Monopolistic Competition with Two-Part Tariffs*

by

Nicholas Economides
Stern School of Business, NYU
New York, NY 10012
(212) 998-0864, FAX (212) 995-4218
neconomi@stern.nyu.edu
http://www.stern.nyu.edu/networks

and

Steven S. Wildman, Director
Quello Center for Telecommunication Management & Law
409 Communication Arts & Sciences
Michigan State University
East Lansing, MI 48824-1212
(517) 432-8004, FAX (517)432-8065
swildman@msu.edu
http://www.quello.msu.edu

Revised July 2005

Abstract

Non-uniform pricing equilibria are shown to dominate uniform pricing equilibria in free entry, monopolistically competitive markets with identical consumers. The non-uniform pricing equilibrium is welfare optimal. Comparisons of Cournot and non-uniform pricing equilibria in terms of the equilibrium number of firms and sales per firm show that the positioning of Cournot equilibria relative to the welfare optimal configuration of firms and outputs depends on the relative curvatures of inverse demand and average cost functions, entry-induced rotation of inverse demand functions, and the relative price effects of changes in own and other firms outputs. The choice between the non-uniform and uniform pricing interpretations of equilibria in differentiated product markets may have important implications for policy analysis.

Key words: two part tariffs, monopolistic competition, variety

JEL Classification: L1, D4

* We thank Jim Refalo, Chris Stefanadis, Larry White, and participants at the applied microeconomics seminar at Northwestern for helpful comments and suggestions.

Monopolistic Competition with Two-Part Tariffs

I. Introduction

While non uniform prices (NUPs) would seem to be an obvious response to the triangle of deadweight loss associated with uniform prices (UP) and downward-sloping demand functions, models of competition in differentiated product industries have, for the most part, assumed uniform prices. Thus, unexploited gains from trade is an implicit assumption of these models. The reliance on the UP assumption may have been dictated more by the presumed intractibility of NUP models than by a belief that UP models better describe real-world markets.

In this paper we show that, for the commonly employed assumption of identical consumers (or, equivalently, a single consumer), it is possible to model NUP competition among sellers of differentiated products in a way that is a straightforward and intuitive analogue of UP models. When firms can use NUPs to extract all of the surplus under their demand functions, a monopolistic competition equilibrium in NUPs can be described by substituting average surplus functions[1] for inverse demand functions and substituting inverse demand functions for marginal revenue functions in the usual UP model of monopolistic competition. With this approach, it is easy to show that the NUPs dominate UPs in a free entry equilibrium in differentiated product industries because access to the triangle of deadweight loss associated with UPs enables sellers to offer better deals to buyers.

This conclusion is opposite to the findings of recent work on multifirm, competitive equilibria in markets for homogeneous goods.[2] Mandy (1992) demonstrates that the uniform pricing assumption is theoretically sound for multi-firm, competitive markets for homogenous goods as long as free entry drives profits to zero.[3] While earlier work by Mandy (1991) and models by Hayes

---

[1] For a firm selling output x, the average surplus associated with x is the area under its inverse demand function up to x divided by x.

[2] However, Panzar and Postlewaite (1984) and Shaffer (1987) have shown that NUPs dominate UPs in contestable natural monopolies.

[3] Mandy also shows that if entry barriers limit the number of firms in a market to less than the fee

(1987) and Locay and Rodriguez (1992) showed that there are circumstances in which two-part tariffs (TPTs) dominate uniform prices in competitive, homogeneous good markets, each of these studies dealt with special cases in which TPTs facilitate transfers among individuals (or among potential states of the world for a given individual) that are not possible with uniform prices. In the absence of a demand for such transfers, Mandy (1992) shows that non-uniform prices can always be undercut by a uniform price set equal to minimum average cost.

Casual empiricism suggests that product differentiation is common and probably predominates in consumer good industries. Our analysis suggests that absent transactional barriers to its implementation, non-uniform pricing should be just as common. While the prevalence of non-uniform prices can only be determined empirically, we would suggest that they are probably more common than has heretofore been recognized. We show below that NUPs may superficially appear very much like UPs. Therefore, it is likely that NUPs are frequently not recognized for what they are because economists have been conditioned to think in terms of uniform-pricing models. This has probably resulted in mistaken analyses of the nature of competition and inappropriate policy prescriptions for a number of industries.

This paper is organized as follows. In the next section we show that non-uniform prices dominate uniform prices in monopolistically competitive markets with identical consumers and describe the NUP pricing as a non-cooperative equilibrium.[4] We show that NUP equilibria are characterized by an average surplus curve-average cost curve tangency that is similar in appearance to the demand curve-average cost curve tangency of a Chamberlinian equilibrium. The NUP equilibrium is also welfare-optimal. UP and NUP equilibria are compared in Section III in terms of numbers of firms and output per firm. The shapes of the cost functions, as well as the shapes of the

---

entry number, two-part tariffs dominate if competition is Bertrand.

[4] In the NUP equilibrium, firms employ TPTs or other pricing schemes that are equivalent to TPTs from a representative consumer's perspective

demand functions and the manner in which they shift in response to the entry or increased outputs of competitors, are critical to determining the relative outputs and numbers of firms for the two equilibria. The analysis is illustrated in Section IV through an example that uses a specific class of utility functions. Section V discusses non-Cournot competitive responses. Section VI presents our concluding remarks.

II. Non-Uniform Price Dominance and Competitive Equilibrium

Consider a market for differentiated products where firms compete in quantities under the assumption that competitors outputs remain constant. Figure 1 depicts a representative firm in a zero-profit Chamberlinian equilibrium with its demand curve tangent to its average cost curve. The price p_1 and quantity x_1 corresponding to the tangency also maximize profits because the tangency implies marginal revenue equal to marginal cost. Product differentiation is implicit in the downward sloping demand curve.

As a simplifying assumption, let D represent the demand of a single individual who accounts for all of the sales of all products in the market. That is, the demand curves for all other products in this market represent the demands of this same individual.[5] The analysis would be the same if D and the demands for all other products were the aggregated demands for a large number of identical consumers, each of whom prefers some diversity in his or her consumption of this class of products.[6]

[5] The representative consumer assumption is common in modelling exercises reported in the industrial organization literature. See Spence (1976), Dixit and Stiglitz (1977) and Chamberlin (1962) for applications to monopolistic competition.

[6] The assumption that the single consumer is representative of many homogeneous consumers means that issues relating to monopsony can be ignored.

