WHAT'S NEW ABOUT THE NEW ECONOMIC GEOGRAPHY?

PAUL KRUGMAN
Massachusetts Institute of Technology

Since 1990 a new genre of research, often described as the 'new economic geography', has emerged. It differs from traditional work in economic geography mainly in adopting a modelling strategy that exploits the same technical tricks that have played such a large role in the 'new trade' and 'new growth' theories; these modelling tricks, while they preclude any claims of generality, do allow the construction of models that—unlike most traditional spatial analysis—are fully general-equilibrium and clearly derive aggregate behaviour from individual maximization. The new work is highly suggestive, particularly in indicating how historical accident can shape economic geography, and how gradual changes in underlying parameters can produce discontinuous change in spatial structure. It also serves the important purpose of placing geographical analysis squarely in the economic mainstream.

I. INTRODUCTION

The study of spatial economics—of the location of production—has a long if somewhat thin history. Von Thünen's (1826) analysis of land rent and use around an isolated city was roughly contemporaneous with Ricardo's statement of comparative advantage; the location analysis of Weber (1909), the central-place theory of Christaller (1933) and Lösch (1940), the regional science of Isard (1956), and the urban systems theory of Henderson (1974) are all old and well-established ideas.

None the less, the simple model developed in Krugman (1991) is widely regarded as having given birth to something called the 'new economic geography', and has certainly stimulated the emergence of a new wave of theorizing and (to a lesser extent) empirical work. This approach inevitably has much in common with older approaches. Nevertheless, it also has a number of distinctive features that do qualify as a new departure. The purpose of this article is to review briefly the distinctive aspects of the 'new economic geography' as a theoretical construct, describe the main lines of work within

the genre, and assess its contribution to economic theory.

II. THE NEW ECONOMIC GEOGRAPHY: DISTINCTIVE ASPECTS

(i) Modelling Strategy

Many economic activities are markedly concentrated geographically. Yet we do not all live in one big city, nor does the world economy concentrate production of each good in a single location. Obviously there is a tug of war between forces that tend to promote geographical concentration and those that tend to oppose it—between 'centripetal' and 'centrifugal' forces.

What are these forces? We may represent them in terms of a menu, of the type shown in Table 1. The menu should not be viewed as comprehensive; it is a selection of *some* forces that may be important in practice. It shows two columns: one of centripetal forces, one of centrifugal forces.

The centripetal forces listed on the left side of Table 1 are the three classic Marshallian sources of external economies. A large local market creates both 'backward linkages'—sites with good access to large markets are preferred locations for the production of goods subject to economies of scale—and 'forward linkages'—a large local market supports the local production of intermediate goods, lowering costs for downstream producers. An industrial concentration supports a thick local labour market, especially for specialized skills, so that employees find it easier to find employers and vice versa. And a local concentration of economic activity may create more or less pure external economies via information spillovers. (In Marshall's words:

'The mysteries of the trade become no mystery, but are, as it were, in the air.')

The centrifugal forces listed on the right-hand side of the table are a bit less standard, but represent a useful breakdown. Immobile factors—certainly land and natural resources, and, in an international context, people as well—militate against concentration of production, both from the supply side (some production must go to where the workers are) and from the demand side (dispersed factors create a dispersed market, and some production will have an incentive to locate close to the consumers). Concentrations of economic activity generate increased demand for local land, driving up land rents and thereby providing a disincentive for further concentration. And concentrations of activity can generate more or less pure external diseconomies such as congestion.

In the real world not only agglomeration in general, but any particular example of agglomeration, typically reflects *all* items on the menu. Why is the financial services industry concentrated in London? Partly because the sheer size of London itself makes it an attractive place to do business, and the concentration of the financial industry itself means that many clients and many ancillary services are located there; but a thick market for special skills, such as securities lawyers, and the general importance of being in the midst of the buzz are also important. Why doesn't all financial business concentrate in London? Partly because many clients are *not* there, partly because renting office space in London is expensive, and partly because dealing with the city's traffic, crime, etc. is such a nuisance.

To conduct analytical work on economic geography, however, it is necessary to cut through the complexities of the real world and focus on a more

Table 1
Forces Affecting Geographical Concentration

Centripetal forces	Centrifugal forces
Market-size effects (linkages)	Immobile factors
Thick labour markets	Land rents
Pure external economies	Pure external diseconomies

limited set of forces. In fact, the natural thing is to pick one from the first column and one from the second: to focus on the tension between just one centripetal and one centrifugal force. One way to summarize the distinctive character of the 'new economic geography' is in terms of its choice, namely the first item in each column: linkages as the force for concentration, immobile resources creating the tension necessary to keep the model interesting.

