Spatial structure in the retail trade: a study in product differentiation with increasing returns

Geoffrey Heal*

This article develops a model to analyze the implications of economies of scale in transportation for the spatial distribution of retail outlets and for the structure and pricing of public transportation systems. Consumers (or workers) are located around the circumference of a circle with a single producer (or employer) at the center. Consumers may buy directly from the center (workers may drive directly to the center), or they may buy from shops on the circumference which have bought in bulk from the center (workers may drive to stations on the circumference and commute by public transportation). The article examines the socially optimal pattern of outlets, the pricing policy which supports this and the effects of alternative pricing policies, and the patterns of outlets resulting from various forms of competition. Among the results derived, it is shown that the socially optimal pattern cannot be sustained by nondiscriminatory pricing that covers average costs, and that attempts to cover average costs may lead to large distortions. In addition, without price discrimination, competition usually leads to an excessive number of outlets if the market is large and to an insufficient number if the market is small.

1. Introduction

■ The classic article on problems of spatial location and their implications for product differentiation is the work of Hotelling (1929). His analysis of the location decisions of two firms competing to serve consumers distributed along a finite straight line led him to enunciate the much quoted principle of minimum differentiation. He argued that the locational choices of the firms would converge towards the center of the line, and argued by analogy that, in general, competitive forces would tend to create "a socially uneconomical system of prices, leading to needless shipment of goods and kindred deviations from optimum activities, and an undue tendency for competitors to imitate each other in quality of goods, in location, and in other essential ways." In some other early and instructive contributions, Lewis (1945) and Lerner and Singer (1937) developed these and related ideas further.

---

* University of Essex.

I have benefited greatly from the perceptive comments made on an earlier draft by my colleagues Peter Holmes, Pat Rice, Manik Sen, and Tony Venables. Discussions with Graciela Chichilnisky and Partha Dasgupta and comments of a referee and the Editorial Board were also of great value in clarifying my ideas.

More recently Eaton and Lipsey (1975), Shaked (1975), and Stern (1972) have returned to this general class of problem. They have extended the analysis to consider numbers of firms in excess of two, nonuniform distributions of consumers, and unbounded, nonlinear, and two-dimensional market areas. Their conclusions suggest that the bounded linear market area considered by Hotelling was rather special, and that his principle of minimum differentiation does not generally carry over to more complex situations. Although the present paper is clearly within the tradition that these earlier writers have established, it is directed towards a rather basic issue which is not directly tackled in the earlier work. My concern is primarily with the determination of the number of outlets, something which is taken as exogenous in the earlier work. Most of the results that one finds there are of the form: "*If* there are n firms, *then* the equilibrium locational pattern is such and such, which is (or is not) socially optimal." Typically, there is no attempt to explain how many firms there will be, or whether this number of firms itself (as opposed to the pattern in which they are distributed) is correct. There are, of course, good technical reasons why this question is avoided: in these models technology is always assumed to display constant or diminishing returns to scale, in which case the number of enterprises is typically indeterminate or infinite.

Another related difficulty with the earlier literature is that, although it is tempting to see it as modeling the retail trade, the models used usually have no explicit reference to production. Retailers, however, are by definition intermediaries between consumers and producers. Therefore, to analyze retailing convincingly one needs a model containing both consumers and producers, so that there is a genuine role for a middleman. Such a model should explain why there is a need for retailers as middlemen.

My basic hypothesis here is that one cannot explain either the number of retail outlets or the role of retailers as middlemen without recognizing the importance of increasing returns in retailing. There are probably many causes of this, though my main concern here will be with economies of scale in transportation. However, the basic activities of storing, retrieving, and selling also display some measure of increasing returns. It is also true that an important aspect of the retailer's role is the acquisition and dissemination of information—information about the quality, range, and prices of products available: Wilson (1975) has shown that activities involving the acquisition and use of information may also be characterized by increasing returns. Hence, there are several reasons for believing that economies of scale are important, and indeed this will come as no surprise to anyone familiar with current trends in concentration in the retail sector.

As I have indicated, I am concerned here with the effect of economies of scale in transportation, and to this effect will consider a very simple model with consumers located around the periphery of a circle and a single producer at the center. I shall assume that the cost of transporting a unit of weight a unit of distance is smaller, the larger the total load. Each consumer may buy either by going directly to the center or by going to his local shop, which is located on the circumference. This shop will, in turn, have bought from the center, but in bulk and thus at lower transportation costs. Within this simple framework I examine:

(1) the pattern of shops which minimizes total transportation costs—this is referred to as the socially optimal pattern—and the pricing policies which will support this pattern;

(2) the patterns of shops resulting from profit maximization, either by single independent firms or by all firms acting together as a retail chain to maximize joint profits;
(3) the pattern that results if shops practice price discrimination between consumers at different locations by paying some or all of the transportation costs from the shop;
(4) the patterns resulting from free entry and from imperfect competition in this retail market.

In summary form, the conclusions that emerge are:

(1) The socially optimal pattern of locations cannot be sustained by nondiscriminatory pricing that covers average costs, and attempts to charge uniform prices covering average costs may lead to major distortions.
(2) If a profit-maximizing retail chain is allowed to exercise price discrimination by offering free delivery, the outcome will be optimal.
(3) Without the possibility of price discrimination, profit maximization usually leads to an excessive number of shops (in general, excessive product differentiation) in large markets, but an insufficient number of shops in small markets. In other words, with increasing returns and product differentiation, large markets tend to be overserved and small markets underserved, relative to the optimum.

Although the model presented here is intended primarily as one of the spatial structure of retail outlets, it can also be interpreted in terms of alternative transportation systems or in terms of product differentiation in general. In this sense, it has similarities to recent works by Spence (1976) and Salop (1979). These alternative interpretations are all presented in the next section.

2. The models

■ Consider a group of consumers of a single good. They live uniformly distributed around the circumference of a circle of radius r. The good is manufactured at the center of the circle. Each consumer has an inelastic demand of density q for the good, and each occupies a length of circumference of δl, subtending an angle $\delta\theta = \delta l/r$ radians at the center.

