NEW ECONOMIC GEOGRAPHY
FOR LOW DENSITY PLACES:
INSIGHTS FROM KALDOR AND LANCASTER

Maureen Kilkenny, Assistant Professor
Department of Economics
Iowa State University


This symposium has touched on rural capital in all its varieties and on the usefulness of understanding land allocation. Castle and Oakerson pointed out that rural capital has "important time and space specificities," i.e., it varies widely. A point made by Jim Hite is that Von Thunen's nineteenth century model of rural land use remains useful and applicable to day. The model, however, takes the location of urban agglomerations of population and industry as exogenously predetermined. The evolution of the new economic geography explains positive feedback and local growth as being driven by preferences for variety, and how the locations of urban agglomerations are endogenously determined.

Most new economic geography models, however, are suitable only for modeling urban, rather than rural, economic development. Nevertheless, students of rural development should not dwell on the pessimistic implications that low density places dependent on immobile natural resource industries can never grow. New economic geography models that operationalize Kaldor's (1935) and Lancaster's (1979) critiques of monopolistic competition are suitable for modeling and do predict how rural economic development might be obtained.

Fundamentals of economic geography

Broadly, economic geography is the study of what happens where. Why do firms and people tend to concentrate in a few locations (cities) rather than spread out evenly across the countryside? How can we rationalize the clustering of similar firms in the same place? What explains the number and variety of economic activities in a region? Why aren't both small and large firms in the same industry equally competitive? What are the economic development prospects for remote, low density (i.e., rural) places? The answers to these questions depend on the trade-offs between scale (fixed costs and externalities) and the cost of distance (transport costs).

We have to start with a historical overview of economic geography, because for a while, modern economists actually forgot some of the most relevant fundamentals. Economists of my generation often ignore the fixed costs of being in business. But fixed, or sunk, costs are the most obvious bases of economies of scale. In contrast, those of us who mastered neoclassical microeconomics since the 1970s were taught to think of all inputs as variable (in the long run, anyway!), with technology specified by Cobb-

1 See Kilkenny (1996b) for a detailed explanation of why most new economic geography models are incapable of simulating rural development.
Douglas production functions. In that constant-returns-to-scale (CRS) world, average and marginal costs are equal; so "marginal cost pricing" is sufficient to sustain firms in long-run industry equilibrium. Many of us never even tried to reconcile the strictly linear Cobb-Douglas cost functions with the intuitively appealing (and, much more realistic) U-shaped cost curves that still abound in our undergraduate textbooks.

In any event, as long as scale has no economic payoff, we expected to see large and small firms co-existing in the same industry. We also expected population and economic activity to be widely dispersed. And we couldn't explain the existence of urban agglomerations within the CRS paradigm. Those of us interested in regional development should simply quit dwelling on the CRS, Cobb-Douglas, paradigm.

Between David Ricardo's first publications in the 1800s and the 1970s, however, economists (especially those interested in the location of economic activity) paid a lot of attention to the fact that industries are subject to increasing returns to scale. The larger the fixed costs, the larger the range in output levels over which average costs decline. This suggests a puzzle: why aren't all goods provided by single, very large firms, capturing the most benefits for society from these scale economies? This puzzle was easy to solve: transport costs from a single location could easily outweigh the costs of opening new plants to serve the markets farther away. Thus, given a dispersed agricultural population and technology with minimum efficient scale (the level of output at which all economies of scale are exploited and average costs of production are lowest), there is a socially optimal number of firms dispersed across space that minimize both fixed and transport costs paid by society. This model also rationalizes the dispersion of economic activity; it still cannot explain urban agglomerations.

That's because one can't explain all urban agglomerations within the paradigm of purely internal increasing returns to scale. The widely observed clustering of industries in one location can only be rationalized on the basis of external increasing returns to scale. Economists used to pay a lot of attention to these. Adam Smith noted that "the division of labor is limited by the extent of the market" (Stigler, 1951). In this case, as an industry's market increases, the opportunities to achieve economies of scale in intermediate input supplies increase. Even if the industry in question is subject to CRS, average costs of production can decline as the intermediate input suppliers achieve their minimum efficient scales. Marshall (1890) argued that matching firms and workers is more efficient (lower cost) the larger the labor pool in the location. Most recently, Rivera-Batiz and Romer (1991) argue that technology transfer (knowledge) is subject to spatial decay; implying that productivity growth rates can be highest in the locations that already do the most of it. This is one basis for what is now called endogenous growth. The existence of external scale economies rationalizes urban agglomeration.