Both sellers and consumers (the consumer) might expect to do better than at the tangency in Figure 1 if they could negotiate a more complex contract. For example, assuming the quantity produced by other firms is held constant, consumer surplus and seller profits could both be increased by a two-part tariff with a per unit price of p_2 and a fixed tariff E equal to $(p_1 - p_2)x_1$. Consumer surplus would increase by the area of triangle A, and the seller's profits would increase by the area of B. Of course, if consumers were price takers, the firm would set E equal to the area under the demand curve above p_2. Therefore a uniform price would never be employed when the seller could employ a two-part tariff (TPT) instead.[7]

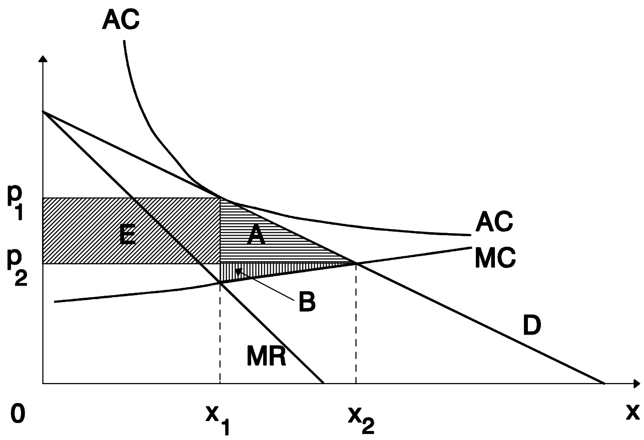

Figure 1

In the presence of TPTs, Chamberlinian demand-average cost curve tangencies cannot be a feature of a free entry equilibrium. Positive profits would attract entry, which would shift demand curves inward. Even without entry, demand curves would shift inward as all firms tried to increase output to the point of the demand curve-marginal cost intersection. Therefore, demand curves must lie inside of average cost curves in an equilibrium with two-part tariffs.

It is straightforward to show that the zero-profit equilibrium with two-part tariffs is characterized by the tangency of an average surplus curve with the average cost curve that looks very much like the Chamberlinian average revenue-average cost tangency, where average surplus for quantity x, AS(x), is defined to be the area under the inverse demand curve up to x divided by x, i.e., $AS(x) = [\int_0^x p(y)dy]/x$. The two-part tariff equilibrium is illustrated in Figure 2. The average

[7] In the discussion that follows, we use TPTs as representative of full surplus extracting NUPs generally.

surplus curve lies above the demand curve since it averages the higher willingness to pay of earlier units.[8] Note that the inverse demand function bears the same relationship to AS as marginal revenue does to the inverse demand function and lies completely inside of the average cost curve.[9] The TPT can be thought as consisting of a marginal fee $p(x^*)$ equal to $MC(x^*)$, and a fixed fee equal to the shaded area $T = [p_{as} - p(x^*)]x^*$. Since $p_{as} = AS(x^*)$, it is easy to show that T is also equal to the area under the demand and above $p(x^*)$ from quantity zero up to x^*.

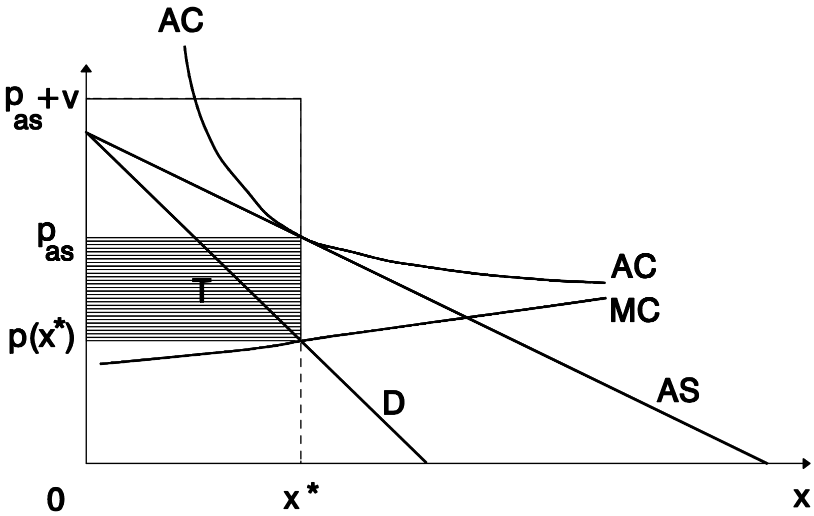

Figure 2

To see that this is an equilibrium, let total cost for output x, TC(x), be composed of a fixed cost, F, and variable cost V(x), with marginal cost $MC(x) = \partial V(x)/\partial x$. Let $i = 1, ..., n$ firms offer two-part tariff contracts. We will establish the AS-AC tangency configuration as a non-cooperative equilibrium of this game. Let the ith firm offer T_i as the fixed fee, and p_i as the marginal fee. This implies a revenue for firm i of $R_i(x) = T_i + xp_i$, and an average revenue function $AR_i(x) = p_i + T_i/x$. If, for any x, $AR_i(x) > AC(x)$, another firm j, selling the same variant of the product as i, can undercut firm i by offering a contract with $AR_j(x)$ below $AR_i(x)$ but

---

[8] $AS(x) > p(x) \Leftrightarrow S(x) = \int_0^x p(y)dy > xp(x)$, which is true for any downward-slopping demand curve since $p(y) > p(x)$ for all $y < x$.

[9] Marginal revenue is the partial derivative of total revenue, and price is the partial derivative of total surplus with respect to x.

above or at average cost, $AR_i(x) > AR_j(x) \geq AC(x)$. Therefore, competition between firms will force each firm to offer an average revenue function below AC or at most at AC; i.e., $AR_i(x) \leq AC(x)$ for all x. Further, at the operating output x^*, a firm must cover costs to stay in business, i.e., $AR_i(x^*) = AC(x^*)$. Clearly, two firms offering the same variety cannot coexist.

The consumer who buys quantity x is willing to pay up to the total surplus of x. Therefore the consumer is willing to pay an average price of $AS(x)$. Consumers buy x if $AS(x) \geq AR_i(x)$. Combining this with earlier results, we have $AS(x^*) \geq AR_i(x^*) = AC(x^*)$.

If $AS(x^*) > AR_i(x^*) = AC(x^*)$, and firm i is the only one providing this brand, it has an incentive to offer a higher average revenue function. Such an action cannot be part of an equilibrium as explained above (because firm i can then be undercut by firm j). Therefore at equilibrium we must have $AS(x^*) = AR_i(x^*) = AC(x^*)$ for produced quantity x^*. For all other levels of production the average revenue schedule of firm i lies below average cost. Therefore it is tangent to average cost at x^*. Equality of average surplus to average cost, together with tangency of AS and AC, implies equality of unit price to marginal cost. To see this, note that $dAC/dx \equiv AC' = (MC - AC)/x$ and $dAS/dx \equiv AS' = (p - AS)/x$, so that $AC' = AS'$ and $AC = AS$ implies $p = MC(x^*)$.[10]

An arbitrary number of firms, with each firm producing one variety, will result in an $AS(x)$ function that may lie above or below $AC(x)$. Further, as the number of substitutes increases, the demand and average surplus function of each variety shifts inwards. Therefore, there exists a number of varieties that makes $AS(x)$ and $AC(x)$ just touch at x^*. Since every product that generates consumer benefits as great as its costs is produced at the optimal level of production, the TPT (NUP) equilibrium is also welfare-optimal (Spence 1976).

How important is the equilibrium depicted in Figure 2 as a description of empirical reality? It is not hard to identify "competitive" markets with two-tariffs: for example, taxi meters start with a

[10] For constant marginal cost, the fixed fee is exactly equal to the lump-sum part of the two-part tariff.

fixed fee to which is added a constant per mile charge, bars and movies theaters change fixed admission fees and sell food and beverages on a per unit basis, and mortgages agreements usually involve an up front payment of points in addition to the monthly interest charge; but most goods and services clearly are not priced in this manner, perhaps because it is too expensive to enforce prohibitions on resale. The administrative costs of collecting both parts of the tariff may be another transactional barrier to employing two-part tariffs.