The good must be transported to the consumers and there are economies of scale in transportation. In particular, the cost of transporting Q units a distance d is

$$C(Q)Qd, \qquad C'(Q) < 0, \qquad C''(Q) > 0. \tag{1}$$

$C(Q)$ is thus the unit cost for transportation—the cost of moving one unit of the good one unit of distance. This cost falls, at a decreasing rate. Particular forms to be considered for $C(Q)$ are:

$$C(Q) = \frac{1}{Q}, \tag{2}$$

where all costs are fixed, and

$$C(Q) = \frac{\alpha}{Q} + \beta, \tag{3}$$

which contains a fixed cost and an element of unit cost linearly related to the amount moved.

We assume each consumer may, if he wishes, arrange for the good to be sent directly from the center to him at a cost per unit of the good of $C(q\delta l)r$. This is the average cost per unit of sending $q\delta l$ units a distance r. However, because of the economies of scale in transportation, there may be scope for shops on the circumference to buy in bulk from the center, thereby incurring relatively low transport costs, to serve the consumers near to them. Even if the shop makes a profit, the total of transportation costs from the center to the shop, plus profit, plus transportation from the shop to the consumer, may, because of these economies, be less than the cost of transportation from the center to the consumer directly. We shall be concerned with analyzing the optimal configuration of shops, i.e., the configuration which minimizes the total cost of meeting the given inelastic demand, and with comparing this with the configurations that would be established by market forces.

It is well worth noting that this model can be given another interesting interpretation. Consider a community of commuters who live around the periphery of a circular city and work at the center: they need to travel to the center every day, and there are again economies of scale in transportation. There are two modes of transportation available: private transport, which enables the individual to proceed directly along a radius from his home to the center, and public transportation. The use of the latter requires the individual to travel by private transportation around the circumference to one of a finite number of stations on the periphery and then to go directly to the center along a radius. Although this involves greater total mileage, it may nevertheless be less expensive because public transportation can take advantage of economies of scale. In the context of such a model one can examine the optimal provision of public transportation and the effects of different pricing policies. It should be clear that, formally, the two models are identical: all that differs is whether one speaks of stations or shops. In each, what is at issue is a comparison of costs from small scale but direct journeys to the center with those of larger scale but less direct journeys, given the existence of scale economies.

In the transportation interpretation, $C(Q)$ can be interpreted as giving cost per person per mile as a function of the number of people carried (a continuous variable). Then $q\delta l$ is the value of the argument Q when the number of people carried is one, and if n is the number carried,

$$nq\delta l = Q,$$

so that we can alternatively write $C(nq\delta l)$ for the cost per person, where $q\delta l$ is a parameter of the transportation system. It is convenient to absorb this into the functional form and just write $\hat{C}(n)$ as the cost per person per mile.

As mentioned in the introduction, our model can also be interpreted as having some bearing on the problem of product differentiation. At one level this is obvious because the shops are supplying services differentiated only by location, so that the variable under consideration is just the degree of differentiation. There is also a less direct but perhaps more important relationship, since it is possible to see the product differentiation problem in its most general setting as essentially a problem in patterns of location. Suppose one adopts a model which regards consumer demand as a function of a number of basic characteristics embodied in products, so that the natural setting of the analysis is a characteristics rather than a goods space. Each ray through the origin represents a potential commodity defined by possession of the various characteristics in certain proportions, and one can clearly choose the units of commmodities so that potential commodities are represented by points on the surface of the unit sphere—or, in the case of only two characteristics, on the perimeter of the unit circle. In such a setting the choice by firms of a particular pattern of product differentiation is equivalent to the selection of a set of points on this surface.

This analogy between problems of product differentiation and of spatial location will not be developed further here, because it is by no means complete. It should, however, be noted that two recent papers on monopolistic competition and product differentiation (Spence, 1976; Salop, 1979) consider issues very close to those with which I am concerned. Both consider models with fixed costs and so, in effect, increasing returns in production.

In particular, the present model is very close to Salop's, which also has consumers arranged around the circumference of a circle. In his model there is competition from an "outside good" which is not necessary identical with the good under consideration: in my model the possibility of buying from the center plays the role of the outside good. Although there is a striking similarity between Salop's model and the present one, the questions to which they are oriented are somewhat different. Salop's concern is mainly with comparative static issues and the nature of market equilibrium, whereas mine is mainly with welfare issues relating to pricing policies and the optimal degree of differentiation.

In terms of the questions asked, Spence's (1976) paper is close to the present one: he, too, is concerned with comparing optimal and actual levels of differentiation. His conclusions, however, are different: he tends to find that a market equilibrium produces an insufficient number of products, each at an excessive scale, relative to the social optimum. The present model cannot investigate the scale of total production, as demand is inelastic: it does, however, find biases in both possible directions in the range of products available.

3. The socially optimal solution

■ In this section we analyze the configuration of shops which minimizes the total cost involved in transporting the good from the center to the consumers. Two assumptions are basic to the discussion: the number of shops is a continuous variable, and the solution is symmetric. Given the completely symmetric statement of the problem, the latter seems eminently reasonable and could readily be given a formal justification. The former simplifies the analysis by replacing inequalities with equations, without apparently changing the salient features of a solution.

We shall suppose each retailer has a market area which subtends an angle of θ radians at the center, calculate the total associated transportation costs, and then minimize these with respect to θ. The total quantity retailed by a shop with market area θ is $r\theta q$, so that the cost of transportation from the center to the shop is

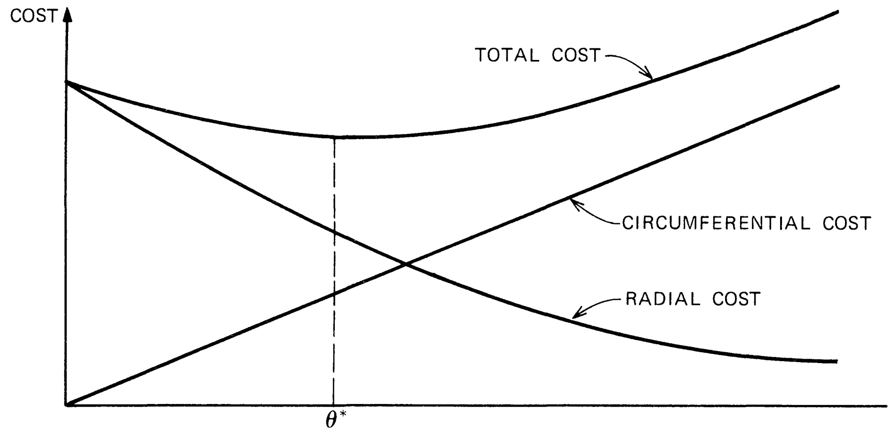
$$C(r\theta q)r\theta qr. \tag{4}$$

There is, in addition, a cost of transporting goods from the retailer to the consumer. For a consumer situated an angular distance α radians from a shop, this is

$$C(q\delta l)r\alpha q\delta l, \tag{5}$$

as $q\delta l$ is the amount he buys, and αr the distance it is moved. Note that these

