MONOPOLISTIC COMPETITION AND OPTIMUM PRODUCT DIVERSITY

Avinash K. Dixit
and
Joseph E. Stiglitz

NUMBER 64

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

MONOPOLISTIC COMPETITION AND OPTIMUM PRODUCT DIVERSITY

Avinash K. Dixit
and
Joseph E. Stiglitz

NUMBER 64

February 1975

1. Introduction

This paper began as an attempt to formalize a simple general equilibrium version of the Chamberlinian monopolistic competition model, in order to see whether there is any validity in the common assertion that monopolistic competition leads to too much product diversification. Chamberlin himself was careful not to make any such assertion; as a matter of fact, he saw the central issue very clearly, and expressed it succinctly as follows: "It is true that the same total resources ... may be made to yield more units of product by being concentrated on fewer firms. The issue might be put as efficiency versus diversity".[1/] Kaldor, too, saw that excess capacity was not the same as excessive diversity. He said that if economies of scale were exploited to a greater extent, "the public would be offered finally larger amounts of a smaller number of commodities; and it is impossible to tell how far people prefer quantity to diversity and vice versa".[2/] What this really means is that a model must be specified in greater detail to determine conditions under which the one or the other might be expected.

---

*The authors are Professors of Economics at the University of Warwick, Coventry, England and at Standford University respectively. The research was initiated while Dixit was at Balliol College, Oxford, and Stiglitz was Visiting Fellow at St. Catherine's College, Oxford. Stiglitz's research was supported in part by National Science Foundation Grant SOC74-22182 at the Institute for Mathematical Studies in the Social Sciences, Standford University. The authors are indebted to Michael Spence for comments and suggestions on an earlier draft.

The one example which has been worked out in some detail, the Hotelling spatial location model, has led many economists to the presumption that there is in fact excessive diversity associated with monopolistic competition.[3/] The results of our analysis throw considerable doubt on this presumption.

Further, it soon becomes apparent that the issue of diversity is only part of the more general question of a comparison between the market equilibrium and the optimum allocation: not only may the numbers produced be incorrect, but the choice of which commodities to produce, and how much of each to produce, may differ. There are a number of effects at work. Whether a commodity is produced depends on revenues relative to total costs. Social profitability depends, on the other hand, on a number of factors. In deciding whether to produce a commodity, the government would look not only at the profitability of the project, but also at the consumer surplus (the profitability it could attain if it were acting as a completely discriminating monopolist), and the effect on other industries and sectors (on the consumer surplus, profitability and viability). The effects on other sectors result both from substitution and income effects.

The whole problem hinges crucially on the existence of economies of scale. In their absence, it would be possible to produce infinitesimal amounts of every conceivable product that might be desired, without any additional resource cost. Private and social profitability would coincide given the other conventional assumptions, and the repercussions on other sectors would become purely pecuniary externalities. With non-convexities, however, we shall see that all these considerations are altered.

Moreover, given economies of scale in the relevant range of output, market realisation of the 'unconstrained' or first-best optimum, i.e. one subject to constraints of resource availability and technology alone, requires pricing below average cost, with lump sum transfers to firms to cover losses. The conceptual and practical difficulties of doing so are clearly formidable. It would therefore appear that perhaps a more appropriate notion of optimality is a constrained one, where each firm must operate without making a loss. The government may pursue conventional regulatory policies, or combinations of excise and franchise taxes and subsidies, but the important restriction is that lump sum subsidies are not possible.

The permissible output and price configurations in such an optimum reflect the same constraints as the ones in the Chamberlinian equilibrium. The two solutions can still differ because of differences implicit in the objective functions.

Consider first the manner in which the desirability of variety can enter into the model. Some such notion is already implicit in the convexity of indifference surfaces of a conventional utility function defined over quantities of all the varieties that might exist. Thus, a person who might be indifferent between the combinations of quantities (1,0) and (0,1) of two product types would prefer the combination $(\frac{1}{2},\frac{1}{2})$ to either extreme. If this is the only relevant consideration, we shall show that in one central case the Chamberlinian equilibrium and the constrained optimum coincide. In the same case, we shall also show that the first best optimum has firms of the same size as in the other two solutions, and a greater number of such firms. These results undermine much of the conventional wisdom concerning excess capacity as well as excessive diversity.

However, it is conceivable that the range of products available is by itself an argument of the utility function, over and above what is taken into account through the amounts actually consumed. This may reflect the desirability of accommodating a sudden future change of tastes, or of retaining one's identity by consuming products different from those consumed by one's neighbours, or some such consideration. Variety then takes on some aspects of a public good, and this raises the usual problems for the optimal provision of such goods in a market system.

Even if variety is not a public good, its private and social desirability can still differ on account of the failure to appropriate consumers' surplus as noted above. In the large group case, it so happens that if the elasticity of demand is constant and the same for all products, the consumers' surplus is proportional to the revenue, with the same factor of proportionality for all goods. The difference in the objectives of firms and of welfare maximization then does not matter. Otherwise, we expect the equilibrium outcome to be biased against those varieties for which the ratio of consumers' surplus to revenue is large. However, this simple principle does not yield much direct insight. A change in the output of a commodity, or the introduction of a new commodity, affect the demands for all other goods. With possible changes in the levels as well as the elasticities of all demands, the consumers' surpluses and revenues can change in complicated ways. Therefore the answers to the questions of the equilibrium and the optimum levels of output, including possibly considerations of the viability of these commodities, involve a very large range of possibilities. We need an explicit model with a detailed formulation of demand, in order to isolate and analyse the various questions. The rest of the paper attempts to provide such analyses. In the next section we discuss the problems of modelling the demand for variety, and set up the model of the special case mentioned above. In Section 3, this case is analysed in detail. Sections 4-6 consider the various generalisations mentioned above. In each case, we compare various features of the Chamberlinian equilibrium with those of the two types of optima, with particular regard to (1) the number and mix of products, (2) their prices and quantities, and (3) the total resource allocation for this group of products.

We focus on the allocation problems that are of interest here, and neglect two other issues. The first is that of income distribution. We assume utility to be a function of market aggregate quantities. This is justified if the consumers have identical tastes, and either identical incomes or linear Engel curves; alternatively we can assume that lump sum redistributions take place to maximise an individualistic social welfare function, thus yielding Samuelsonian social indifference curves. Also, we assume that the consumers' preferences are exogenous, thus excluding considerations of advertising and its welfare implications. We feel that prevailing thinking has overstressed this aspect at the expense of some basic allocative issues, and that the qualitative effects of adding these considerations to our model should in any case be fairly evident.

Our model differs from the spatial location model in one important respect. There, each consumer purchases only one of the products in the industry. Increasing product differentiation leads to the consumer being able to purchase a commodity closer to his likeing, i.e. to go to a store closer to his residence. Our model includes such considerations in its interpretation with heterogeneous consumers and social indifference curves. But in addition, it can allow each consumer to enjoy product diversity directly. There are numerous examples where this formulation is clearly more appropriate than one modelled on location. The ability to diversity a portfolio by spreading one's wealth over a large number of assets was one of the instances that provided the original motivation for formulating this kind of model, and is discussed in more detail elsewhere.[4/] Clothes suited to different climatic conditions, or flavours of ice-cream, are other examples of this type.

