NBER WORKING PAPER SERIES

MARKUPS AND ENTRY IN A CIRCULAR HOTELLING MODEL

Robert J. Barro

Working Paper 32660
http://www.nber.org/papers/w32660

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024, Revised December 2024

I have benefited from comments by Ed Glaeser, Elhanan Helpman, Casey Mulligan, Steve Salop, Chad Syverson, and Jonathan Vogel. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Markups and Entry in a Circular Hotelling Model
Robert J. Barro
NBER Working Paper No. 32660
July 2024, Revised December 2024
JEL No. L1, L12, L13

ABSTRACT

The circular version of Hotelling's locational model is extended by incorporating a continuum of consumers with constant-elasticity demand functions along with stores that have constant marginal costs of production. The stores are evenly spaced in equilibrium. The model implies that the markup of price over marginal cost depends on the spacing between stores and a transportation-cost parameter but is, as an approximation, independent of the elasticity of demand. This result reflects pricing decisions by stores that factor in the threat of losing business entirely at the borders with neighboring stores. This model provides a theory of price markups that substitutes for, or at least supplements, the familiar Lerner approach, which puts all the weight on the elasticity of demand. Moreover, the pricing results apply even when the magnitude of the elasticity of demand is less than one. The price markups determine the extent of entry into the market and, thereby, the efficiency of market outcomes. Entry is excessive when the approximate markup formula is accurate.

Robert J. Barro
Department of Economics
Littauer Center 218
Harvard University
Cambridge, MA 02138
and NBER
rbarro@harvard.edu

Hotelling (1929) constructed a model of monopolistic competition in which stores were located along a finite straight line and consumers purchased goods from their nearest store. Because of locational advantages, stores had some degree of monopoly power, and this power entered into their pricing decisions. The circular version of the Hotelling model, usually labeled the Salop model and attributed to Salop (1979),[1] has major technical advantages because it avoids issues related to the end points of a finite straight line.[2] The present analysis extends this model by allowing for a continuum of consumers with constant-elasticity demand functions along with stores that have constant marginal costs of production.

The finite straight line is an attractive framework in some contexts because the end points represent extremes of characteristics such as preferences of political parties about the size of government (Downs [1957]) or of religions about the degree of strictness (Barro and McCleary [2005]). However, for most applications in industrial organization, these end-point features do not apply, and the circle model provides a more useful setting.

A central feature of the model is that each store's pricing behavior involves competition from neighboring stores. From the perspective of customers, the closest substitutes for each store are the adjacent ones. This interaction between stores means that the threat of losing business at borders, rather than the elasticity of demand, is the key force in the determination of price markups. These margins of competition provide a theory of price markups that substitutes for, or at least supplements, the familiar Lerner formula, which is based on the elasticity of

---

[1]The circular framework was used earlier by Vickrey (1964, Ch. 8), who referred to "… a simple loop (which may or may not be circular) …" Schmalensee (1978) also used a circular setting.

[2]Salop (1979, p. 142) says: "The product space of the industry is taken to be an infinite line or the unit-circumference of a circle. While neither assumption is realistic, both allow the 'corner' difficulties of the original Hotelling model to be ignored and an industry equilibrium with identical prices by equally-spaced firms to obtain." Hay (1976, p. 243) relied on a very long straight line: "For most of the analysis we assume that the market is very long in comparison with the market size of the firm. This enables us to avoid the analytic complications of market intervals at the end points of the market."

demand. Moreover, the pricing results apply even when the magnitude of the elasticity of demand is less than one. The price markups ultimately determine the extent of entry into the market and, thereby, the efficiency of market outcomes.

I. Circular Hotelling Model

Consumers are distributed uniformly around a circle with circumference H, as shown in Figure 1. A number N of stores service the consumers. Each store j produces goods at the same constant marginal cost, c.[3] (Hotelling [1929, p. 45] assumed that this marginal cost was zero.) Store j prices at P_j. In addition to paying P_j for each unit of goods, a customer pays an amount per unit that increases linearly with the distance, z, from its chosen store, as in Hotelling (1929, p. 45). The transportation cost per unit is tz, where $t>0$ is a transport-cost parameter that depends on transportation technology and the value of customer time. As stressed by Hotelling (1929, p. 45), "transportation cost" should be viewed as a metaphor for many differentiating characteristics of goods and stores that cause customers to prefer one seller over another for a given price charged.[4]

As in Salop (1979), the stores are evenly spaced around the circle in Figure 1. This even spacing is an equilibrium outcome in the present model. The spacing between stores, denoted $2h$ in the figure, will satisfy the condition

$$2h = \frac{H}{N}. \tag{1}$$

[3]These costs can be measured in units of a market basket of final products. Transportation costs, fixed costs, and prices of individual goods can be measured in the same units.
[4]Hotelling (1929, p. 45) says: "… there will be many causes leading particular classes of buyers to prefer one seller over another, but the ensemble of such consideration is here symbolized by transportation costs."

Store 1 is adjacent to stores 2 and N, as shown. Going to the right from its location, store 1 services customers out a distance that will ultimately equal h, the mid-point of the distance from store 2. For now, this market distance is denoted by h_{12}. Similarly, going to the left, store 1 services customers out a distance denoted h_{1N}, which will ultimately equal h. In the initial analysis, the number of stores, N, is given. However, in a full equilibrium, N and, hence, $2h$ in Eq. (1) will be determined from a free-entry condition.

Store j chooses the price, P_j, of its good for given prices and locations of the other stores (as in a Bertrand analysis). Without loss of generality, take $j=1$ and consider only the market going to the right in Figure 1, where store 2 is the adjacent alternative. (The results that include movements to the left, toward store N, will be analogous.) An assumption is that the market border between stores 1 and 2, located at distance h_{12} from store 1, is in the interior between the two stores; that is $0 < h_{12} < 2h$ applies. This assumption is later demonstrated to be valid.

