NBER WORKING PAPERS SERIES

FIRST NATURE, SECOND NATURE, AND METROPOLITAN LOCATION

Paul R. Krugman

Working Paper No. 3740

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 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 #3740
June 1991

FIRST NATURE, SECOND NATURE, AND METROPOLITAN LOCATION

ABSTRACT

This paper develops models of spatial equilibrium in which a central "metropolis" emerges to supply manufactured goods to an agricultural hinterland. The location of the metropolis is not fully determined by the location of resources: as long as it is not too far from the geographical center of the region, the concentration of economic mass at the metropolis makes it the optimal location for manufacturing firms, and is thus self-justifying. The approach in this paper therefore helps explain the role of historical accident and self-fulfilling expectations in metropolitan location.

Paul R. Krugman
Department of Economics
M.I.T
50 Memorial Drive
Cambridge, MA 02139
and
NBER

In his justly acclaimed recent book Nature's Metropolis: Chicago and the Great West, the historian William Cronon documents the extraordinary 19th century rise of Chicago as the central city of the American heartland. As Cronon points out, what made this rise particularly remarkable was the absence of any distinctive natural advantages of Chicago's site. The city stood on a flat plain; the river that ran through the city was barely navigable; the city's lakeside harbor was inadequate and tended to silt up. Whatever natural advantages the site did have proved transitory. Initially Chicago seemed the natural terminus of a canal linking the watershed of the Mississippi with the Great Lakes, but when a canal was finally built it had only a few years of economic importance before being overshadowed by the railroads. Chicago's harbor on the Great Lakes was not unique, and in any case lake transportation became relatively unimportant by the 1870s as compared with rail links. Yet once Chicago had become established as a central market, as a focal point for transportation and commerce, its strength fed on itself. As Cronon puts it, the advantages that "first nature" failed to provide the city were more than made up for by the self-reinforcing advantages of "second nature": the concentration of population and production in Chicago, and the city's role as a transportation hub, provided the incentive for still more concentration of production there, and caused all roads to lead to Chicago.

In Cronon's interpretation, then, the rise of Chicago was a striking example of what David (1985) has called "path dependence": historical accident, which led people to expect a central role for

Chicago, led them into decisions that justified that expectation.

The purpose of this paper is to develop an approach to modeling urban location that makes sense, albeit in a highly simplified way, of the kind of history-driven process of metropolitan growth that Cronon describes. The paper develops models of spatial equilibrium in which a central "metropolis" emerges to supply manufactured goods to an agricultural hinterland. This metropolis could be at the agricultural region's geographical center. However, it need not be, because the metropolis is part of its own market (and also supplies part of what its own residents consume). Because of the feedback from the location of the metropolis to the geography of demand and supply, there is a range of potential metropolitan sites. As Cronon would put it, the "second nature" that the existence of the metropolis creates drags the optimal location of firms with it.

In addition to being of some realistic interest, particularly as an aid in thinking about urban history, this approach contains some interesting echoes of a number of intellectual traditions. There are aspects of both Hotelling and von Thünen to the model; there is some common ground as well with the central place theory of Lösch and Christaller; and the models also provide a rigorous justification for the commonly used geographical concept of "market potential".

The paper is in four parts. The first part sets out a basic modeling approach. The second analyzes spatial equilibrium in a one-dimensional region, in which rural population is spread along a line. The third part analyzes a two-dimensional case, in which population is distributed across a circular plain. The final section suggest some directions for extension.

1. <u>A modeling approach</u>

The intuition behind the approach in this paper is simple: firms that have an incentive to concentrate production at a limited number of locations prefer, other things equal, to choose locations with good access to markets; but access to markets will be good precisely where large number of firms choose to locate. This positive feedback loop drives the formation of urban centers; it also implies that the location of such centers is not wholly determined by the underlying natural geography, but can also be influenced by history and self-fulfilling expectations.

In order to capture this intuition, a formal model must have three features. First, there must be some costs of transportation, so that location matters. Second, there must be some immobile factors of production, providing some form of "first nature" that constrains the possible spatial structure of the economy. Finally, there must be economies of scale in the production of at least some goods, so that there is an incentive for concentration.

The framework that will be used here is based on the two-location model in Krugman (1991), which in turn relies heavily on the monopolistic competition model of Dixit and Stiglitz (1977). We envision an economy with two sectors: constant-returns agriculture and increasing-returns manufacturing. Agricultural output is produced by geographically immobile factors, which are spread across space. Manufactures -- which come in many differentiated varieties -- are produced by mobile factors, which move to wherever they can achieve the highest return. Cities emerge when manufacturing firms clump together to be near the markets they provide for one another.