2. The demand for variety

Consider a potentially infinite range of related products, numbered[5/] 1, 2, ... n, ... A competitive sector labelled 0 aggregates the rest of the economy. Good 0 is chosen as the numéraire, and the amount of the economy's endowment of it is normalized at unity; this can be thought of as the time at the disposal of the consumers.

If the amounts of the commodities consumed are x_0 and $\underline{x} = (x_1, x_2, .. x_n ..)$. we define a utility function

$$u = U(x_0, x_1, x_2, \ldots x_n, \ldots) \qquad (2.1)$$

This function, assumed to have convex indifference surfaces, considers variety as a private good in the sense defined before. If a subset S of commodities is actually being produced, i.e. $x_i > 0$ for $i \in S$ and $x_i = 0$

for $i \notin S$, then the public good case can be modelled by allowing u to depend explicitly on S, i.e.

$$u = U(x_0, x_1, x_2, \ldots x_n, \ldots ; S) \qquad (2.2)$$

We shall take up this case in Section 4.

It is clear that at this level of generality, nothing specific or interesting could be said. We proceed to impose some structure on U in order to isolate issues for sharper focus. First, we assume that the group of products in question is separable from the aggregated sector, i.e.

$$u = U(x_0, V(x_1, x_2, \ldots x_n \ldots)) \qquad (2.3)$$

For most of this article, we assume that V is a symmetric function. This, combined with an assumption about the symmetry of costs, removes the issue of the product mix. The number of products is still a relevant consideration, but given this number n, it does not matter what labels they bear. Then we may as well label them $1, 2, \ldots n$, and 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 sense of order along a spectrum, and two products closer together on this spectrum are better substitutes than two products farther apart. This makes V asymmetric, and the actual labels of products available become important. This is naturally recognised in the spatial context, but the Chamberlin tradition where the nature of the products in the group is left unspecified has implicitly assumed symmetry. We shall follow this tradition, but in Section 6 we shall return to the question of the product mix.[6/]

The next simplification is to consider an additively separable form for the function $V(\underline{x})$, i.e.

$$u = U(x_0, \; \Sigma_i \; v(x_i) \;) \tag{2.4}$$

We take up this case in Section 5. In the next section, we consider an even more special form where $V(\underline{x})$ has a constant elasticity of substitution, i.e.

$$u = U(\, x_0, \; [\Sigma_i \; x_i^{\rho}]^{1/\rho} \;) \tag{2.5}$$

For concavity, we need $\rho < 1$. Further, since we wish to allow a situation where several of the x_i are zero, we need $\rho > 0$.

Finally, we assume that U is homogeneous of degree one in x_0 and $V(\underline{x})$. Then, with unit income elasticities, we can study substitution between the sectors without the added complication of unequal income effects.

In the remainder of this section we shall derive the demand functions for the special case (2.5), and comment on their properties. Suppose products 1, 2, .. n are being produced, and write the budget constraint as

$$x_0 + \Sigma_{i=1}^{n} \; p_i \, x_i = I \tag{2.6}$$

where I is income in terms of the numéraire, i.e. the endowment which has been normalized at 1, plus the profits of firms distributed to consumers, or minus the lump sum transfers to firms, as the case may be.

We omit the details of utility maximization. The interesting feature is that a two-stage budgeting procedure is applicable.[7]

Thus we can define a quantity index $y = V(\underline{x})$, and a price index $q = Q(\underline{p})$, such that (x_0, y) maximize $U(x_0, y)$ subject to $x_0 + q\,y = I$, and then $\underline{x}$ maximises $V(\underline{x})$ subject to $\Sigma_i\, p_i\, x_i = q\,y$. Moreover, with the quantity index of a constant elasticity form, so is the price index. Thus, when

$$y = \left[\Sigma_{i=1}^{n} \quad x_i^{\rho} \right]^{1/\rho} \qquad (2.7)$$

we have

$$q = \left[\Sigma_{i=1}^{n} \quad p_i^{-1/\beta} \right]^{-\beta} \qquad (2.8)$$

where $\beta = (1-\rho)/\rho$. From the conditions imposed on ρ, we know that β is positive.

Now consider the first stage of budgeting. Since U is homogeneous of degree one, x_0 and y are each proportional to I, and the budget shares are functions of q alone. Let $s(q)$ be the budget share of y, i.e.

$$y = I\, s(q)/q \qquad (2.9)$$

The ratio x_0/y is a function of q alone, and its elasticity is defined as the intersectoral elasticity of substitution, which we shall write as $\sigma(q)$. The behaviour of budget shares depends on the relation between $\sigma(q)$ and 1 in the standard manner; thus we have the elasticity

$$\theta\,(q) = q\, s'(q)/s(q) = \left[1 - \sigma(q)\right] \; \left[1 - s(q)\right] \qquad (2.10)$$

We see at once that

$$\theta\,(q) < 1 \qquad (2.11)$$

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

$$x_i = y \left[q/p_i \right]^{1/(1-\rho)} \qquad (2.12)$$

where y is defined by (2.9). Consider the effect of a change in p_i alone. This affects x_i directly, and also through q, and thence through y as well. Now from (2.8) we have the elasticity

$$\partial \log q / \partial \log p_i = (q / p_i)^{1/\beta} \qquad (2.13)$$

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 and thus the indirect effects on x_i. This leaves us with the elasticity

$$\partial \log x_i / \partial \log p_i = -1/(1-\rho) = (1+\beta)/\beta \qquad (2.14)$$

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

$$y = x\,n^{1/\rho} = x\,n^{1+\beta} \tag{2.15}$$

$$q = p\,n^{-\beta} = x\,n^{-(1-\rho)/\rho} \tag{2.16}$$

and then, from (2.8) and (2.12)

$$x = I\,s(q) / (p\,n) \tag{2.17}$$

The elasticity of this is easy to calculate; we find

$$\partial \log x / \partial \log p = -[1 - \theta\ (q)] \tag{2.18}$$

Then (2.11) shows that the DD-curve slopes downward. The conventional condition that the dd-curve be more elastic is seen from (2.14) and (2.18) to be

$$1/\beta + \theta\ (q) > 0 \tag{2.19}$$

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

$$x_i / x_j = [p_j / p_i]^{1/(1-\rho)} \tag{2.20}$$

Thus $1/(1-\rho)$ is the elasticity of substitution between any two products within the group. This calls for some comment. A constant intra-sectoral elasticity of substitution has some undesirable features in a model of product diversity. Some problems of assuming symmetry were pointed out earlier. For a spectrum of characteristics, we would expect the elasticity to depend on the distance betwen i and j. In addition, the total number of products being produced may be thought to influence the elasticities. If the total conveivable range of variation is finite, then products have to crowd closer together as their number increases, and thus the elasticity of substitution should on the whole increase and tend to infinity in the limit. However, it is often the case that the total range is very large, and most practicable product ranges can only hope to cover a negligible fraction of it. This is particularly true if there are several relevant characteristics, and therefore several dimensions to the spectrum. Since this is a very likely situation, we think it interesting to have a model where there is an infinity of conceivable products but only a finite number are ever produced, and the elasticities of substitution are all bounded above, thus always leaving some monopoly power in existence. With fresh apologies for symmetry, the assumption of constancy then offers some simplicity and an interesting result. In Section 5, we shall relax constancy to some extent.