FIGURE 1



circumferential costs are so formulated that they do not have the advantage of economies of scale. Each shop-consumer journey is a separate one: there is no delivery man moving from the center to the edge of the market area delivering *en route*. This seems the most appropriate assumption in the context of an analysis of suburban shopping or commuting habits. Given this assumption, total circumferential transportation costs are

$$2\int_0^{\theta/2} C(q\delta l) r \frac{\alpha}{\delta\theta} q\delta l d\alpha = C(q\delta l) q \frac{\delta l}{\delta\theta} r \frac{\theta^2}{4}. \tag{6}$$

Total costs per retail outlet are thus given by (4) plus (6). There are $2\pi/\theta$ outlets, giving a total transportation cost over the whole market of

$$2C(r\theta q) r^2 q\pi + C(q\delta l) q \frac{\delta l}{\delta\theta} r \frac{\theta}{2}\pi. \tag{7}$$

It follows that the optimal market area θ^* must satisfy

$$C'(r\theta^* q) = -\frac{1}{4}\frac{\delta l}{r^2 q}\frac{C(q\delta l)}{\delta\theta}. \tag{8}$$

We shall investigate the solution to (8) for functional forms such as (2) and (3). It may, however, be useful to approach these via a graphical analysis, as in Figure 1. This shows both components of (7) and their sum, total cost, as a function of θ. The circumferential element of costs increases linearly in market size θ, whereas the radial element decreases with θ because of economies of scale, but does so at a decreasing rate. The optimal market size θ^* strikes a balance between these tendencies.

For the functional form $C(Q) = 1/Q$, representing the extreme case where all costs are overhead, the solution to (8) is

$$\theta^* = 2(\delta\theta)^{1/2}. \tag{9}$$

More generally, the cost function (3) with both fixed and variable costs gives

$$\theta^* = \left(\frac{4\alpha\delta\theta}{\alpha + \beta q\delta l}\right)^{1/2}. \tag{10}$$

Equation (9) is striking because the solution is independent of the level of demand. In (10) the level of demand appears explicitly, with θ^* a decreasing function of this level. Partial differentiation of (8) shows that for general cost functions $\partial\theta^*/\partial q$ may have either sign. Neglecting terms in $\delta\theta\delta l$, we have

$$\frac{\partial\theta^*}{\partial q} = -\frac{\theta^*}{q} + \frac{\delta l C(q\delta l)}{4\delta\theta r^3 q^3 C''(r\theta^* q)} .$$

An increase in q lowers the per-unit transportation costs at any given θ because of scale economies, but raises the number of units to be moved. Which effect is more important depends on the functional forms involved. For the cost function in (3), $\partial\theta^*/\partial q < 0$, and an increase in demand density leads to a decrease in the optimal market area. The increase in density allows sellers to take advantage of scale economies in a smaller market.

It is worth recalling that as $\delta l = r\delta\theta$, the radius can be made to appear explicitly in (10). Whether this is useful depends on whether one regards it as more appropriate to hold δl or $\delta\theta$ constant in the face of variations in r. Almost certainly, it makes more sense to hold δl, the consumer's "length," constant, ensuring a number of consumers proportional to the perimeter of the circle. In this case the solutions θ^* for both (2) and (3) are independent of r. More generally, θ^* does depend on r, but in a fashion which is difficult to sign:

$$\frac{\partial\theta^*}{\partial r} = -\frac{\theta^*}{r} + \frac{C(q\delta l)}{2r^3 q^2 C''(r\theta^* q)} .$$

4. Pricing policy

■ Suppose that an optimal number of shops (or stations) is established. What, then, is the cost of providing service to a consumer on the edge of a market area? Clearly, the cost per unit is

$$C(r\theta^* q)r + \frac{\theta^*}{2} rC(q\delta l), \tag{11}$$

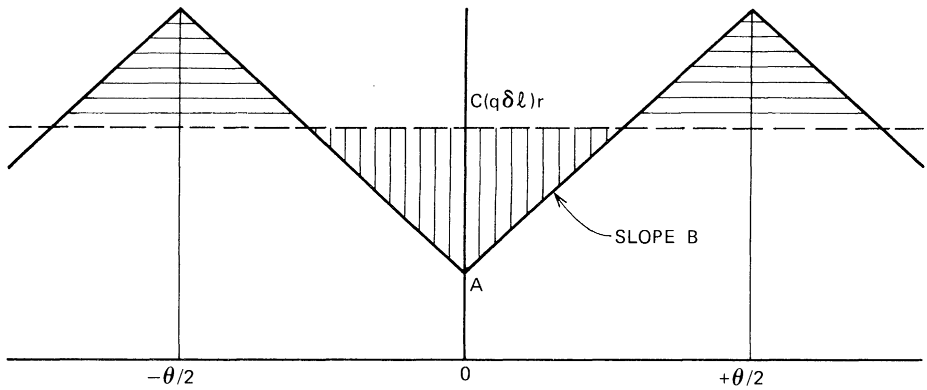
whereas the cost per unit of buying directly from the center is

$$C(q\delta l)r. \tag{12}$$

Obviously, (11) will exceed (12) for $\theta^* = 2$, and indeed for a range of θ^* "near enough" to 2. This means that average cost per unit of serving someone on the edge of an optimal market area via a shop may exceed the average cost of serving him directly from the center.

Although slightly surprising, this property does not contradict the optimality of the chosen market area. To see this it helps to consider switching from a situation where people serve themselves from the center to one where each consumer is served by a shop of market area θ^*. Clearly, the switch leads to an increase in the cost of providing for a marginal consumer (i.e., one on the margin of a market area), but at the same time, because of economies of scale, it leads to decreases in the costs of providing for those located near the shops. Figure 2 illustrates this: it shows the costs of serving consumers along a segment of the perimeter. A shop with market area θ^* is located at zero, and the cost of serving a consumer at $\pm\alpha$ via the shop consists of the radial transportation cost $0A$

FIGURE 2



plus circumferential costs αB, where

$$0A = C(r\theta^* q)r$$

$$B = C(q\delta l)rq\delta l.$$

The cost to a consumer of buying directly from the center is $C(q\delta l)r$, which exceeds $0A$. The switch clearly leads to a reduction in total costs, provided that the areas striped horizontally are less in total than those striped vertically, but does, of course, lead to an uneven distribution of costs, which were previously the same for all.