How do industries subject to internal and external scale economies achieve an equilibrium size? Obviously, the available supply of necessary factors of production, and the effective demand for the industry's product;  

1 See Kilkenny and T Bitte (1990) for an in-depth survey of geographic firm location theory.  
2 If there were only internal returns to scale, such as in transportation activity, to minimize the costs to society, only a few ideal port locations would be chosen. Labor would concentrate there. And, those sites would grow as other industries moved near to minimize their transport costs to the markets. But this engine of growth cannot explain the growth of non-port cities.
will define the level of output of the industry. These factors will also determine the market price, but this market price will not equal the marginal cost of production. It must cover the average cost of production, which is everywhere higher than marginal costs of production throughout the relevant range of output. The full price also reflects transport costs, borne either by the producer who ships or by the consumer who shops.

**Insights from Kaldor**

In his seminal work of the 1930s, Chamberlin formalized an explanation for the observed multiplicity of similar firms in a single economy with his model of monopolistic competition (Chamberlin, 1933). In his model, each firm in an industry produces one variety of the good (monopolistic). Pricing is such that revenues are only sufficient to cover average costs at any time, because there are no barriers to entry (competition). People prefer to consume as many varieties as possible, so there will be demand for each firm’s product. Chamberlin was not trying to explain the clustering of similar firms in a single location, however, as he reasonably posited that the number of firms in the industry would be so large that the entry of a single firm would have a negligible effect on the residual demand, and thus the prices, consumers would pay for the products of all the other firms in the industry. This is the inframarginality assumption. In Chamberlin’s world, competition between firms within an industry is non-localized.

In contrast to Chamberlin’s assumption of inframarginality is Hotelling’s model of “stability in competition” (Hotelling, 1929). Hotelling described the locational choices of firms in an industry in which all firms are rival in the sense that each captures the market closer to them. The analysis is relevant in explaining the locations of firms supplying “shipping” or “shopping” goods. “Shipping” goods are those for which the producer bears the cost of transport (e.g., agricultural produce; manufactures). “Shopping” goods are those for which the consumer bears the cost of transport (e.g., personal services, related goods). In either case, space and the density between and among firms and consumers play an important role. These are important considerations for rural areas, which are remote and low density places. The full price of a shopping good includes production plus transport costs, which rise with the distance between the point of supply and the location of the consumer. The full price of a shopping good rises the lower the density of the suppliers, and the fewer the locally available varieties. Rural customers in low-density areas incur higher search costs for shopping goods, and may pay higher prices for shopping goods, too.

Hotelling argued that (any type of) firms will concentrate in the center of the market (i.e., in the central place (see Mulligan, 1994) or the urban core of a Composite Economic Area (REA, 1986)). With respect to shopping goods, they will divide the market equally, and price at their marginal cost of production (Hotelling abstracts from fixed costs). Hotelling’s result (reformulated correctly by d’Aspremont, Calszowicz and Thiasse, 1979) is known as the Principle of Minimum Differentiation. That firms in a duopoly fail to be able to raise their prices above marginal costs is known as Bertrand’s Paradox (see Trolle, 1988). The point, however, is that suppliers of undifferentiated products cannot afford to extract anything other than competitive prices; and they cannot afford to locate farther from the bulk of the market than their rivals do.
Alternatively, Kaldor (1935) argued that competition is localized. Not all firms are rivals (as in Hotelling), nor are all firms infra marginal (as in Chamberlin). Kaldor noted that the number of firms supported in any one location is too low for infra marginality to be a valid assumption. Furthermore, a new entrant can affect sales of other firms in the neighborhood. He also argued that new entrants cannot affect the sales of more remote firms (Targetti and Thrilwall, 1989). The analysis of firm location in the context of Kaldor’s market structure, however, requires the use of the tools of non-cooperative game theory, at best. In the least, the local market price must be endogenous to the entrance of a new firm, which means the price-location problem does not have a closed-form solution, and must be solved using computable general equilibrium (CGE) techniques (Kokkeny, 1996a).

