Monopolistic Competition and Optimum Product Diversity

By AVINASH K. DIXIT AND JOSEPH E. STIGLITZ*

The basic issue concerning production in welfare economics is whether a market solution will yield the socially optimum kinds and quantities of commodities. It is well known that problems can arise for three broad reasons: distributive justice; external effects; and scale economies. This paper is concerned with the last of these.

The basic principle is easily stated.[1] A commodity should be produced if the costs can be covered by the sum of revenues and a properly defined measure of consumer's surplus. The optimum amount is then found by equating the demand price and the marginal cost. Such an optimum can be realized in a market if perfectly discriminatory pricing is possible. Otherwise we face conflicting problems. A competitive market fulfilling the marginal condition would be unsustainable because total profits would be negative. An element of monopoly would allow positive profits, but would violate the marginal condition.[2] Thus we expect a market solution to be suboptimal. However, a much more precise structure must be put on the problem if we are to understand the nature of the bias involved.

It is useful to think of the question as one of quantity versus diversity. With scale economies, resources can be saved by producing fewer goods and larger quantities of each. However, this leaves less variety, which entails some welfare loss. It is easy and probably not too unrealistic to model scale economies by supposing that each potential commodity involves some fixed set-up cost and has a constant marginal cost. Modeling the desirability of variety has been thought to be difficult, and several indirect approaches have been adopted. The Hotelling spatial model, Lancaster's product characteristics approach, and the mean-variance portfolio selection model have all been put to use.[3] These lead to results involving transport costs or correlations among commodities or securities, and are hard to interpret in general terms. We therefore take a direct route, noting that the convexity of indifference surfaces of a conventional utility function defined over the quantities of all potential commodities already embodies the desirability of variety. Thus, a consumer who is indifferent between the quantities (1,0) and (0,1) of two commodities prefers the mix $(1/2, 1/2)$ to either extreme. The advantage of this view is that the results involve the familiar own- and cross-elasticities of demand functions, and are therefore easier to comprehend.

There is one case of particular interest on which we concentrate. This is where potential commodities in a group or sector or industry are good substitutes among themselves, but poor substitutes for the other commodities in the economy. Then we are led to examining the market solution in relation to an optimum, both as regards biases within the group, and between the group and the rest of the economy. We expect the answer to depend on the intra- and intersector elasticities of substitution. To demonstrate the point as simply as possible, we shall aggregate the rest of the economy into one good labeled 0, chosen as the numeraire. The economy's endowment of it is normalized at unity; it can be thought of as the time at the disposal of the consumers.

*Professors of economics, University of Warwick and Stanford University, respectively. Stiglitz's research was supported in part by NSF Grant SOC74-22182 at the Institute for Mathematical Studies in the Social Sciences, Stanford. We are indebted to Michael Spence, to a referee, and the managing editor for comments and suggestions on earlier drafts.

[1]See also the exposition by Michael Spence.

[2]A simple exposition is given by Peter Diamond and Daniel McFadden.

[3]See the articles by Harold Hotelling, Nicholas Stern, Kelvin Lancaster, and Stiglitz.

The potential range of related products is labeled 1,2,3,.... Writing the amounts of the various commodities as x_0 and $x = (x_1, x_2, x_3 \ldots)$, we assume a separable utility function with convex indifference surfaces:

$$u = U(x_0, V(x_1, x_2, x_3 \ldots)) \tag{1}$$

In Sections I and II we simplify further by assuming that V is a symmetric function, and that all commodities in the group have equal fixed and marginal costs. Then the actual labels given to commodities are immaterial, even though the total number n being produced is relevant. We can thus label these commodities $1,2, \ldots, n$, where the potential products $(n+1), (n+2), \ldots$ are not being produced. This is a restrictive assumption, for in such problems we often have a natural asymmetry owing to graduated physical differences in commodities, with a pair close together being better mutual substitutes than a pair farther apart. However, even the symmetric case yields some interesting results. In Section III, we consider some aspects of asymmetry.

We also assume that all commodities have unit income elasticities. This differs from a similar recent formulation by Michael Spence, who assumes U linear in x_0, so that the industry is amenable to partial equilibrium analysis. Our approach allows a better treatment of the intersectoral substitution, but the other results are very similar to those of Spence.

We consider two special cases of (1). In Section I, V is given a *CES* form, but U is allowed to be arbitrary. In Section II, U is taken to be Cobb-Douglas, but V has a more general additive form. Thus the former allows more general intersector relations, and the latter more general intrasector substitution, highlighting different results.

Income distribution problems are neglected. Thus U can be regarded as representing Samuelsonian social indifference curves, or (assuming the appropriate aggregation conditions to be fulfilled) as a multiple of a representative consumer's utility. Product diversity can then be interpreted either as different consumers using different varieties, or as diversification on the part of each consumer.

I. Constant-Elasticity Case

A. *Demand Functions*

The utility function in this section is

$$u = U\left(x_0, \left\{\sum_i x_i^\rho\right\}^{1/\rho}\right) \tag{2}$$

For concavity, we need $\rho < 1$. Further, since we want to allow a situation where several of the x_i are zero, we need $\rho > 0$. We also assume U homothetic in its arguments.

The budget constraint is

$$x_0 + \sum_{i=1}^{n} p_i x_i = I \tag{3}$$

where p_i are prices of the goods being produced, and I is income in terms of the numeraire, i.e., the endowment which has been set at 1 plus the profits of the firms distributed to the consumers, or minus the lump sum deductions to cover the losses, as the case may be.