Denote by $q(z)$ the quantity purchased of store 1's goods by customers located at distance z to the right of the store. The effective price per unit faced by a customer at z, where $0 \leq z \leq h_{12}$, adds the linear transportation cost, tz, to the price set by store 1:

$$P_1^*(z) = P_1 + tz. \tag{2}$$

The parameter t is the same for each store. An assumption is that store 1 charges each customer the same price regardless of the customer's location (perhaps because the customer's residence is unknown to the store).

Transportation costs can be reinterpreted as costs from consuming a good with characteristics that deviate from a customer's ideal type, as in Lancaster (1966) and Baumol (1967). As noted by Schmalensee (1978 p. 309), " … the formal correspondence between Lancastrian models with two characteristics and one-dimensional spatial models is almost exact." Helpman (1981, part 2) applied the Lancaster-type model to the Salop-style circle framework, with arc distances around the circle representing differences in product characteristics.

Equation (2) implies that total transportation expenses are proportional to the quantity transported, *q(z)*. In some contexts, such as driving to and from a grocery store, these charges might be independent of *q(z)* over some range. However, it seems reasonable for most applications to view transportation charges, including the analogous costs in Lancaster-type models, as proportional to quantity.[5]

The range of store 1's territory to the right of its location applies for z going from 0 to h_{12}. Within this range, the quantity demanded of store 1's goods by a customer at distance z depends only on the effective price, $P_1^*(z)$, and is assumed to take a constant-elasticity form:[6]

(3) $$q_1(z) = A \cdot [P_1^*(z)]^{-\eta},$$

where $\eta \geq 0$.[7] The standard analysis requires $\eta > 1$, but this restriction is ultimately unnecessary in the present model. The parameter $A > 0$ represents the scale of each consumer's demand. The parameters A and η are the same for each store.

---

[5]In contrast, Anderson and de Palma (2000) and Gu and Wenzel (2009) assume that transport charges do not depend on the quantity transported. This assumption means that the effective price paid per unit of goods is independent of distance, z.

[6]Anderson and de Palma (2000) and Gu and Wenzel (2009) show that this form of demand function is implied by a utility function of the form $U = \frac{\eta}{\eta-1} q(z)^{(\eta-1)/\eta} + \hat{q}$, where $\hat{q}$ is the quantity of a numeraire good and gross income is $Y = P^*(z) \cdot q(z) + \hat{q}$. If all consumers have the same Y, customers located at higher z will effectively have lower net income because they spend more on transportation costs for a given *q(z)*. However, this net income difference would be exactly offset by lower residential rents if the only dimension of residential location that matters is its distance from the nearest store. This result accords with the standard insight in urban economics, summarized in Brueckner (1987), whereby differences in rents represent compensating differentials for differences in commuting costs.

[7]Hotelling (1929, pp. 45, 56) assumed $\eta = 0$. He says (p. 56): "The problem … might be varied by supposing that each consumer buys an amount of the commodity in question which depends on the delivered price. If one tries a particular demand function the mathematical complications will now be considerable … " Smithies (1941) and Hay (1976) extended the Hotelling analysis to allow for elastic consumer demand, though in the form of linear demand functions. Anderson and de Palma (2000) and Gu and Wenzel (2009) allowed for a constant elasticity of demand, as in Eq. (3), but their restriction to $\eta < 1$ limits the applicability of their setting. See also n.5.

A modeling device that is convenient for aggregating heterogeneous buyers treats $q_1(z)$ as contributing an infinitesimal amount to the total quantity of goods demanded, Q_1, from store 1; that is,

$$\frac{dQ_1}{dz} = q_1(z).$$

The total quantity, Q_1, sold by store 1 to customers to the right of the store's location equals the integral of $q_1(z)$ for z going from 0 to the market border, h_{12}:

$$Q_1 = \int_0^{h_{12}} q_1(z)dz.$$

Using the demand curve from Eq. (3) and making a change of variable from z to $P^*(z)$, the integral can be evaluated to get

(4) $$Q_1 = \frac{A}{t(\eta-1)}\left[P_1^{1-\eta} - (P_1 + th_{12})^{1-\eta}\right],$$

where $\eta>1$ is assumed at this stage but is ultimately not required.

The marginal effect of P_1 on Q_1 follows from Eq.(4) as

(5) $$\frac{\partial Q_1}{\partial P_1} = -\left(\frac{A}{t}\right)\cdot\left[P_1^{1-\eta} - (1 + t\frac{\partial h_{12}}{\partial P_1})(P_1 + th_{12})^{-\eta}\right].$$

The essence of the Hotelling analysis is the term $\frac{\partial h_{12}}{\partial P_1}$, which indicates how an increment in P_1 affects the location of the market border between stores 1 and 2. However, to see the relation to the standard Lerner formula for markup pricing, it is useful to begin with the counter-factual assumption that $\frac{\partial h_{12}}{\partial P_1} = 0$.

The profit flow for store 1 (for sales to the right in Figure 1) is

(6) $$\pi_1 = (P_1 - c)Q_1 - f,$$

where $f > 0$ is the fixed cost of operating a store. The parameter f is the same for each store. The fixed cost could correspond to the rent paid on the store's premises. In the subsequent analysis, this fixed cost also covers sales to the left of store 1 in Figure 1.