Other pricing schemes for which resale is not a problem can be employed to accomplish exactly the same end, however. For example, the good in Figure 2 might be sold at a per unit price of $p_{as} + v$ with the understanding that consumers purchasing x^* receive a rebate of vx^*. As long as $p_{as} + v$ was greater than the price intercept of the demand curve, the outcome would be the same as for the two-part tariff illustrated in Figure 2. This might account for ostensible promotional practices such as two-for-the-price-of-one sales and the quantity-based ties to other products, such as the glassware that gas stations formerly gave to customers who filled their tanks during price wars and the frequent practice of mail order book services of allowing customers to select an additional title free when their orders exceed a certain dollar threshold. Quantity discounts, normally considered to be second degree price discrimination, could serve the same purpose. Full surplus would be extracted by setting price above the price intercept for purchases of less than x^* and charging p_{as} for purchases of x^* or greater.

The simplest alternative to two-part pricing that is equivalent in surplus extracted, and one that is no more difficult to administer than uniform pricing, may be for firms to package their products in fixed-quantity sales units of x^* sold at a per sales unit "price" of $p_{as}x^*$. Many, if not most, products are sold in this manner. A 20 oz. box certainly is no more a natural unit for measuring Cheereos than a Ford Taurus is a natural unit for measuring transportation services. Economies of scale in packaging undoubtedly dictate that most products be offered in a limited number of different-sized packages; but absent transaction cost barriers, profit maximization requires that package size be determined by the logic of the non-uniform pricing model developed here when products are differentiated. This theoretical necessity, combined with its consistency with a broader range of price and packaging strategies, and a common tendency to interpret sales of fixed-quantity packages at a single per package price as uniform pricing, are the basis of our claim in the introductory section that we probably encounter non-uniform pricing in a wide variety of markets of common goods and services, but fail to recognize it for what it is due to a training-induced bias toward the uniform price interpretation.

III. Comparing UP and NUP equilibria

Perhaps the most frequently recurring question in the literature on monopolistic competition, at least since the "excess capacity controversy,"[11] is whether competitive markets provide the welfare optimal number of differentiated products.[12] A closely related but less intensively investigated question is how closely output per product (or per firm) approaches optimal levels.

[11] See articles by Barzel (1970), Demsetz (1959, 1964, 1968), and Schmalensee (1972).

[12] Work with various models of consumer demand has shown that product diversity at a competitive equilibrium may exceed or fall short of the optimum. In their survey of the work on the diversity question, Besanko, Perry and Spady (1990) observe that, while there are exceptions, findings of too little variety are generally associated with representative consumer models, such as that presented in the previous section, while findings of too much variety are generally produced with spatial and characteristics models. They hypothesize that the strong association of too little diversity with representative consumer models and too much diversity with spatial and characteristics models is a reflection of the generalized nature of competition in the former and the localized nature of competition in the latter. All products are equally good substitutes for each other in representative consumer models, so competition is generalized across firms; but a product competes directly only with near neighbors in spatial and characteristics models. Thus, new brands "crowd" the characteristics space in the characteristics models, whereas, they create a new dimension to the preference space in the representative consumer models. This suggests that the social value of an additional firm will be lower in characteristics models than in representative consumer models. The logit model of competition with an extreme value representation of consumer demand described in Besanko *et al*., which does not fit either the representative consumer demand or the spatial/characteristics interpretation of consumer demand, generates too little diversity. Each firm is in direct competition with all other firms in this model, however.

In this section we show that comparisons of UP equilibria with the optimum can be described in terms of three basic relationships: (1) The shape of the inverse demand function facing the individual firm; (2) The degree of substitutability between different firms' outputs; and (3) The shape and location of the average cost curve. Knowledge of the first relationship is sufficient to determine whether output per firm at a UP equilibrium exceeds of falls short of its value at the optimum. Consideration of the other two relationships is necessary to answer the diversity question.[13]

III.A Comparison of UP and NUP Equilibrium Outputs

The importance of the shape of the inverse demand function and entry- or output- induced changes in slopes in answering the relative output question is evident in Figure 3, where x_c is equilibrium sales per firm in the standard Cournot equilibrium with uniform prices, x_t is equilibrium per firm output for Cournot competitors with two-part tariff equivalent non-uniform prices, and D_c, D_t and AS_t are the associated inverse demand and average surplus functions when prices are uniform and non-uniform respectively. Whether x_c is greater than or less than x_t (the value of x at the global optimum) depends on whether the D_c-AC tangency lies to the right or to the left of the AS_t-AC tangency.[14] With declining average cost, this is determined entirely by the relative slopes of the inverse demand and average surplus functions at the two equilibria. If D_c is steeper than AS_t, x_c is greater than x_t and vice versa. The case illustrated in Figure 3 is that of linear inverse demand functions derived from a quadratic utility function, which is a special case of the example of Section IV.

---

[13] Whether consumers are representative or competition is local plays no necessary role in this analysis.

[14] The absolute difference between x_c and x_t also depends in part on the slope of the average cost function. It is curious that while economists have worked with different demand representations, they haven't investigated the impact of variation in cost functions.