In this case, a two-stage budgeting procedure is valid.[4] Thus we define dual quantity and price indices

$$y = \left\{\sum_{i=1}^{n} x_i^\rho\right\}^{1/\rho} \qquad q = \left\{\sum_{i=1}^{n} p_i^{-1/\beta}\right\}^{-\beta} \tag{4}$$

where $\beta = (1-\rho)/\rho$, which is positive since $0 < \rho < 1$. Then it can be shown[5] that in the first stage,

$$y = I\frac{s(q)}{q} \qquad x_0 = I(1 - s(q)) \tag{5}$$

for a function s which depends on the form of U. Writing $\sigma(q)$ for the elasticity of substitution between x_0 and y, we define $\theta(q)$ as the elasticity of the function s, i.e., $qs'(q)/s(q)$. Then we find

$$\theta(q) = \{1 - \sigma(q)\}\{1 - s(q)\} < 1 \tag{6}$$

but $\theta(q)$ can be negative as $\sigma(q)$ can exceed 1.

[4]See p. 21 of John Green.

[5]These details and several others are omitted to save space, but can be found in the working paper by the authors, cited in the references.

Turning to the second stage of the problem, it is easy to show that for each i,

$$x_i = y\left[\frac{q}{p_i}\right]^{1/(1-\rho)} \tag{7}$$

where y is defined by (4). Consider the effect of a change in p_i alone. This affects x_i directly, and also through q; thence through y as well. Now from (4) we have the elasticity

$$\frac{\partial \log q}{\partial \log p_i} = \left(\frac{q}{p_i}\right)^{1/\beta} \tag{8}$$

So long as the prices of the products in the group are not of different orders of magnitude, this is of the order $(1/n)$. We shall assume that n is reasonably large, and accordingly neglect the effect of each p_i on q; thus the indirect effects on x_i. This leaves us with the elasticity

$$\frac{\partial \log x_i}{\partial \log p_i} = \frac{-1}{(1-\rho)} = \frac{-(1+\beta)}{\beta} \tag{9}$$

In the Chamberlinian terminology, this is the elasticity of the *dd* curve, i.e., the curve relating the demand for each product type to its own price with all other prices held constant.

In our large group case, we also see that for $i \neq j$, the cross elasticity $\partial \log x_i/\partial \log p_j$ is negligible. However, if all prices in the group move together, the individually small effects add to a significant amount. This corresponds to the Chamberlinian *DD* curve. Consider a symmetric situation where $x_i = x$ and $p_i = p$ for all i from 1 to n. We have

$$\begin{aligned} y &= xn^{1/\rho} = xn^{1+\beta} \\ q &= pn^{-\beta} = pn^{-(1-\rho)/\rho} \end{aligned} \tag{10}$$

and then from (5) and (7),

$$x = \frac{Is(q)}{pn} \tag{11}$$

The elasticity of this is easy to calculate; we find

$$\frac{\partial \log x}{\partial \log p} = -[1-\theta(q)] \tag{12}$$

Then (6) shows that the *DD* curve slopes downward. The conventional condition that the *dd* curve be more elastic is seen from (9) and (12) to be

$$\frac{1}{\beta} + \theta(q) > 0 \tag{13}$$

Finally, we observe that for $i \neq j$,

$$\frac{x_i}{x_j} = \left[\frac{p_j}{p_i}\right]^{1/(1-\rho)} \tag{14}$$

Thus $1/(1-\rho)$ is the elasticity of substitution between any two products within the group.

B. *Market Equilibrium*

It can be shown that each commodity is produced by one firm. Each firm attempts to maximize its profit, and entry occurs until the marginal firm can only just break even. Thus our market equilibrium is the familiar case of Chamberlinian monopolistic competition, where the question of quantity versus diversity has often been raised.[6] Previous analyses have failed to consider the desirability of variety in an explicit form, and have neglected various intra- and intersector interactions in demand. As a result, much vague presumption that such an equilibrium involves excessive diversity has built up at the back of the minds of many economists. Our analysis will challenge several of these ideas.

The profit-maximization condition for each firm acting on its own is the familiar equality of marginal revenue and marginal cost. Writing c for the common marginal cost, and noting that the elasticity of demand for each firm is $(1+\beta)/\beta$, we have for each active firm:

$$p_i\left(1 - \frac{\beta}{1+\beta}\right) = c$$

Writing p_e for the common equilibrium price for each variety being produced, we have

$$p_e = c(1+\beta) = \frac{c}{\rho} \tag{15}$$

[6]See Edwin Chamberlin, Nicholas Kaldor, and Robert Bishop.

The second condition for equilibrium is that firms enter until the next potential entrant would make a loss. If n is large enough so that 1 is a small increment, we can assume that the marginal firm is exactly breaking even, i.e., $(p_n - c)x_n = a$, where x_n is obtained from the demand function and a is the fixed cost. With symmetry, this implies zero profit for all intramarginal firms as well. Then $I = 1$, and using (11) and (15) we can write the condition so as to yield the number n_e of active firms:

$$\frac{s(p_e n_e^{-\beta})}{p_e n_e} = \frac{a}{\beta c} \tag{16}$$

Equilibrium is unique provided $s(p_e n^{-\beta})/p_e n$ is a monotonic function of n. This relates to our earlier discussion about the two demand curves. From (11) we see that the behavior of $s(pn^{-\beta})/pn$ as n increases tells us how the demand curve DD for each firm shifts as the number of firms increases. It is natural to assume that it shifts to the left, i.e., the function above decreases as n increases for each fixed p. The condition for this in elasticity form is easily seen to be

$$1 + \beta\theta(q) > 0 \tag{17}$$

This is exactly the same as (13), the condition for the dd curve to be more elastic than the DD curve, and we shall assume that it holds.