This observation has substantial implications for the sort of pricing policy which is capable of sustaining the optimum. In particular, it should immediately be clear that if shops price at or above the average cost, the resulting equilibrium will not be optimal. Clearly, consumers for whom

$$B\alpha + C(r\theta^* q)r > C(q\delta l)r \tag{13}$$

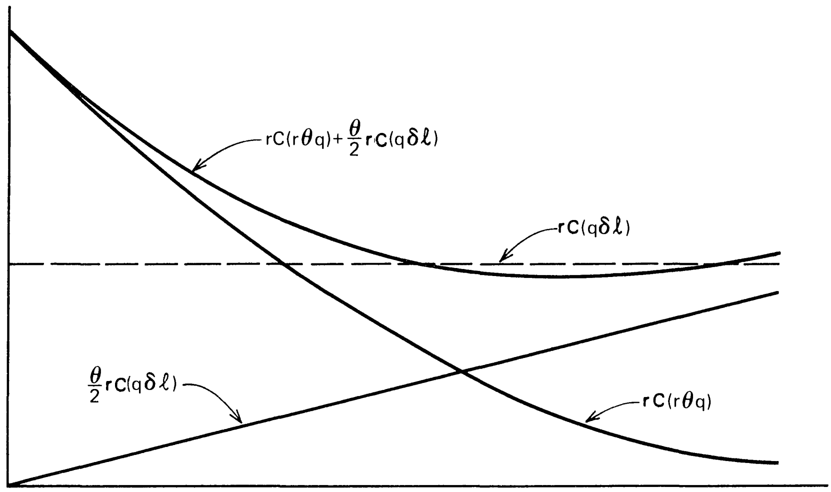
will buy directly from the center, thereby reducing the firm's effective market size. This will have the effect of raising the radial transportation costs $0A$ of a shop, thereby reducing further the value of α at which an inequality such as (13) holds. This process may continue until a new set of smaller equilibrium market areas is established, and in some cases it may continue until the shops close completely, and everyone buys from the center. We now examine these possibilities more closely.

If retail outlets price at average cost, then resulting market area $\hat{\theta}$ must satisfy

$$C(r\hat{\theta}q)r + \frac{\hat{\theta}}{2}rC(q\delta l) = rC(q\delta l). \tag{14}$$

Here the first term on the left-hand side is the shop price, and the second is the cost of transportation from the shop to the marginal consumer. An equilibrium will occur when the cost per unit to the marginal consumer just equals the cost to the consumer of buying directly from the center; that is, when (14) holds. The constituents of (14) are displayed graphically in Figure 3. The left-hand side is a U-shaped function of θ, and the right-hand side is independent of θ. There will therefore be two solutions (as shown in Figure 3), or no solution (if $rC(q\delta l) < \min\ (rC(r\theta q) + rC(q\delta l)\theta/2)$, or, in a singular intermediate case, one solution.

FIGURE 3



If there is no solution, this corresponds to a situation where a shop pricing at or above average cost can never retain a positive market. If there are two solutions, then average cost pricing may lead to either a small number of big shops or a large number of small shops. Which particular outcome is achieved will presumably depend on the historically given initial conditions from which the system begins. It seems clear that, of the two possible equilibria, the one with larger market areas is stable, and the other is unstable. To see this, let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the solutions, with $\hat{\theta}_2 > \hat{\theta}_1$. Then for $\theta > \hat{\theta}_2$, the marginal consumer finds it cheaper to buy from the center, so the shop's market area falls. For $\hat{\theta}_1 < \theta < \hat{\theta}_2$, the converse is true. Therefore, $\hat{\theta}_2$ is locally stable, and similar arguments establish the opposite for $\hat{\theta}_1$.

The optimality of solutions to (14) has not so far been discussed. The matter clearly merits investigation, and Table 1 shows θ^*, $\hat{\theta}_1$, and $\hat{\theta}_2$ for three different cost functions. It is clear that for $C(Q) = 1/Q$, $\hat{\theta}_1$ and $\hat{\theta}_2$ tend to 0 and 2, respectively, as $\delta l/r$ tends to zero. Hence, one of the two average cost pricing equilibria corresponds in this limit to the social optimum, which calls for an unlimited number of infinitesimal shops. If an equilibrium is established at $\hat{\theta}_1$, then the

TABLE 1

SOLUTION	COST FUNCTION		
	$1/Q$	$\alpha/Q + \beta$	$\frac{1}{\gamma + Q}$
θ^*	$2(\frac{\delta\ell}{r})^{\frac{1}{2}}$	$(\frac{4\alpha\delta\theta}{\alpha+\beta q\delta\ell})^{\frac{1}{2}}$	$\frac{-\gamma}{rq} + 2(\frac{\gamma}{q\delta\ell}+1)^{\frac{1}{2}}(\frac{\delta\ell}{r})^{\frac{1}{2}}$
$\hat{\theta}_1, \hat{\theta}_2$	$1 \pm (1-\frac{2\delta\ell}{r})^{\frac{1}{2}}$	$\frac{\alpha}{\alpha+\beta\delta\ell} \times (1 \pm \sqrt{1-\frac{2\delta\ell}{r}-\frac{2\delta\ell^2\beta q}{\alpha r}})$	$1-\frac{\gamma}{2rq} \pm \sqrt{(\frac{\gamma}{2rq}-1)^2-\frac{2\delta\ell}{r}}$

welfare loss will be negligible in the case of δl "very small," which is obviously the interesting case. Similar statements hold for $C(Q) = \alpha/Q + \beta$: in this case θ^* tends to 0 as δl tends to zero, and the corresponding limits of $\hat{\theta}_1$ and $\hat{\theta}_2$ are 0 and 2, respectively. Hence, in both cases there is an average cost pricing equilibrium which involves larger market areas than are optimal, and another which tends to the optimal market area as the size of an individual consumer shrinks to insignificance. In both cases it is reasonable to claim that an equilibrium at $\hat{\theta}_1$ will involve little welfare loss relative to θ^*.