The first-order condition for choosing P_1 to maximize π_1 follows from Eq. (6) as

$$(7) \qquad \frac{\partial \pi_1}{\partial P_1} = Q_1 + (P_1 - c)\frac{\partial Q_1}{\partial P_1} = 0.$$

Substituting into Eq. (7) for Q_1 from Eq. (4) and for $\frac{\partial Q_1}{\partial P_1}$ from Eq, (5) (assuming $\frac{\partial h_{12}}{\partial P_1} = 0$) leads, if $\eta>1$, to the first-order maximization condition:

$$(8) \qquad \left(\frac{1}{\eta-1}\right) \cdot \left[P_1^{1-\eta} - (P_1 + th_{12})^{1-\eta}\right] = (P_1 - c)\left[P_1^{1-\eta} - (P_1 + th_{12})^{-\eta}\right].$$

The results simplify if the maximal operative transport cost, th_{12}, is small compared to the price, P_1; that is, if the main component of the effective price, $P_1^*(z)$ in Eq. (2), is always the price charged, P_1, rather than the transport cost, tz. This approximation, which is explored in detail in the next section, allows for simplifications of the terms involving $P_1 + th_{12}$ in Eq. (8) by means of first-order Taylor approximations. These approximations deliver the standard result, described in Lerner (1934), whereby P_1 is a constant markup on marginal cost, c:[8]

$$(9) \qquad P_1 \approx c \cdot \left(\frac{\eta}{\eta-1}\right),$$

so that the markup ratio is

$$(10) \qquad \frac{P_1 - c}{c} \approx \frac{1}{\eta-1}.$$

Therefore, when the effect of P_1 on the location of the market border is neglected, the markup ratio is determined in the usual way by the elasticity of demand, η, with a higher elasticity

[8]The second-order condition for profit maximization is satisfied if $\eta>1$.

implying a lower markup ratio. Finite price, P_1 and profit, π_1, require $\eta>1$. This restriction on the elasticity of demand is familiar but not necessarily realistic.

The price, P_1, shown in Eq. (9) is the one chosen by store 1. However, the same price applies for each store around the circle in Figure 1 as long as η and c are the same in each territory. The analysis now turns to the crucial influence on pricing behavior from the interactions at the market borders between stores.

II. Hotelling Effect at the Borders

At the market border between stores 1 and 2, where $z=h_{12}$, customers are indifferent between buying goods from store 1 or store 2. An assumption is that, aside from differences in "transportation costs," the goods offered by stores 1 and 2 (and the other stores) are viewed as identical by all consumers. Therefore, in order for customers to be indifferent to patronizing store 1 or store 2 at the market border, the prices inclusive of transportation costs, P_1^* and P_2^*, must be equal at the border. That is,

$$(11) \qquad P_1 + th_{12} = P_2 + t \cdot (2h - h_{12}),$$

where $2h - h_{12}$, already assumed to be positive, is the distance of the market border from store 2. Starting from a position where Eq. (11) holds, an increment in P_1, for given P_2, requires a compensating adjustment of the border position, h_{12}. Specifically, Eq. (11) implies that the connection of the border position to an increment in the price is[9]

$$(12) \qquad \frac{\partial h_{12}}{\partial P_1} = -\frac{1}{2t}.$$

Substituting the result from Eq. (12) into the expression for $\frac{\partial Q_1}{\partial P_1}$ in Eq. (5) implies

[9]This analysis requires $N \geq 2$, so that $h \leq H/4$. Otherwise, there is no border competition, and the solution is the standard one given in Eqs. (9) and (10).

(13) $$\frac{\partial Q_1}{\partial P_1} = -\left(\frac{A}{t}\right) \cdot [P_1^{-\eta} - \tfrac{1}{2}(P_1 + th_{12})^{-\eta}].$$

Substitution of this result into the expression for $\frac{\partial \pi_1}{\partial P_1}$ in Eq. (7) yields the new first-order maximization condition:

(14) $$\left(\frac{1}{\eta-1}\right) \cdot \left[P_1^{1-\eta} - (P_1 + th_{12})^{1-\eta}\right] = (P_1 - c) \cdot [P_1^{-\eta} - \tfrac{1}{2}(P_1 + th_{12})^{-\eta}].$$

The results again simplify if the maximal operative transport cost is small compared to the price; that is, if $th_{12} \ll P_1$ holds. In this case, Eq. (14) simplifies using first-order Taylor approximations for the terms involving $P_1 + th_{12}$ to

(15) $$P_1 \approx c + 2th_{12}.$$

This result turns out, as explored below, also to require η not be too large. The second-order condition for profit maximization is satisfied if $\eta \geq 0$. Hence, the solution for the price in Eq. (15) holds even when $\eta \leq 1$. In the standard, Lerner-type analysis, this range for η has to be excluded to avoid infinite prices and profits by producers.[10]

It is now possible to relate the market border position, h_{12}, to the spacing, $2h$, between stores 1 and 2. Equation (15) holds for pricing by store 1 and an analogous condition holds for pricing by store 2:

(16) $$P_2 \approx c + 2t \cdot (2h - h_{12}),$$

where $2h - h_{12}$ is the (positive) distance from store 2's position to the market border with store 1. Combining Eqs. (15) and (16) with Eq. (11) yields the intuitive result that the market

[10]The key mechanism in the Hotelling model is that a rise in P_1 contracts the border, h_{12}, with h_{12} reaching zero for sufficiently high P_1. When $\eta \leq 1$, this mechanism prevents a store from achieving infinite profit by charging an infinite price.

border is equidistant between stores 1 and 2; that is, $h_{12}=h$. Correspondingly, the prices charged by the two stores are the same:

$$P_1 = P_2 = P \approx c + 2th. \quad (17)$$

The analysis thus far assumed that the distance from store 1 to the market border with store 2, h_{12}, is in the interior between the stores, so that $0<h_{12}<2h$ applies. Another possibility is that store 1 would undercut store 2 by pricing so as to move the market border, h_{12}, up to or beyond $2h$.[11] To be attractive to buyers at store 2's location (and at locations further beyond store 2), P_1 would have to be set below P_2 by at least the extra transport cost, *2th*. But if $P_2 = c + 2th$, as in Eq. (17), this undercutting requires P_1 to be set at or below marginal cost, *c*, an outcome that would be unprofitable for store 1. Therefore, it was satisfactory to assume an interior equilibrium where $0<h_{12}<2h$.