The condition can be violated if $\sigma(q)$ is sufficiently higher than one. In this case, an increase in n lowers q, and shifts demand towards the monopolistic sector to such an extent that the demand curve for each firm shifts to the right. However, this is rather implausible.

Conventional Chamberlinian analysis assumes a fixed demand curve for the group as a whole. This amounts to assuming that $n \cdot x$ is independent of n, i.e., that $s(pn^{-\beta})$ is independent of n. This will be so if $\beta = 0$, or if $\sigma(q) = 1$ for all q. The former is equivalent to assuming that $\rho = 1$, when all products in the group are perfect substitutes, i.e., diversity is not valued at all. That would be contrary to the intent of the whole analysis. Thus, implicitly, conventional analysis assumes $\sigma(q) = 1$. This gives a constant budget share for the monopolistically competitive sector. Note that in our parametric formulation, this implies a unit-elastic DD curve, (17) holds, and so equilibrium is unique.

Finally, using (7), (11), and (16), we can calculate the equilibrium output for each active firm:

$$x_e = \frac{a}{\beta c} \tag{18}$$

We can also write an expression for the budget share of the group as a whole:

$$s_e = s(q_e) \tag{19}$$

where $$q_e = p_e n_e^{-\beta}$$

These will be useful for subsequent comparisons.

C. *Constrained Optimum*

The next task is to compare the equilibrium with a social optimum. With economies of scale, the first best or unconstrained (really constrained only by technology and resource availability) optimum requires pricing below average cost, and therefore lump sum transfers to firms to cover losses. The conceptual and practical difficulties of doing so are clearly formidable. It would therefore appear that a more appropriate notion of optimality is a constrained one, where each firm must have nonnegative profits. This may be achieved by regulation, or by excise or franchise taxes or subsidies. The important restriction is that lump sum subsidies are not available.

We begin with such a constrained optimum. The aim is to choose n, p_i, and x_i so as to maximize utility, satisfying the demand functions and keeping the profit for each firm nonnegative. The problem is somewhat simplified by the result that all active firms should have the same output levels and prices, and should make exactly zero profit. We omit the proof. Then we can set $I = 1$, and use (5) to express utility as a function of q alone. This is of course a decreasing function. Thus the problem of maximizing u becomes that of minimizing q, i.e.,

$$\min_{n,p} pn^{-\beta}$$

subject to

$$(20) \qquad (p - c)\frac{s(pn^{-\beta})}{pn} = a$$

To solve this, we calculate the logarithmic marginal rate of substitution along a level curve of the objective, the similar rate of transformation along the constraint, and equate the two. This yields the condition

$$(21) \qquad \frac{\frac{c}{p-c} + \theta(q)}{1 + \beta\theta(q)} = \frac{1}{\beta}$$

The second-order condition can be shown to hold, and (21) simplifies to yield the price for each commodity produced in the constrained optimum, p_c, as

$$(22) \qquad p_c = c(1 + \beta)$$

Comparing (15) and (22), we see that the two solutions have the same price. Since they face the same break-even constraint, they have the same number of firms as well, and the values for all other variables can be calculated from these two. Thus we have a rather surprising case where the monopolistic competition equilibrium is identical with the optimum constrained by the lack of lump sum subsidies. Chamberlin once suggested that such an equilibrium was "a sort of ideal"; our analysis shows when and in what sense this can be true.

D. *Unconstrained Optimum*

These solutions can in turn be compared to the unconstrained or first best optimum. Considerations of convexity again establish that all active firms should produce the same output. Thus we are to choose n firms each producing output x in order to maximize

$$(23) \qquad u = U(1 - n(a + cx), xn^{1+\beta})$$

where we have used the economy's resource balance condition and (10). The first-order conditions are

$$(24) \qquad -ncU_0 + n^{1+\beta}U_y = 0$$

$$(25) \qquad -(a + cx)U_0 + (1 + \beta)xn^{\beta}U_y = 0$$

From the first stage of the budgeting problem, we know that $q = U_y/U_0$. Using (24) and (10), we find the price charged by each active firm in the unconstrained optimum, p_u, equal to marginal cost

$$(26) \qquad p_u = c$$

This, of course, is no surprise. Also from the first-order conditions, we have

$$(27) \qquad x_u = \frac{a}{c\beta}$$

Finally, with (26), each active firm covers its variable cost exactly. The lump sum transfers to firms then equal an, and therefore $I = 1 - an$, and

$$x = (1 - an)\frac{s(pn^{-\beta})}{pn}$$

The number of firms n_u is then defined by

$$(28) \qquad \frac{s(cn_u^{-\beta})}{n_u} = \frac{a/\beta}{1 - an_u}$$

We can now compare these magnitudes with the corresponding ones in the equilibrium or the constrained optimum. The most remarkable result is that the output of each active firm is the same in the two situations. The fact that in a Chamberlinian equilibrium each firm operates to the left of the point of minimum average cost has been conventionally described by saying that there is excess capacity. However, when variety is desirable, i.e., when the different products are not perfect substitutes, it is not in general optimum to push the output of each firm to the point where all economies of scale are exhausted.[7] We have shown in one case that is not an extreme one, that the first best optimum does not exploit economies of scale beyond the extent achieved in the equilibrium. We can then easily conceive of cases where the equilibrium exploits economies of scale too far from the point of view of social optimality. Thus our results undermine the validity of the folklore of excess capacity, from the point of view of the

[7]See David Starrett.