The third case is rather more complex: the cost function $1/(\gamma + Q)$ differs from the others primarily in that the unit cost does not tend to infinity as the quantity moved tends to zero, but is bounded above by $1/\gamma$. The only point worth noting in this case is that if

$$\gamma/2rq = 1,$$

$$\hat{\theta}_1, \hat{\theta}_2 = \pm\sqrt{-2\delta l/r},$$

and both are imaginary. This is clearly still true for a range of parameter values near those for which $\gamma/2rq = 1$, and it demonstrates that in certain cases there may be no equilibrium.

The emphasis has so far been on pricing at average cost, and the extent to which this causes departures from the optimum. In principle, of course, pricing at marginal cost would ensure the decentralization of an optimum, but would lead to a loss because of the nonconvex production possibilities. Indeed, the reader can readily verify that for the functions $1/Q$ and $1/(\gamma + Q)$, variable and hence marginal costs are zero, so that the optimum can clearly be sustained by selling at marginal cost, leaving the total cost to be covered by taxation. It is interesting that in such an extreme case, pricing at average cost produces so little distortion.

It is particularly worth emphasizing the implications of this section for the pricing of public transportation services. It has often been remarked that requiring these to break even (i.e., price at average cost) leads to their being caught in a vicious cycle of declining demand and consequently rising cost and price. Interpreting the model as one of public and private transportation, we see that this is, in fact, precisely what happens in cases in which $\hat{\theta}_2$ is substantially less than θ^* or in which there is no equilibrium. People on the edge of a station's intended service area find it cheaper to travel privately, and the fall in demand leads to increases in average costs, further falls in demand, and so on. *The intuitively reasonable idea that a vicious cycle of falling demand and rising costs may force a cost-covering increasing-returns industry to contract to a significantly suboptimal level can be modeled rather simply.*

5. Individual profit maximization

■ It is natural to proceed now to a discussion of the market structure that would result from profit maximizing behavior. The easiest approach is to consider first a single firm facing no competition other than the possibility of consumers' buying from the center and to analyze the firm's profit-maximizing strategy. Suppose that it serves a market area of θ: the cost per unit of buying from the center is then $C(r\theta q)r$. The selling price P determines the market area

θ through the equation

$$P + \frac{\theta}{2} rC(q\delta l) = rC(q\delta l). \tag{15}$$

The left-hand side gives the cost a consumer on the edge of the market would incur by buying from the shop. The right-hand side gives the cost to him of buying directly: the edge of the market is by definition the point where these are equal. The shop's total profit is

$$(P - C(r\theta q)r)r\theta q. \tag{16}$$

Substituting for P from (15) and maximizing with respect to θ give

$$C(q\delta l)(1 - \theta) - r\theta q C'(r\theta q) - C(r\theta q) = 0. \tag{17}$$

Equation (17) is not easily analyzed in this general form, but for the function $C(Q) = \alpha/Q + \beta$ it yields a solution

$$\theta' = \frac{\alpha}{\alpha + \beta q\delta l} \leqq 1.$$

For small $\delta\theta$, this solution is larger than θ^* for the same function, so that an *individual profit-maximizing shop would seek to control an area greater in size than the optimal market area*. Perhaps it is worth pointing out that although the individuals' demands are inelastic, the demand faced by a firm is elastic because its market area declines as the price rises. It is this elasticity, whose magnitude is determined by the cost to the consumers of providing the service themselves, which determines the profit-maximizing price.

6. Joint profit maximization

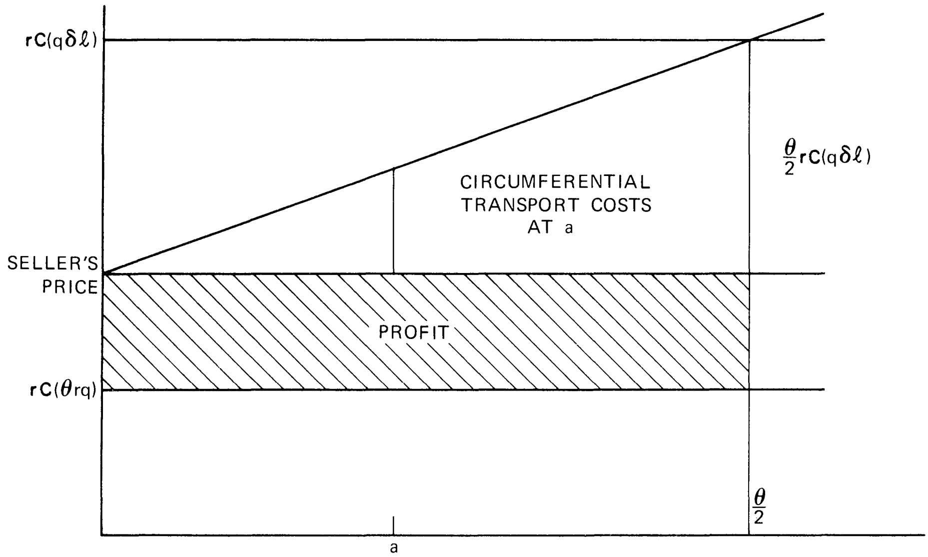
■ A natural extension is to consider the policy of a firm's owning a chain of retail outlets which it may position as it wishes around the perimeter. We assume the firm positions these and chooses pricing policies so as to maximize the total profits earned in the whole market. It can once again reasonably be assumed that the solution is symmetric with each shop's having a market area θ''. The boundary of a market area may be determined either by the now standard condition that at this point a consumer is indifferent between buying from the shop or from the center or by the condition that the consumer is indifferent between buying from two equidistant shops. In the latter case the common cost to the consumer of buying from the equidistant shops cannot exceed the cost of buying from the center (or the point under consideration would not be in either shop's area), but at the same time will clearly not be less if profits are being maximized. Hence, the boundary of the shops' market areas can once again be defined by the indifference of a consumer between buying from a shop or from the center.

Market areas are therefore again determined by (15), the profit of each firm by (16), and total profits are given by $2\pi/\theta''$ times (16), that is:

$$2\pi rq\left(C(q\delta l)r - \frac{\theta}{2} rC(q\delta l) - C(r\theta q)r\right).$$

This has an obvious interpretation: $2\pi rq$ is the total amount sold, and the

FIGURE 4



expression in brackets is the difference between price and cost. Differentiating with respect to θ gives as a first-order condition for a maximum

$$C'(r\theta''q) = -\frac{1}{2rq}C(q\delta l). \tag{18}$$

Recalling (8),

$$C'(r\theta^*q) = -\frac{\delta l C(q\delta l)}{4r^2 q\delta\theta},$$

it is clear that

$$\frac{C'(r\theta''q)}{C'(r\theta^*q)} = 2,$$

so that in general θ'' is much less than θ^*.