As regards production, we assume for most of the paper that each firm has the same fixed cost, a, and a constant marginal cost, c, also equal for all firms. All our results remain valid if the variable cost of production is allowed to depend on output, but the algebra is considerably more complicated. In Section 6 we consider a case where different firms have different values of a and c, and in the concluding remarks we mention some other problems.

3. The constant elasticity case.

Market Equilibrium

In this section we study the consequences of the utility function (2.5) and the associated demand functions derived in Section 2. Let us begin with the Chamberlinian group equilibrium. The profit-maximization condition for each firm is the familiar equality of marginal revenue and marginal cost. With a constant elasticity of demand and constant marginal cost for each firm, this becomes

$$p_i \left\{ 1 - \frac{1}{1/(1-\rho)} \right\} = c \quad \text{for } i = 1, 2, \ldots n$$

Write p_e for the common equilibrium price for each variety being produced. Then we have

$$p_e = c/\rho = c\,(1 + \beta) \tag{3.1}$$

The second condition of equilibrium is that firms enter until the next potential entrant would make a loss, i.e. n is defined by

$$\left.\begin{aligned} (p_n - c)\, x_n &\geq a \\ (p_{n+1} - c)\, x_{n+1} &< a \end{aligned}\right\}$$

We shall assume that n is large enough that 1 can be regarded as a small increment. Then we can treat n as if it were a continuous variable, and write the condition approximately as an equality,

$$(p_n - c)\, x_n = a \tag{3.2}$$

With symmetry, this implies zero profit for all other firms as well. Then we have $I = 1$, and using (2.12) and (3.1) we can write the condition in a way that defines the number of firms in the equilibrium, n_e :

$$s(p_e n_e^{-\beta}) / (p_e n_e) = a/ (\beta c) \tag{3.3}$$

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 (2.17) we see that the behaviour of $s(p n^{-\beta})/(pn)$ 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

$$1 + \beta \ \theta \ (q) \ > \ 0 \tag{3.4}$$

This is exactly the same as (2.19), 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.x is independent of n, i.e. that $s(p\,n^{-\beta})$ 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, (3.4) holds and so equilibrium is unique.

Finally, using (2.12) and (3.3), we can calculate the equilibrium output for each active firm:

$$x_e = a/(\beta c) \tag{3.5}$$

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

$$s_e = s(q_e) \text{ where } q_e = p_e n_e^{-\beta} \tag{3.6}$$

These will be useful for subsequent comparisons.

Constrained optimum

Turning to the constrained optimum, we wish to find an n and the corresponding p_i and x_i for the active firms so as to maximise utility subject to the constraint that no firm makes a loss. There now arises the question of whether the basic symmetry of our model is preserved. There is some unavoidable asymmetry, as some firms are active and the others are not. It would still simplify the problem greatly if we could know in advance that all active firms would have the same price and output.

Fortunately, this is so in the large group case. Suppose two firms are producing unequal but positive outputs x_1 and x_2, each without making a loss. By (2.12), we see that each firm's revenue $p_i\, x_i$ is proportional to x_i^{ρ}, where the factor of proportionality is the same for both, and depends on q, thus being independent of each firm's decisions to order (1/n). Since $\rho < 1$, revenue is a concave function of output. With constant marginal cost, therefore, any output between x_1 and x_2 would also yield non-negative profit. Thus it would be feasible to have each firm produce $(x_1 + x_2)/2$ instead, and by the convexity of the indifference surfaces it would be preferable to do so.

Complications arise with few firms, and also if marginal cost can vary with output, and declines fast enough to offset the concavity of revenue. Note also why the argument cannot be applied with one active firm and one inactive firm : the fixed cost presents a basic non-convexity at zero.

Finally, it is easy to show that moving all prices proportionately towards the corresponding marginal costs will increase utility, provided marginal costs are non-decreasing functions of the respective outputs.[8/] In the present situation, this implies that the optimum price should lie on the boundary of the feasible set, i.e. each active firm should make exactly zero profit.

Thus we have $I = 1$, and indirect utility is a (decreasing) function of q alone. The constrained optimum problem can then be written as

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

subject to

$$(p - c)\quad s(p\, n^{-\beta}) / (p\, n) = a \qquad (3.7)$$

Consider the curve defined by (3.7) in the (p,n) space. We have assumed that the left hand side is a decreasing function of n for each fixed p. Thus (3.7) defines n as a single-valued function of p, for all values of p above p_{min} defined by

$$(p_{min} - c)\ \lim_{n \to \infty}\ s(p_{min}\, n^{-\beta}) / (p_{min}\, n) = a \qquad (3.8)$$

In particular, if this limit is infinite, $p_{min} = c$.

Differentiating along the constraint logarithmically, we evaluate the elasticity

$$\frac{d \log n}{d \log p} = \frac{\frac{c}{p - c} + \theta(q)}{1 + \beta\, \theta(q)} \qquad (3.9)$$

The denominator has been assumed positive. The numerator will always be positive if $\theta(q)$ is positive. Even if $\theta(q)$ is negative, conditions (2.19) ensures that the numerator will be positive for $p \leq c(1+\beta)$. It may become negative for higher values of p, and may even fluctuate in sign if $\theta(q)$ fluctuates. Thus the constraint curve may have alternatively rising and falling portions. However, we will show that only the initial rising portion matters. Figure 1 shows the case where θ (q) is negative; the other case is even simpler.

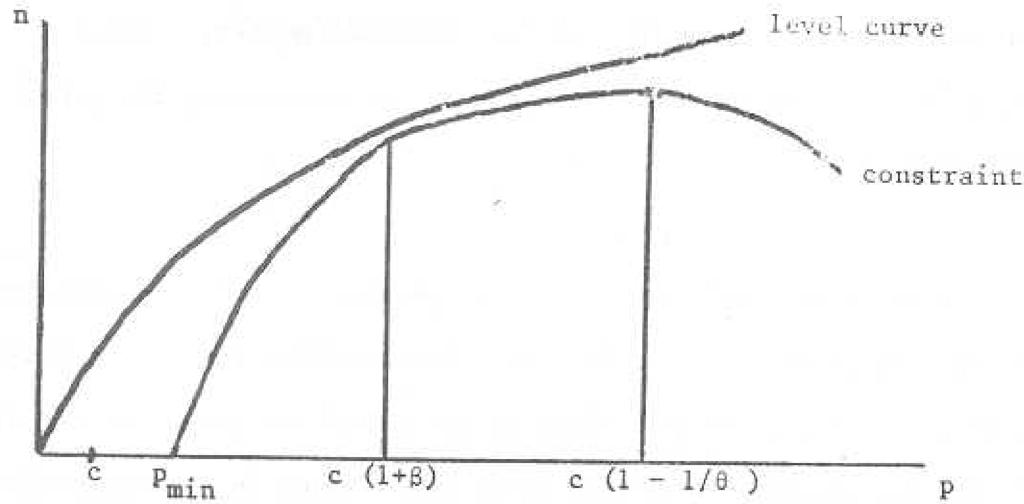

Figure 1

The contours of the objective function have equations

$$n = \text{constant} \cdot p^{1/\beta}$$

and the first order condition for optimality is the equality between the slopes or elasticities of (3.7) and (3.10). Equating the right hand side of (3.9) to $(1/\beta)$ yields a unique solution p_c :

$$p_c = c \ (1 + \beta) \qquad (3.11)$$

We show that this unique solution satisfies the second order condition for a maximum, and therefore that p_c is the constrained optimum price. Note that the level curves of the objective function have the constant elasticity $(1/\beta)$. First suppose θ (q) is constant. Then the right hand side of (3.9) is a decreasing function of p, i.e. the constraint curve has an elasticity greater than $(1/\beta)$ to the left of

p_c and less than that to its right. Since $q = p\, n^{-\beta}$ is stationary at that point, variations in $\theta\,(q)$ have only a second order effect, and thus neglecting them does not matter. Since utility increases as p decreases and as n increases, the proof is complete.