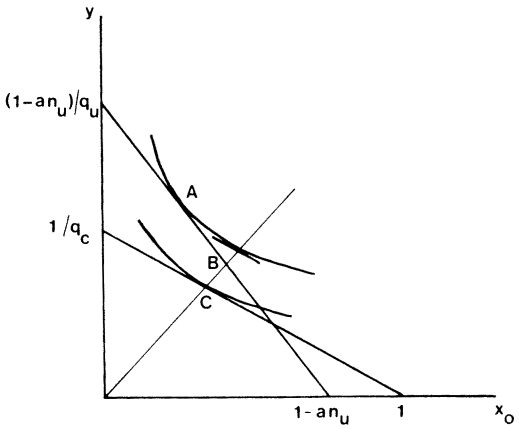

FIGURE 1

unconstrained optimum as well as the constrained one.

A direct comparison of the numbers of firms from (16) and (28) would be difficult, but an indirect argument turns out to be simple. It is clear that the unconstrained optimum has higher utility than the constrained optimum. Also, the level of lump sum income in it is less than that in the latter. It must therefore be the case that

$$q_u < q_c = q_e \tag{29}$$

Further, the difference must be large enough that the budget constraint for x_0 and the quantity index y in the unconstrained case must lie outside that in the constrained case in the relevant region, as shown in Figure 1. Let C be the constrained optimum, A the unconstrained optimum, and let B be the point where the line joining the origin to C meets the indifference curve in the unconstrained case. By homotheticity the indifference curve at B is parallel to that at C, so each of the moves from C to B and from B to A increases the value of y. Since the value of x is the same in the two optima, we must have

$$n_u > n_c = n_e \tag{30}$$

Thus the unconstrained optimum actually allows more variety than the constrained optimum and the equilibrium; this is another point contradicting the folklore on excessive diversity.

Using (29) we can easily compare the budget shares. In the notation we have been using, we find $s_u \gtrless s_c$ as $\theta(q) \gtrless 0$, i.e., as $\sigma(q) \gtrless 1$ providing these hold over the entire relevant range of q.

It is not possible to have a general result concerning the relative magnitudes of x_0 in the two situations; an inspection of Figure 1 shows this. However, we have a sufficient condition:

$$x_{0u} = (1 - an_u)(1 - s_u) < 1 - s_u \leq 1 - s_c = x_{0c} \text{ if } \sigma(q) \geq 1$$

In this case the equilibrium or the constrained optimum use more of the numeraire resource than the unconstrained optimum. On the other hand, if $\sigma(q) = 0$ we have L-shaped isoquants, and in Figure 1, points A and B coincide giving the opposite conclusion.

In this section we have seen that with a constant intrasector elasticity of substitution, the market equilibrium coincides with the constrained optimum. We have also shown that the unconstrained optimum has a greater number of firms, each of the same size. Finally, the resource allocation between the sectors is shown to depend on the intersector elasticity of substitution. This elasticity also governs conditions for uniqueness of equilibrium and the second-order conditions for an optimum.

Henceforth we will achieve some analytic simplicity by making a particular assumption about intersector substitution. In return, we will allow a more general form of intrasector substitution.

II. Variable Elasticity Case

The utility function is now

$$u = x_0^{1-\gamma}\left\{\sum_i v(x_i)\right\}^{\gamma} \tag{31}$$

with v increasing and concave, $0 < \gamma < 1$. This is somewhat like assuming a unit intersector elasticity of substitution. However, this is not rigorous since the group utility $V(\underline{x}) = \sum_i v(x_i)$ is not homothetic and therefore two-stage budgeting is not applicable.

It can be shown that the elasticity of the *dd* curve in the large group case is

$$(32) \quad -\frac{\partial \log x_i}{\partial \log p_i} = -\frac{v'(x_i)}{x_i v''(x_i)} \quad \text{for any } i$$

This differs from the case of Section I in being a function of x_i. To highlight the similarities and the differences, we define $\beta(x)$ by

$$(33) \quad \frac{1 + \beta(x)}{\beta(x)} = -\frac{v'(x)}{xv''(x)}$$

Next, setting $x_i = x$ and $p_i = p$ for $i = 1, 2, \ldots, n$, we can write the DD curve and the demand for the numeraire as

$$(34) \quad x = \frac{I}{np}\omega(x), \qquad x_0 = I[1 - \omega(x)]$$

where

$$(35) \quad \omega(x) = \frac{\gamma\rho(x)}{[\gamma\rho(x) + (1 - \gamma)]}$$

$$\rho(x) = \frac{xv'(x)}{v(x)}$$

We assume that $0 < \rho(x) < 1$, and therefore have $0 < \omega(x) < 1$.