It is also important to note that the expression for total profit can be written

$$2\pi r^2 q\left(C(q\delta l)\left(1-\frac{\theta}{2}\right) - C(r\theta q)\right),$$

which makes it clear that a necessary condition for the attainment of positive profits is that $\theta < 2$. But with θ thus bounded above, $C(r\theta q)$ must rise as r falls to zero, and indeed for a cost function with fixed costs $C(r\theta q)$ will rise without limit as r falls. This suggests that there will typically be some minimum value of r below which the maximum profit level achievable is always negative, and below which the market will not be served. Figure 4 illustrates this point graphically. A retailer is located at position 0 and obtains goods at a cost of $rC(\theta rq)$. The consumer can obtain them directly for $rC(q\delta l)$, and θ is determined so that at $\theta/2$ the seller's price plus transportation costs along the circumference equal this direct cost. As r is reduced, $rC(q\delta l)$ and $(\theta/2)rC(q\delta l)$ both fall, but $rC(\theta rq)$ may rise, thus squeezing profits from both sides and causing them to vanish at a positive radius. At such a point, the market will cease to be

TABLE 2

	θ^*	θ''
$\frac{1}{Q}$	$2(\delta\theta)^{\frac{1}{2}}$	$(\frac{2\delta\ell}{r})^{\frac{1}{2}}$
$\frac{\alpha}{Q}+\beta$	$(\frac{4\alpha\delta\theta}{\alpha+\beta q\delta\ell})^{\frac{1}{2}}$	$(\frac{2\alpha\delta\ell}{(\alpha+\beta q\delta\ell)r})^{\frac{1}{2}}$

served by profit-maximizing firms, even though the existence of a retail system will still be socially optimal.

To illustrate these points, we present the values of θ^* and θ'' for two specific functions in Table 2. In each case the ratio θ^*/θ'' equals $(2)^{1/2}$, with θ^* much larger than θ''. For the function $C(Q) = 1/Q$, it is also relatively easy to confirm that for $r \leqq 1/2$, the maximum attainable profits under the joint profit-maximizing solution are negative, thus ensuring that the market would not be supplied.

It is clear that for each of the cost functions discussed,

$$\theta'' \to 0 \qquad \text{as} \qquad \delta\theta = \delta l/r \to 0.$$

That is, the profit-maximizing market area shrinks towards zero as the significance of individual consumers does likewise—or as the radius rises without limit. In either case the density of demand per radian is rising, so that it is perhaps not surprising that the angular density of shops also rises. In terms of Figure 4, the vertical distance $rC(q\delta l)$ rises, as does the slope of the line showing circumferential transportation costs, and the balance is such that their intersection moves to the left.

The following proposition summarizes the results of this section.

Proposition 1. Under the joint profit-maximizing regime, there are two possibilities. If the market is big, there will be more shops with smaller market areas than is socially optimal. In the case of a small market, however, there will be no shops, even though optimality requires a positive number.

Heuristically, one might say that in large markets there is excessive product differentiation via location, and in small markets product differentiation is insufficient. That such an outcome should result from economies of scale is not intuitively surprising: scale economies obviously increase the attractiveness of large relative to small markets.

7. Price discrimination via free delivery

■ In the model under discussion, the results above show that joint profit maximization does not lead to the socially optimal outcome. It has often been conjectured that in such situations allowing the monopolist to practice price discrimination will restore efficiency, and it will be shown below that in the context of the present model this is certainly true. Spence (1976) also demonstrates this result for his model. A more general set of results on the efficiency of discriminatory pricing in a general equilibrium model with increasing returns in production appears in Brown and Heal (1980). The results of that paper and of Brown and Heal (1979) imply that efficiency cannot always be attained by discriminatory pricing in a general equilibrium model with nonconvexities. In general, of course, this type of result is of only limited interest, because effective and complete price discrimination is a very complex form of behavior, requiring a great deal of information about the demand pattern of each consumer. Spence quite reasonably describes it as unrealistic in the context of his model. However, in the present model matters are simpler.

In the current model, complete price discrimination can be achieved by setting the price to each consumer at slightly less than the cost to him of buying directly from the center, and this, in turn, can be done by the retail chain's quoting a uniform delivered price at this level to all consumers and meeting all transportation costs itself. It should be intuitively very clear that this will lead to the efficient outcome. Once it has been decided to sell to each consumer at fractionally less than the cost to him of buying from the center, the chain's total revenue is fixed, and maximizing profits is equivalent to minimizing the transportation costs incurred in meeting total consumer demand—and this is precisely the problem that was used to define the socially optimal outcome in Section 3.

There are two stages involved in demonstrating the result at issue. The first is to show that a policy of price discrimination such as that outlined is indeed profit-maximizing, and the second is to confirm that such a policy leads to the socially optimal market areas. Although demonstrating that the suggested policy is profit-maximizing among the set of all conceivable policies involves technicalities that do not seem worth tackling, it is routine to show that it is best among the set of policies which take the form of charging consumers at an angular distance α from the shop a shop price P_s plus some fixed fraction K of transport costs:

$$P(\alpha) = P_S + KC(q\delta l)r\alpha.$$

By computing total profits from this rule it is easy to verify that the firm should set $K = 0$ and $P_s = C(q\delta l)r$, as suggested.

It remains to show that this policy leads to the correct market areas. This is again simple: the total revenue to the retail chain is $2\pi rq(C(q\delta l) - \epsilon)$, where ϵ is the amount by which it undercuts direct buying from the center. Under the assumption that deliveries to separate customers are separate and independent journeys, so that there are no circumferential economies of scale, the total costs per shop are

$$C(r\theta q)r\theta qr + 2\int_0^{\theta/2} C(q\delta l)r\alpha\,\frac{\delta l}{\delta\theta}\,d\alpha,$$

giving total costs to the chain of

$$2C(r\theta q)r^2q\pi + C(q\delta l)q\,\frac{\delta l}{\delta\theta}\,r\,\frac{\theta}{2}\,\pi.$$

As revenues are independent of θ, maximizing profit is equivalent to minimizing this term, which is identical with equation (7), whose minimum defines the socially optimal market area θ^*, thus establishing the desired result.