Table A1 in the appendix examines the accuracy of the approximations that underlie Eq. (17). Equation (14) was solved numerically to get the solution for P_1—that is, for *P*—given a specification of the underlying parameters, which can be expressed as η and *2th/c*. The latter term is the ratio of markup to marginal cost in the formula in Eq. (17). Note from Eq. (14) that Eq. (17) holds exactly if $\eta=0$. Equation (17) is also exact if $th=0$. More generally, as shown in Table A1, part 1, Eq. (17) more closely approximates the true solution for *P* when η and *2th/c* are smaller. Quantitatively, for ranges of the parameters that seem "reasonable"— such as $\eta\leq5$ and $2th/c \leq 0.1$ or $\eta\leq2$ and $2th/c\leq0.2$—the solution for *P* falls short of $c+2th$ by less than 3%. Therefore, in plausible ranges for the parameters η and *2th/c*, the formula for *P* in Eq. (17) yields a close approximation to the true solution.

[11] Vogel (2008) analyzes this type of undercutting in a model of spatial competition.

Table A1, part II compares the model's numerical solution for P with that from the Lerner formula in Eq. (9), $P = c \cdot \left(\frac{\eta}{\eta-1}\right)$, which is well defined only if $\eta > 1$. The Lerner formula provides a good approximation to the model's results only for very large η. This formula also tends to fit better the larger *2th/*c. However, in the ranges of "reasonable" parameter values noted above—$\eta \leq 5$ and $2th/c \leq 0.1$ or $\eta \leq 2$ and $2th/c \leq 0.2$—the Lerner formula gives poor results.

Another way to look at the pricing result in Eq. (17) is that, for a given value of *2th*, a change in marginal cost, *c,* passes through one-to-one to *P*.[12] In the language of Sangani (2024), the model predicts 100% pass-through in levels. The empirical analysis in his paper (Tables 4 and 5) confirms this relationship for retail gasoline and an array of retail food prices. In contrast, the Lerner formula in Eq. (9) implies that P is proportional to c for given η. Sangani (2024) refers to this case as 100% pass-through in logs and shows that it is inconsistent with his data.

In the literature on price markups, many studies mention the Lerner formula but make no use of it empirically. Examples of this research strategy include Hall (2018, p. 2), Bond, et al. (2021, p. 4), and Syverson (2024, p. 5). Given the poor performance of the Lerner formula in Hotelling-type models (as in Table A1, part II), it makes sense that this formula would not be relied on for empirical analyses of price markups. Instead, the recent applied literature, exemplified by de Loecker, Eeckhout, and Unger (2020), focuses on how to use available data to measure price markups, following the approach pioneered by Hall (1988). Although these results are useful, they do not relate the measured markups to "fundamentals," which include differences in the elasticity of demand and in factors that affect the substitutability of

[12]Profit declines because the rise in P implies a decrease in quantity, Q.

neighboring products. In the Hotelling-type model, the latter group comprises influences on the spacing between stores (or products), *2h*, and the transportation-cost parameter, *t*.

There is a sense in which the results for the markup ratio in the present version of the Hotelling model accord with the standard formula based on the elasticity of demand in Eq. (10). The standard result holds if η is interpreted not as the parameter in the individual demand curve in Eq. (3) but rather as the full magnitude of the elasticity of *Q* with respect to *P*—including the extensive margin whereby a store loses all of the business at its market borders by raising its price. In this setting, the micro-level elasticity of demand, η—effectively the intensive margin for quantity demanded—does not affect the markup ratio as an approximation for a range of reasonable parameter values.

Another way to view the results is in terms of the model's prediction about the size of the full magnitude of the elasticity of *Q* with respect to *P*. When the approximate formula for *P* in Eq. (17) is valid, the equation for *Q* from Eq. (4) can be approximated as

(18) $$Q \approx AhP^{-\eta}.$$

The overall elasticity of *Q* with respect to *P* then involves the usual $-\eta$ term, along with an additional negative effect coming from the reduction in the market border, *h*, based on Eq. (12) (combined with the condition $h_{12}=h$). The result is that the full elasticity is given by

(19) $$\frac{\partial Q}{\partial P} \cdot \frac{P}{Q} \approx -\eta - \frac{c+2th}{2th}.$$

The standard Lerner analysis, which ignores the reduction in *h*, requires $\eta>1$ to get meaningful results. However, since the magnitude of the final term on the right-hand side of Eq. (19) exceeds one, the magnitude of the full elasticity exceeds one for any $\eta \geq 0$. In other words, the incorporation of the market-border effect from the Hotelling model avoids having to make the assumption $\eta>1$, which is likely to be unrealistic.

Return now to the assumption that stores are evenly spaced around the circle in Figure 1. Suppose that store 1 takes as given the positions of stores 2 and *N*. Suppose, starting from a position equidistant between stores *N* and 2, that store 1 considers moving its location, say by a distance *X* to the right. The market border with store *N* would then shift from a distance *h* to a distance *h+X/2* from store 1's location. The border with store 2 would shift from a distance *h* to a distance *h-X/2*. Therefore, the total distance covered by store 1 would remain at *2h.* However, there is a shift toward customers who are relatively far from the store (in the region toward store *N*) and away from those who are relatively close (in the region toward store 2). Because of the downward-sloping demand curve in Eq. (3), the more distant customers buy a smaller quantity and are less profitable for store 1. Hence, on net, store 1's profit declines. It follows that store 1 would not move and is best off remaining equidistant between stores 2 and *N*. That is, the equal-spacing pattern is an equilibrium outcome.[13] Note that this conclusion depends on the downward-sloping demand curve—the result would not apply under the common assumption that each household buys exactly one unit of the good (for example, in the version of the Salop model described in Tirole [1988, Section 7.1.2]). The result also depends on the assumption that transportation charges are proportional to the quantity bought, *q(z),* rather than being independent of this quantity (see n.5).