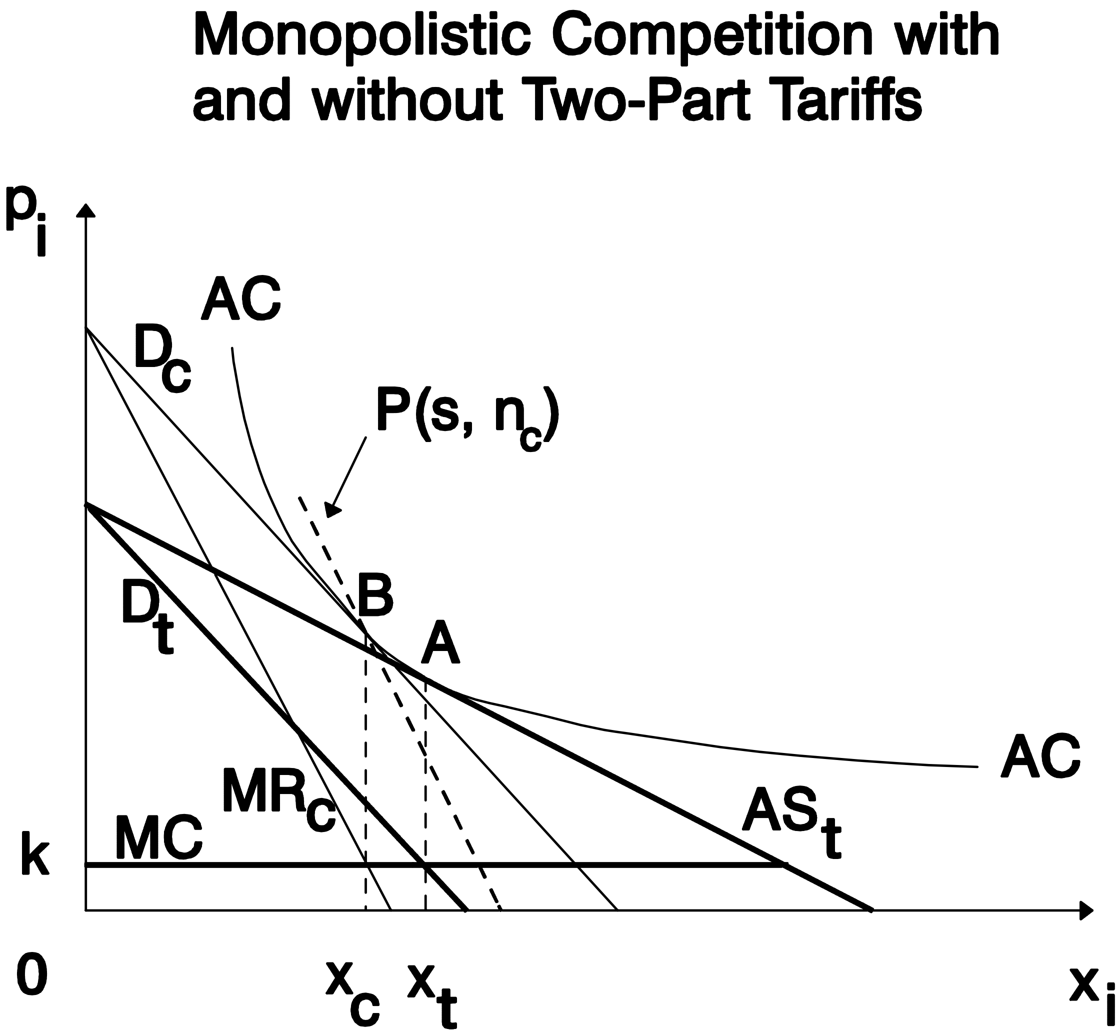

Figure 3

To move from the D_c-AC tangency to the AS_t-AC tangency, the inverse demand function must shift inward. In the process, the slopes of D and AS may change, which will affect the relative values of x_t and x_C. To isolate the effect of the difference in the relative values of the two slopes from the effects of shifts in these values, we assume in this subsection that D and AS shift parallel to themselves in response to entry or increases in the outputs of other firms, before considering the effects of changes in slope. Inverse demand functions that are separable in own output and the outputs of other firms have this property. Inverse demand functions that are linear in own and other firms' outputs are a common example.

Intuition suggests that, in general, the average surplus function (AS) is flatter than the underlying inverse demand function, but that is not always the case. In fact we can show that linear or concave inverse demand functions imply an average surplus function that is flatter than the inverse demand function, and exceptions occur only for very convex inverse demand functions. This is stated in the following Lemma. Its proof is in the appendix.

Lemma 1: (a) For a weakly concave demand curve, the average surplus curve is flatter than the demand curve, $|AS'(x)| < |p'(x)|$.

(b) The average surplus function is concave (convex, linear) if and only if the demand function is concave (convex, linear).

If the average surplus function is flatter than the underlying demand, and the demand curve for variety i shifts parallel to itself when a larger amount of other varieties is produced, we can clearly see that A lies to the right of B on the AC curve in Figure 3. We prove this in the following theorem.

Theorem 1: For weakly concave demand curves, if the demand curve for variety i shifts parallel to itself when the outputs of other varieties increase, then a larger amount of each variety is produced at the NUP equilibrium than at the UP equilibrium.

Proof: Suppose otherwise, i.e., that $x_c > x_t$. By the tangency at the monopolistic competition equilibrium we have

$$|AC'(x_c)| = |p_c'(x_c)|.$$

By the weak concavity of demand curve, we have

$$|p_c'(x_c)| \geq |p_c'(x_t)|.$$

By the parallel shift property,

$$|p_c'(x_t)| = |p_t'(x_t)|.$$

Using Lemma 1(a) at point x_t we have

$$|p_t'(x_t)| > |AS_t'(x_t)|.$$

Finally, by tangency at the TPT equilibrium we have

$$|AS_t'(x_t)| = |AC'(x_t)|.$$

Combining these inequalities we get

$$|AC'(x_c)| > |AC'(x_t)|.$$

But given the assumption in the beginning of the proof that $x_c > x_t$, and the convexity of AC we have

$$|AC'(x_t)| > |AC'(x_c)|,$$

which contradicts the last inequality. Therefore we must have $x_c \leq x_t$.

Equal production at the two equilibria, $x_c = x_t$, is immediately ruled out since if it were true we would have

$$|AC'(x_c)| = |p_c'(x_c)| > |AS_c'(x_c)| = |AS_t'(x_c)| =$$
$$= |AS_t'(x_t)| = |AC'(x_t)| = |AC'(x_c)|,$$

a contradiction. Therefore $x_c < x_t$. QED.

If we relax the parallel shift assumption, then the slopes of D and AS may change as other firms enter or change their outputs. If the slope of AS increases (in absolute value) as it shifts in, this increases the likelihood that $x_c > x_t$ and vice versa if the slope of AS decreases. This should be apparent from inspection of Figure 3.

III.B Comparing UP and NUP Equilibrium Numbers of Varieties

To examine factors influencing the relative numbers of UP and NUP equilibrium product varieties, it is necessary to take explicit account of the behaviors of other firms. Therefore, let $p_i = p_i(x_i, x, n)$ be the inverse demand function for firm i in a market with n firms, where x is the common value of output for all firms except firm i. Let $AS_i(x_i, x, n) = [\int_0^{x_i} p_i(y, x, n)dy)]/x_i$ be the associated average surplus function. With imperfect substitutes, p_i and AS_i are declining in all three arguments.

Let n be the equilibrium number of firms in a symmetric equilibrium and set $x_i = x = s$, a common level of output for all firms in the market. Let $P(s, n) \equiv p_i(s, s, n)$ be the industry inverse demand function scaled to the size of a representative firm for an n-firm equilibrium. Clearly, $dP(s)/ds < \partial p_i/\partial x_i$. That is, the industry inverse demand function is steeper than the individual firm's inverse demand function due to substitutability among products.

Given F and k, the fixed and marginal costs common to all firms, the industry inverse demand function for the UP equilibrium number of firms passes through point B in Figure 3, where $P(x_c, n_c) = AC(x_c) = k + F/x_c$. The industry inverse demand function for the NUP equilibrium number of firms also passes through the point (x_t, k), since at the non-uniform pricing equilibrium, the marginal consumer is willing to pay marginal cost, i.e., $P(x_t, n_t) = k$. Therefore the difference in prices (the markup at B) is equal to the average fixed cost at B:

$$P(x_c, n_c) - P(x_t, n_t) = F/x_c.$$

Using a linear approximation on the LHS we have

$$(dP/ds)(x_t - x_c) + (dP/dn)(n_t - n_c) = - F/x_c \Leftrightarrow n_t - n_c = - [F/x_c + (dP/dx)(x_t - x_c)]/(dP/dn).$$

Since $dP/ds < 0$, $dP/dn < 0$, $n_t > n_c \Leftrightarrow F/x_c > (x_t - x_c)|dP/ds|$. In the "standard" case of Lemma 1 and Theorem 1, $x_t > x_c$, and the inequality is equivalent to the line $P(s, n_c)$ passing through B in Figure 3, which is the industry inverse demand function for n_c firms, intersecting MC to the right of x_t.