Comparing (3.1) and (3.11), 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 of all other variables can be calculated from these two. Thus in this case the monopolistically competitive equilibrium is the optimum constrained by the lack of lump sum subsidies. Chamberlin once called this equilibrium "a sort of ideal";[9/] our analysis gives some precision to that concept, and establishes when it is valid.

Unconstrained optimum.

These solutions may in turn be compared to the unconstrained (first best) optimum. Considerations of convexity once again establish that all active firms should produce the same output. Thus we want to choose n firms each producing output x in order to maximise

$$u = U\left(1 - n(a+cx\,),\;\; x\, n^{1+\beta} \right) \qquad (3.12)$$

where we have used (2.15) and the economy's resource balance constraint.

The first order conditions are

$$- n\, c\, U_0 + n^{1+\beta}\, U_y = 0 \tag{3.13}$$

$$- (a + c x)\, U_0 + (1+\beta)\, x\, n^{\beta}\, U_y = 0 \tag{3.14}$$

From the first stage of the budgeting problem, we know that $q = U_y / U_0$. Using (3.13) and (2.16), we find the price charged by each active firm in the unconstrained optimum, p_u, to equal marginal cost

$$p_u = c \tag{3.15}$$

This, of course, is no surprise. Next, from the first order conditions, we have $(a + c x)/(n c) = (1+\beta) x / n$, which gives the output of each active firm, x_u. We have

$$x_u = a/(c\beta) \tag{3.16}$$

Finally, with (3.15), each active firm covers its variable cost exactly. The lump sum transfers to firms then equal $a\, n$, and therefore $I = 1 - a\, n$, and

$$x = (1 - a\, n)\; s(p\, n^{-\beta})/(pn)$$

The number of firms n_u is then defined by

$$s(c\, n_u^{-\beta})/ n_u = (a/\beta)/(1-an_u) \tag{3.17}$$

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.[10/] 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 unconstrained optimum as well as the constrained one.

A direct comparison of the numbers of firms from (3.3) and (3.17) would be difficult, but an indirect argument turns out to be easy. The one clear thing about the unconstrained optimum is that it 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

$$q_u < q_c = q_e \qquad (3.18)$$

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 2. Let C be the constrained optimum, A the unconstrained one, and let B be the point where the line joining the origin to C meets the budget constraint in the unconstrained case. By homotheticity,

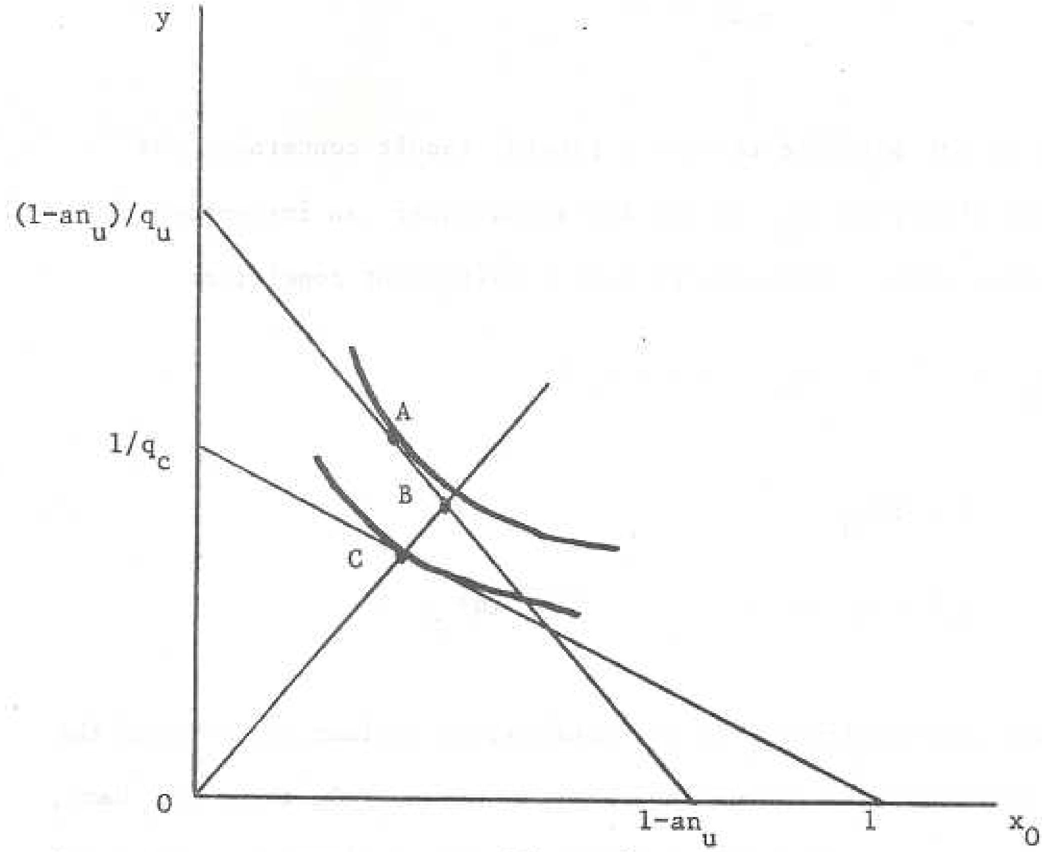

Figure 2

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 increase the value of y. Since the value of x is the same in the two optima, we must have

$$n_u > n_c = n_e \tag{3.19}$$

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

Using (3.18) we can easily compare the budget shares. In the notation we have been using, we find

$$s_u \gtrless s_c \quad \text{as} \quad \theta(q) \lessgtr 0, \text{ i.e. as } \sigma(q) \gtrless 1 \tag{3.20}$$

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 2 shows this. However, we have a sufficient condition:

$$x_{0u} = (1 - a\, n_u)(1 - s_u)$$

$$< 1 - s_u$$

$$\leq 1 - s_c = x_{0c} \qquad \text{if } \sigma(q) \geq 1$$

In this case the equilibrium or the constrained optimum use more of the numéraire resource than the unconstrained optimum. On the other hand, if $\sigma(q) = 0$, we have L-shaped isoquants and in Figure 2, points A and B coincide giving the opposite conclusion.

4. Diversity as a public good

In this section we consider the consequences of having the range of products actually produced as a direct argument of utility, over and above the effect through the amounts consumed. The general formulation of this public good probolem was given in (2.2). Once again, it is too general to be useful, and we specialise it in several ways. In particular, we retain symmetry, so that the only feature of the set S of goods produced that is relevant is the number of elements in it, i.e. the number n of goods being produced. Next, we assume the separable constant elasticity of substitution form for utility as a function of the amounts consumed. Finally, we assume that the direct argument n is separated with the products in the group, and at this stage much is gained in analytic convenience without changing the qualitative features further if we assume a multiplicative power form. Thus the utility function is

$$u = U\left(x_0, \left[\Sigma_{i=1}^{n} \; x_i^{\rho} \right]^{1/\rho} \; n^{\delta} \right) \qquad (4.1)$$

While we shall normally speak of the public good case, the formal analysis allows δ to be positive or negative, i.e. variety to be a public good or a public bad. However, we will need $(\beta+\delta)$ positive, where β is as before.