The result in Eq. (17) is reminiscent of models of contestable markets, as described by Baumol, Panzar, and Willig (1982). In those models, an individual firm's markup is limited by the cost of production for potential entrants into a market, and marginal-cost pricing prevails under assumptions that include zero costs of entry and exit and equal access of all firms to all technologies. A difference in the present, Hotelling context is that the relevant competition

[13]This conclusion accords with Vickrey (1964, p. 330).

depends on actual participants in the market—neighboring stores—rather than hypothetical entry by potential participants. Moreover, the Hotelling-type model does not predict marginal-cost pricing but rather price markups that depend on the distances between active stores in the market.

III. Free-Entry Condition

The analysis of entry in this section applies when Eq. (17) is an accurate formula for the price; that is, when $P \approx c + 2th$. The total quantity sold, Q, by store 1 (or any store) can then be determined from Eq. (4) to be:

$$Q \approx 2Ahc^{-\eta}, \tag{20}$$

where the 2 reflects store 1's sales on both sides of store 1's location in Figure 1. The associated profit is given from Eq. (6) by

$$\pi \approx 4tAh^2c^{-\eta} - f. \tag{21}$$

If $\pi > 0$, there is an incentive for new stores to enter the market, thereby raising N and lowering $h = \frac{1}{2}\frac{H}{N}$.[14] The decrease in h lowers π in Eq (21). This process continues as long as $\pi > 0$. If the integer constraint on N can be neglected, as will be satisfactory if N is large,[15] the free-entry condition will be $\pi \approx 0$. Equation (21) then implies

$$2h \approx \sqrt{\frac{f}{tAc^{-\eta}}}. \tag{22}$$

Therefore, under free entry, the spacing between stores, $2h$, follows a square-root rule, whereby spacing is larger the lower the transport cost per unit, t, the higher the fixed cost f of operating a

[14] The assumption here is that firms immediately relocate to preserve their equal spacing around the circle in Figure 1. More realistically, the model applies to a long-run, steady-state situation, rather than to the dynamics of entry and exit.

[15] For example, for New York City, with a central city population of 8 million in 2019, the New York State Comptroller reports that the number of retail businesses at that time was about 33,000, each servicing on average 242 domestic residents. In this context, where N is *33,000*, the integer constraint would be of no consequence. Mankiw and Whinston (1986) discuss integer constraints for a related model.

store, and the (approximate) quantity sold to each buyer, $Ac^{-\eta}$. The last two results apply because the fixed cost is effectively scaled by the quantity sold to each buyer. Note that the spacing does not depend on *H*, which relates to the overall size of the economy in Figure 1.

The solution for spacing, *2h,* in Eq. (22) implies that the number of firms is given from Eq. (1) by

$$N = \frac{H}{2h} \approx H \cdot \sqrt{\frac{tAc^{-\eta}}{f}}. \quad (23)$$

Therefore, *N* is proportional to *H* and rises with the square root of the transport cost per unit, *t,* and the quantity sold to each buyer, $Ac^{-\eta}$. The number *N* varies inversely with the square root of the fixed cost, f, of operating a store.

The free-entry result in Eq. (22) implies from Eq. (17) that the price charged will be

$$P \approx c + \sqrt{\frac{tf}{Ac^{-\eta}}}. \quad (24)$$

That is, the markup on marginal cost, *c*, is increasing with the transport-cost parameter, *t*, and the fixed cost f of operating a store and decreasing with the quantity sold to each buyer, $Ac^{-\eta}$. The markup does not depend on *H*. The markup also does not depend (as an approximation) on the elasticity of demand, η, except for a positive effect that involves the quantity sold, $Ac^{-\eta}$. Moreover, the results hold for any $\eta \geq 0$, including Hotelling's case where $\eta = 0$.

Equation (24) extends the result from Eq. (17) with regard to the predicted response of *P* to a change in marginal cost, *c*. Because an increase in *c* raises *P* and, thereby, lowers quantities purchased (in accordance with the term $Ac^{-\eta}$), profit declines, and a restoration of zero profit requires exits of stores; that is, *h* rises. This increase in *h* leads to a rise in the price markup, corresponding to the square-root term on the right-side of Eq. (24). In other words, *P* rises by more than one-to-one with *c*. However, under the conditions assumed before—where *2th/c* and η

are not too large—the predicted pass-through in levels is only a small amount above unity. In this sense, the extended model that allows for entry is likely still consistent with the empirical findings on pass-through in levels in Sangani (2024, Tables 4 and 5).

The theoretical results can be related to the empirical analysis of Chevalier, Kashyap, and Rossi (2003) concerning pricing patterns of a major supermarket chain. Their key finding is that markups for affected categories of products are relatively low at times of peak demand, notably during major holidays and events such as Christmas, Thanksgiving, and Lent. Their empirical results rule out an explanation for this pricing pattern based on the elasticity of demand being unusually high at these times of peak demand. This finding accords with the present model in that the price markup is not predicted to fall when the demand elasticity, η, rises in Eq.(17) for a given market size, $2h$, or in Eq. (24), where $2h$ is determined by a free-entry condition. In the model, peak demand would be represented by a temporarily high value of the parameter A, which enters into the demand function in Eq. (3). High A has no effect on the price markup in Eq. (17) but reduces this markup in Eq. (24), which factors in the free-entry condition. An application of this last result to the setting of Chevalier, Kashyap, and Rossi (2003) depends on there being significant entry during peak demand periods. Possibly this entry can take the form not of new stores but of longer store hours or increased advertising (a margin emphasized by Chevalier, Kashyap, and Rossi).

