decisions by individual firms leads to the emergence of urban concentrations. It also shows that the location of a city is indeterminate within some range, with the width of that range a function of the model's underlying parameters.

The second model examines the factors affecting the number and size of cities. It shows how the maximum size of cities is limited by economies of scale, transportation costs, population density, and the share of the population employed in manufacturing.

Finally, the third model illustrates the possibility of multiple equilibria in the degree of urbanization itself.

It should be emphasized that the approach presented here is intended as a complement to other approaches to the issue of urbanization, not as a substitute. In particular, there is a rich literature in urban economics, both theoretical and empirical, that draws on the general concept of external economies and diseconomies to explain the emergence of an urban system; this short paper is not intended as competition for such rich and detailed analyses as those of Henderson (1988), for example. The point here is instead to show how much insight can be gained from strikingly simple models.

1. The basic framework

In order to tell the kind of location stories presented in this paper, it is necessary to have a model with four key elements. First, there must be transportation costs -- otherwise location is irrelevant. Second, there must be increasing returns in the production of at least some goods -- otherwise there is no incentive for concentration. Third, the location of demand must depend on the location of production, to generate the essential circularity. Finally, there must be at least some factors of production, such as land, that are not mobile -- otherwise the model will have the trivial outcome that everything always concentrates in one place.

In this paper I capture these elements via a somewhat ad hoc framework that sacrifices both realism and rigor in the interests of simplicity. Instead of explicitly modeling the role of land, I simply assume that there is an agricultural labor force that is exogenously distributed across space. Increasing returns at the level of firms must lead to imperfect competition, but this framework is vague about the specifics. I simply assume that firms choose locations to minimize the sum of production and transportation costs. Finally, demand at a given location is treated as simply proportional to employment at that location -- more rigorous treatment of both income and substitution effects is possible (see Krugman 1991b), but it greatly complicates the story.

We suppose, then, the following: there are two kinds of goods, agricultural and manufactures. Agricultural employment and production is spread uniformly along a line, of unit length. There are many symmetric manufactured goods. Each one is subject to a fixed cost of production F per production site, and a cost t per unit of output shipped one unit of distance. We let x represent the demand of the economy as a whole for a typical manufactured good, and treat it as fixed (i.e., we ignore income and substitution effects). We assume that manufacturing employment in any location is proportional to manufacturing production (which leaves blurry the question of the nature of fixed costs). We assume that demand for manufactured goods at any location is proportional to employment, manufacturing plus agricultural, at that location (fudging income and substitution again), with a share π of demand coming from manufacturing workers, $1-\pi$ from agricultural. Finally, we suppose that the location of production of manufactures is chosen so as to minimize the sum of production and transportation costs.

The problems with this framework are obvious. Aside from the lack of realism, the framework does not quite hang together as a piece of microeconomics -- while it basically makes sense, microfoundations and adding-up constraints are not quite respected. The point, of course, is that there is a compensating payoff in simplicity and insight. The framework should thus be seen in a "macroeconomic" spirit -- like the IS-LM model, or the Lucas supply function, it is inspired by microeconomic arguments without being securely grounded in them, to help us get insight that would be denied us if we insisted on complete rigor.

2. Model I: Urban location

We consider an economy in which the agricultural population is distributed along a line, normalized to be of unit length. We assume that economies of scale are sufficiently strong that each manufacturing producer wants to have only one production site (we will examine the conditions under which this happens in the next two sections). Let z, $0<z<1$, represent the location of a manufacturing producer along the line, with $z=0$ at the "west" end and $z=1$ at the "east" end.

As we will see in a moment, the equilibrium spatial distribution of manufacturing production will be very simple: all of it will be concentrated at a single point, which we may call a city. Let z_c be the location of this city; each firm will sell πx units of output to the residents of the city, and $(1-\pi)x$ units to the agricultural population.

What each firm does is to choose a location that minimizes transportation cost, which in turn consists of two parts. The cost of shipping goods to the agricultural sector depends only on the firm's location z. A fraction z of the farmers lie to the "left" of a firm that locates at z; they purchase $(1-\pi)xz$ from the firm; their average distance from the firm is z/2. Similarly, a fraction (1-z) of farmers lie to the "right", at an average distance (1-

z)/2, buying $(1-\pi)x(1-z)$. The total cost of shipping to farms is therefore

$$T_A = \frac{1-\pi}{2} tx[z^2+(1-z)^2] \tag{1}$$

This cost is shown in Figure 1; it is minimized for z = 0.5.