Specifically, we begin by assuming that everyone in the economy shares a utility function into which both the agricultural good and a manufactures aggregate enter:

$$U = C_M^{\mu} C_A^{1-\mu} \tag{1}$$

where μ is the share of manufactured goods in expenditure.

The manufactures aggregate in turn is a CES function of a large number of potential varieties, not all of which will actually be available:

$$C_M = \left[\sum_i c_{Mi}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{2}$$

Agricultural goods are produced by a sector-specific factor, agricultural labor. (Ideally, we would introduce land as an explicit additional factor of production. For simplicity, however, we do not; the role that should be played by land is proxied by making agricultural labor immobile). They are produced under constant returns; without loss of generality we assume that the unit labor requirement is one. Manufactured goods are similarly produced by a specific factor, manufacturing labor. In the manufacturing sector, however, there are increasing returns; we introduce these by assuming that for any variety that is actually produced, there is a fixed labor input required independent of the volume of output:

$$L_{Mi} = \alpha + \beta q_{Mi} \quad (3)$$

Both kinds of labor will be assumed to be fully employed. Thus

$$L_M = \sum_i L_{Mi} \quad (4)$$

and

$$L_A = C_A \quad (5)$$

Aside from the sector-specificity of the two kinds of labor, this is essentially the Dixit-Stiglitz model. The main innovation here is to make the model spatial. This is done through two assumptions. First, agricultural labor will be assumed to be distributed across space: in the next section we will assume that it is evenly spread along a line, in the section following that it is evenly spread across a disk. Second, we assume that although agricultural goods can be costlessly transported (an assumption made purely for analytic convenience), manufactured goods are costly to transport. Specifically, we follow Samuelson's "iceberg" assumption, under which goods "melt" in transit, so that transportation costs are in effect incurred in the same goods that are shipped. The proportional rate of melting is assumed to be constant per unit of distance, implying that if a single unit of a manufactured good is shipped a distance D, the quantity that arrives is only $e^{-\tau D}$, where τ is the transport cost.

This setup is incomplete unless we specify a particular spatial structure. We may, however, quickly review several familiar features of this kind of model, well-known both from the original Dixit-Stiglitz paper and from the extensive derivative literature in the field of international trade.

First, we note that given the absence of any transportation costs for agricultural goods, all agricultural workers will receive the same nominal (although not real) wage.

Second, we note that manufacturing will have a monopolistically competitive market structure, in which the price of each manufactured good at the factory gate will be a constant proportional markup on the wage rate, and in which all profits will be competed away by entry.

Third, we note that a fraction μ of total expenditure will fall on manufactured goods (including those that "melt" in transit); since profits are competed away, manufacturing workers will receive a share μ of total income, agricultural workers $1-\mu$.

Finally, we note that the elasticity of substitution between any two products is σ.

The minimalism of this framework is apparent. Yet it is sufficient to generate some interesting insights into spatial equilibrium.

2. A one-dimensional model

We begin by considering a one-dimensional region. We assume that agricultural workers are distributed evenly along a line of unit length. What we will show is that if transport costs are not too high, there is an equilibrium in which all manufacturing is concentrated at a single point along that line. This "metropolis" could be at the region's center, but it need not be: in general there is a range of potential locations, whose width depends in an economically meaningful way on the model's parameters.

The method we will use is to posit an initial situation in which all manufacturing workers are concentrated at a single location, then ask whether a small group of workers will find it advantageous to move to any other location. If not, then concentration of manufacturing at that location is indeed an equilibrium.

Suppose, then, that all manufacturing is concentrated at the location x_C along the unit interval. We need to ask whether it to the advantage of a small group of these workers to relocate to some other site x_A. To do this, we need to calculate the real wage that the relocated workers could earn at x_A relative to that which they can earn at x_C.