The analysis of demand is almost unchanged from Section 2. The two-stage budgeting property still holds, and we define the quantity index

$$y = \left[\Sigma_{i=1}^{n} \; x_i^{\rho} \right]^{1/\rho} \; n^{\delta} \qquad (4.2)$$

and the associated price index

$$q = \left[\Sigma_{i=1}^{n} \; p_i^{-1/\beta} \right]^{-\beta} n^{-\delta} \tag{4.3}$$

In the symmetric situation, with $x_i = x$ and $p_i = p$ for all i, we have

$$y = x\, n^{(\delta + 1/\rho)} \tag{4.4}$$

$$q = p\, n^{-(\beta + \delta)} \tag{4.5}$$

These can be contrasted with (2.7), (2.8), (2.15) and (2.16).

The first stage demand function $y(I,q)$ is exactly as before, i.e. (2.9) continues to hold, while, at the second stage, we have

$$x_i = I \; \frac{s(q)}{q} \left[\frac{q\, n^{\delta\rho}}{p_i} \right]^{1/(1-\rho)} \tag{4.6}$$

In the symmetric situation, this simplifies to

$$x = I \; \frac{s(\, p\, n^{-(\beta+\delta)} \,)}{p\, n} \tag{4.7}$$

These define the dd-curve and the DD-curve respectively. The break-even constraint is

$$(\, p - c\,) \; \frac{s(\, p\, n^{-(\beta+\delta)} \,)}{p\, n} = a \tag{4.8}$$

The solutions for the equilibrium and the two types of optima can be found by the same methods as before, and we shall only state the results.

Once source of difference should be evident : the elasticity of the dd-curve and therefore the equilibrium price-cost margin is unaffected by δ, but q and therefore the objective in the constrained optimum, depends on δ.

In the equilibrium, as before,

$$p_e = c \left(1 + \beta \right) \tag{4.9}$$

$$x_e = a/(c\beta) \tag{4.10}$$

while in the constrained optimum

$$p_c = c \left(1 + \beta + \delta \right) \tag{4.11}$$

$$x_c = a/\left(c \left(\beta + \delta \right) \right) \tag{4.12}$$

In each case, the number of firms is defined by (4.8) with the appropriate value of p.

The conditions for the dd-curve to be more elastic than the DD-curve, and for the latter to shift to the left as n increases, ensure that both the equilibrium and the optimum lie on the initial rising portion of the constraint curve like that in Figure 1. Thus we have

$$n_c \gtrless n_e \quad \text{as} \quad p_c \gtrless p_e \quad \text{as} \quad x_c \lessgtr x_e \quad \text{as} \quad \delta \gtrless 0 \tag{4.13}$$

The government can achieve the constrained optimum by imposing a specific tax of $c\delta/(1+\beta)$ on each product in the group, and using the proceeds to finance a franchise subsidy.

Comparisons between the constrained and the unconstrained optima are even easier. The former minimises q given by (4.5) subject to (4.8), while the latter maximises $u = U(\ 1 - n(a+cx),\ \ x\, n^{(1+\beta+\delta\ )}\ )$. This problem is the same as that of the previous section with β replaced by $(\ \beta+\delta\ )$ everywhere. Thus we have

$$p_u = c < p_c \tag{4.14}$$

$$x_u = a/(c(\beta+\delta\ )\ ) = x_c \tag{4.15}$$

and n_u is defined by

$$s(c\, n_u^{-\beta+\delta\ })/n_u = (a/(\beta+\delta\ ))\ /\ (1-an_u) \tag{4.16}$$

with

$$n_u > n_c \tag{4.17}$$

It is easy to compare the unconstrained optimum with the equilibrium as far as the output and price of each active firm are concerned. The number of firms is somewhat harder. If δ is positive, we know $n_u > n_c > n_e$. If δ is negative, the situation is not clear. An argument similar to that accompanying Figure 2 will show that the quantity index y is higher in the unconstrained optimum, but so is the output of each firm, and thus a higher y could be consistent with a lower n. In the special case where $\theta(q)$ is always zero, i.e. $s(q)$ is constant and equal to $\bar{s}$, say, we can make an explicit calculation which shows that

$$n_u \gtrless n_e \quad \text{as} \quad \bar{s} + 1/(\beta+\delta\ ) \lessgtr 1 + 1/\beta \tag{4.18}$$

Then, provided the monopolistic sector is a small part of the economy, and

δ is not too large in absolute value, we can expect the unconstrained optimum to have more firms.

We conclude this section with some reflections on the public good problem. This, too is related to the existence of a fixed cost. The existence of a product type is a feature of the economy that is available in common to all consumers. While there is a cost associated with the introduction of an additional product type, the marginal cost of an additional individual availing himself of this feature is zero. This way of thinking about a pure public good is in the Dupuit-Hotelling tradition, and somewhat different from the newer Samuelson approach. There are two distinct but related questions involved. The first is whether to undertake a particular project, which entails an infra-marginal calculation, and the second is the level at which to carry on the activity of a chosen project, which is a marginal calculation. In the same way and for the same reason, the choice of the output of a commodity that is being produced involves a marginal calculation, but the choice of whether to produce it at all involves an infra-marginal one, of gains from its provision.

Unlike many public goods, exclusion is feasible here, in the form of fixed charges for the right to purchase certain commodities. If all the individuals are identical, this is easy to implement by means of a two-part tariff. If there is diversity of tastes, the full optimum will require the fixed charge to be different for different individuals, which is much harder to arrange. Moreover, projections of demand for a new variety based on offer-price schedules for quantities of it will fail to capture that part of the gain which is associated with the common good of making it available.

If, in the manner treated in this section, the _set_ of private goods actually produced can itself be a public good, then the distinction that is conventionally made between the activities that "ought" to be in the public sector and those that "ought" to be in the private sector becomes somewhat blurred. But these are questions that will have to be pursued on another occasion.

5. Variable elasticity utility functions

In this section we revert to considering variety as a private good, but remove the assumption of a constant elasticity of substitution within the monopolistic sector.[11] We retain separability as in (2.4); in fact we consider a somewhat more restrictive form

$$u = x_0^{1-\gamma} \left[\Sigma_{i=1}^{n} \quad v(x_i) \right]^{\gamma} \qquad (5.1)$$

This is somewhat like assuming a unit inter-sectoral elasticity of substitution. However, since the group utility $V(\underline{x}) = \Sigma \, v(x_i)$ is not in general homothetic, two-stage budgeting is inapplicable and such an elasticity does not have any rigorous meaning.

Considering demand functions in this case, we have the first order conditions

$$(1-\gamma)/x_0 = \lambda \quad , \quad \gamma \, v'(x_i) \, / \, V(\underline{x}) = \lambda \, p_i \qquad (5.2)$$

where λ gives the marginal effect of income on log u. As before, if the number of products is sufficiently large, we can take each of them to be a negligible fraction of expenditure, and then the second set of equations in (5.2) will define the dd-curves with $V(\underline{x})$ and λ held constant. The demand elasticities are

$$-\frac{\partial \log x_i}{\partial \log p_i} = -\frac{v'(x_i)}{x_i \, v''(x_i)} \qquad (5.3)$$

Clearly we will need v to be a concave function.