Unfortunately, neither technique was sufficiently developed to apply to economic problems until the late 1980s, so that economists could not operationalize Kaldor’s localized competition model. Rather, Chamberlin’s theory of monopolistic competition was widely applied, as formalized by Dixit and Stiglitz (1977), to explain macroeconomic geography. In macroeconomic geography, the relevant regions are entire countries. Venables (1975), for example, applied Dixit-Stiglitz’s monopolistic competition model to explain international specialization and trade as arising from increasing returns to scale. A country with a relatively larger number of firms in an industry, producing a wider variety, can be more competitive in international markets. This breakthrough helped explain the high volume of trade between similarly endowed and similarly productive countries. Neither Ricardo’s model of comparative advantage based on relative productivity, nor the Heckscher-Ohlin-Samuelson factor proportions model could rationalize those most significant trade flows in the world.

Krugman (1981) popularized application of the Dixit-Stiglitz version of Chamberlinian, non-localized, monopolistic competition to topics in international economics. Meanwhile, urban systems modelers applied the Dixit-Stiglitz version of monopolistic competition to explain urban agglomeration (Abdel-Rahman, 1988; Fujita, 1988). The urban systems modelers used the concept that product differentiation (distance in characteristics space) is analogous to distance in physical space. By supplying slightly different products, firms insulate themselves from their competition in the same way that transport costs insulate one competitor from the next (see Anderson, de Palma, and Thisse, 1992). The more homogeneous the varieties, the stronger the local competition, and the lower market prices, profits, and thus firm numbers in the location. The more heterogeneous, the higher the prices, the more firms.

**New economic geography**

A ‘later treatment’ of the urban systems application of Chamberlinian monopolistic competition to explain urban concentration and specialization (Quigley, 1998) was also presented by Krugman (1991). Krugman’s intuitively appealing explanations and broadly applicable stylization significantly popularized regional economics. Basically, Krugman did for regional science what he had done a decade earlier for international economics: he popularized the use of Chamberlinian (Dixit-Stiglitz) monopolistic competition market structure assumptions to close models of regional eco-
nomic systems. He called one the "New Trade Theory," and the other "New Economic Geography" (Krugman, 1996).

The prototypical new economic geography model has just two regions; two industries or goods (food and manufactures), and two primary factors of production (farmers and workers). Manufacturing is subject to internal increasing returns to scale. Marginal-cost pricing is incompatible with long-run equilibrium, since a price equal to marginal cost would cover average costs only at zero output. Thus, a more-geographically competitive market structure is assumed. Each firm produces a differentiated product, and each consumer demands a positive quantity of every variety. Firms do not need to disperse across space to serve the same markets profitably, but they do incur delivery costs to serve consumers in the other region. The degree of product differentiation is represented by the parametric elasticity of substitution (ς) characterizing consumer preferences.

Each household consumes some agricultural product, F, and a bundle of manufactures, M, to maximize satisfaction according to a Cobb-Douglas utility function:

\[ U = MF^\mu \]

where \( \mu \) also measures the budget share on manufactures. The manufactures are a constant elasticity of substitution (CES) aggregate of the varieties produced by \( n \) firms (i = 1,...,n):

\[ M = \sum m_i^{1/\sigma} \]

where the CES exponent \( \sigma = 1/(1/\omega) \). Abstracting from any technological differences, all manufacturing firms will have the same optimal scale, so that

(2) \[ M \approx n^{1/\mu} \]

Consumer utility (1) is thus increasing in the number of firms or varieties:

(4) \[ U = n^{1/\mu} F^\mu \]

Analysis of the consumer expenditure minimization problem concerning manufactures consumption reveals that \( s \) is also the price elasticity of demand. Facing this elasticity of demand, industrial firms maximize profits by equating marginal revenue to marginal cost:

(5) \[ P \cdot (1-s/\sigma) = W \]

(variety and region subscripts dropped for ease of exposition) so that the profit-maximizing delivered (and mill) price to local markets is a parametric markup \( (1/\rho) \) over the local wage. Within each region each industrial firm charges the same delivered price to local residents. The optimal delivered price would be lower the more substitutable products are (the larger is \( \rho \), and \( \sigma \)). The optimal delivered price is higher if preferences for variety or the degree of differentiation is higher (lower \( \rho \) and \( \sigma \)), or the higher local wages. To compete with another region's suppliers, therefore, producers must either pay lower wages or differentiate their product more.