Let w be the ratio of the <u>nominal</u> wage rate at x_A to that at x_C. Given the monopolistically competitive market structure, the ratio of the f.o.b. price of a good manufactured at x_A to that of a typical good manufactured at x_C will also be w. The price ratio to a consumer at location x will reflect both this f.o.b. price ratio and transport costs, which depend on the consumer's relative distance from x_A and x_C. Let p_x be this relative price to a consumer at location x: given our assumption about transport costs, it is simply

$$p_x = we^{\tau(|x_A - x| - |x_C - x|)} \tag{6}$$

Next consider the ratio of sales of a product manufactured at x_A to that of a typical good manufactured at x_C. Given the elasticity of substitution of σ, the ratio of consumption by a consumer at location x is

$$c = p_x^{-\sigma} \tag{7}$$

The ratio of value of sales to the consumer at x, however, is less sensitive to the price, because volume effects are offset by valuation effects; thus we have for the value ratio

$$s = p_x^{-(\sigma-1)} \tag{8}$$

To calculate the overall ratio of sales of a product at x_A to that of a typical product from the metropolis, we note that manufacturing workers, who account for a fraction μ of demand, are all concentrated at x_C; while agricultural workers, who account for the rest, are spread evenly along the unit interval. This implies that the overall sales ratio is

$$S = \mu\left[we^{-\tau|x_C - x_A|}\right]^{\sigma-1} + (1-\mu)\int_0^1\left[we^{\tau(|x_C - x| - |x_A - x|)}\right]^{\sigma-1}dx \tag{9}$$

or, rearranging,

$$S = w^{-(\sigma-1)}\left[\mu\left[e^{-\tau|x_C-x_A|}\right]^{\sigma-1} + (1-\mu)\int_0^1\left[e^{\tau(|x_C-x|-|x_A-x|)}\right]^{\sigma-1}dx\right] \tag{10}$$

Equation (10) determines relative sales as a function of the relative wage rate. We can, however, turn it around to determine the relative wage rate by invoking the zero-profit condition. First, note that by assumption all firms at x_C are earning zero profits, with their operating surpluses just covering their fixed costs. A firm at x_A must do the same. But the operating surplus of a firm in the Dixit-Stiglitz model is proportional to its sales, while the fixed costs are incurred in manufacturing labor, which at x_A receives a relative wage w. It follows, then, that if there are to be zero profits we must have

$$S = w \tag{11}$$

Putting (10) and (11) together, we have our expression for the nominal wage rate at x_A:

$$w = \left[\mu\left[e^{-\tau|x_C-x_A|}\right]^{\sigma-1} + (1-\mu)\int_0^1\left[e^{\tau(|x_C-x|-|x_A-x|)}\right]^{\sigma-1}dx\right]^{\frac{1}{\sigma}} \tag{12}$$

To determine whether a concentration of manufactures at x_C is an equilibrium, however, we need to compare not the relative nominal wage but the relative real wage. The difference between the two comes from the fact that manufactured goods produced at x_C are part of workers' consumption basket, with a weight μ. Taking this into account, we note that the relative real wage rate is

$$\omega = we^{-\mu\tau|x_A-x_C|} \tag{13}$$

Equations (12) and (13) determine the real wage rate that a small group of workers would receive if they were to locate at a site x_A when all other manufacturing workers are concentrated at x_C. One way to think about (13) is that it is a kind of index of "market potential" of alternative sites. Such indices are widely used by geographers as a way to help think about plant location. In the standard calculation of market potential, the potential of a site is measured by a weighted sum of the purchasing power of all available markets, with the weights inversely proportional to distance from that site. In this case the weights are derived from an explicit model of profit maximization, and there is also a "forward linkage" due to the role of metropolitan products in consumption. At a broad level, however, the idea is similar -- a similarity that will become even more evident when we turn to the two-dimensional model of the next section.

Returning to (13), we immediately note that workers will choose to locate at the value of x_A that maximizes their real wage. We thus have a simple definition of an equilibrium metropolitan location. A metropolis at x_C is an equilibrium if, given that location, the maximum of (13) is also at x_C.

Figure 1 illustrates such an equilibrium with the most obvious case, a metropolis located at $x_C = 0.5$, at the exact center of the line. For the purposes of this example we set $\sigma = 4$, $\tau = .5$, $\mu = .2$.[1] We see that the real wage is indeed maximized at $x_A = 0.5$, so

[1]For the purposes of the calculation the agricultural labor force, instead of being continuously spread along the line, was placed at 11 discrete locations (that is, at intervals of 0.1).

that this is an equilibrium metropolitan location.

It might seem that a metropolis at the center of the region is always a possible equilibrium. Unfortunately, this is not quite right, because there may exist no equilibrium with only a single metropolis. If transportation costs are high enough, then even if one posits a concentration of all manufacturing at the center, workers will find it advantageous to move away from the center to get better access to the rural market. Figure 2 illustrates this point, by calculating market potential with a hypothetical metropolis at x_c=0.5, but with a transport cost of τ = 1.5. The central metropolis is not an equilibrium.