Thus, while $x_t \gtrless x_c$ is determined by the relative slopes of the firm's inverse demand and average surplus functions, the slope of the industry inverse demand function must also be considered to determine whether $n_t \gtrless n_c$, that is, whether the UP equilibrium has more or less than the optimum variety. The flatter is the industry inverse demand function, the more likely is $n_t > n_c$. Further, the industry inverse demand function is steeper than the representative firm's inverse demand function, $dP/ds < \partial p_i/\partial x_1$, because increases in competitors' outputs also suppress its price. The price-suppressing effects of competitors' products are greater the more substitutable they are for the firm's own product. Therefore, $n_t < n_c$ is more likely the greater the degree of substitutability between the products of competing sellers. The intuition for this result is straight forward. The higher the degree of substitutability among products, the more will the individual firm's inverse demand function shift inward as firms expand their outputs with a shift from UP to NUP pricing. For a sufficiently high degree of substitutability, the inward shift of firm inverse demand functions will be so large that average surplus functions will end up interior to the average cost function and some firms will have to leave the market if the break-even conditions is to be satisfied.

Clearly, cost plays a role in determining the relative values of n_t and n_c. Consider, for example, a unidimensional spatial market with products evenly distributed throughout the product space in equilibrium.[15] For any seller, the degree of substitutability between its product and the products of it closest competitors is greater the more tightly sellers are packed in the product space. Increasing fixed costs would reduce the number of firms in a UP equilibrium and increase the

[15] See, for example, Salop (1979) and Economides (1989).

average distance between those who remain. This would reduce the substitutability between competitors' products, which would increase the likelihood that $n_t > n_c$.

IV. An Example

We illustrate this analysis by examining symmetric UP and NUP equilibria for inverse demand functions of the form

$$p_i = a - bx_i^{\alpha} - (n - 1)x_c^{\gamma}, \tag{2}$$

where $\alpha, \gamma, a, b, c > 0$, x_i is the representative firm's output, n is the number of firms in the market, and x is the level of output of any firm other than i. This functional form gives us considerable flexibility in examining the independent effects of the curvature of the own and firm-proportionate industry inverse demand functions. When $\alpha = \gamma = 1$, both the firm and industry demands are linear; in that case the demand can be generated by a representative consumer with quadratic utility function

$$U = v + \Sigma_i ax_i - (\Sigma_i bx_i^2 + 2\Sigma_i \Sigma_{j\neq i} cx_ix_j)/2,$$

with $i, j = 1, ..., n$, and $v > 0$.[16]

The cost function $C(x) = F + kx$ is also common to all firms. At the NUP equilibrium, price is equal to marginal cost, $p = k$ and, since entry drives profits to zero, F equals the area above k under the representative firm's inverse demand function up to quantity x_t. Solving for x_t gives

---

[16] Firm inverse demand functions are concave iff $\alpha > 1$. When $\alpha = 1$, the industry inverse demand function is concave if $\gamma > 1$.

$$x_t = [(\alpha + 1)F/(\alpha b)]^{1/(\alpha+1)}. \tag{3}$$

The number of varieties is

$$n_t = \{(a - k)[(\alpha + 1)F/(\alpha b)]^{-\gamma/(\alpha+1)} - b[(\alpha + 1)F/(\alpha b)]^{(\alpha - \gamma)/(\alpha+1)}\}/c + 1. \tag{4}$$

In contrast, the free-entry UP Cournot equilibrium is characterized by two conditions:

$$\partial\Pi_i/\partial x_i = 0 \;\Leftrightarrow\; p_i - k = \alpha b x_i^{\alpha},$$

and

$$\Pi_i = 0 \;\Leftrightarrow\; (p_i - k)x_i = F.$$

Solving for the Cournot UP equilibrium values of x_i and n, we have,

$$x_c = (F/\alpha b)^{1/(\alpha+1)}, \tag{5}$$

and

$$n_c = \{(a - k)[F/(\alpha b)]^{-\gamma/(\alpha+1)} - b(\alpha + 1)[F/(\alpha b)]^{(\alpha - \gamma)/(\alpha+1)}\}/c + 1. \tag{6}$$

It is immediate by inspection that $x_t > x_c$, as long as $\alpha > 0$. Note that this result holds for both concave and convex demand functions, whereas Theorem 1 guaranteed the result only for concave functions.

Comparing the equilibrium numbers of products in UP and NUP competition, we have

$$n_t - n_c = (F/ab)^{-\gamma/(\alpha+1)}\{b(F/ab)^{\alpha/(\alpha+1)}[\alpha+1 - (\alpha+1)^{(\alpha-\gamma)/(\alpha+1)}] - (a - k)[1 - (\alpha+1)^{-\gamma/(\alpha+1)}]\}/c. \tag{7}$$

This comparison makes clear the importance of fixed costs in determining whether product diversity at the UP equilibrium exceeds or falls short of the optimum, which is provided by the NUP equilibrium. The terms in both square brackets in (7) are positive for $\gamma > 0$. The first square brackets is multiplied by an increasing function of fixed cost. Thus, when fixed costs are high, the free entry UP equilibrium number of varieties exceeds the optimum (NUP) number, n_t - n_c. Specifically, let F_1 be the critical fixed cost that makes diversity the same across regimes.[17] $n_t > n_c$ iff $F > F_1$. We can also show that as γ increases, the critical value of F_1 increases;[18] therefore for large γ, it is more likely to have $n_t > n_c$.

V. <u>Equilibria with Non-Zero Competitive Responses</u>

Up to this point, the analysis has assumed that competition is Nash in quantities. It is fairly easy to modify the framework set out in Section III so that Nash equilibria in other strategic variables, such as price or market share can be compared with the UP and NUP Nash quantity equilibria and the optimum. These comparisons provide additional insights into competitive NUP pricing strategies, including the possibility of competitive two-part tariffs with per unit charges less than marginal cost. Facilitating these comparisons is the fact that Nash competition in another strategic variable can always be expressed in terms of the amount by which a seller's competitors' quantities must change to hold their values of the strategic variable constant in the face of its own changes in this variable.[19] For example, the Bertrand assumption that a seller believes that its competitors will not change their prices in response to a cut in its own price is equivalent to the belief that they will reduce their outputs enough to hold their prices constant.

Key to the comparisons of Nash equilibria in different strategy spaces is the fact that all zero-

---

[17] $F_1 = \alpha b\{(a - k)[1 - (\alpha+1)^{-\gamma/(\alpha+1)}]/b[\alpha+1 - (\alpha+1)^{(\alpha-\gamma)/(\alpha+1)}]\}^{(\alpha+1)/\alpha}$.

[18] $dF_1/d\gamma$ is proportional and of the same sign as

$$(a-k)(1+\alpha)^{-1+\gamma/(1+\alpha)}[1+\alpha-(1+\alpha)^{\alpha/(1+\alpha)}]\mathrm{Log}[1+\alpha]/\{b[(1+\alpha)^{\gamma/(1+\alpha)} -(1+\alpha)^{\alpha/(1+\alpha)} + \alpha(1+\alpha)^{\gamma/(1 +\alpha)}]^2\}$$

which is positive since the first square brackets and the logarithm are positive.