Now consider the Chamberlinian equilibrium. The profit-maximization condition for each active firm yields the common equilibrium price p_e in terms of the common equilibrium output x_e as

$$(36) \quad p_e = c[1 + \beta(x_e)]$$

Note the analogy with (15). Substituting (36) in the zero pure profit condition, we have x_e defined by

$$(37) \quad \frac{cx_e}{a + cx_e} = \frac{1}{1 + \beta(x_e)}$$

Finally, the number of firms can be calculated using the DD curve and the break-even condition, as

$$(38) \quad n_e = \frac{\omega(x_e)}{a + cx_e}$$

For uniqueness of equilibrium we once again use the conditions that the dd curve is more elastic than the DD curve, and that entry shifts the DD curve to the left. However, these conditions are rather involved and opaque, so we omit them.

Let us turn to the constrained optimum. We wish to choose n and x to maximize u, subject to (34) and the break-even condition $px = a + cx$. Substituting, we can express u as a function of x alone:

$$(39) \quad u = \gamma^{\gamma}(1 - \gamma)^{(1-\gamma)} \frac{\left[\frac{\rho(x)v(x)}{a + cx}\right]^{\gamma}}{\gamma\rho(x) + (1 - \gamma)}$$

The first-order condition defines x_c:

$$(40) \quad \frac{cx_c}{a + cx_c} = \frac{1}{1 + \beta(x_c)} - \frac{\omega(x_c)x_c\rho'(x_c)}{\gamma\rho(x_c)}$$

Comparing this with (37) and using the second-order condition, it can be shown that provided $\rho'(x)$ is one-signed for all x,

$$(41) \quad x_c \gtrless x_e \text{ according as } \rho'(x) \lessgtr 0$$

With zero pure profit in each case, the points (x_e, p_e) and (x_c, p_c) lie on the same declining average cost curve, and therefore

$$(42) \quad p_c \lessgtr p_e \text{ according as } x_c \gtrless x_e$$

Next we note that the dd curve is tangent to the average cost curve at (x_e, p_e) and the DD curve is steeper. Consider the case $x_c > x_e$. Now the point (x_c, p_c) must lie on a DD curve further to the right than (x_e, p_e), and therefore must correspond to a smaller number of firms. The opposite happens if $x_c < x_e$. Thus,

$$(43) \quad n_c \lessgtr n_e \text{ according as } x_c \gtrless x_e$$

Finally, (41) shows that in both cases that arise there, $\rho(x_c) < \rho(x_e)$. Then $\omega(x_c) < \omega(x_e)$, and from (34),

$$(44) \quad x_{0c} > x_{0e}$$

A smaller degree of intersectoral substitution could have reversed the result, as in Section I.

An intuitive reason for these results can be given as follows. With our large group assumptions, the revenue of each firm is proportional to $xv'(x)$. However, the contribution of its output to group utility is $v(x)$. The ratio of the two is $\rho(x)$. Therefore, if $\rho'(x) > 0$, then at the margin each firm finds it more profitable to expand than what would be socially desirable, so $x_e > x_c$.

Given the break-even constraint, this leads to there being fewer firms.

Note that the relevant magnitude is the elasticity of utility, and not the elasticity of demand. The two are related, since

$$(45) \qquad x\frac{\rho'(x)}{\rho(x)} = \frac{1}{1+\beta(x)} - \rho(x)$$

Thus, if $\rho(x)$ is constant over an interval, so is $\beta(x)$ and we have $1/(1+\beta) = \rho$, which is the case of Section I. However, if $\rho(x)$ varies, we cannot infer a relation between the signs of $\rho'(x)$ and $\beta'(x)$. Thus the variation in the elasticity of demand is not in general the relevant consideration. However, for important families of utility functions there is a relationship. For example, for $v(x) = (k + mx)^j$, with $m > 0$ and $0 < j < 1$, we find that $-xv''/v'$ and xv'/v are positively related. Now we would normally expect that as the number of commodities produced increases, the elasticity of substitution between any pair of them should increase. In the symmetric equilibrium, this is just the inverse of the elasticity of marginal utility. Then a higher x would correspond to a lower n, and therefore a lower elasticity of substitution, higher $-xv''/v'$ and higher xv'/v. Thus we are led to expect that $\rho'(x) > 0$, i.e., that the equilibrium involves fewer and bigger firms than the constrained optimum. Once again the common view concerning excess capacity and excessive diversity in monopolistic competition is called into question.

The unconstrained optimum problem is to choose n and x to maximize

$$(46) \qquad u = [nv(x)]^{\gamma}[1 - n(a + cx)]^{1-\gamma}$$

It is easy to show that the solution has

$$(47) \qquad p_u = c$$

$$(48) \qquad \frac{cx_u}{a + cx_u} = \rho(x_u)$$

$$(49) \qquad n_u = \frac{\gamma}{a + cx_u}$$

Then we can use the second-order condition to show that

$$(50) \qquad x_u \lesseqgtr x_c \text{ according as } \rho'(x) \gtreqless 0$$

This is in each case transitive with (41), and therefore yields similar output comparisons between the equilibrium and the unconstrained optimum.

The price in the unconstrained optimum is of course the lowest of the three. As to the number of firms, we note

$$n_c = \frac{\omega(x_c)}{a + cx_c} < \frac{\gamma}{a + cx_c}$$

and therefore we have a one-way comparison:

$$(51) \qquad \text{If } x_u < x_c, \text{ then } n_u > n_c$$

Similarly for the equilibrium. These leave open the possibility that the unconstrained optimum has both bigger and more firms. That is not unreasonable; after all the unconstrained optimum uses resources more efficiently.