Presumably in the case of high transport costs equilibrium must take the form of several metropolitan centers. The lower are transport costs, the fewer and larger the metropolises can be. In this paper, however, I will not try to pursue that line of inquiry. Instead, we will simply assume that transport costs are sufficiently low that equilibria with a single urban center do exist, and restrict ourselves to examining that center's potential location.[2]

Let us now return to Figure 1. Notice that in that figure the market potential line has a "kink" at the metropolitan location. This kink reflects the concentration of economic mass at the

---

[2]One might expect that the ability to sustain a single metropolis would depend on the geographic extent of the region as well as one transportation costs. In this model, however, these are essentially the same thing. The size of the region (the length of the line) is normalized at unity; lengthening the line and increasing τ have exactly the same effects.

metropolis. Stepping slightly outside the model, the incentive to be at this concentration of mass makes the existence of a metropolis robust to small amounts of "noise". Suppose, for example, that some workers have a small preference to be a little left of center, while others have a small preference to be to the right. The kink in market potential at the metropolis implies that these workers will nonetheless all clump together at the same place.[3]

Returning to the model, the kink also implies that the exact center of the line is not the only possible site for the metropolis. For suppose that the metropolis lies a little bit to the left or right of the center. There will then be some incentive to move away from the metropolis toward the center. But given the kink, as long as the metropolis is not too far from the center the market potential line will still slope down in both directions. In effect, if we move the metropolis we drag the point of maximum market potential along with it.

Figure 3 illustrates this point. All parameters are the same as in Figure 1, but this time we suppose that the metropolis is located at $x_c = 0.4$, that is, somewhat left of center. Nonetheless, as we see from the figure, the point of maximum market potential is still at x_c. Thus this is also an equilibrium metropolitan

---

[3]For example, in Cronon's case of Chicago in the 19th century, the market potential map for wheat marketing would presumably have looked different from that for the slaughterhouses, which in turn would have looked different from that for lumber, and different yet again for different manufacturing industries. Yet the gravitational attraction of Chicago meant that the location of peak market potential for all of these sectors was in the same place.

location.

Not all locations, however, are necessarily suitable for a metropolis. If the hypothetical metropolis is too far from the center, it may be advantageous for workers to move away. In Figure 4, again with the same parameters, we hypothesize a metropolis at $x_c = 0.1$. In this case, workers can achieve a higher real wage at sites to the right. Thus 0.1 is <u>not</u> an equilibrium metropolitan location.

It seems apparent that there is a range of potential metropolitan locations, including the center of the region but also extending some distance to either side. We can solve analytically for this range by applying a criterion of local stability: a necessary condition for a metropolitan site to be an equilibrium if that given a hypothetical metropolis at that site, the market potential has a local maximum there. Given the symmetry of the problem, we need only consider locations to the left of center; obviously in that case a more desirable alternative site, if it exists, will lie to the right. So the defining criterion for the range of potential sites is: if we posit a metropolis at some $x_c < 0.5$, then $d\omega/dx_A$ for x_A slightly greater than x_c must be negative.

This local stability criterion simplifies the algebra massively: all the absolute value terms in (12) and (13) become unambiguously signed, and when the expression is evaluated in the vicinity of $x_A = x_c$ all of the exponential terms become unity. Thus letting R be the derivative of the relative real wage with respect to x_A, when x_A is just slightly greater than x_c, we have

$$R = \frac{\tau}{\sigma}[-\mu-\mu(\sigma-1)+(1-\mu)(\sigma-1)(1-2x_c)] \qquad (14)$$

Bearing in mind that for an equilibrium metropolitan location we must have $R<0$, we note that (14) contains two negative terms and one positive. Broadly, the negative and positive effects may be seen respectively as the "centripetal" forces tending to hold a metropolis together and the "centrifugal" forces tending to pull it apart -- or, as Myrdal (1957) put it, "backwash" and "spread" effects. More specifically, by examining (12) and (13) one can place direct economic interpretations on the three terms. The first term in (14) comes from the role of metropolitan goods in consumption, introduced in (13); it therefore represents, in Hirschman's (1958) terms, a "forward linkage". The second term comes from the role of the metropolis as a source of demand for manufactured goods; it therefore represents a Hirschman-type "backward linkage". The third term, finally, represents the incentive to move away from the metropolis to be closer to the rural market.

The criterion $R<0$ defines the range of potential metropolitan sites. For a metropolis to the left of center, we have

$$1-2x_c < \frac{\mu}{1-\mu}\frac{2\sigma-1}{\sigma-1} \qquad (15)$$

The larger is the right hand side of (15), the wider the range of potential sites; if the right hand side exceeds 1, any point on the line can accommodate a metropolis.