The analysis will be similar to that in Section 3, but some magnitudes that were constant there will now be functions of the x_i. For recognition and comparison, we will denote these functions by the same symbols as were used for the parameters in Section 3. Thus we define $\beta(x)$ by

$$1 + 1/\beta(x) = - v'(x) / (x v''(x)) \qquad (5.4)$$

Finally, solving for λ using the budget constraint and reducing to the symmetric situation, i.e. one with $x_i = x$ and $p_i = p$ for all i, we have the DD-curve defined implicitly by

$$x = \frac{I}{n p} \quad \frac{\gamma \cdot \rho(x)}{\gamma\rho(x) + (1-\gamma)} \qquad (5.5)$$

where

$$\rho(x) = x v'(x) / v(x) \qquad (5.6)$$

As was the case when ρ was constant, we shall assume that $\rho(x)$ lies between 0 and 1.

It can be verified that if $\rho(x)$ is constant, we have $\beta(x)$ also constant and the two are related as in Section 3. Otherwise, the relationship between the two is different; it is easy to verify that

$$x \rho'(x)/\rho(x) = 1/ [1 + \beta(x)] - \rho(x) \qquad (5.7)$$

Also, we have the demand for numéraire

$$x_0 = I \frac{1-\gamma}{\gamma\rho(x) + (1-\gamma)} \qquad (5.8)$$

Now consider the Chamberlinian equilibrium. The profit-maximisation condition yields, for each active firm,

$$p = c \left[1 + \beta (x) \right] \qquad (5.9)$$

Substituting this in the zero-pure-profit condition, we have x_e defined by

$$x \ \ \beta(x) = a/c \qquad (5.10)$$

Finally, the number of firms can be calculated using the DD-curve and the breakeven condition, yielding

$$n = \frac{1}{a + c x} \quad \frac{\gamma\rho \ (x)}{\gamma\rho(x) + (1-\gamma)} \qquad (5.11)$$

Evaluating this for $x = x_e$ yields n_e.

For uniqueness of equilibrium once again we need conditions relating to the shift of the DD-curve, relative elasticities etc. However, these conditions are now rather involved and not transparent. We shall omit them to save space, and indicate where they are used in the subsequent discussion.

Let us turn to the constrained optimum. We wish to choose n and x to maximise u, with x_0 defined by (5.8), and subject to the constraint that each firm make zero profit while choosing a point on its DD-curve. This condition is precisely (5.11), and we can make explicit substitutions to obtain a maximand in terms of x alone. This finally becomes

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

Choosing x to maximise this, we find the condition

$$\frac{c\ x}{a + c\ x} = \frac{1}{1 + \beta(x)} - \frac{\rho\ (x)}{\gamma\rho(x) + (1-\gamma)} \frac{x\ \rho'(x)}{\rho\ (x)} \qquad (5.12)$$

The corresponding condition for equilibrium could be written

$$\frac{c\ x}{a + c\ x} = \frac{1}{1 + \beta\ (x)} \qquad (5.9')$$

Now the left hand side as a function of x increases from 0 to 1 as x increases from 0 to ∞. If we draw the right hand sides in each case as functions of x, and use the second order conditions and conditions for the uniqueness of equilibrium, we find that, provided ρ' is one-signed,

$$x_c \gtrless x_e \quad \text{as} \quad \rho'(x) \lesseqgtr 0 \qquad (5.13)$$

Comparison of the numbers of firms uses (5.11), but the algebra is in this instance more easily understood from a diagram. This is Figure 3. Both the equilibrium and the constrained optimum have each firm's price and output combination on the average cost curve, and also on the appropriate DD-curve. The actual point is determined by some other consideration; the tangency of the average cost curve and the dd-curve for the equilibrium, and something not geometrically obvious for the optimum. However, if in the equilibrium each firm that is active produces more output, the price-output point must be further down the average cost curve than it is for the optimum, i.e. it must lie on a DD-curve further to the right. Given our assumption, this can only result with fewer firms, thus yielding the result

$$n_c \gtrless n_e \quad \text{as} \quad \rho'(x) \gtrless 0 \tag{5.14}$$

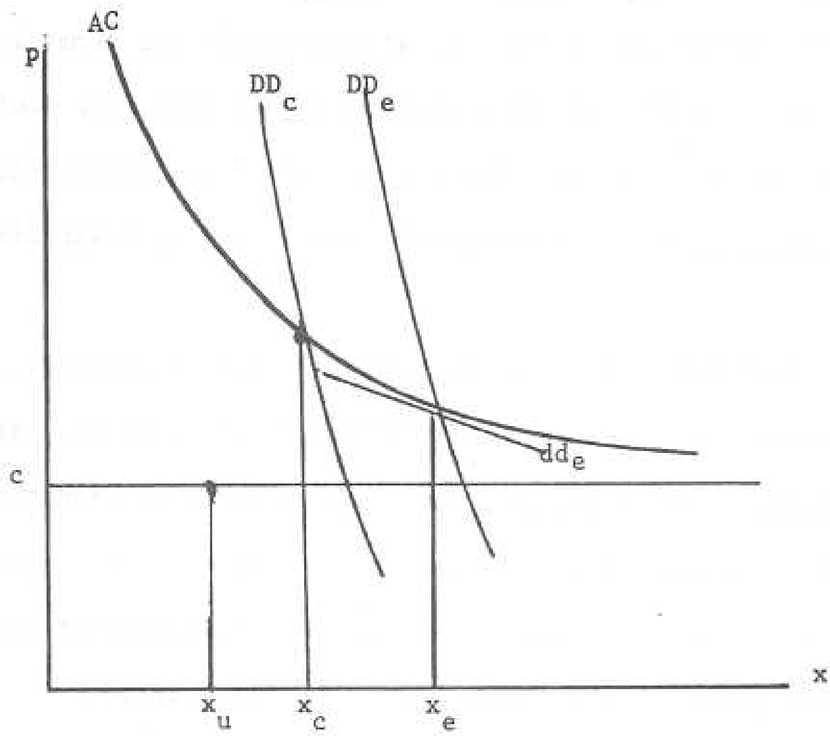

Figure 3

Finally, (5.13) shows that $\rho(x_c) < \rho(x_e)$, and then from (5.8).

$$x_{0c} > x_{0e} \tag{5.15}$$

A different degree of inter-sectoral substitution could yield the opposite. This is an opposition of income and substitution effects as in Figure 2.

An intuitive reason for these results can be given as follows. With our large group assumptions, the revenue of each active firm is proportional to $x\,v'(x)$. However, the contribution of its output to group utility is $v(x)$. The ratio of the two is $\rho(x)$. Therefore, if $\rho'(x) > 0$, at the margin the firms in equilibrium find it more profitable to expand than what would be desirable in the optimum. Given the break-even constraint, this has to happen at the infra-marginal cost of having fewer firms.

Note that if $\rho(x)$ is constant over an interval, the right hand side of (5.7) is zero, and on differentiating it, $\beta(x)$ is also constant. However, if $\rho(x)$ is non-zero, we cannot infer a relationship between the signs of $\rho'(x)$ and $\beta'(x)$. Thus the relevant consideration here is not how the elasticity of demand varies with output, but how the elasticity of utility varies.