IV. Socially-Optimal Entry

Heuristically, there are two elements in the model that may cause the free-entry choices of spacing, $2h$, and number of stores, N, from Eqs. (22) and (23) to deviate from socially-optimal values. The first distortion is the markup, approximated by $2th$ in Eq. (17), which generates an excess of the effective price, P^*, over social marginal cost, $c+tz$, in Eq. (3). This excessive price leads to quantities of goods consumed that fall short of socially-optimal values. That is, at the existing quantity sold, each household's willingness to pay, P^*, for an additional unit—corresponding to the inverse-demand function implied by Eq. (3)—exceeds the social marginal cost, $c+tz$. Therefore, the profit flow in Eq. (6), which is the signal for entry, is too low in the sense of falling short of the amount that corresponds to the socially-optimal quantity of goods produced. On this ground, one would expect the free-entry choice of the number of stores, N, to be too low, corresponding to the choice of spacing, $2h$, being too high.

The second distortion is that an entering firm's profit includes revenue that is transferred from incumbent firms. That is, there is a "business-stealing effect,"[16] which is a private reward for entry that has no social benefit. (This distortion can be viewed as generated by the lack of property rights for incumbent firms in their profit flows.) On this ground, one would expect the free-entry choice of the number of stores, N, to be too high, corresponding to a choice of spacing, $2h,$ that is too low.

Tirole's (1988, Ch. 7, p. 284) analysis of the Salop-Hotelling model finds that entry is excessive in that model: "… we compare the free-entry equilibrium with the allocation selected by a social planner. We already know that the price charged by the firms is greater than the marginal cost. However, in this case, where consumers all receive the same utility from the good and each consumes only one unit, this price introduces no distortion." Hence, Tirole's

[16]This term appears in Mankiw and Whinston (1986, p. 49) and also in Tirole (1988, p. 284). The business-stealing effect plays a large role in growth models that feature creative destruction, as in Aghion and Howitt (1992). The term creative destruction was used by Schumpeter (1942, p. 83): "This process of Creative Destruction is the essential fact about capitalism. It is what capitalism consists in and what every capitalist concern has got to live in." Schumpeter's analysis of creative destruction was heavily influenced in Schumpeter (1942, part I) by his reading of Karl Marx. It is unclear whether Schumpeter's vision was preceded by Picasso's famous quote: "Every act of creation begins with an act of destruction."

conclusion that the free-entry equilibrium features too much entry applies because the analysis considers only the distortion from business-stealing.

Suppose in the present model that the social planner can dictate the spacing between stores, $2h$, and the quantity of goods purchased, $q(z)$, by each household. However, the planner is assumed to accept the uniform distribution of consumers around the circle in Figure 1. Optimality entails the equation of each customer's willingness to pay, corresponding to the demand price from the inversion of Eq. (3), to the social marginal cost, $c+tz$. The socially-optimal value of spacing, $2h$ can then be derived, following the approach of Mankiw and Whinston (1986, p. 50), by maximizing an expression for aggregate consumer surplus (netting out costs of producing and transporting goods and the fixed costs of operating stores). The result, derived in the appendix, is that the social planner's choice of store spacing, $2h$, exceeds the free-entry choice (in Eq. [22]). Correspondingly, the social planner's choice of number of stores, N, falls short of the free-entry number (in Eq. [23]). In this sense, the model accords with the excess-entry result of Tirole (1988).

Note that Tirole's calculations correspond to assuming $\eta=0$, which eliminates the distortion associated with inefficiently low quantities of goods. In the present model, η is allowed to be positive, but the results on entry rely on approximations that involve η and $2th/c$ not being too large. These conditions generate a quantity distortion that is nonzero but quantitatively minor. Therefore, the excess-entry result still obtains. For very high values of η or $2th/c$, the results from the approximate solution become unsatisfactory, as shown in Table A1, part I, in the appendix. In this range, the distortion from inefficiently low quantity becomes more important, and the excess-entry result need not apply.

Free entry and socially-optimal entry have also been studied in non-spatial models of monopolistic competition, which include Spence (1976), Dixit and Stiglitz (1977), and Mankiw and Whinston (1986). In this literature, free entry sometimes results in the socially optimal amount of entry, but this result is not general. However, an important feature of these models is that the various products are treated symmetrically, with each good equally substitutable with each alternative good. In contrast, the central feature of Hotelling-type models and related Lancaster-type models is the strong substitution with "neighboring" producers and products. This property underlies the markup-pricing result in Eq. (17), which is driven by the threat of losing business at the market borders with adjacent stores. This formula for markup pricing is, in turn, a key ingredient in the analysis of free entry and socially-optimal entry.

V. Summary

In a circular Hotelling model, customers are uniformly distributed around a circle with circumference H and number of stores N. Equal spacing of stores applies in equilibrium. A customer's transportation cost per unit is tz, where z is the distance from the nearest store. The effective price per unit is $P^*=P+tz$, where P is the price at the store's location. The marginal cost of production is the constant c, and the magnitude of the elasticity of each customer's quantity demanded with respect to P^* is the constant η.

Stores serve customers out a distance $h=H/2N$. If production costs dominate transport costs and the elasticity of demand is not extremely large, each store's equilibrium price approximates the markup $2th$ over marginal cost, independent of the elasticity of demand. The key element in a store's markup is the threat of losing all of the business at its market borders. In a free-entry equilibrium, h is larger the lower t, the higher the fixed cost of operating a store, and the smaller the scale of quantity demanded by each customer. Given these factors, the number of stores, *N*, is proportional to *H*, and the markup is independent of *H*.

Two distortions affect the equilibrium: the excess of price over marginal cost implies inefficiently low quantities of goods produced, and a business-stealing effect implies that a new entrant's profit includes revenue transferred from incumbent stores. In plausible ranges for the parameters η and $2th/c$, the business-stealing effect dominates, and entry is excessive.

From the perspective of understanding price markups, the biggest contribution from Hotelling-type models is the emphasis on factors that influence the substitutability among neighboring producers and products. This focus contrasts with the standard Lerner formula, which puts all the weight on the elasticity of demand.

References

Aghion, Philippe and Peter Howitt (1992). "A Model of Growth through Creative Destruction," *Econometrica*, 60, 323-351.