The cost of shipping goods to the city depends both on z and on the location of the city:

$$T_C = \pi tx|z-z_c| \tag{2}$$

This is a V-shaped line that touches zero at z_c.

A natural guess might be that z_c=0.5 -- i.e., that the city is located at the exact center. This situation is illustrated in Figure 1, which shows T_A, T_C, and total transport costs for a representative firm under the working assumption that manufacturing production is in fact concentrated at 0.5. Evidently, in this case z=0.5 is the location that minimizes transport costs, so firms will in fact choose to produce at that point. A city in the center of the line is therefore an equilibrium locational structure.

But it is not the only equilibrium. There is a range of potential equilibrium sites for the city. Figure 2 illustrates the point. It shows T_A, T_C, and total transport costs for a representative firm when z_c is located somewhat to the left of 0.5. In spite of the fact that this city location does not minimize the cost of selling to the rural market, from the point of view of any individual firm that takes the city's location as given locating at z_c still minimizes overall transportation costs. Intuitively, the weight of the urban market in firms' decisions is sufficiently large that moving the city a little bit to the left or the right from its collectively optimal location will "drag" the individually optimal locations of firms along with it. More technically, the concentration of mass at the city site creates a discontinuity in the derivative of transportation costs with respect to location; this "kink" causes a clustering of firms at the same location.

Not all locations for the city are necessarily equilibria: if the city were too far from the center of the line, firms might find it optimal to locate somewhat closer to the center. We can establish the range of potential city sites as follows. Consider a hypothetical city somewhere to the left of z_c=0.5 (the case to the right is symmetric). Would it pay a firm to move its plant away from the city? Clearly a location still further to the left would not be desirable, since the costs of servicing both the rural and urban markets would be higher. A move toward the center, however, would reduce the costs of shipping to rural customers. The city site will only be an equilibrium if the rise in costs of shipping to the city as one moves to the right is at least as large as the fall in rural costs. That is, we must have

$$\frac{dT_A}{dz} + \left(\frac{dT_C}{dz}\right)_+ \geq 0 \qquad (3)$$

at $z = z_c$, where the second term represents the derivative as we move to the right. Figure 3 illustrates how this sets a range of potential city sites. The leftmost potential city location, z_{min}, is where the slope of T_A equals negative dT_C/dz, and there is a corresponding z_{max}.

Algebraically, we note that

$$\frac{d(T_C+T_A)}{dz} = (1-\pi)\,tx(2z-1)-\pi\,tx \tag{4}$$

if $z < z_c$, and that

$$\frac{d(T_C+T_A)}{dz} = (1-\pi)\,tx(2z-1)+\pi\,tx \tag{5}$$

if $z > z_c$. Consider the case $z_c < 0.5$: such a city is an equilibrium iff

$$(1-\pi)\,tx(2z_C-1)+\pi\,tx \geq 0 \tag{6}$$

implying

$$z_C \geq \frac{1-2\pi}{2(1-\pi)} \tag{7}$$

Employing the same reasoning for cities to the right of center, we find that the potential range of city sites is

$$\frac{1-2\pi}{2(1-\pi} \leq z_C \leq \frac{1}{2(1-\pi)} \tag{8}$$

The width of this range is $\pi/(1-\pi)$ -- i.e., the range of indeterminacy in city location depends on the share of demand that is generated by the city itself. Evidently a sufficiently high share of urban demand in total demand, in this case $\pi>0.5$, allows any location to be an equilibrium.

This model, then, allows us to see in a very simple way why production ends up concentrated in an urban center, and why the location of that urban center is to at least some degree indeterminate. A weakness of the model, however, is that it makes a strong assumption about scale economies, namely that they are strong enough that each manufacturer always wants to have only a single production site. The payoff to this strong assumption is obvious, but the cost is that it does not allow us to examine the role of economic and technological parameters in determining the number and size of cities. The remainder of the paper is concerned with remedying this oversight.