It is perhaps worth observing that the assumption that circumferential deliveries are independent journeys yielding no economies of scale is almost certainly unreasonable in the present context. One would expect the retail chain to organize deliveries so as to minimize the circumferential costs for any given market size, thereby leading to a lower total than the optimum reported in Section 3, though if the assumption of that section about circumferential transport arrangements were appropriately revised, the two outcomes would again coincide.

One final point that is worth emphasizing about the discussion of this section is the interpretation of price discrimination via free delivery when the model is applied to the transportation problem discussed earlier. Uniform pricing in this case means that everyone has to pay the same fare to reach the center, whatever the distance of his residence from a station. Such a policy could, for example, be implemented by a fixed-fare circumferential public transportation system, so that the total fare for a journey to the center is independent of its point of origin.

8. Free entry

■ In the model under consideration the products of sellers are necessarily differentiated by location, and the number of sellers is necessarily finite. Hence, there is little meaning to a perfectly competitive solution. The most appropriate analogue to this is almost certainly a free-entry solution, i.e., the outcome of a process in which firms continue to establish new shops until the profits of all are driven down to zero. Let $\hat{\theta}$ be the market area of a representative shop in such an equilibrium. Then its selling price must be $C(r\hat{\theta}q)r$, and its market area must be given by

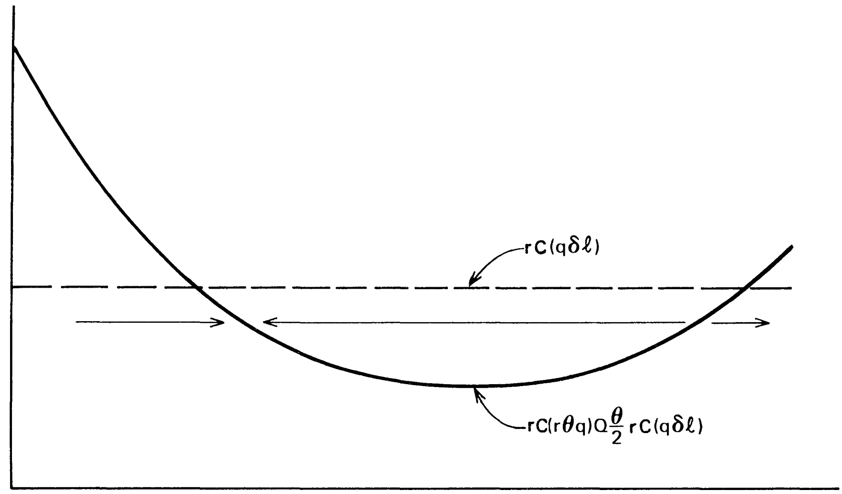
$$C(r\hat{\theta}q)r + r\frac{\hat{\theta}}{2}C(q\delta l) = rC(q\delta l). \tag{14}$$

The reader will note that this is in fact simply equation (14), used earlier to define the equilibrium market areas in a regime of average cost pricing. It is, of course, clear that the two equilibria will be the same, as both involve the same pricing rule.

The analysis of Section 4 now gives us complete information about the equilibrium possibilities in the present case. Typically, there will either be two equilibrium market sizes or none (see Table 1). If there are two possible equilibria, then for the cost functions analyzed, one of these may be quite similar to the optimal outcome; if there are none, then it is clear from Figure 3 that a firm competing with the possibility of consumers' buying directly from the center can never make a profit ($rC(q\delta l)$ is always less than average cost), and no firm will enter. As the parameters of the system are varied, the two distinct equilibria may fuse and vanish—an interesting example of a bifurcation.

In Section 4 it was suggested that of the two possible equilibria, the one with the larger market area was stable. The reasoning was that a firm pricing at average cost with an initial market area in excess of $\hat{\theta}_2$ would find its market contracting, and that firms whose areas were initially between $\hat{\theta}_1$ and $\hat{\theta}_2$, and who competed only with the possibility of buying from the center, would likewise find their market growing. In the present case it is no longer clear that these arguments are valid. Although firms will be pricing at average cost at a zero-profit equilibrium, it is not clear that they will do so out of equilibrium. For example, they could price at $rC(q\delta l)$, the price of buying from the center. In this case $\hat{\theta}_2$ is likely to be unstable and $\hat{\theta}_1$ stable. The point is that under such a policy, a market area between $\hat{\theta}_1$ and $\hat{\theta}_2$ would generate positive profits, attracting new entrants and driving down market size to $\hat{\theta}_1$. On the other hand, areas

FIGURE 5



less than $\hat{\theta}_1$ or greater than $\hat{\theta}_2$ would generate losses, forcing firms out of business and increasing market areas. This is shown schematically in Figure 5.

An alternative assumption would be that in the situation under consideration, with shops competing with both existing and potential entrants for markets, prices would be set at or near average cost, partly to deter potential entrants by a form of limit pricing and partly to realize the competitive advantages of economies of scale. In this case the outcome will depend on initial conditions, but if the initial number of shops is less than $2\pi/\hat{\theta}_2$, the outcome will be an equilibrium with areas of $\hat{\theta}_2$. The need to recognize the role of initial conditions arises because if the number of shops initially present exceeds $2\pi/\hat{\theta}_2$, an equilibrium with areas $\hat{\theta}_2$ cannot be established without the elimination of some firms. In such a process firms are likely to react to each others' policy choices, and it will be more appropriate to consider a game-theoretic equilibrium concept such as that of Nash. As will be shown in the next section, this raises problems of its own.

The discussion of free entry can be summarized in the following terms. *If the market is small, no firms will enter. If it is large, there are two possible equilibrium configurations, involving either large numbers of small shops or small numbers of large shops.* In the case of the cost functions considered in detail, both equilibria will involve market sizes less than the optimal, although the large-number equilibrium may be nearly optimal. It is not clear which of these equilibrium configurations is the more stable: this depends on the adjustment process and initial conditions assumed.

9. Nash equilibria

■ We have already indicated that it may be appropriate to consider the outcome of strategic interactions between competing shops. This clearly opens up a very wide range of possibilities, and we shall not attempt a comprehensive exploration here. Instead, we consider one possible outcome, a Nash equilibrium. Indeed, we shall only focus on symmetric Nash equilibria; it is not at present clear whether this represents a real restriction.