[19] See Economides (1995) for a formal proof of this equivalence.

profit equilibria are characterized by tangencies of firm-perceived average revenue and average surplus schedules to a common average cost function. Changes in the average revenue and average surplus schedules that firms believe constrain their profits shift these tangencies with fairly obvious implications for equilibrium outputs and numbers of firms.

Let Δ_i represent firm i's belief regarding $\partial x_c/\partial x_i$. Up to this point we have worked with the assumption that $\Delta_i = 0$, i.e., that firms set output on the belief that their competitors' outputs are fixed. Let $r_i(x_i)$ be the schedule of prices for product i that incorporates i's beliefs regarding its competitors' output responses to variations in its own output. r_i has slope $\partial p_i/\partial x_i + \Delta_i(\partial p_i/\partial x_c)$, and is flatter or steeper than p_i for Δ_i negative or positive, respectively. Because firm i attempts to maximize $(r_i - k_i)x_i - F$, a zero profit symmetric equilibrium is characterized by conditions

$$p_i(x_c) = AC(x_c)$$

and

$$p_i(x_c, x_c, n) + (\partial p_i/\partial x_i + \Delta_i \partial p_i/\partial x_c)x_i - k = 0,$$

where k is marginal cost as before.

These two conditions imply $dAC/dx = \partial r_i/\partial x_i = \partial p_i/\partial dx_i + \Delta_i(\partial p_i/\partial x_c)$; thus, firms' perceived inverse demand functions are tangent to their average cost functions, $dAC/dx_i = (k - AC)/x_i$. Substituting from the first order condition for k and setting $AC = p$ gives $dAC/dx = \partial p_i/\partial dx_i + \Delta_i(\partial p_i/\partial x_c)$. Otherwise firms would not be satisfied with price equal to average cost. Let r_i^z be r_i for $\Delta_i = 0$ and let r_i^n be r_i for some $\Delta_i < 0$, as would be the case with Bertrand competition for example. Consider the Cournot (fixed quantities) UP equilibrium described in Section III. The UP equilibrium with $\Delta_i = 0$ is depicted in Figure 4 by the tangency of $r_i^z(x_c^*, x_c^*, n_c^*)$ with $AC(x)$ at x_c^*. $r_i^n(x_i, x_c^*, n_c^*)$ also passes through this point of tangency. (x_c^*, n_c^*) cannot be a UP equilibrium if $\Delta_i < 0$ as reflected in r_i^n because each of the n_c^* competitors will lower its price in an effort to move into the region above AC to the right of the point of the tangency with r_i^z. Thus, with Bertrand competition equilibrium UP outputs would be larger than the Nash quantity equilibrium output.

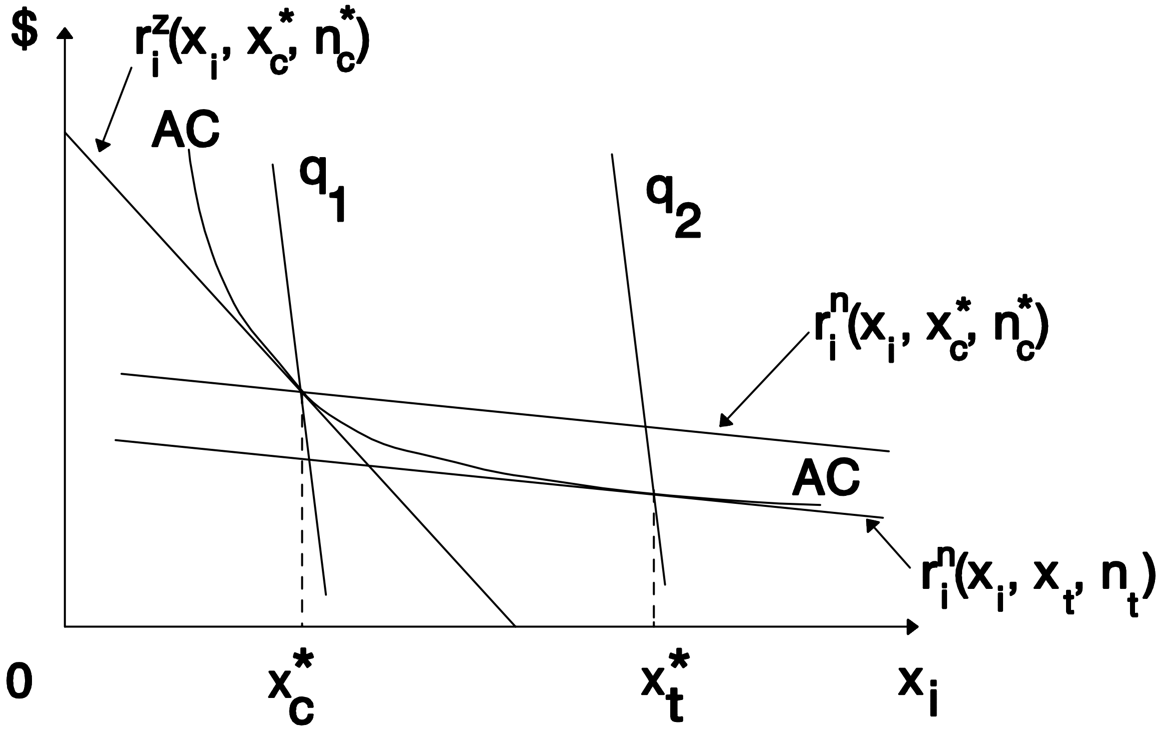

Figure 4

Larger equilibrium outputs also imply fewer firms for a Bertrand equilibrium than for a Cournot equilibrium. Each equilibrium tangency must lie on a firm-scaled industry inverse demand function. We showed in Section III.B that industry inverse demand functions are steeper than firms' inverse demand functions when $\Delta_i = 0$. Therefore, as illustrated in Figure 4, q_2, the industry inverse demand function passing through the r_i^n-AC tangency must lie to the right of q_1, the industry inverse demand function passing through the r_i^z-AC UP equilibrium tangency. The fact that q_2 is to the right of q_1 implies fewer firms in the negative conjecture UP equilibrium.

Clearly the conclusions of this analysis are reversed if $\Delta_i > 0$. r_i will be steeper than p_i so that equilibrium outputs will be smaller than x_c^* and the equilibrium number of firms will be greater than n_c^*.

The analysis of NUP equilibria for Bertrand and other beliefs that are non-Nash in quantities parallels that for UP equilibria. Each firm captures all of the surplus associated with its own output, so $\Pi_i = \int_0^{x_i} p(y, x_c, n)dx_i - kx_i - F$. Let δ_i be the amount by which the representative firm expects its competitors' outputs to change in response to a small positive change in its own output when prices are non-uniform. Then the NUP equilibrium is described by the following conditions:

$$AS_i = AC$$

and

$$p_i + \delta_i \int_0^{x_i} \partial p_i/\partial x_c dx_i - k = 0.$$

When $\delta_i = 0$, we have $p_i = k$, the standard result for monopolists able to perfectly discriminate in price and the constant quantities (zero conjecture) NUP competitive solution of Section II. This is also the welfare optimum. However, the equation of non-uniform pricing with optimum variety and output breaks down if competitors are not expected to hold their outputs constant because $p_i \neq k$ in equilibrium.