The width of the range depends in an economically sensible way on the parameters. It is increasing in μ, the share of manufactures in spending: this makes sense, because it is the importance of manufactures in consumption and of manufacturing workers' income in demand that gives rise to the forward and backward linkages that attract production to the metropolis. The range is decreasing in σ: the less monopoly power firms have (and hence also the smaller the degree of increasing returns in equilibrium), the less powerful the forces for metropolitan concentration.

For the case of a one-dimensional region, then, we have been able to show that if transport costs are not too high there will be an equilibrium in which all manufacturing is concentrated in a single metropolis; that this equilibrium is not unique, because there is a range of potential metropolitan sites; and that the width of this range depends in an economically meaningful way on the model's parameters.

The tradition of economic geography, however, contains a strong cartographic component, in which one tries to relate theoretical concepts to actual maps. And maps, unfortunately, have two dimensions rather than one. So it is natural to ask whether the insights gained from this one-dimensional model still apply in a two-dimensional world. In the next section we show that they do.

3. A two-dimensional model

The two-dimensional analogue of a line is a disk. We now imagine, then, a region in which the agricultural labor force is spread evenly across a disk; we normalize the radius of the disk to unity, and let the center be x=y=0. All the other assumptions are the same as in the one-dimensional model.

The analogy with the one-dimensional model immediately suggests what we are going to find. Provided that transportation costs are not too high, there will be an equilibrium with all manufacturing concentrated at a central metropolis. This metropolis could be in the geometric center of the disk, but it need not be: because changing the location of the metropolis itself changes the map of market potential, there will be a range of potential metropolitan locations. In the one-dimensional case this range was a central portion of the line segment; in the two-dimensional case it will be a central disk within the regional disk. Figure 5 illustrates schematically what we will find: the larger disk represents the region as a whole, the shaded interior disk the set of potential metropolitan locations.

The analytics of this model are very similar to the one-dimensional model, but complicated by the need to measure distances in two dimensions. Let us posit a metropolis at x_C, y_C; without loss of generality (since one can always rotate the disk), assume $y_C=0$. We want to consider the attractiveness of an alternate location x_A, y_A. To calculate this, we need to know three distances. Let D_{AC} be the distance between the alternative location and the metropolis; we have

$$D_{AC} = \sqrt{(x_A - x_C)^2 + y_A^2} \tag{16}$$

Let $D_C(x,y)$ be the distance from the metropolis to some other location (x,y):

$$D_C(x,y) = \sqrt{(x_C - x)^2 + y^2} \tag{17}$$

And let $D_A(x,y)$ be the distance from the alternate location to (x,y):

$$D_A(x,y) = \sqrt{(x - x_A)^2 + (y - y_A)^2} \tag{18}$$

By analogy with the one-dimensional case, the sales of a firm at the alternate location relative to those of one in the metropolis are

$$S = \mu\left[we^{-\tau D_{AC}}\right]^{\sigma-1} + \frac{1-\mu}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\left[we^{\tau(D_I(x,y)-D_A(x,y))}\right]^{\sigma-1} dy\, dx \tag{19}$$

The zero-profit condition once again requires that S=w. Thus we have

$$w = \left[\mu e^{-\tau(\sigma-1)D_{AC}} + \frac{1-\mu}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} e^{(\sigma-1)\tau(D_C(x,y)-D_A(x,y))} dy\, dx\right]^{\frac{1}{\sigma}} \tag{20}$$

And the relative real wage at the alternate location is

$$\omega = we^{-\mu\tau D_{AC}} \tag{21}$$

Equations (16)-(21) can be used to construct a market potential map, given the location of a hypothetical metropolis. And as in the one-dimensional case, if the location of the posited metropolis is also the point of greatest market potential, then that is an equilibrium location.

Figure 6 shows the case of a metropolis located at the geometric center of the region, under the assumptions $\tau = 0.5$, $\mu = 0.3$, $\sigma = 4$.[4] (Since the diagram is symmetric, only the upper half of the disk is shown). The market potential is represented by contour lines, loci of equal real wages that are .95, .9, and .8 of the real wage at the metropolis; in this case, of course, the contour lines are simply circles around the metropolitan center. Since the central bulls-eye is also the point of peak market potential, a central metropolis is an equilibrium.