Consider a shop located at position 0, with nearest neighbors located at angular distances $\pm\,\psi$, each charging a price P. A consumer located at angular distance ϕ from the shop ($\phi < \psi$) will face a cost of

$$P + (\psi - \phi)rC(q\delta l)$$

in buying from the nearest competitor. It follows that if the original firm charges a price P_0, its market angle θ will be determined by

$$P_0 + \frac{\theta}{2} rC(q\delta l) = P + \left(\psi - \frac{\theta}{2}\right) rC(q\delta l) \tag{19}$$

or

$$\theta = \frac{P - P_0}{rC(q\delta l)} + \psi, \tag{20}$$

which has the reassuring property that $\theta = \psi$ when $P = P_0$. The total profit of the firm under consideration, given the location of the nearest neighbors and their prices, will be

$$(P_0 - C(r\theta q)r)r\theta q = (P + \psi rC(q\delta l) - \theta rC(q\delta l) - C(r\theta q)r)r\theta q.$$

Maximizing with respect to θ produces the first-order condition:

$$\frac{P}{r} + C(q\delta l)(\psi - 2\theta) - C(r\theta q) - rC'(r\theta q)q = 0. \tag{21}$$

Equation (21) defines a value of θ (and hence via (20) a value of P_0) which represents a profit-maximizing response to the price P charged by neighbors located at $\pm\psi$. For the function $C(Q) = 1/Q$, (21) implies that

$$\theta = \frac{Pq\delta l}{2r} + \frac{\psi}{2}, \tag{22}$$

and by restricting attention to symmetric equilibria where $\theta = \psi$, we find that

$$\theta = \psi = \frac{Pq\delta l}{r}. \tag{23}$$

The analysis has now given us details of a Nash equilibrium that satisfies a symmetry condition. The equilibrium consists of shops with market areas θ charging prices P related to θ by (23). It should be noted that there is nothing in (21), or its particular solution (22), which requires symmetry and $\theta = \psi$: this is an additional condition being imposed because it seems *a priori* reasonable that the equilibrium locations for technologically identical firms in a symmetric space with uniform demand should also be symmetric—or at least that the symmetric equilibria should be of particular interest. Given the symmetry restriction, and given a market described by values of r, q, δl and a cost function $C(Q) = 1/Q$, it follows that corresponding to a market size θ, there exists a price P, given by (23), such that shops charging price P with market area θ constitute a Nash equilibrium.

This clearly implies that the symmetric Nash equilibrium is far from unique. Nevertheless, there are limits within which θ must lie if there is to be a price P, given by (23), such that θ and P form a Nash equilibrium. An upper bound on θ exists because firms have to compete with the possibility of a consumer's buying directly from the center, and this places an upper limit on the market area sustainable. There is a lower limit because unless θ is at least a certain size, it is not possible to make a positive profit.

For the particular cost function considered above, we can easily derive these limits. The requirement that a consumer on the edge of the market must not prefer to buy from the center can be stated as:

$$C(q\delta l)\,\frac{\theta}{2}\,r + P \leqq C(q\delta l)r,$$

and substituting from (23) and $C(Q) = 1/Q$, we have:

$$\theta \leqq 2/3.$$

A shop's profit is given by

$$\pi = (P - C(r\theta q)r)r\theta q,$$

so that $\pi \geqq 0$ requires

$$\theta^2 \geqq \delta l/r.$$

It is now clear that for the cost function $C(Q) = 1/Q$, there is an infinity of symmetric Nash equilibria having market areas θ and prices $r\theta/q\delta l$, for values of θ satisfying $\theta \leqq 2/3$ and $\theta^2 \geqq \delta l/r$. *These bounds of course imply that all of these equilibria involve market areas considerably below the optimum. They also imply that imperfectly competitive firms whose interactions are represented by symmetric equilibria will not serve small markets.* It should be noted that profits will differ across these Nash equilibria, and will be greatest for those with the largest market area and price. If one imposes the constraint that because of the possibility of entry, firms in the industry wish to hold the profit level below some preassigned level, then this may impose an additional degree of determinancy on the solution. A limiting case of this would be the zero-profit condition.

10. Conclusions

■ We have applied the model introduced in this paper to a wide range of issues, and we shall make no attempt here to summarize all that has emerged. However, it is clear that one can easily adapt it to analyze certain efficiency issues relating to retailing, transportation, and product differentiation in general.

There are two conclusions that I would like to highlight. One is that with increasing returns, whether there is excessive or insufficient product differentiation depends on the scale of the market. In terms of variety, small markets are underserved and large markets overserved. Anyone whose tastes or needs are at all nonstandard will no doubt find that this is substantiated by casual empiricism. The other, rather different, conclusion that appears to merit emphasis is that the impact of enforcing a cost-covering policy may be very destructive. It can lead to an equilibrium very different from the optimum or even to the nonexistence of an equilibrium.

Obviously, the clarity with which these conclusions emerge has much to do with the very simple structure of the model. However, it seems likely that these results would be robust to considerable relaxations of the assumptions.

References

BROWN, D.J. AND HEAL, G.M. "Equity, Efficiency, and Increasing Returns." *Review of Economic Studies* (October 1979).

——— AND ———. "Two-Part Tariffs, Marginal Cost Pricing, and Increasing Returns in a General Equilibrium Framework." *Journal of Public Economics* (1980).

EATON, B.C. AND LIPSEY, R.G. "The Principles of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition." *Review of Economic Studies* (January 1975), pp. 27–50.

HOTELLING, H. "Stability in Competition." *Economic Journal* (March 1929), pp. 41–57.

LEWIS, W.A. "Competition in Retail Trade." *Economica* (1945), pp. 202–234.

LERNER, A.P. AND SINGER, H.W. "Some Notes on Duopoly and Spatial Competition." *Journal of Political Economy* (April 1937).

SALOP, S. "Monopolistic Competition with Outside Goods." *Bell Journal of Economics* (Spring 1979).

SHAKED, A. "Nonexistence of Equilibrium for the 2-Dimensional 3-Firm Location Problem." *Review of Economic Studies* (January 1975), pp. 51–56.

SPENCE, A.M. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* (June 1976).

STERN, N. "The Optimal Size of Market Areas." *Journal of Economic Theory* (1972).

WILSON, R. "Informational Economies of Scale." *Bell Journal of Economics* (Spring 1975).