Anderson, Simon P. and Andre de Palma (2000). "From Local to Global Competition," *European Economic Review*, 44, 423-448.

Barro, Robert J. and Rachel M. McCleary (2005). "Which Countries Have State Religions?" *Quarterly Journal of Economics*, 130, 1331-1370.

Baumol, William J. (1967). "Calculation of Optimal Product and Retailer Characteristics: The Abstract Product Approach," *Journal of Political Economy,* 75, 674-685.

Baumol, William J., John C. Panzar, and Robert D. Willig (1982). *Contestable Markets and the Theory of Industrial Structure*, New York, Harcourt Brace Jovanovich.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch (2021). "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," *Journal of Monetary Economics*, 121, 1-14.

Brueckner, Jan K. (1987). "A Unified Treatment of the Muth-Mills Model," in Edwin S. Mills, ed., *Handbook of Regional and Urban Economics*, Amsterdam, Elsevier.

Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi (2003). "Why Don't Prices Rise during Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 93, 15-37.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger (2020). "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561-644.

Dixit, Avinash K. and Joseph E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297-308.

Downs, Anthony (1957). "An Economic Theory of Political Action in a Democracy," *Journal of Political Economy*, 65, 135-150.

Gu, Yiquan and Tobias Wenzel (2009). "A Note on the Excess Entry Theorem in Spatial Models with Elastic Demand," *International Journal of Industrial Organization*, 27, 567-571.

Hall, Robert E. (1988). "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96, 921-947.

Hall, Robert E. (2018). "Using Empirical Marginal Cost to Measure Market Power in the US Economy," NBER working paper no. 25251, November.

Hay, Donald A. (1976). "Sequential Entry and Entry-Deterring Strategies in Spatial Competition," O*xford Economic Papers*, 28, 240-257.

Helpman, Elhanan (1981). "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition," *Journal of International Economics*, 11, 305-340.

Hotelling, Harold (1929). "Stability in Competition," *Economic Journal*, 39, 41-57.

Lancaster, Kelvin (1966). "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 132-157.

Lerner, Abba P. (1934). "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies*, 1, 157-175.

Mankiw, N. Gregory and Michael D. Whinston (1986). "Free Entry and Social Inefficiency," *Rand Journal of Economics*, 17, 48-58.

Salop, Steven C. (1979). "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 10, 141-156.

Sangani, Kunal (2024). "Pass-Through in Levels and the Incidence of Commodity Shocks," unpublished, Stanford University, September.

Schmalensee, Richard (1978). "Entry Deterrence in the Ready-to-Eat Cereal Industry," *Bell Journal of Economics*, 9, 305-327.

Schumpeter, Joseph A. (1942). *Capitalism, Socialism, and Democracy*, New York, Harper & Brothers Publishers.

Smithies, Arthur A. (1941). "Optimum Location in Spatial Competition," *Journal of Political Economy*, 49, 423-439.

Spence, Michael (1976). "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, 43, 217-235.

Syverson, Chad (2024). "Markups and Markdowns," NBER working paper no. 32871, August.

Tirole, Jean (1988). *The Theory of Industrial Organization,* Cambridge MA, MIT Press.

Vickrey, William S. (1964). *Microstatics*, New York, Harcourt, Brace and World.

Vogel, Jonathan (2008). "Spatial Competition with Heterogeneous Firms," *Journal of Political Economy*, 116, 423-466.

Appendix

I. The Approximation in the Markup Formula

The result for the price markup, $P \approx c + 2th$ in Eq. (17), is an approximation to the solution for *P* from Eq. (14). Recall that the price *P* is determined exactly from

$$\text{(A1)} \quad \left(\frac{1}{\eta-1}\right) \cdot [P^{1-\eta} - (P+th)^{1-\eta}] = (P-c) \cdot [P^{-\eta} - \tfrac{1}{2}(P+th)^{-\eta}].$$

Table A1, part I reports numerical solutions for *P* from Eq. (A1) determined for an array of underlying parameters, which can be expressed as η and $2th/c$. The latter term is the ratio of the markup to marginal cost in the approximate solution in Eq. (17). The table shows the model's numerical results for *P* expressed as a ratio to $c+2th$, which is the approximate expression for *P* in Eq. (17). The approximate solution is exact (that is, the ratio equals 1) if $\eta=0$ or $th=0$, and the approximation works better the smaller η and $2th/c$. For example, if $\eta \leq 5$ and $2th/c \leq 0.1$, the true solution for *P* falls short of $c+2th$ by less than 3%. Further, if $\eta \leq 2$ and $2th/c \leq 0.2$, the true solution for *P* again falls short of $c+2th$ by less than 3%. If η is very large, *P* can be below $c+2th$ by substantial amounts for reasonable values of $2th/c$. However, the bottom line, based on $\eta \leq 5$ and $2th/c \leq 0$ or $\eta \leq 2$ and $2th/c \leq 0.2$ as plausible conditions, is that the approximation in Eq. (17) will be good for a range of reasonable parameters.

Table A1, part II compares the solution for *P* with that from the Lerner formula in Eq. (9), $P = c \cdot \left(\frac{\eta}{\eta-1}\right)$. The table shows the model's numerical results for *P* expressed as a ratio to this Lerner value, which is well defined only if $\eta > 1$. The Lerner formula tends to provide a better approximation to the model's results (that is, the ratio is closer to 1) the larger η and the larger $2th/c$. However, in the ranges noted above—$\eta \leq 5$ and $2th/c \leq 0.1$ or $\eta \leq 2$ and $2th/c \leq 0.2$—the Lerner formula gives poor results (that is, the ratio is well below 1).