2. Model II: The number and size of cities

Neither the number nor the size of cities is, in general, determinate in models of the kind considered here. Indeed, the next section will suggest that there may even be multiple equilibria in the degree of urbanization itself. Nonetheless, the range of possible and likely outcomes is surely a function of such parameters as economies of scale and transportation costs. We would like to have at least a partial model of that function.

Here I take the approach of asking the determinants of the minimum possible number (and hence maximum possible size) of cities. There is no particular reason to expect that this minimum number will actually be the outcome of the dynamic evolution of a system of cities. Nonetheless, by studying this case we get at least some insight into what might happen.

Consider, then, the same basic setup as in the previous section, with two modifications. First, we suppose that the unit line along which agricultural population is spread is in fact the circumference of a circle, so that there are no end points. Second, we no longer assume that there is only a single city. Instead, we will look at potential equilibria in which there are N cities, symmetrically located around the circle. Thus each city is a distance 1/N from its neighbors on either side, and producers in each city sell both to the local urban consumers and to a market area that stretches half way to each neighbor.

Our question is now the following: how small can N be and still be an equilibrium?

An urban system of the form described will not be an equilibrium if producers find it in their interest to move away from the assumed production sites. If N is too small, it will pay firms to establish additional plants outside the existing urban centers, in which case the system is not an equilibrium. (If N is too large, the opposite would occur: producers would abandon some cities in order to consolidate production. This sounds as if it should be possible to establish an upper as well as a lower bound on N. The upper bound case presents technical difficulties, however, and is not pursued in this paper).

A production facility in an urban center incurs transport costs to serve the surrounding rural market area. If x is the total sales of a typical manufactured good, the sales of the plant in a given city are x/N, of which $\pi x/N$ are sales to urban consumers and $(1-\pi)x/N$ are sales to the rural market. The average distance of rural consumers from the nearest city is 1/4N. Thus the transport costs of serving a rural market are

These transport costs can be reduced by establishing

$$T_A = \frac{1-\pi}{4N^2}xt \qquad (9)$$

additional production locations. The optimal site for such a location is halfway between existing locations, which will cut transportation costs in half. On the other hand, in order to establish such a production site a firm must incur a fixed cost F.

Clearly, then, the lower bound on the number of cities is set by the requirement that the fixed cost of establishing a new plant be at least as large as the savings in transportation cost. That is, N must be sufficiently large so that

$$F \geq \frac{1-\pi}{8N^2}xt \qquad (10)$$

or, equivalently,

$$N \geq \sqrt{\frac{1-\pi}{8F}xt} \qquad (11)$$

The inequality (11) suggests several sensible things about the forces limiting the size of cities. A system with a few, large cities can emerge only if F is large (strong economies of scale), π is large (a large urban population), and t is small (low transportation costs). Geographers analyzing the rise of large cities in late 19th century America -- notably Pred (1966) -- have stressed precisely these factors. This approach offers a formal justification.

We might also note some implications about the role of population density and overall urbanization. An increase in the size of the overall population would mean increased total sales of each good, that is, higher x. We note that the minimum number of cities would rise, but only as the square root of the rise in population. Thus a more densely populated country would in general support more but also bigger cities. At the same time, an increase in the urban share of the population, other things equal, would actually lead to <u>fewer</u> cities (or more precisely, would make it possible to concentrate population in fewer cities). The reason is that the greater concentration of mass in the cities would reduce the temptation for producers to move out to serve rural markets. To illustrate the point, imagine that in some country π were to rise from 0.2 to 0.8, while the overall population remained constant. Then the number of cities could fall in half, implying an eightfold rise in the population of a typical city.

This analysis is, however, only suggestive, because we are only describing possible equilibria. There is no guarantee that these would be the equilibria that would emerge. There will normally be a range of possible city systems. And with a plausible modification of the basic framework, we may argue that the degree of urbanization itself is subject to multiple equilibria.

3. Model III: The degree of urbanization

We return to the setup of Model I, where the unit line has two ends instead of being a circle, and in which there will be a maximum of one city. Now, however, we introduce the possibility that there will be no city at all.

To do this, we follow a route already explored in the recent work of Murphy, Shleifer, and Vishny (1989). There are assumed to be two technologies for producing manufactured goods: a "traditional" technique that produces goods under constant returns at a unit cost c_1, and a "modern" technique with a marginal cost c_2 lower than c_1, but that involves a fixed cost F per production site.