These choices are dictated less by empirical judgement than by two strategic modelling considerations. First, it is desirable to put some distance between the assumptions and the conclusions—to avoid something that looks too much like the assertion that agglomeration takes place because of agglomeration economies. This is especially true because much of the analysis we will want to undertake involves asking how a changing economic environment alters economic geography. This will be an ill-defined task if the forces producing that geography are inside a black box labelled 'external effects'. So the pure external economies and diseconomies are put on one side, in favour of forces that are more amenable to analysis.

Second, if location is the issue, it is helpful to be able to deal with models in which distance enters in a natural way. Linkage effects, which are mediated by transportation costs, are naturally tied to distance; so is access to immobile factors. On the other hand, thickness of the labour market, while it must have something to do with distance, does not lend itself quite so easily to being placed in a spatial setting. And land rents as a centrifugal force turn out to pose a tricky conceptual issue: why don't we simply get an immense conurbation, in which the suburbs of each city blend into that of the next? (I sometimes refer to this as the 'infinite Los Angeles problem'.)

The central thrust of the 'new economic geography' work to date, in short, has been driven by considerations of modelling strategy toward an approach that concentrates on the role of market-size effects in generating linkages that foster geographical concentration, on one side, and the opposing force of immobile factors working against such concentration on the other.

(ii) Modelling Principles

If one had to define the philosophical difference between the 'new' economic geography and the location theory that preceded it, it would be this: the new literature insists on models that are *general equilibrium*, and in which spatial structure emerges from *invisible-hand* processes. Both of these aspects need some discussion.

First, most traditional analysis in location theory is partial equilibrium, or not even that. Weber (1909) framed the problem of location in terms of an individual producer who took not only the locations of other producers but all prices (including his own) as given. Subsequent work has enlarged on this, notably by letting prices be endogenous, and by considering strategic interaction among location decisions of different firms. But typically the geographical distribution of demand, the location of input sources outside the industry in question, and so forth were taken as given. And adding-up constraints for the economy as a whole tended to be ignored. By contrast, the new economic geography consists of full general-equilibrium models, in which budget constraints on both money and resources are carefully specified and honoured; the geographical distributions of population, demand, and supply are endogenous, and it is, indeed, the two-way feedback between location decisions by individual agents and these distributions that is the main source of interesting stories.

Second, some of the most influential strains of thought in location theory—especially the central place theory of Christaller (1933) and Lösch (1940)—appear to be describing planning solutions rather than market outcomes. Or at least they fail to explain how the spatial structures they describe would be either created or maintained by the actions of self-interested individuals. Again by contrast, the new economic geography is all about what spatial equilibria might exist—and perhaps emerge through a dynamic process—when individuals are choosing locations to maximize their welfare given what other individuals are doing.

Of course, the failure of traditional location theory to be based on internally consistent maximization-and-equilibrium models was not the result of a moral

failing on the part of the theorists; most location theorists have understood quite well that they were engaged in some kind of partial exercise. What makes it possible for the newer models to do something different is the development of a useful menu of modelling tricks.

(iii) Modelling Tricks

The idea that there may be a circular process, in which individual producers choose locations with good access to markets and suppliers, but in which the decision of each individual producer to choose a location improves the market and/or supply access of other producers in that location, is hardly a new one. Indeed, it was the central theme of well-known (among geographers) studies by Harris (1954) and Pred (1966). Why, then, did this idea not become widely known in economics until the 1990s?

The most likely answer is that underlying all such stories is the implicit assumption that there are substantial economies of scale at the level of the plant. In the absence of such scale economies, producers would have no incentive to concentrate their activity at all: they would simply supply consumers from many local plants. And expansion of a regional market would not predictably lead to any increase in the range of goods produced within that region. Increasing returns, in other words, are central to the story.