The new economic geography models highlight that a preference for variety is a basis for increasing returns to scale. The higher the preference for variety, the higher the price premium, and the larger the external economies of scale for producers of luxury goods in the region. The more producers will locate in the region, the more workers, the wider the variety produced, and the higher the real utility of the population there. Lancaster (1979, page 54) showed that economies of scale (\( \beta \)), expressed as the inverse elasticity of cost (percentage change in cost with respect to percentage increase in quantity)
are an inverse function of the elasticity of substitution between inside (differentiated) and outside (homogeneous) goods (d). Taking the ratio of average cost (AC) to marginal cost (MC) cost:

\[ \frac{AC}{MC} = \frac{Wg/W(1-\theta)}{(1-\theta)} \]

In long run equilibrium, the market price (shown in 5) will equal average cost. The local wage is the marginal cost. In general, the production process is homogeneous of degree 0, which is inversely related to \( \theta \). All new economic geography models include this demand-side engine of agglomeration economies of scale.

Note also that since the local labor supply constrains the size of the industry, the higher the preference for variety (smaller \( \theta \)), the more numerous and smaller the firms will be. In fact, the external scale economies far outweigh the internal returns to scale, so in ‘new economic geography’ models, the optimal firm size can be as small as a proprietorship (single employee). Thus, there is no particular reason why rural industry, if differentiated from the urban competition, couldn’t profit from the same preference for variety. (Think of French wine or cheese industries.) The problem with the existing new economic geography models, however, is that goods are not qualitatively differentiated according to their origin of production. Iserman (1996) critiques the new economic geography model as ‘oversimplifying both urban and rural places.’

“Rural areas, however, produce food, natural resources, and other goods and services that are tied to place-specific attributes. Many of these goods cannot be produced in cities. Leaving these goods out of the model becomes a problem when drawing conclusions about the effects of trade on urbanization...” (Iserman, page 39)

Most Krugman-style models also assume mill-pricing, which implies that both prices and firm size are parametric (Krugman, 1993, 1996; Fujita and Krugman, 1995; Calvetti and Le Pottier 1995; Walz, 1996). Only real wages are endogenous; and, to close their systems, other models assume that the price level is unaffected by the entry or exit of a firm (infra marginality). As discussed above the infra marginality assumption seems most egregious for modeling new business openings in remote, low density places. In sympathy with Calder’s localized competition critique, in my 1998(a) paper I assumed uniform delivered pricing and endogenized market prices and firm size. Firm profits depend explicitly on the spatial distribution of the population. I did not invoke infra marginality to close my system, but solved it as a fixed point problem (a true general equilibrium system). With that system I show the existence of stable equilibria in which all regions are occupied in various densities. Krugman-style models only have one stable equilibrium: full concentration in some region.

A recent paper by Helpman (1998) also shows that with the proper assumptions, concentration is not the only stable equilibrium. Helpman shows that all regions are occupied when “housing” (the outside or homogeneous good) has the largest budget share, and people have sufficiently low preferences for variety. I found that all regions are occupied even if the share on the outside good (“food”) is low.

**Insights from Lancaster**

If there is no feedback from the opening of a new firm in the ‘hinterland,’ then our models can only simulate urban concentration. If there is
feedback, we can simulate the entire range of relative population densities observed in the real world. What pattern is optimal? There is no single socially desirable spatial allocation. Desirable patterns vary with community preferences. The Tiebout Hypothesis (Tiebout, 1956) is that people choose their desired mix of public or non-market goods by migrating, where the spatial configuration of the community is, in effect, one of those non-market goods. The work of Kelvin Lancaster may help us model remote, low density populations more effectively.

In 1979, Lancaster made an "informal guess" about how the distribution of population (location in geographic space) may arise from the distribution of preferences (location in characteristics space).