III. Asymmetric Cases

The discussion so far imposed symmetry within the group. Thus the number of varieties being produced was relevant, but any group of n was just as good as any other group of n. The next important modification is to remove this restriction. It is easy to see how interrelations within the group of commodities can lead to biases. Thus, if no sugar is being produced, the demand for coffee may be so low as to make its production unprofitable when there are set-up costs. However, this is open to the objection that with complementary commodities, there is an incentive for one entrant to produce both. However, problems exist even when all the commodities are substitutes. We illustrate this by considering an industry which will produce commodities from one of two groups, and examine whether the choice of the wrong group is possible.[8]

Suppose there are two sets of commodities beside the numeraire, the two being perfect substitutes for each other and each having a constant elasticity subutility function. Further, we assume a constant budget share

[8]For an alternative approach using partial equilibrium methods, see Spence.

for the numeraire. Thus the utility function is

(52)
$$u = x_0^{1-s}\left\{\left[\sum_{i_1=1}^{n} x_{i_1}^{\rho_1}\right]^{1/\rho_1} + \left[\sum_{i_2=1}^{n_2} x_{i_2}^{\rho_2}\right]^{1/\rho_2}\right\}^{s}$$

We assume that each firm in group i has a fixed cost a_i and a constant marginal cost c_i.

Consider two types of equilibria, only one commodity group being produced in each. These are given by

(53a)
$$\bar{x}_1 = \frac{a_1}{c_1\beta_1},\ \bar{x}_2 = 0$$
$$\bar{p}_1 = c_1(1+\beta_1)$$
$$\bar{n}_1 = \frac{s\beta_1}{a_1(1+\beta_1)}$$
$$\bar{q}_1 = \bar{p}_1\bar{n}_1^{-\beta_1} = c_1(1+\beta_1)^{1+\beta_1}\left(\frac{a_1}{s}\right)^{\beta_1}$$
$$\bar{u}_1 = s^s(1-s)^{1-s}\bar{q}_1^{-s}$$

(53b)
$$\bar{x}_2 = \frac{a_2}{c_2\beta_2},\ \bar{x}_1 = 0$$
$$\bar{p}_2 = c_2(1+\beta_2)$$
$$\bar{n}_2 = \frac{s\beta_2}{a_2(1+\beta_2)}$$
$$\bar{q}_2 = \bar{p}_2\bar{n}_2^{-\beta_2} = c_2(1+\beta_2)^{1+\beta_2}\left(\frac{a_2}{s}\right)^{\beta_2}$$
$$\bar{u}_2 = s^s(1-s)^{1-s}\bar{q}_2^{-s}$$

Equation (53a) is a Nash equilibrium if and only if it does not pay a firm to produce a commodity of the second group. The demand for such a commodity is

$$x_2 = \begin{cases} 0 & \text{for } p_2 \geq \bar{q}_1 \\ s/p_2 & \text{for } p_2 < \bar{q}_1 \end{cases}$$

Hence we require

$$\max_{p_2}(p_2 - c_2)x_2 = s\left(1 - \frac{c_2}{\bar{q}_1}\right) < a_2$$

or

(54)
$$\bar{q}_1 < \frac{sc_2}{s - a_2}$$

Similarly, (53b) is a Nash equilibrium if and only if

(55)
$$\bar{q}_2 < \frac{sc_1}{s - a_1}$$

Now consider the optimum. Both the objective and the constraint are such as to lead the optimum to the production of commodities from only one group. Thus, suppose n_i commodities from group i are being produced at levels x_i each, and offered at prices p_i. The utility level is given by

(56)
$$u = x_0^{1-s}\{x_1 n_1^{1+\beta_1} + x_2 n_2^{1+\beta_2}\}^s$$

and the resource availability constraint is

(57)
$$x_0 + n_1(a_1 + c_1x_1) + n_2(a_2 + c_2x_2) = 1$$

Given the values of the other variables, the level curves of u in (n_1, n_2) space are concave to the origin, while the constraint is linear. We must therefore have a corner optimum. (As for the break-even constraint, unless the two $q_i = p_i n_i^{-\beta_i}$ are equal, the demand for commodities in one group is zero, and there is no possibility of avoiding a loss there.)

Note that we have structured our example so that if the correct group is chosen, the equilibrium will not introduce any further biases in relation to the constrained optimum. Therefore, to find the constrained optimum, we only have to look at the values of $\bar{u}_i$ in (53a) and (53b) and see which is the greater. In other words, we have to see which $\bar{q}_i$ is the smaller, and choose the situation (which may or may not be a Nash equilibrium) defined in (53a) and (53b) corresponding to it.