The same may be said of spatial economics in general. Almost all of the interesting ideas in location theory rely implicitly or explicitly on the assumption that there are important economies of scale enforcing the geographic concentration of some activities. Thus Weber's (1909) analysis of the location decisions of an individual producer trying to minimize the combined costs of producing and delivering his product assumes that there can be only one production site; Christaller's (1933) suggestion that cities form a hierarchy of central places depends on the assumption that larger cities can support a wider range of activities; Lösch's (1940) famous demonstration that an efficient pattern of central places would imply hexagonal market areas assumes that there are economic activities that can be undertaken only at a limited number of sites. (The main example of a location model that does not rely on some form of scale economies, the land-rent analysis of von Thünen (1826), in effect hides the role of increasing returns by simply assuming the existence of a central city.) But unexhausted economies of scale at the level of the firm necessarily undermine perfect competition.

The key reason for the emergence of the new economic geography is therefore technical: imperfect competition is no longer regarded as impossible to model, and so stories that crucially involve unexhausted scale economies are no longer out of bounds. Indeed, the new interest in geography may be regarded as the fourth (and final?) wave of the increasing returns/imperfect competition revolution that has swept through economics over the past two decades. First came the 'new industrial organization', which created a tool-box of tractable if implausible models of imperfect competition; then the 'new trade theory', which used that tool-box to build models of international trade in the presence of increasing returns; then the 'new growth theory', which did much the same for economic growth. Like these earlier movements, the new economic geography might be best described as a 'genre': a style of economic analysis which tries to explain the spatial structure of the economy using certain technical tricks to produce models in which there are increasing returns and markets are characterized by imperfect competition. These tricks are summarized in Fujita *et al.* (forthcoming) with the slogan 'Dixit–Stiglitz, icebergs, evolution, and the computer'. Let us consider each part of that slogan in turn.

Dixit–Stiglitz
The remarkable model of monopolistic competition developed by Dixit and Stiglitz (1977) has become a workhorse in many areas of economics. In the new economic geography, it has one especially appealing feature: because it assumes a continuum of goods, it lets the modeller respect the integer nature of many location decisions—no fractional plants allowed—yet analyse the model in terms of the behaviour of continuous variables such as the share of manufacturing in a particular region. In effect, Dixit–Stiglitz lets us have our cake and cut it into arbitrarily small pieces, too.

Icebergs
This is a less familiar technical trick. In location theory, transportation costs are of the essence; yet any attempt to develop a general-equilibrium model

of economic geography would be substantially complicated by the need to model the transportation as well as the goods-producing sectors. Worse yet, transportation costs can undermine the constant-demand elasticity that is one of the crucial simplifying assumptions of the Dixit–Stiglitz model. Both problems can be sidestepped with an assumption first introduced by Paul Samuelson (1954) in international trade theory: that a fraction of any good shipped simply 'melts away' in transit, so that transport costs are in effect incurred in the good shipped. (In the new geography models, melting is usually assumed to take place at a constant rate per distance covered—e.g. 1 per cent of the cargo melts away per mile.) In terms of modelling convenience, there turns out to be a spectacular synergy between Dixit–Stiglitz market structure and 'iceberg' transport costs: not only can one avoid the need to model an additional industry, but because the transport cost between any two locations is always a constant fraction of the free-on-board price, the constant elasticity of demand is preserved.

Evolution
Interesting stories about economic geography often seem to imply multiple equilibria. Suppose, for example, that producers want to locate where other producers choose to locate; this immediately suggests some arbitrariness about where they actually end up. But which equilibrium does the economy select? New economic geography models typically assume an *ad-hoc* process of adjustment in which factors of production move gradually toward locations that offer higher current real returns. This sort of dynamic process was initially proposed apologetically, since it neglects the role of expectations. But it is possible to regard models of geography as games in which actors choose locations rather than strategies—or rather in which locations *are* strategies—in which case one is engaged not in old-fashioned static expectations analysis but rather in state-of-the-art evolutionary game theory! (To middlebrow modellers like myself, it sometimes seems that the main contribution of evolutionary game theory has been to re-legitimize those little arrows we always wanted to draw on our diagrams.)

The computer
Finally, despite the best efforts of the theorist, all but the simplest models of economic geography usually turn out to be a bit beyond the reach of paper-and-pencil analysis. As a result, the genre relies to an unusual extent on numerical examples—on the exploration of models using both static calculations and dynamic simulations.

(iv) Dynamics of Geographical Change

Suppose that, for some reason, some economic activity has a slightly larger initial concentration in one location than in another. Will that concentration be self-reinforcing, with a growing disparity between the locations, or will they tend back toward a symmetric state? The answer presumably depends on the relative strength of centripetal and centrifugal forces.