"...the degree of preference variation over the population may vary from one society to another. This degree of variation is itself an important economic parameter, as is shown in the economic simplifications that would result from a total lack of variation. It might be expected that societies with a large degree of ethnic and cultural homogeneity (Scandinavian countries, for example) would have less variation than societies with more diverse populations (like the United States or the Soviet Union), but this is merely an informal guess." (page 9)

Lancaster’s exploration of the equity and efficiency problems of modern economies was based on three key elements: variety among individual preferences, variety among products, and economies of scale. In contrast with Chamberlin (1962), who did not attempt to quantify qualitative differences between products, Lancaster says that differences between products can be decomposed into differences in measurable characteristics (such as the degree of sweetness of cider). Consumers derive utility from the characteristics. As such, the product demand is derived from consumer demand for characteristics. Lancaster classifies products of common characteristics as being in a group, and further specifies that consumer utility is separable between the characteristics of the group and those outside the group. Finally, Lancaster assumes that goods in the same group are also similarly produced. He modeled production by displaying constant marginal and declining average cost (never lower than marginal cost due to a fixed cost). Lancaster also assumed the most competitive market structure compatible with the aforementioned economies of scale: monopolistic competition.

Lancaster pointed out that fewer different products (but possibly many plants) and smaller market areas are to be expected when agglomeration scale economies are low. If preferences are such that no variety is desired in the consumption of necessities, but it is desired for luxuries, the opportunities to exploit agglomeration economies will rise as income increases. This happens over time. Population accrues and capital is accumulated. This growth may support the positive feedback we saw above; as Lancaster says: "an increase in population will (1) increase the degree of product differentiation if the range of diversity is unchanged and the degree of economies of scale varies with output, but have no effect if it is constant, (2) increase the number of goods (and the degree of product differentiation in that sense) if the range of diversity is increased, and (3) lower the per capita resources required for a given level of welfare per capita, whether or not the degree of economies of scale varies with output." (1979, page 79)
Conclusion

The problem with some rural areas in the USA may be that communities do not differ one from the next. Rural places in the corn belt, for example, are not only sparsely settled, they are also homogeneous: soy and corn fields as far as the eye can see; no architectural variety; and the same franchise fast food/gas shops in all the towns. In the old days, high transport costs justified the proliferation of fixed costs, and insulated one rural place from the next. Now that people can travel across the state within hours, only a few rural places offering homogeneous services and amenities, are needed. Rural places compete with each other in the classic hotelling sense.

A preference for variety, however, can motivate customers and residents to ‘go the extra mile.’ All new economic geography models highlight the fact that if consumers value diversity, those who supply it can capture supra-normal profits. Thus, rural areas that offer desirable, unique, or distinctive products or natural resource amenities can capitalize on the premium prices they are able to charge for their fare. If rural citizens and rural tourists, however, actually prefer crowded places and homogeneous products, consolidation is the only hope for survival of a few, relatively accessible, rural towns. Workforces in such towns face the prospect of competing with workers in other countries on the basis of lower wages. They can work in large factories (which capture internal economies of scale), and they mass-produce items in the final (homogeneous) stage of their product cycles. They may also face cyclical profligacy problems, as the old products become obsolete.

All ‘new economic geography’ papers also show that regional development in the context of agglomeration economies depends on raising real rural wages. Higher real wages attract population and thus support cumulatively cause growth. My own ‘new economic geography’ work has shown that a market-decreasing policy may be an effective rural development strategy. In general, a higher real wage for rural people can be obtained in four ways: higher nominal rural wages, lower rural prices, subsidies, or more positive rural externalities. Unfortunately, high nominal rural wages and low rural prices would repel firms. Our models also show that government subsidies cannot substitute for the positive production externalities (external to firms but internal to an industry) of urban agglomeration in rural places. As should be expected, subsidies ultimately go the owners of the relatively fixed factor of production: urban land. That leaves externalities.

One way to raise real wages in rural areas without repelling firms or raising urban land rents is to improve the quality of rural life. Increasing the array of amenities enjoyed by rural people may be a more effective strategy than attempting to alter private incentives facing businesses.

References


Chamberlin, Edward Hastings (1962) The Theory of Monopolistic Competition