We can immediately see that there are two qualitatively different kinds of equilibrium possible in this model. In one equilibrium, manufactures are produced using the traditional technology; production of manufactures is geographically dispersed, and no transportation costs are incurred. In the other type equilibrium, production of manufactures is concentrated in a city, and transported to rural consumers.

Serving the market via traditional production incurs total costs of c_1x per manufactured good. Serving the market via modern production involves production costs of $F + c_2x$, plus transportation costs.

But transportation costs themselves depend on the location of population. If manufacturing is dispersed, an optimally located modern plant will be a distance of 1/4 from its average consumer, and will thus incur transport costs $tx/4$. On the other hand, if all manufacturing were concentrated at $z=0.5$, an urban plant located at the same point could serve a fraction π of consumers at zero transport cost, and incur transport costs of only $(1-\pi)tx/4$.

Traditional manufacturing, then, is an equilibrium as long as it is not cheaper to concentrate production of an individual manufactured good, i.e.,

$$F+c_2x+\frac{tx}{4} > c_1x \tag{12}$$

Modern manufacturing is an equilibrium as long as it is not cheaper to disperse production, i.e.,

$$F+c_2x+\frac{1-\pi}{4}tx < c_1x \tag{13}$$

These criteria are not mutually exclusive. There will be multiple equilibria in the degree of urbanization as long as

$$F+c_2x+\frac{tx}{4} > c_1x > F+c_2x+(1-\pi)\frac{tx}{4} \tag{14}$$

This story bears an obvious resemblance to the Big Push story of Rosenstein-Rodan, as formalized by Murphy, Shleifer, and Vishny. Here, however, the key element is spatial: industrialization does not create a larger market, but rather a more compact one, and that is what makes it self-sustaining.

5. Conclusions

This paper has offered a minimalist approach to the formation of urban centers in a spatial framework. The approach is neither realistic nor rigorous, merely suggestive. It is intended as a complement rather than a substitute for other lines of inquiry. What it shows, however, is that some interesting stories arise out of even very simplistic models. Once one introduces the basic circularity in which location of production and location of demand are interdependent, a series of results emerge. Agglomeration -- the formation of cities -- follows immediately, as does some indeterminacy about urban location. The number of cities is bounded in an economicly meaningful way by underlying parameters. And under some conditions fundamental aspects of the economy, such as the degree of urbanization itself, can be shown to be subject to multiple equilibria.

There have been numerous appeals over the history of economic thought to take the spatial dimension seriously. It is disputable why these calls have not had more effect; but a likely reason is the perception that spatial modelling involves a high ratio of effort to insight. I hope that this paper helps demonstrate that this need not be the case.

REFERENCES

Henderson, J.V. (1988), Urban Development: Theory, Fact, and Illusion, New York: Oxford.

Krugman, P. (1991a), "History and industry location: the case of the US manufacturing belt", American Economic Review, May.

Krugman, P. (1991b), "Increasing returns and economic geography", Journal of Political Economy, forthcoming, June.

Murphy, K., Shleifer, A. and Vishny, R. (1989), "Industrialization and the big push", Journal of Political Economy, 97: 1003-1026.

Pred, A. (1966), The Spatial Dynamics of US Urban-Industrial Growth, 1800-1914, Cambridge: MIT Press.