Suppose, on the other hand, that a concentration of economic activity already exists, but that some of that activity for some reason moves elsewhere. Will the activity move back, or will the concentration unravel? The answer to this question similarly depends on the relative strength of centripetal and centrifugal forces.

As these generic questions suggest, models of economic geography will typically exhibit a pattern in which the qualitative behaviour of the model changes abruptly when the quantitative balance of forces passes some critical level. That is, the models are characterized by bifurcations. And bifurcation diagrams are therefore a central analytical tool in this literature.

The typical form of these bifurcations may be illustrated by Figure 1, which summarizes the dynamics implied by the model originally introduced in Krugman (1991). That paper was, in effect, an attempt to formalize the story suggested by Harris and Pred. The model envisaged an economy consisting of two symmetric regions with two industries: immobile, perfectly competitive agriculture and mobile, imperfectly competitive (Dixit–Stiglitz) manufacturing. The backward and forward linkages in manufacturing generate centripetal forces; the pull of the immobile farmers the centrifugal force.

In that model, the difference in real wages between regions depends on the allocation of manufacturing between those regions—but the nature of that dependence in turn depends on the parameters of

Figure 1
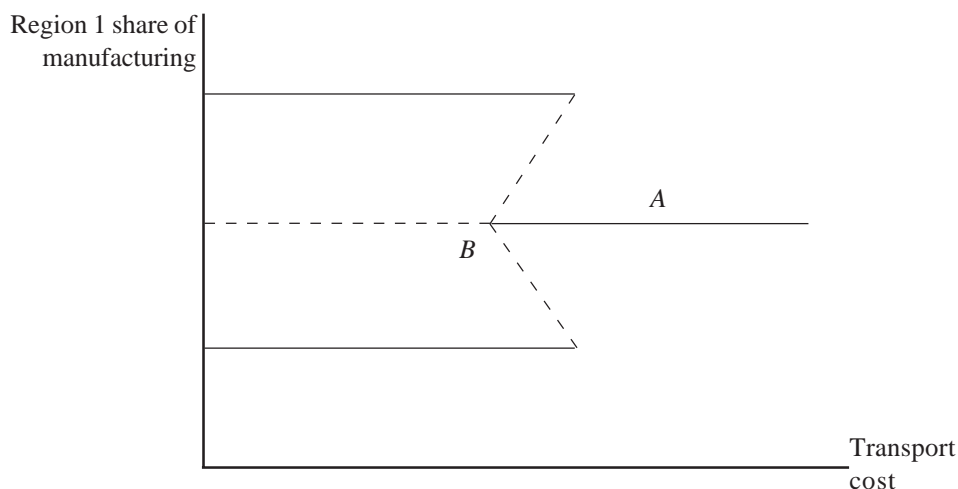
the model, including the level of transport costs. The rough intuition behind this dependence runs as follows. In the case of high transport costs, there is relatively little inter-regional trade; so the wages workers can earn depend mainly on the amount of local competition, and are thus decreasing in the number of other workers in the same region. On the other hand, when transport costs are low, a typical firm sells extensively in both regions; but since it has better access to markets if it is located in the region with the larger population, it can afford to pay higher wages—and the purchasing power of these wages is also higher because workers have better access to consumer goods. So in that case real wages are increasing in a region's population.

Since workers are assumed to move to whichever region offers the higher real wage, in the case of high transport costs there is a unique equilibrium with workers evenly divided between the regions. In the case of low transport costs there are three equilibria—an unstable equilibrium with workers evenly divided, and two with workers concentrated in either region. And it turns out (in what is essentially an artifact of the particular functional forms) that there is an intermediate range in which there are five equilibria: a locally stable equilibrium with equal division, two unstable flanking equilibria in which there is some manufacturing in one region but more in the other, and finally two in which all manufacturing is concentrated in one region.

These equilibria are summarized in Figure 1, which shows how the set of equilibria (as measured by the share of the manufacturing labour force in region 1) depend on transport costs, with solid lines indicating stable and broken lines unstable equilibria. The figure illustrates nicely one of the appealing features of the new economic geography: it easily allows one to work through interesting 'imaginary histories'. Suppose, for example, that we imagine an economy that starts with high transport costs and therefore with an even division of manufacturing between regions, a situation illustrated by the point labelled *A*. Then suppose that transport costs were gradually to fall. When the economy reached *B*, it would begin a cumulative process, in which a growing concentration of manufacturing in one region would lead to a still larger concentration of manufacturing in that region. That is, the economy would spontaneously organize itself into a core–periphery geography.