Normally, we would 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. Further, we can expect

$-xv''/v'$ and xv'/v to be positively related (e.g. for the family $(k+mx)^j$ with $m>0$, $0<j<1$). Then a higher x would correspond to a lower n, and so lower substitution, higher $-xv''/v'$, and higher xv'/v, i.e. $\rho'(x)>0$. Then the equilibrium would have fewer and bigger firms than the constrained optimum. Once again the common views concerning excess capacity and excessive diversity are called into question.

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

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

This yields the conditions

$$\gamma\left[1 - n(a+cx)\right] = (1-\gamma)\ n(a+cx) \tag{5.16}$$

$$\gamma\rho(x)\left[1 - n(a+cx)\right] = (1-\gamma)\ n\ c\ x \tag{5.17}$$

Using these and (5.5) and recalling that $I = 1-an$ in the unconstrained optimum, we find

$$p_u = c \tag{5.18}$$

and x_u is defined by

$$cx/(a+cx) = \rho\ (x) \tag{5.19}$$

Subtracting the right hand side of this from the right hand side of (5.12) yields the expression

$$\frac{1}{1+\beta(x)} - \rho(x) - \frac{\rho(x)}{\gamma\rho(x) + (1-\gamma)} \frac{x\rho'(x)}{\rho(x)}$$

$$= \left[1 - \frac{\rho(x)}{\gamma\rho(x) + (1-\gamma)}\right] \frac{x\,\rho'(x)}{\rho(x)}$$

$$= \frac{(1-\gamma)\;(1-\rho(x))}{\gamma\rho(x) + (1-\gamma)} \frac{x\,\rho'(x)}{\rho(x)}$$

This has the same sign as $\rho'(x)$. Then, using second order conditions, we find

$$x_u \gtrless x_c \qquad \text{as} \qquad \rho'(x) \lessgtr 0. \tag{5.20}$$

This is in each case transitive with (5.13) to yield output comparisons between the equilibrium and the unconstrained optimum.

Even though the unconstrained and the constrained optima have the same objective, roughly speaking, the break-even constraint forces the latter to pay more attention to revenue. Therefore, the consideration of the ratio of revenue to utility helps us to understand the output comparisons in (5.20).

The DD-curve on which the price-output combination in the unconstrained optimum lies differs from that in the other two situations, because of differences in the lump sum incomes as well as the numbers of firms. The latter cannot therefore be compared using an argument like that of Figure 3. However, we have from (5.16) that

$$n_u = \gamma/(a + c\,x_u) \tag{5.21}$$

and

$$n_c = \frac{\gamma}{a+c\,x_c}\ \frac{\rho\,(x)}{\gamma\rho\,(x) + (1-\gamma)} < \frac{\gamma}{a + c\,x_c}$$

This yields a one-way comparison:

$$\text{If } x_u < x_c\text{, then } n_u > n_c \tag{5.22}$$

We also have a similar result comparing the unconstrained optimum with 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.

6. Asymmetric cases

The next important modification is to remove the assumption of symmetry. We can then ask the broader question : will the right set of commodities be produced in monopolistically competitive equilibrium? And if not, can we say anything about the nature of the biases?[12/] Not surprisingly, the answer to the first question is that a wrong commodity bundle may result. The determination of the set of commodities produced depends on a number of factors : the fixed cost of establishing each firm, the marginal cost of producing the commodity, the elasticity of the demand schedule, the level of the demand schedule, and the cross-elasticities of demand. The following simple example illustrates the fact that there may be multiple equilibria, in one of which everyone is better off than in the other. Assume we have four commodities, coffee, tea, sugar and lemons. Coffee and sugar are strong complements, as are tea and lemons. But coffee-sugar and tea-lemon are strong substitutes. Then there might exist an equilibrium in which coffee and sugar are produced, but tea and lemons are not, and conversely, Given that no tea is produced, the demand for lemons is so low that it cannot meet fixed costs, and conversely, given that no lemons are produced, the demand for tea is equally low. But everyone might prefer a tea-lemon equilibrium to a coffee-sugar one.

This anecdote illustrates the kinds of interactions that are relevant, but does not provide insight into the determinants of the bias possible. Further, it is open to the objection that with complementary commodities, the availability of one increases the demand for the other, so that there is an incentive for one entrant to produce both. In the above example, an entrepreneur who believes that consumers prefer tea-lemon to coffee-sugar will expect a profit from joint production of tea and lemons. However, the problem remains even when there is no complementarity. We illustrate this by means of an example.

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

$$u = x_0^{1-s}\left\{\left[\sum_{i_1=1}^{n_1} x_{i\, i_1}^{\rho_1}\right]^{1/\rho_1} + \left[\sum_{i_2=1}^{n_2} x_{2\, i_2}^{\rho_2}\right]^{1/\rho_2}\right\}^{s} \tag{6.1}$$

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, in each of which only one commodity groups is being produced. These are given by

$$\begin{array}{ll} \text{I} & \text{II} \\ \bar{x}_1 = a_1/(c_1\beta_1)\ , \quad \bar{x}_2 = 0 & \bar{x}_2 = a_2/(c_2\beta_2)\ , \quad \bar{x}_1 = 0 \\ \bar{p}_1 = c_1\ (1+\beta_1) & \bar{p}_2 = c_2\ (1+\beta_2) \\ \bar{n}_1 = s\beta_1/\left[a_1(1+\beta_1)\right] & \bar{n}_2 = s\beta_2/\left[a_2(1+\beta_2)\right] \\ \bar{q}_1 = \bar{p}_1\ \bar{n}_1^{-\beta_1} & \bar{q}_2 = \bar{p}_2\ \bar{n}_2^{-\beta_2} \\ \bar{u}_1 = s^s\ (1-s)^{1-s}\ \bar{q}_1^{-s} & \bar{u}_2 = s^s\ (1-s)^{1-s}\ \bar{q}_2^{-s} \end{array} \tag{6.2}$$

The first 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

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

Hence we require

$$\max_{p_2} \ (p_2 - c_2)\, x_2 = s \ (1 - c_2/\bar{q}_1) < a_2$$

or

$$\bar{q}_1 \quad < \quad s\ c_2/\ (s-a_2) \tag{6.3}$$

Similarly, the second is a Nash equilibrium if and only if

$$\bar{q}_2 \quad < \quad s\ c_1/\ (s-a_1) \tag{6.4}$$

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

$$u = x_0^{1-s} \ \{ \ x_1 n_1^{1+\beta_1} \ + \ x_2\, n_2^{1+\beta_2} \ \}^s \tag{6.5}$$

and the resource availability constraint is

$$x_0 + n_1\ (a_1 + c_1\ x_1) + n_2\ (a_2 + c_2 x_2) = 1 \tag{6.6}$$

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 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 (6.2) 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 (6.2) corresponding to it.

Figure 4 is drawn to depict the possible equilibria and optima. Given all the relevant parameters, we calculate $(\bar{q}_1, \bar{q}_2)$ from (6.2) Then (6.3) and (6.4) 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 non-negative 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 parameter values. Moreover, we can compare the location of the points corresponding to different parameter values and thus do some comparative statistics.