Let $R_i(x_i)$ be the schedule of values for its average surplus that incorporate firm i's beliefs regarding its competitors's output responses to changes in its own output.

$$\partial R_i/\partial x_i = [p_i - AS_i + \delta_i \int_0^{x_i} \partial p_i(y, x_c, n)/\partial x_c dy]/x_i.$$

The two equilibrium conditions imply[20]

$$dAC/dx_i = \partial R_i/\partial x_i = \partial AS_i/\partial x_i + \delta_i \int_0^{x_i} \partial p_i(y, x_c, n)/\partial x_c dy]/x_i,$$

[20] The first inequality is produced by substituting from the first order condition for k in $dAC/dx_i = (k - AC)/x_i$ and then substituting AS_i for AC from the zero profit condition, $dAS_i/dx_i = (p_i - AS_i)/x_i$ gives the second equality.

and, because $AS_i = R_i = AC$ in equilibrium, that R_i is tangent to AC.

Assume $\delta_i < 0$. Then R_i is flatter than AS_i, which implies an equilibrium tangency with the average cost curve that is to the right of the AS_i-AC tangency of the zero conjecture NUP equilibrium. So, if the representative firm believes that its competitors will reduce their outputs in response to an increase in its own output, then equilibrium NUP outputs will exceed the optimal value of x^*_n. Similarly, equilibrium NUP outputs will be less that the optimum if firms believe $\delta_i >$ 0.

Define $Q(s, n) = [\int_0^s p_i(x_i, s, n)dx_i]/s$. Q is a firm-scaled industry average surplus function analogous to the firm-scaled industry inverse demand function $P(s, n)$ which was defined in section III.B.[21] For $n = n^*_n$, $Q(s, n^*_n)$ passes through the AS-AC tangency of the NUP equilibrium for $\delta = 0$, and, because products are substitutes, $Q(s, n^*_n)$ is steeper than the AS_i curve at the tangency, as shown in Figure 5.

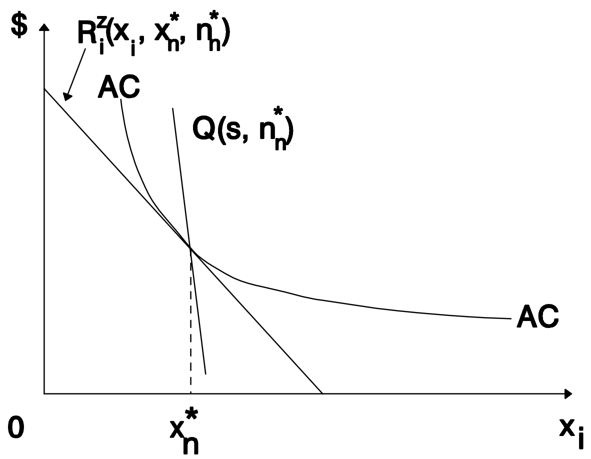

Figure 5

For n continuous, there is a firm-scaled industry average surplus curve through every point on the average cost curve. dAS_i/dx_i, dAS_i/dx_c, and dAS_i/dn are all negative. Therefore industry

[21] Note, however, that $Q(s, s, n)$ is not the area under $P(s, s, n)$ because only x_i is allowed to vary in calculating Q, whereas P is determined by varying x_i and x_c together. x_c is fixed at s in $Q(s, n)$; therefore it must be less than the area under $P(s, s, n)$, which is one nth of the total surplus provided by the industry. This is an alternative way of demonstrating Spence's (1976) result that even when firms practice perfect price discrimination, so that consumers receive no net benefit from the marginal firm, there is surplus leftover that is not captured by consumers in a monopolistically competitive industry.

average surplus schedules that lie to the right of $Q(s, n_n^*)$ must have fewer than n_n^* firms. Because for $\delta_i < 0$ the R_i-AC tangency is to the right of the tangency for $\delta_i = 0$, there will be fewer than the optimum number of firms, n_n^*, if $\delta_i < 0$. A parallel analysis shows that the NUP equilibrium number of firms is greater than n_n^* if $\delta_i > 0$.

An implication of $\delta_i \neq 0$ in the first order condition is that competitive firms employing two-part tariffs will set the per unit charge below marginal cost if $\delta_i < 0$, and above marginal cost if $\delta_i > 0$. The intuition is straight forward. If i's competitors respond to an increase in x_i by reducing their outputs, i's inverse demand function shifts outward, which allows i to increase the size of the fixed component of the tariff. The increase in the fixed component of the tariff compensates i for what would otherwise be losses on a per unit charge below marginal cost.

Similarly, if firms believe that increasing their own outputs will stimulate their competitors to do likewise, the per unit component of a two-part tariff will be set above marginal cost because of the belief that the cost of selling more will include a downward shift in its inverse demand function. Analogous results should hold for other types of full surplus extracting non-linear pricing schemes.

VI. Concluding Remarks

We examined the use of two-part tariffs and more generally non-uniform pricing in monopolistic competition. We found that the non-uniform pricing equilibrium is characterized by a tangency between the average cost and the average surplus function. (The average surplus function is simply the consumers' surplus up to quantity x divided by x.) This tangency is reminiscent of the traditional tangency between average cost and demand characterizing equilibrium for monopolistic competition with uniform pricing.

Non-uniform pricing dominates uniform pricing as a competitive strategy when both are feasible. Consideration of various ways in which full-surplus-extracting, non-uniform prices might be implemented suggests that non-uniform pricing is probably quite common; pricing strategies traditionally viewed as examples of uniform prices are quite likely non-uniform prices instead. Previous attempts to estimate inverse demand functions for differentiated products such as ready-to-eat cereals (Scherer (1979)) and beer (Baker and Bresnahan (1985)) have assumed uniform pricing. Our analysis raises the very strong possibility that the uniform pricing assumption, as well as the implied demand and surplus estimates and policy conclusions of these analyses, is inappropriate for these industries.

There are a number of important differences between the UP and NUP equilibria. The most important difference is that the NUP equilibrium is efficient while the UP equilibrium is inefficient, when firms take their competitors' output as given.[22] Depending on the relative slopes of inverse demand functions and average surplus functions, the UP equilibrium output per firm may be greater or less than the NUP equilibrium output per firm. UP equilibrium output increases relative to NUP equilibrium output the flatter is the representative firm's inverse demand function relative to its average surplus function. Uniform pricing may lead at equilibrium to a greater or smaller number of firms (and varieties) than non-uniform pricing. The greater the degree of substitutability between the products of competing sellers, the larger is the UP equilibrium number of firms relative to the number of firms at a NUP equilibrium. Conversely, increasing fixed cost reduces the number of firms in a UP equilibrium relative to the number for the corresponding NUP equilibrium.

[22] However, the correspondence of non-uniform pricing with efficiency breaks down when firms are not Cournot competitors in quantities. For example, Bertrand equilibria (whether UP or NUP) have fewer firms than Cournot equilibria.

References

Chamberlin, Edward, (1962), *The Theory of Monopolistic Competition*, 8th Edition, Cambridge, MA.