This example is, of course, only illustrative. None the less, it gives a sense of the typical dynamics of new geography models: multiple equilibria; spontaneous self-organization of the economy into some kind of spatial structure, often one with very uneven distribution of activity among locations with more or less identical natural endowments; and qualitative, often discontinuous change as a result of quantitative changes in underlying parameters.

III. MODELLING THEMES

(i) Core and Periphery

What is Figure 1 a model of? Its most natural interpretation is as a model of the spontaneous organization of a single country into a manufacturing 'core' and an agricultural 'periphery', along the lines of the division of the United States into manufacturing belt and farm belt in the middle of the 19th century, or perhaps the emergence of Italy's industrial north and Mezzogiorno some decades later.

Since its original statement in Krugman (1991), this core–periphery model has become to the new economic geography more or less what, say, the two-by-two-by-two model is in international trade: not so much a model that everyone believes, as the simplest model that illustrates all the main principles of the genre, and therefore the model one teaches first to show how this sort of thing works. This is not to say that the evolution of core–periphery patterns within nations is an unimportant question in itself. On the contrary, it is such a striking feature of modern economic history that one must view it as nearly scandalous that economists have ignored it until now. But it remains true that much, perhaps most, of the usefulness of the core–periphery model is that it opens the door to the study of a much wider range of issues.

(ii) Regions, Cities, and Nations

Broadly speaking, post-core–periphery theoretical work in the new economic geography has moved in two directions. One direction has been an effort to build links from the new genre to traditional questions of location theory. The other has been an effort to use the genre as the basis for a new, 'spatial' view of international trade.

If you want to use the new economic geography to bring the grand tradition of location theory into the economic mainstream, you are likely to be unsatisfied with a two-region model, which does not have much spatial content. What you really want is a multi-location model, or even better a model with continuous space. As long as you are willing to rely on numerical examples, however, the new economic geography style of model can easily handle any number of regions, with whatever 'geometry' of transport costs one likes.

An interesting aspect of such multi-regional models is that they provide a justification for a version of the 'market potential' function used by Harris (1954). In this case, market potential can be defined as the real wage manufacturing firms can afford to pay in any given location; it is a function of access to markets, although not quite as simple a function as Harris used. And the dynamics of the economy can, if one likes, be viewed as the co-evolution of two landscapes: a landscape of current distribution of economic activity, which determines a second landscape of market potential, which in turn determines how that first landscape changes over time.

The results of multi-region simulations depend, of course, not only on parameters but on the geometry of the economy. An interesting if artificial special case has turned out to be the 'racetrack' economy: an economy whose regions are laid out around a circle, with transportation possible only along that circle's circumference. In such an economy a uniform distribution of manufacturing is always an equilibrium, referred to in Fujita *et al.* (forthcoming) as the Flat Earth. However, the Flat Earth may be unstable: the circular logic of concentration can cause an even slightly perturbed Flat Earth spontaneously to develop one or more local concentrations of manufacturing. One might expect the resulting structures to have a high degree of arbitrariness; but simulation results show a surprising tendency toward highly regular structures, in which concentrations of economic activity are evenly spaced across the landscape.

The reasons for this regularity, it turns out, can be understood using an approach originally suggested by none other than the mathematician, Alan Turing (1952). Turing was interested in the interacting effects of chemical signals diffusing around a ring of cells, but his approach works equally well for the distribution of manufacturing around a ring of regions. It involves linearizing the model in the vicinity of the Flat Earth, then representing the initial distribution of manufacturing (or whatever) as a Fourier series. It turns out that the components of that Fourier representation are also eigenvectors of the linearized model, so that you can in effect think of

13

the components of that series as growing independently—and of the distribution becoming increasingly dominated by a fluctuation at some 'preferred frequency' that depends on the parameters of the model, but *not* on the initial conditions. And it is this preferred frequency that determines the eventual number of agglomerations that emerge. (This is only one example of the tantalizing affinity that one often finds between the new economic geography and fashionable scientific trends such as 'complexity' theory.)