But a somewhat off-center metropolis may also be an equilibrium. Figure 7 shows the market potential map generated by a metropolis located at $x_c=0.3$. The contour lines are dragged off to the east by the metropolis's economic mass; the metropolis is still the best location, and thus this proposed geography is also an equilibrium.

As in the one-dimensional model, we can analyze the determinants of the range of potential metropolitan sites by examining local stability. Suppose we posit a metropolis at $(x_c, 0)$, with $x_c>0$. If there is a better location for a small group of

[4] As in the one-dimensional case, this example was constructed using a discrete distribution of rural population at 317 points, located at grid intervals of 0.1.

workers, it will be toward the center; thus we need only consider alternative locations with $x_A<_C$ and $y_A=0$. For local stability, the derivative of ω with respect to x_A must be positive for x_A slightly less than x_C. Letting L represent this derivative, we have

$$L = \frac{\tau}{\sigma}\left[\mu(2\sigma-1)+\frac{1-\mu}{\pi}(\sigma-1)\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}(x-x_A)\,[(x_A-x)^2+y^2]^{-1/2}dy\,dx\right] \quad (22)$$

As in the one-dimensional model, this local derivative is simpler than the global equation, but unfortunately not quite as much so. There are two terms inside the brackets. The first term captures the forward and backward linkage effects, and is always positive. The second term is always negative: one can see this by thinking of the term as representing the weighted sum of a series of values of $x-x_A$. Now compare the weight on each positive value $x-x_A$ with that on $-(x-x_A)$; in each case the weight on the negative value is larger. So the second term in (23) captures the centrifugal forces pulling the metropolis apart.

The further the metropolis is from the geometric center of the disk, the stronger are these centrifugal forces. Differentiating with respect to $x_A=x_C$, we find

$$\frac{dL}{dx_C} = -\frac{\tau}{\sigma}\frac{1-\mu}{\pi}(\sigma-1)\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{y^2}{(x-x_C)^2+y^2}\sqrt{(x-x_C)^2+y^2}\,dy\,dx < 0 \quad (23)$$

Thus there may, as suggested in Figure 5, be a maximum distance of the metropolis from the center.

This maximum distance, and thus the area of potential metropolitan sites, depends on μ and σ. In particular, it is straightforward to show that $dL/d\mu>0$, which implies that the radius of the range of potential metropolitan sites is increasing in μ. That is, the larger the share of income spent on manufactures, and hence the stronger the positive feedback of actual to optimal manufacturing location, the less "first nature" determines where manufacturing takes place.

4. Limitations of the analysis

The models developed in this paper offer a kind of "gravitational" analysis of the existence of metropolitan centers. Figure 7 in particular suggests the following metaphor: the concentration of economic mass at the metropolis bends economic space around itself, and it is precisely this curvature of the economic space that sustains the metropolitan concentration.

I would argue that this approach, in spite of the numerous special assumptions needed to yield tractable models, conveys an essentially correct view of metropolitan location. Nonetheless, the models presented here have four serious limitations.

First, the assumption that transportation of agricultural goods is costless is justified only by the (very considerable) analytical simplification it makes possible. For any realistic application of this approach to actual urban history it must be abandoned.

Second, and related, the assumption of an exogenously distributed agricultural work force is ultimately unsatisfactory. In particular, it prevents the models from accommodating von Thünen's key insight about the relationship between distance from the metropolis, land rents, and land use.

Third, the models assume that transportation cost is strictly proportional to distance. Yet in practice -- and above all in the Cronon's story of Chicago, which motivated this paper -- increasing returns to transportation, which lead to the formation of transportation hubs, play a key role in metropolitan concentration. A preliminary effort to model this is in Krugman (1990), but no effort is made to incorporate this analysis here.

Finally, throughout this paper the focus has been on the location of a single metropolis. A realistic analysis will have to take into account the emergence of a system of cities. At present there are two different approaches to modelling urban systems. Central place theory, deriving from Lösch (1940) and Christaller (1933) and widely used by geographers, is a powerful metaphor but lacks satisfactory microfoundations. Meanwhile, urban economists -- notably Henderson (1974, 1988) -- have developed models of urban systems that carefully model behavior, but which lack any spatial content. The approach in this paper is among other things an effort to build a bridge between these traditions, but it does not manage to get beyond the one-city case.

These are, then, only preliminary models. They do, however, offer a new approach that may eventually prove able to accommodate greater realism.

REFERENCES

Christaller, W. (1933), Central Places in Southern Germany, Jena: Fischer. English translation by C.W. Baskin, London: Prentice Hall, 1966.