Figure 2 is drawn to depict the possible equilibria and optima. Given all the relevant parameters, we calculate $(\bar{q}_1, \bar{q}_2)$ from (53a) and (53b). Then (54) and (55) tell us whether either or both of the situations are possible equilibria, while a simple comparison of the magnitudes of $\bar{q}_1$ and $\bar{q}_2$ tells us which is the constrained optimum. In the figure, the nonnegative quadrant is split into regions in each of which we have one combination of equilibria and optima. We only have to locate the point $(\bar{q}_1, \bar{q}_2)$ in this space to know the result for the given

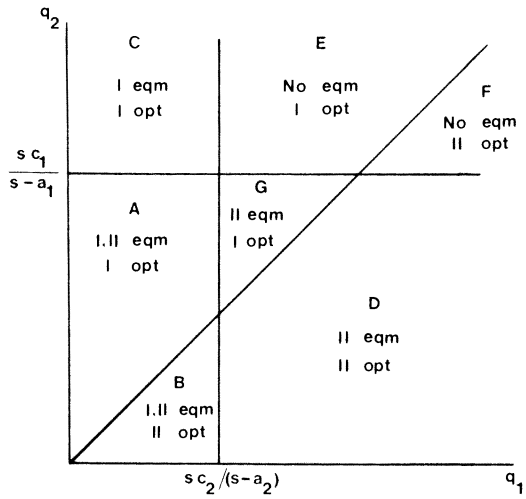

FIGURE 2. SOLUTIONS LABELED I REFER TO EQUATION (53a); SOLUTIONS LABELED II REFER TO EQUATION (53b)

parameter values. Moreover, we can compare the location of the points corresponding to different parameter values and thus do some comparative statics.

To understand the results, we must examine how $\bar{q}_i$ depends on the relevant parameters. It is easy to see that each is an increasing function of a_i and c_i. We also find

$$\frac{\partial \log \bar{q}_i}{\partial \beta_i} = -\log \bar{n}_i \tag{58}$$

and we expect this to be large and negative. Further, we see from (9) that a higher β_i corresponds to a lower own-price elasticity of demand for each commodity in that group. Thus $\bar{q}_i$ is an increasing function of this elasticity.

Consider initially a symmetric situation, with $sc_1/(s - a_1) = sc_2/(s - a_2)$, $\beta_1 = \beta_2$ (the region G vanishes then), and suppose the point $(\bar{q}_1, \bar{q}_2)$ is on the boundary between regions A and B. Now consider a change in one parameter, say, a higher own-elasticity for commodities in group 2. This raises $\bar{q}_2$, moving the point into region A, and it becomes optimal to produce commodities from group 1 alone. However, both (53a) and (53b) are possible Nash equilibria, and it is therefore possible that *the high elasticity group is produced in equilibrium when the low elasticity one should have been.* If the difference in elasticities is large enough, the point moves into region C, where (53b) is no longer a Nash equilibrium. But, owing to the existence of a fixed cost, a significant difference in elasticities is necessary before entry from group 1 commodities threatens to destroy the "wrong" equilibrium. Similar remarks apply to regions B and D.

Next, begin with symmetry once again, and consider a higher c_1 or a_1. This increases $\bar{q}_1$ and moves the point into region B, making it optimal to produce the low-cost group alone while leaving both (53a) and (53b) as possible equilibria, until the difference in costs is large enough to take the point to region D. The change also moves the boundary between A and C upward, opening up a larger region G, but that is not of significance here.

If both $\bar{q}_1$ and $\bar{q}_2$ are large, each group is threatened by profitable entry from the other, and no Nash equilibrium exists, as in regions E and F. However, the criterion of constrained optimality remains as before. Thus we have a case where it may be necessary to prohibit entry in order to sustain the constrained optimum.

If we combine a case where $c_1 > c_2$ (or $a_1 > a_2$) and $\beta_1 > \beta_2$, i.e., where commodities in group 2 are more elastic and have lower costs, we face a still worse possibility. For the point $(\bar{q}_1, \bar{q}_2)$ may then lie in region G, where only (53b) is a possible equilibrium and only (53a) is constrained optimum, i.e., the market can produce only a low cost, high demand elasticity group of commodities when a high cost, low demand elasticity group should have been produced.

Very roughly, the point is that although commodities in inelastic demand have the potential for earning revenues in excess of variable costs, they also have significant consumers' surpluses associated with them. Thus it is not immediately obvious whether the market will be biased in favor of them or against them as compared with an optimum. Here we find the latter, and independent findings of Michael Spence in other

contexts confirm this. Similar remarks apply to differences in marginal costs.

In the interpretation of the model with heterogenous consumers and social indifference curves, inelastically demanded commodities will be the ones which are intensively desired by a few consumers. Thus we have an "economic" reason why the market will lead to a bias against opera relative to football matches, and a justification for subsidization of the former and a tax on the latter, provided the distribution of income is optimum.

Even when cross elasticities are zero, there may be an incorrect choice of commodities to be produced (relative either to an unconstrained or constrained optimum) as Figure 3 illustrates. Figure 3 illustrates a case where commodity A has a more elastic demand curve than commodity B; A is produced in monopolistically competitive equilibrium, while B is not. But clearly, it is socially desirable to produce B, since ignoring consumer's surplus it is just marginal. Thus, the commodities that are not produced but ought to be are those with inelastic demands. Indeed, if, as in the usual analysis of monopolistic competition, eliminating one firm shifts the demand curve for the other firms to the right (i.e., increases the demand for other firms), if the consumer surplus from A (at its equilibrium level of output) is less than that from B (i.e., the cross hatched area exceeds the striped area), then constrained Pareto optimality entails restricting the production of the commodity with the more elastic demand.