An alternative way to go beyond two-region modelling is to use the new economic geography to revisit some of the questions of traditional location theory. In a series of papers, Masahisa Fujita and his students have in essence tried to take the German tradition of urban modelling that began with von Thünen and give it a true microeconomic foundation. (In Fujita's models all labour is mobile; thus the location of agriculture as well as manufacturing is endogenous.) In Fujita and Krugman (1995) a version of the original von Thünen model is offered in which the existence of a central city is no longer simply assumed: instead, manufacturing concentrates in the city because of the forward and backward linkages generated by that very concentration. Or, looking at the issue another way, the concentration of economic mass at the city generates a self-validating cusp at that point in the market potential function that determines where manufacturing locates. Agriculture is then spread around that centre, with land rents declining to zero at the agricultural frontier. Such a monocentric equilibrium, however, turns out to be sustainable only if the population is sufficiently small. Fujita and Mori (1996*a*) take the same basic model, but envision a gradually rising population which leads to the periodic emergence of new cities in a 'long, narrow' economy that gradually spreads along a line; the resulting multi-city spatial economy may be regarded as a (one-dimensional) version of Lösch's central place theory. (Nobody has yet managed to produce a model with Lösch's famous hexagonal market areas.) Fujita *et al.* (1997) consider an economy with multiple manufacturing industries, differing in transport costs and/or scale economies; such an economy spontaneously develops a system of central places that finally provides a justification (again in only one dimension) for Christaller's (1933) hierarchical model of central places. Finally, Fujita and Mori (1996*b*) address

an issue that, astonishingly, appears never to have been formally modelled before: the reasons why ports and other transportation nodes so often become the sites of major cities. What they show is that such transport nodes generate cusps in the market potential function, even in the absence of a city, and therefore tend to serve as the seeds for city growth.

Moving from the local to the global, Anthony Venables and his students have tried to use new economic geography models as the basis for a new style of international trade model. What is the difference between regions and countries? One answer is that factors of production are far less mobile between countries than between regions of the same country; and in Venables-type models they are normally assumed to be completely immobile. None the less, Venables (1996) shows that a circular process leading to economic differentiation between nations can still result if there are intermediate goods produced with economies of scale and subject to transport costs: in that case a country with a large manufacturing sector offers a large market for intermediates; this leads to a concentration of intermediate production in that country, which gives it a cost advantage in downstream production, which further reinforces its advantage, and so on. In Krugman and Venables (1995) this story is used as the basis for a 'history of the world', in which gradually declining transport costs lead first to the spontaneous differentiation of the world into a high-wage industrial 'core' and a low-wage agricultural 'periphery', then to a later convergence of wages as the periphery industrializes. Puga and Venables (1997) offer an alternative version of this story in which the driving force is the growing size of the market rather than growing economic integration. And Krugman and Venables (1997) use the 'race-track' geometry to model global international trade and specialization in a world in which borders are irrelevant, and in which even economic regions are left unspecified; none the less, the world spontaneously organizes itself into manufacturing zones surrounded by agricultural hinterlands.

Most recently, Baldwin and Forslid (1996) have developed an alternative version of geography and trade analysis, this time relying on tools borrowed from the endogenous growth literature. In these models the circular process involves not the move-

ment of factors but their accumulation: countries with large markets invest more, which further enlarges those markets.

It is also possible to mix trade and urban economics using the new genre. Krugman and Livas (1996), for example, develop a model suggested by the relative decline of Mexico City as Mexico has opened itself to trade; the basic idea is that the importance of access to domestic consumers and suppliers, which was crucial as long as Mexico pursued a policy of inward-looking industrialization, has become much less relevant now that it exports more of its output, and imports more of its intermediate goods.

As this partial survey of the theory indicates, the new economic geography has opened the door for analytical discussion of interesting and important issues that were previously pretty much ignored by economists. That in itself is something of an achievement. Moreover, partly because of the novelty, partly because of the inherent sexiness of the stories—geography models naturally produce multiple equilibria, dramatically 'catastrophic' changes in outcomes from small changes in parameters, large effects from small differences in initial conditions, and spontaneous emergence of unexpected order from randomness—these models are a lot of fun to work with.

But are they really relevant?

IV. EMPIRICAL WORK

It has been an unfortunate feature of much of the 'new' theorizing since the 1970s that it has failed to lead to much validating empirical work. The new industrial organization has been notoriously better at creating interesting models than at generating empirical predictions; the new growth theory gave rise to a massive industry of cross-country growth regressions, but with few exceptions these regressions have neither been closely tied to the theory nor provided clear evidence in its support. Under the combined influence of the new growth and new geography movements, there has been a parallel effort to extract insight from cross-sectional regressions on the growth of metropolitan areas; with only a few exceptions like Ades and Glaser (1995),

however, these studies have similarly failed to offer much direct testing of the specifics of the models.