Table A1

Approximations for Price-Markup Solution

I. Ratio of price, P, to Hotelling approximation, $P \approx c + 2th$

	P/(c+2th)							
2th/c	**η=0**	**η=0.1**	**η=0.5**	**η=1**	**η=1.5**	**η=2**	**η=5**	**η=100**
0	1	1	1	1	1	1	1	1
0.02	1	1.000	1.000	1.000	1.000	0.999	0.999	0.988
0.04	1	1.000	0.999	0.999	0.998	0.998	0.995	0.970
0.1	1	1.000	0.997	0.995	0.992	0.989	0.976	0.918
0.2	1	0.998	0.990	0.982	0.973	0.967	0.933	0.842
0.4	1	0.994	0.973	0.949	0.929	0.911	0.841	0.721
1	1	0.984	0.922	0.858	0.806	0.763	0.628	0.505
2	1	0.972	0.866	0.757	0.672	0.604	0.433	0.337
5	1	0.952	0.793	0.621	0.491	0.398	0.221	0.168

II. Ratio of price, P, to Lerner formula, $P = c \cdot (\frac{\eta}{\eta - 1})$

	$\left(\frac{P}{c}\right) \cdot (\frac{\eta - 1}{\eta})$				
2th/c	**η=1.0**	**η=1.5**	**η=2**	**η=5**	**η=100**
0	0	0.33	0.50	0.80	0.990
0.02	0	0.34	0.51	0.82	0.998
0.04	0	0.35	0.52	0.83	0.999
0.1	0	0.36	0.54	0.86	1.00
0.2	0	0.39	0.58	0.90	1.00
0.4	0	0.43	0.64	0.94	1.00
1	0	0.54	0.76	1.01	1.00
2	0	0.67	0.90	1.04	1.00
5	0	0.98	1.20	1.06	1.00

II. Social-Planner Problem

The social planner effectively prices goods at the social marginal cost, which is *c+tz* for a household located at distance *z* from its nearest store. The associated socially optimal quantity of goods is $q(z) = A \cdot (c + tz)^{-\eta}$ from Eq. (3). The consumer surplus, $\Omega(z)$, for an agent at *z* is then

(A2) $$\Omega(z) = \int_0^{q(z)} \left[(\frac{A}{q})^{1/\eta} - c - tz \right] dq,$$

where $(\frac{A}{q})^{1/\eta}$ is the demand price associated with *q*. The integral can be evaluated to get

(A3) $$\Omega(z) = \frac{A(c+tz)^{1-\eta}}{\eta-1},$$

where *η>1* is assumed at this stage (but is ultimately an unnecessary restriction).

The consumer surplus given in Eq. (A3) applies for *z=(0,h)* to the right of store 1 in Figure 1. Including also the analogous term to the left of store 1, the total consumer surplus for all customers of store 1 is

(A4) $$\Omega = 2 \int_0^{h} \frac{A(c+tz)^{1-\eta}}{\eta-1} dz.$$

Using a change of variable from *z* to *c+tz*, the integral can be evaluated to get

(A5) $$\Omega = \frac{2A}{t(\eta-1)(\eta-2)} [c^{2-\eta} - (c + th)^{2-\eta}].$$

The analysis at this stage assumes *η>2*, but this restriction (or *η>1*) is ultimately not needed for the results.

The consumer surplus in Eq. (A5) applies from Eq. (1) to *N=H/2h* markets. The fixed cost of operations is *Nf*. The social planner chooses *h* to maximize the overall net surplus:

(A6) $$Net\ consumer\ surplus = \frac{H}{h}\left\{\frac{A}{t(\eta-1)(\eta-2)}\left[c^{2-\eta} - (c+th)^{2-\eta}\right] - \frac{f}{2}\right\}.$$

Setting the derivative of the expression on the right-hand side of Eq. (A6) with respect to *h* to zero to get the first-order maximization condition leads to

$$\frac{f}{2} = \frac{A}{t(\eta-1)(\eta-2)}\left[c^{2-\eta} - (c+th)^{2-\eta}\right] - \frac{Ah}{\eta-1}(c+th)^{1-\eta}.$$

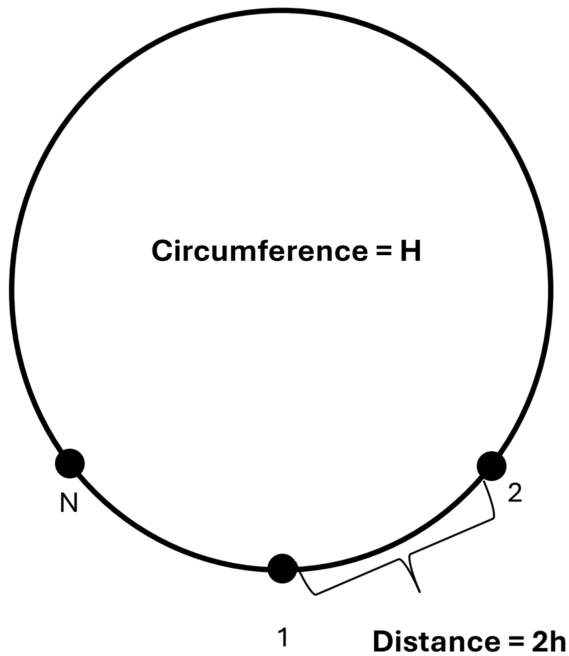
Using the condition from before, $th << c$, the right-hand side simplifies using first-order Taylor approximations for the terms involving $c+th$ to $Ath^2c^{-\eta}$. Therefore, the approximate solution is

(A7) $$2h \approx \sqrt{\frac{2f}{tAc^{-\eta}}}.$$

The social planner's solution for store spacing in Eq. (A6) is the multiple by the square root of 2 of the free-entry solution in Eq. (22). That is, the social planner's spacing exceeds the free-entry one by about 40%, corresponding to having only about 70% of the number of stores, *N*.

Figure 1

Salop-Hotelling Circular City



Note: Consumers are uniformly distributed around a circle with circumference *H*. Stores are located at positions 1, 2, …, *N*, with an equal spacing of *2h.* In the full equilibrium, store 1's market extends out a distance *h* to the right and left of its location.