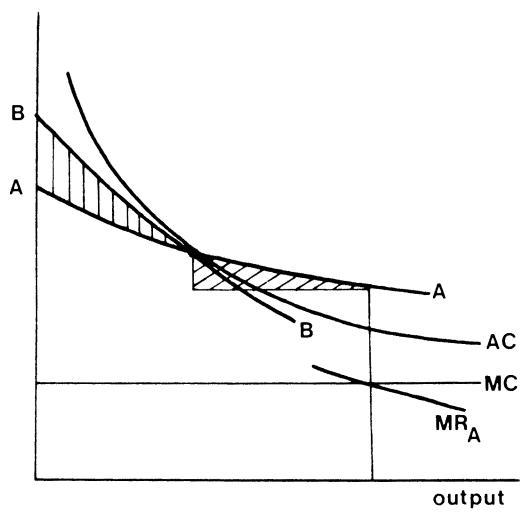

FIGURE 3

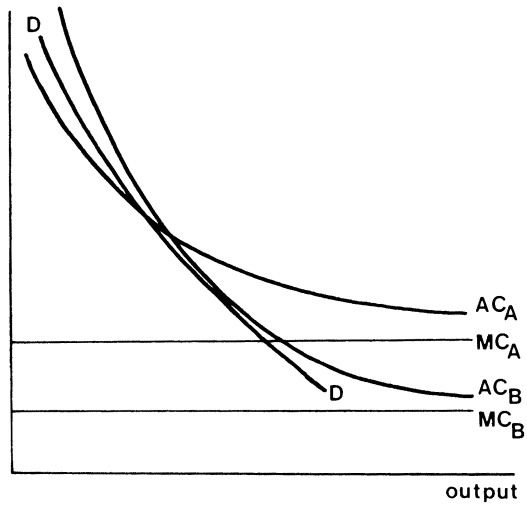

FIGURE 4

A similar analysis applies to commodities with the same demand curves but different cost structures. Commodity A is assumed to have the lower fixed cost but the higher marginal cost. Thus, the average cost curves cross but once, as in Figure 4. Commodity A is produced in monopolistically competitive equilibrium, commodity B is not (although it is just at the margin of being produced). But again, observe that B should be produced, since there is a large consumer's surplus; indeed, since were it to be produced, B would produce at a much higher level than A, there is a much larger consumer's surplus. Thus if the government were to forbid the production of A, B would be viable, and social welfare would increase.

In the comparison between constrained Pareto optimality and the monopolistically competitive equilibrium, we have observed that in the former, we replace some low fixed cost-high marginal cost commodities with high fixed cost-low marginal cost commodities, and we replace some commodities

with elastic demands with commodities with inelastic demands.

IV. Concluding Remarks

We have constructed in this paper some models to study various aspects of the relationship between market and optimal resource allocation in the presence of some nonconvexities. The following general conclusions seem worth pointing out.

The monopoly power, which is a necessary ingredient of markets with nonconvexities, is usually considered to distort resources away from the sector concerned. However, in our analysis monopoly power enables firms to pay fixed costs, and entry cannot be prevented, so the relationship between monopoly power and the direction of market distortion is no longer obvious.

In the central case of a constant elasticity utility function, the market solution was constrained Pareto optimal, regardless of the value of that elasticity (and thus the implied elasticity of the demand functions). With variable elasticities, the bias could go either way, and the direction of the bias depended not on how the elasticity of demand changed, but on how the elasticity of utility changed. We suggested that there was some presumption that the market solution would be characterized by too few firms in the monopolistically competitive sector.

With asymmetric demand and cost conditions we also observed a bias against commodities with inelastic demands and high costs.

The general principle behind these results is that a market solution considers profit at the appropriate margin, while a social optimum takes into account the consumer's surplus. However, applications of this principle come to depend on details of cost and demand functions. We hope that the cases presented here, in conjunction with other studies cited, offer some useful and new insights.

REFERENCES

R. L. Bishop, "Monopolistic Competition and Welfare Economics," in Robert Kuenne, ed., *Monopolistic Competition Theory,* New York 1967.

E. Chamberlin, "Product Heterogeneity and Public Policy," *Amer. Econ. Rev. Proc.,* May 1950, *40,* 85–92.

P. A. Diamond and D. L. McFadden, "Some Uses of the Expenditure Function In Public Finance," *J. Publ. Econ.,* Feb. 1974, *82,* 1–23.

A. K. Dixit and J. E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," econ. res. pap. no. 64, Univ. Warwick, England 1975.

H. A. John Green, *Aggregation in Economic Analysis,* Princeton 1964.

H. Hotelling, "Stability in Competition," *Econ. J.,* Mar. 1929, *39,* 41–57.

N. Kaldor, "Market Imperfection and Excess Capacity," *Economica,* Feb. 1934, *2,* 33–50.

K. Lancaster, "Socially Optimal Product Differentiation," *Amer. Econ. Rev.,* Sept. 1975, *65,* 567–85.

A. M. Spence, "Product Selection, Fixed Costs, and Monopolistic Competition," *Rev. Econ. Stud.,* June 1976, *43,* 217–35.

D. A. Starrett, "Principles of Optimal Location in a Large Homogeneous Area," *J. Econ. Theory,* Dec. 1974, *9,* 418–48.

N. H. Stern, "The Optimal Size of Market Areas," *J. Econ. Theory,* Apr. 1972, *4,* 159–73.

J. E. Stiglitz, "Monopolistic Competition in the Capital Market," tech. rep. no. 161, IMSS, Stanford Univ., Feb. 1975.