Perhaps the closest thing to a direct test of the models has been the recent work of Davis and Weinstein (1996, 1997a,b), who have used international and inter-regional production and consumption data to test for the 'home market effect': the prediction, made by the underlying models of new geography, that a larger demand for the products of an industry in any given region will lead, other things being the same, to a more than one-for-one increase in the regional production of that industry. (Their results are generally negative for international comparisons, but generally positive at the regional level.)

It would not be surprising if it turns out that the market-size effects emphasized by the current generation of new-geography models are a less important source of agglomeration, at least at the level of urban areas, than other kinds of external economies. It is, for example, a well-documented empirical regularity that both plants and firms in large cities tend to be *smaller* than those in small cities (see, for example, Hoover and Vernon, 1959); this suggests that big cities may be sustained by increasing returns that are due to thick labour markets, or to localized knowledge spillovers, rather than those that emerge from the interaction of transport costs and scale economies at the plant level. However, serious empirical work—which will probably require detailed micro studies of particular industries—still remains to be carried out.

Eventually, one might hope to develop 'computable geographical equilibrium' models, which can be used to predict the effects of policy changes, technological shocks, etc. on the economy's spatial structure in the same way that computable general equilibrium models are currently used to predict the effects of changes in taxes and trade policy on the economy's industrial structure. However, preliminary efforts in this direction by several researchers, myself included, have found that such models are not at all easy to calibrate to actual data; in general, the tendency toward agglomeration is stronger in the models than it seems to be in the real economy!

At this point, then, the new economic geography—like its sister genre the new growth theory—has

been more successful at raising questions than at answering them, better at creating a language with which to discuss issues than at creating the tools to resolve those discussions.

V. THE NEW ECONOMIC GEOGRAPHY AND ECONOMIC THEORY

Even the most recent edition of Mark Blaug's *Economic Theory in Retrospect* (1997) speaks of a 'curious disdain of location theory on the part of mainstream economics', and asserts that 'this neglect largely continues to this day'. Blaug is a bit behind the times, and is surely wrong in supposing that the main reason for historical neglect lies in the accident that von Thünen wrote in German. 'Disdain' for location theory surely has mainly to do with the more fundamental issue of market structure—and the not incidental point that too much of the classical tradition in that field seemed not to understand that increasing returns and perfect competition do not mix. Still, Blaug's comments serve to indicate the remarkable extent to which an economic subject of considerable *prima facie* importance has been marginalized.

The most obvious contribution of the new economic geography, then, is that it has helped to end that marginalization. The characteristic modelling tricks of the genre do not represent a general solution to the market structure problems that have historically prevented economists from saying much about spatial structure, but they provide *a* solution—and thus make spatial structure an issue safe for mainstream economic respectability. Since economic geography clearly is important in the real world, this in itself justifies the genre—even if empirical research may eventually lead us to a somewhat different emphasis than that of the currently most popular theoretical models.

In addition, however, one might argue that the new economic geography has some broader implications, implications about how economic theory as a whole is conducted.

The first of these involves increasing returns. It has been a long time since economic theorizing was restricted to constant-returns settings; increasing returns in one form or another are central to modern theory in industrial organization, international trade, and economic growth. However, in each of these cases it is still possible for sceptics to question the importance of increasing returns to the relevant issues: for example, while Romer (1986) may have created a great deal of excitement with his suggestion that increasing returns and growth are intimately linked, constant-returns Solow-type models remain the workhorses of the field. In the case of geography, however, it is impossible even to discuss the important phenomena sensibly without assigning a key role to increasing returns (which is why, as Blaug noted, for the past century and a half most economists have shown a surprising ability *not* to discuss anything involving space or location). By making geography a field that is safe for mainstream economists, then, we also firmly nail down the place of increasing returns in that mainstream.

The second implication is a related one: geography turns out to be perhaps the most naturally 'non-linear' area of economics. Over the decades a series of critics of economics have argued that the field takes too little account of a set of interrelated possibilities, such as the existence of cumulative processes of change involving 'circular causation', the persistent effects of historical accident via 'path dependence', the occasional emergence of discontinuous change (maybe even 'punctuated equilibrium'), and so on. And from time to time mainstream economists have attempted to make allowance for all these exotic possibilities. For the most part, however, such efforts have been forced and relatively unconvincing. In the new economic geography, however, such non-linear phenomena emerge absolutely naturally from the most basic models (indeed, most of them are on view in Figure 1). And they have plausible, if not rigorously proven, empirical relevance. So the new economic geography, in addition to legitimizing the specific issue of location, may help to make economic theory in general a more friendly place for ideas that are exciting but have lacked a natural home in the field.