Baker, Jonathan, and Timothy Bresnahan, (1985), "The Gains from Merger or Collusion in Product Differentiated Industries," *Journal of Industrial Economics*, vol. 33, no. 4, pp. 427-444.

Barzel, Yoram, (1970), "Excess Capacity in Monopolistic Competition," *Journal of Political Economy*, vol. 78, pp. 1142-49.

Beckmann, Martin J., (1976), "Spatial Price Policies Revisited," *Bell Journal of Economics*, vol. 7, pp. 619-630.

Besanko, David, Perry, Martin K., and Spady, R.H., (1990), "The Logit Model of Monopolistic Competition: Brand Diversity," *Journal of Industrial Economics*, vol. 38, pp. 397-416.

Borenstein, Severin, (1985), "Price Discrimination in Free-Entry Markets," *Rand Journal of Economics*, vol. 16, pp. 380-397.

Chamberlin, Edward, (1962), *The Theory of Monopolistic Competition*, 8th Edition, Cambridge, MA.

Demsetz, Harold, (1959), "The Nature of Equilibrium in Monopolistic Competition," *Journal of Political Economy*, vol. 67, pp. 21-30.

Demsetz, Harold, (1964), "The Welfare and Empirical Implications of Monopolistic Competition," *Journal of Political Economy*, vol. 74, pp. 621-41.

Demsetz, Harold, (1968), "Do Competition and Monopolistic Competition Differ?," *Journal of Political Economy*, vol. 76, pp. 146-48.

Dixit, Avinash, (1979), "A Model of Duopoly Suggesting a Theory of Entry Barriers," *Bell Journal of Economics*, vol. 10, pp. 20-32.

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

Economides, Nicholas, (1989), "Symmetric Equilibrium Existence and Optimality in Differentiated Products Markets," *Journal of Economic Theory,* vol. 47, no. 1, (1989), pp. 178-194.

Economides, Nicholas, (1995), "The Choice of Strategy Space in Duopoly," mimeo.

Goldman, M. Barry, Hayne Leland and David Sibley, (1984), "Optimal Nonuniform Pricing," *Review of Economic Studies*, vol. LI, pp. 305-319.

Hayes, Beth (1987), "Competition and Two-Part Tariffs," *Journal of Business*, vol. 60, pp. 41-54.

Locay, Luis and Alvaro Rodriguez, (1992), "Price Discrimination in Competitive Markets," *Journal of Political Economy*, vol. 100, no. 5, pp. 954-965.

Mandy, David M., (1991), "Competitive Two-part Tariffs as a Response to Differential Rates of Time Preference, *Economica*, vol. 58, pp. 377-389.

Mandy, David M., (1992), "Nonuniform Bertrand Competition," *Econometrica*, vol. 60, no. 6, pp. 1293-1330.

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

McAfee, Preston, John McMillan, and Michael Whinston, (1989), "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics*, pp. 372-383.

Mussa, Michael and Sherwin Rosen, (1978), "Monopoly and Product Quality," *Journal of Economic Theory*, vol. 18, pp. 301-317.

Panzar, John C. and Andrew Postlewaite (1984), "Sustainable Outlay Schedules," CMSEMS Discussion Paper #6265, Northwestern University.

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

Scherer, F. M., (1979), "The Welfare Economics of Product Variety: An Application to the Ready-To-Eat Cereals Industry," *Journal of Industrial Economics*, vol. 28, pp. 113-134.

Schmalensee, Richard, (1972), "A Note on Monopolistic Competition and Excess Capacity," *Journal of Political Economy*, vol. 76, pp. 568-91.

Shaffer, Sherrill (1987), "Equilibrium Two-Part Tariffs without Market Power or Uncertainty," *Economica*, vol. 54, pp. 315-316.

Spence, Michael, (1976), "Product Selection, Fixed Costs and Monopolistic Competition," *Review of Economic Studies*, vol. 43, pp. 217-236.

Appendix

Lemma 1: (a) For a weakly concave demand curve, the average surplus curve is flatter than the demand curve, $|AS'(x)| < |p'(x)|$.

(b) The average surplus function is concave (convex, linear) if and only if the demand function is concave (convex, linear).

Proof: By definition, $AS(x) = \int_0^x p(y)dy)/x$. Therefore $AS'(x) = [p(x) - AS(x)]/x < 0$, $|AS'(x)| = [AS(x) - p(x)]/x$, $|p'(x)| = - p'(x)$. Now,

$$|p'(x)| > |AS'(x)| \Leftrightarrow p(x) - xp'(x) > AS(x). \qquad (1)$$

For linear demand, average surplus is the price at x/2, i.e., the willingness to pay at the mid-point of the triangle under the demand up to point x. That is, $AS(x) = p(x) - xp'(x)/2$, and therefore (1) holds. For a concave demand function, surplus (and therefore average surplus) at x is smaller than for the linear demand passing through (x, p(x)). Therefore for a concave demand, $AS(x) < p(x) - xp'(x)/2$, and (1) holds again.

To prove part (b), note that

$$AS''(x) = (xp' - 2p + 2AS)/x^2 < 0 \Leftrightarrow AS < p - xp'/2.$$

As we have argued above, the inequality on the RHS holds for concave demand, holds as an equality for linear demand, and is reversed for convex demand. QED.

Remark: The proof shows that, for a linear inverse demand function, AS is strictly flatter than D. Therefore, the same will be true for some convex functions. However, there exist convex demand functions such that AS is parallel to or even steeper than D.[23] Cobb-Douglas demand functions result in steeper AS curves than demand functions.[24]

[23] For example, inverse demand $p(x) = c_1 + c_2 + c_2\log(x)$, $c_2 < 0$, results in total surplus $S(x) = x[c_1 + c_2\log(x)]$ and average surplus $AS(x) = c_1 + c_2\log(x)$. Then the slope of the demand is *equal* to the slope of the average surplus, $p'(x) = AS'(x) = c_2/x$. For this demand function, the average surplus curve is an upward parallel shift of the demand curve. Modifying this example slightly, we can find a demand curve such that its AS curve is steeper than the demand. Consider $p(x) = c_1 + c_2 + c_2\log(x) + 2c_3x$, $c_2 < 0$, $c_3 > 0$. Then $S = x[c_1 + c_2\log(x)] + c_3x^2$, and $AS = c_1 + c_2\log(x) + c_3x$. Now $p'(x) = c_2/x + 2c_3$, while $AS'(x) = c_2/x + c_3$. For small $c_3 > 0$, we have $|AS'(x)| > |p'(x)|$.

[24] Say $p_1 = x_1^{-a}x_2^{b}$, with $a > 0$, so that $x_1 = p_1^{-1/a}x_2^{b/a}$, and the elasticity of demand is $-1/a$. Then $|p_1'(x_1)| = ax_1^{-a-1}x_2^{b}$, and $AS(x_1) = x_1^{-a}x_2^{b}/(1 - a)$, $|AS'(x_1)| = ax_1^{-a-1}x_2^{b}/(1 - a)$. Therefore $|AS'(x_1)| > |p'(x_1)| \Leftrightarrow 1/(1 - a) > 1$, true for $1 > a > 0$.