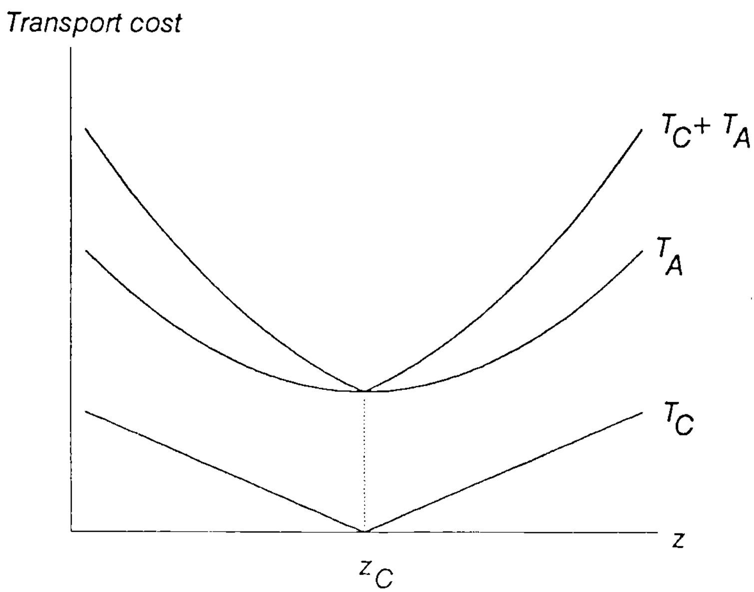

Figure 1

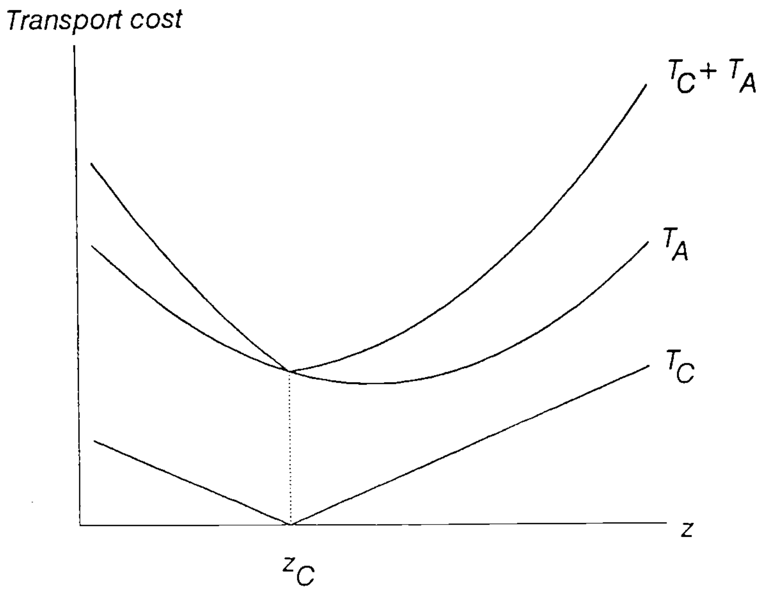

Figure 2

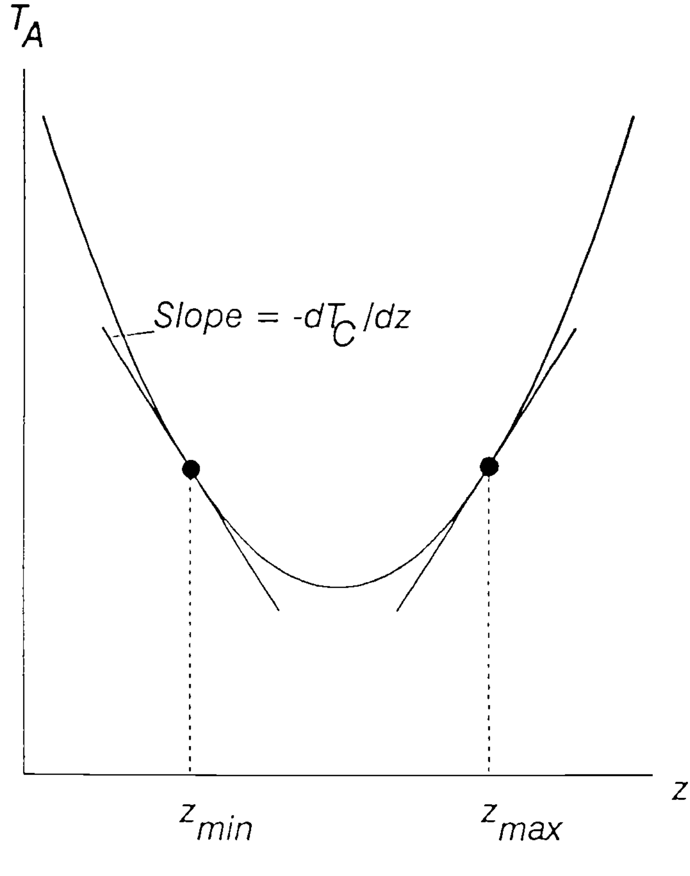

Figure 3