In the end, of course, while the achievements of new economic geography to date certainly justify the work involved, a theory must survive or be discarded based on its empirical relevance. So empirical and quantitative work is clearly the next geographical frontier.

REFERENCES

Ades, A., and Glaser, E. (1995), 'Trade and Circuses: Explaining Urban Giants', *Quarterly Journal of Economics*, **110**, 195–227.

Baldwin, R., and Forslid, R. (1996), 'Trade Liberalization and Endogenous Growth: A q-Theory Approach', Centre for Economic Policy Research, Discussion Paper 1397.

Blaug, M. (1997), *Economic Theory in Retrospect*, 5th edn, Cambridge, New York, and Melbourne, Cambridge University Press.

Christaller, W. (1933), *Central Places in Southern Germany*, Jena, Fischer.

Davis, D. R., and Weinstein, D. E. (1996), 'Does Economic Geography Matter for International Specialization?', NBER Working Paper No. 5706, August.

— — (1997*a*), 'Increasing Returns and International Trade: An Empirical Confirmation', mimeo, Harvard University, October.

— — (1997*b*), 'Empirical Testing of Economic Geography: Evidence from Regional Data', mimeo, Harvard University.

Dixit, A., and Stiglitz, J. (1977), 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review*, **67**, 297–308.

Fujita, M., and Krugman, P. (1995), 'When is the Economy Monocentric: Von Thünen and Christaller Unified', *Regional Studies and Urban Economics*, **25**, 505–28.

— Mori, T. (1996*a*), 'Structural Stability and Evolution of Urban Systems', *Regional Science and Urban Economics*.

— — (1996*b*), 'The Role of Ports in the Making of Major Cities: Self-agglomeration and Hub-effect', *Journal of Development Economics*, **49**, 93–120.

— Krugman, P., and Venables, A. J. (forthcoming), *The Spatial Economy; Cities, Regions, and International Trade*, Cambridge, MA, MIT Press.

— Mori, T., and Krugman, P. (1997), 'On the Evolution of Hierarchical Urban Systems', mimeo, Kyoto.

Harris, C. D. (1954), 'The Market as a Factor in the Localization of Production', *Annals of the Association of American Geographers*, **44**, 315–48.

Henderson, J. V. (1974), 'The Sizes and Types of Cities', *American Economic Review*, **64**(4), 640–56.

Hoover, E., and Vernon, R. (1959), *Anatomy of a Metropolis*, Cambridge, MA, Harvard.

Isard, W. (1956), *Location and Space-economy*, Cambridge, MA, MIT Press.

Krugman, P. (1991), 'Increasing Returns and Economic Geography', *Journal of Political Economy*, **99**, 483–99.

— Livas Elizondo, R. (1996), 'Trade Policy and the Third World Metropolis', *Journal of Development Economics*, **49**, 137–50.

— Venables, A. J. (1995), 'Globalization and the Inequality of Nations', *Quarterly Journal of Economics*, **110**(4), 857–80.

— — (1997), 'The Seamless World: A Spatial Model of International Specialization and Trade', mimeo, MIT.

Lösch, A. (1940), *The Economics of Location*, Jena, Fischer.

Pred, A. R. (1966), *The Spatial Dynamics of US Urban-industrial Growth*, *1800–1914*, Cambridge, MA, MIT Press.

Puga, D., and Venables, A. J. (1997), 'The Spread of Industry: Spatial Agglomeration in Economic Development', CEPR Working Paper No. 1354.

Romer, P. M. (1986), 'Increasing Returns and Long-run Growth', *Journal of Political Economy*, **94**(5), 1002–37.

Samuelson, P. (1954), 'The Transfer Problem and Transport Costs', *The Economic Journal*, **64**, 264–89.

von Thünen, J. (1826), *The Isolated State*, English edn, London, Pergamon Press.

Turing, A. (1952), 'The Chemical Basis of Morphogenesis', *Philosophical Transactions of the Royal Society of London*, **237**(37).

Venables, A. J. (1996), 'Equilibrium Locations of Vertically Linked Industries', *International Economic Review*, **37**, 341–59.

Weber, A. (1909), *Theory of the Location of Industries*, Chicago, IL, University of Chicago Press.