Cronon, W. (1991), Nature's metropolis: Chicago and the Great West, New York: Norton.

David, P. (1985), "Clio and the economics of QWERTY", American Economic Review 75, 332-37.

Dixit, A. and Stiglitz, J. (1977), "Monopolistic competition and optimum product diversity", American Economic Review

Henderson, J.V. (1974), "The sizes and types of cities", American Economic Review 64, 640-56.

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

Hirschmann, A. (1958), The Strategy of Economic Development, New

Haven: Yale University Press.

Krugman, P. (1990). "The hub effect: or, threeness in interregional trade", mimeo, MIT.

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

Lösch, A. (1940), The Economics of Location, Jena: Fischer. English translation, New Haven: Yale University Press, 1954.

Myrdal, G. (1957), Economic Theory and Underdeveloped Regions, London: Duckworth.

von Thünen, J.H. (1826), The Isolated State, Hamburg: Perthes. English translation, Oxford: Pergamon, 1966.

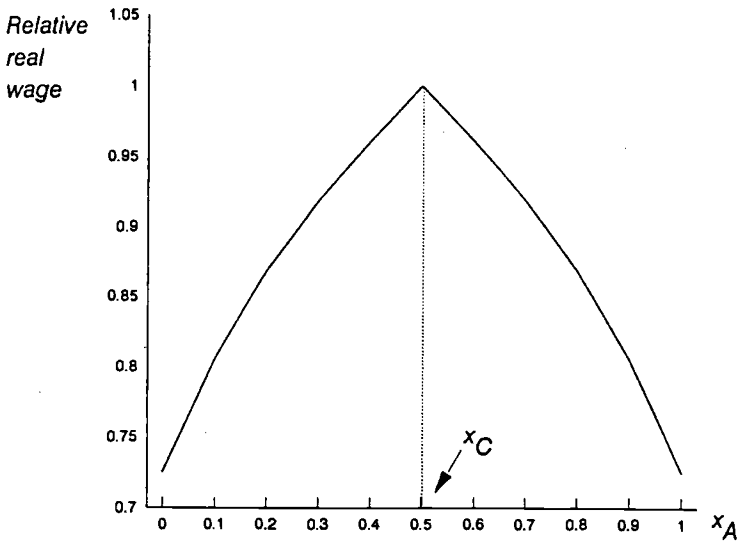

Fig. 1: An equilbrium central metropolis

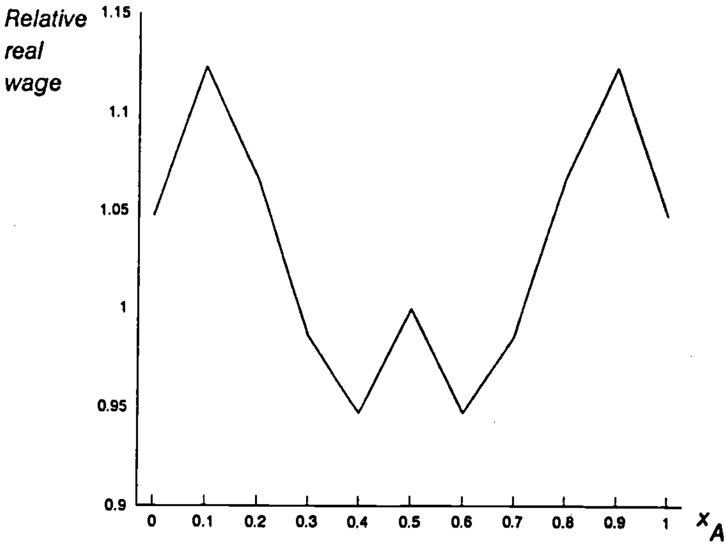

Fig.2: Absence of an equilibrium metropolis

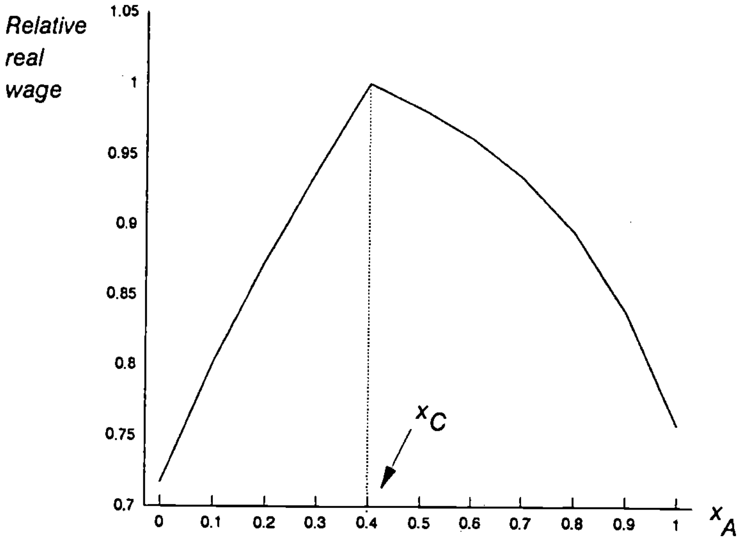

Fig. 3: An off-center metropolis

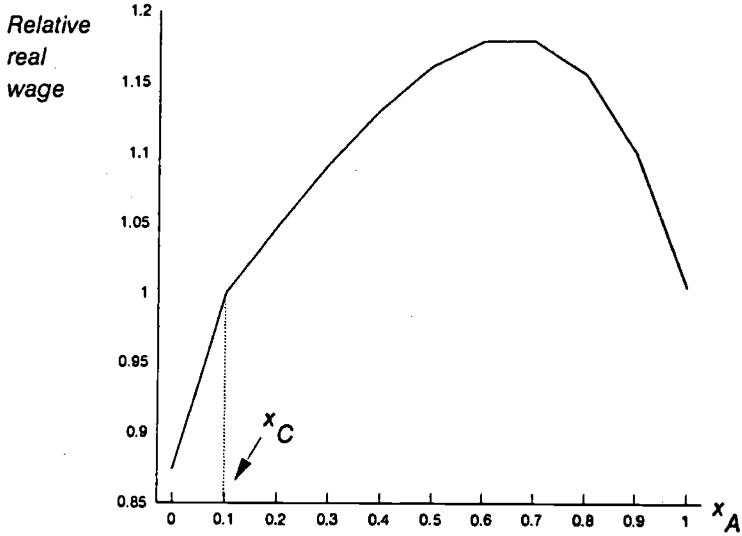

Fig.4: The metropolis cannot be too far off-center

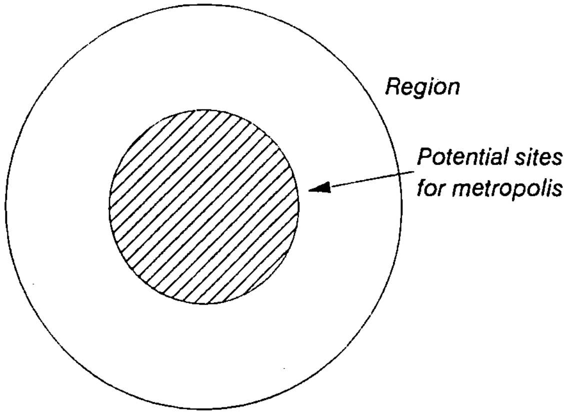

Fig.5: Metropolitan location in two dimensions

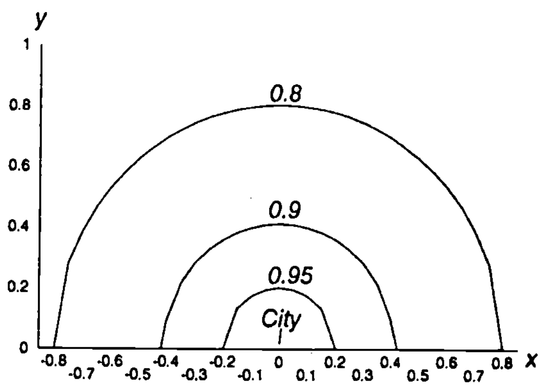

Fig. 6: Market potential with a central metropolis

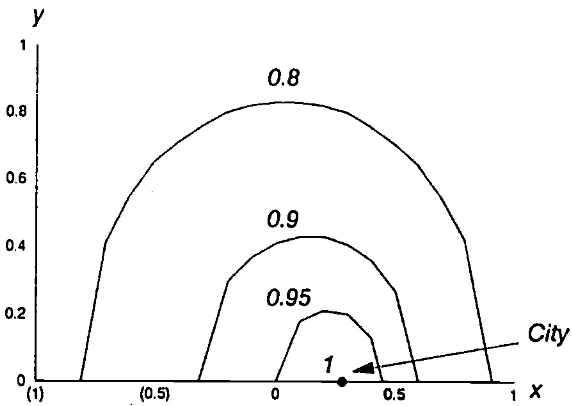

Fig. 7: Market potential with an off-center metropolis