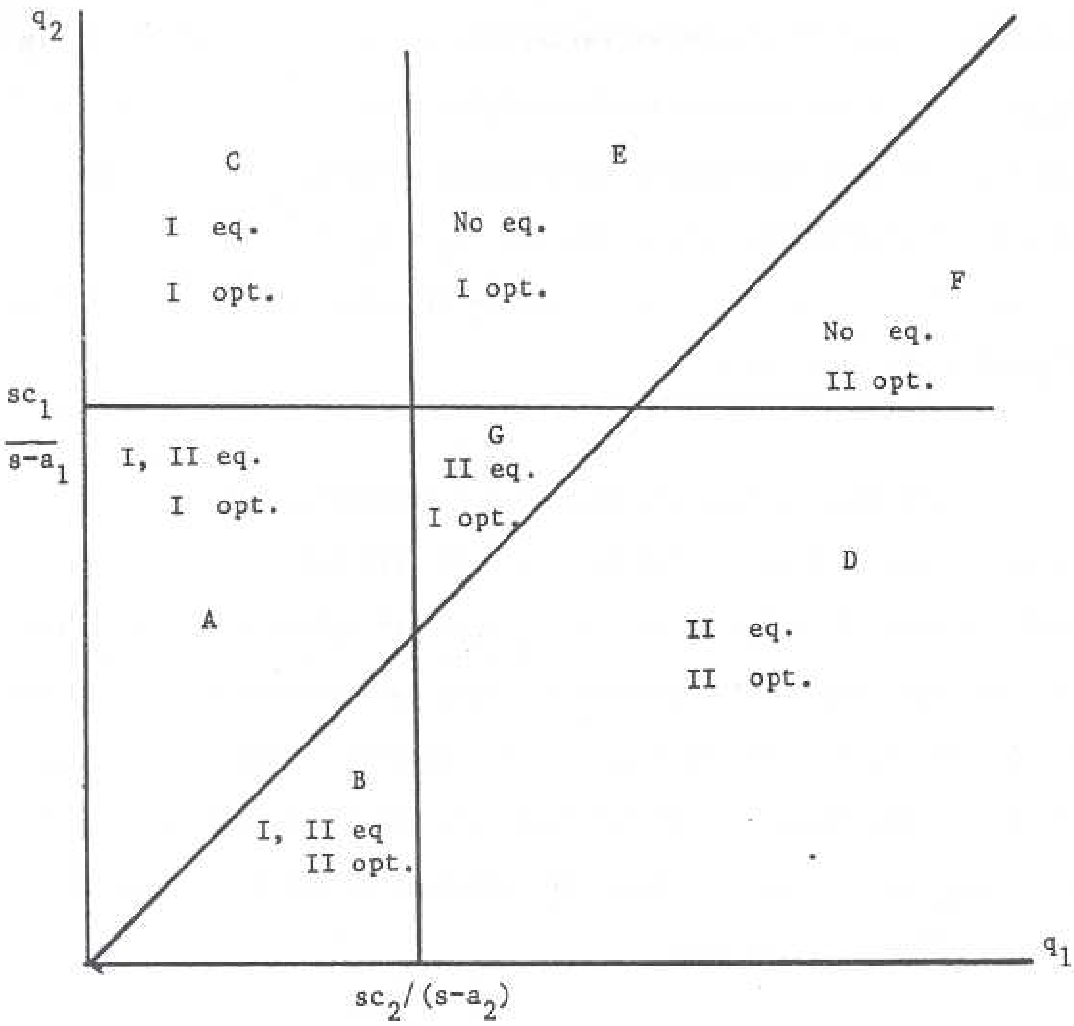

Figure 4

To understand the results, we must examine how much $\overline{q}_i$ 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

$$\partial \log \overline{q}_i/\partial\beta_i = - \log \overline{n}_i \tag{6.7}$$

and we expect this to be large and negative. Further, we see from (2.14) that a higher β_i corresponds to a lower own price elasticity of demand for each commodity in that group. Thus $\overline{q}_i$ is an increasing function of this elasticity.

Consider initially a symmetric situation, with $sc_1/(s-a_1) = sc_2/(s-q_2)$, $\beta_1 = \beta_2$ and when the region G vanishes, and suppose the point $(\overline{q}_1,\overline{q}_2)$ is on the boundary between regions A and B. Now consider a change in one parameter, e.g. a higher own-elasticity for commodities in group 2. This raises $\overline{q}_2$, moving the point into region A, and it becomes optimal to produce commodities from group 1 alone. However, both I and II 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 II 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 $\overline{q}_1$ and moves the point into region B, making it optimal to produce the low-cost group alone while leaving both I and II 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 II is a possible equilibrium and only I 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.

The basic principle underlying the analysis of biases in the choice of commodities is that while the viability of a firm in monpolistically competitive equilibrium depends on the ability to earn sufficient revenues in excess of variable costs to pay for the fixed costs, the desirability of having a firm operate from a social viewpoint depends on the magnitude of revenue <u>plus</u> consumer surplus relative to total costs. Thus, although low own-elasticity commodities would appear to have the potential of earning large revenues in excess of variable costs, they may not be able to do so if there is a high cross-elasticity with a commodity with a high own-elasticity, and low own price-elasticity commodities also tend to have large consumer surpluses associated with their production. In the above example, the inefficient equilibrium is the one in which the high demand elasticity commodity group is produced, when the other commodity group 'ought' to have produced.

In the interpretation of the model with heterogeneous 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 subsidisation 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 5 illustrates. Figute 5a 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 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

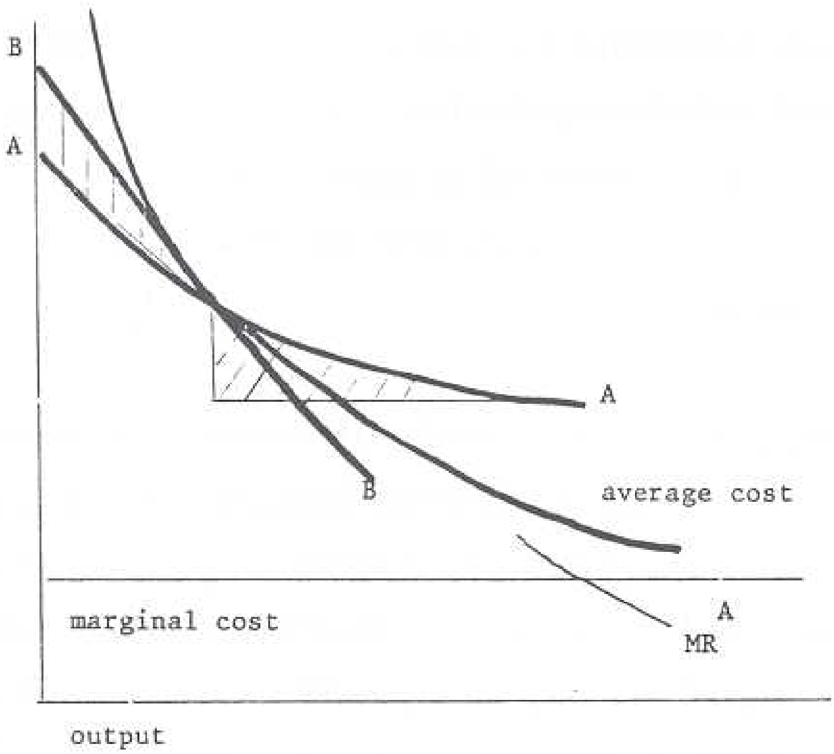

Figure 5 a

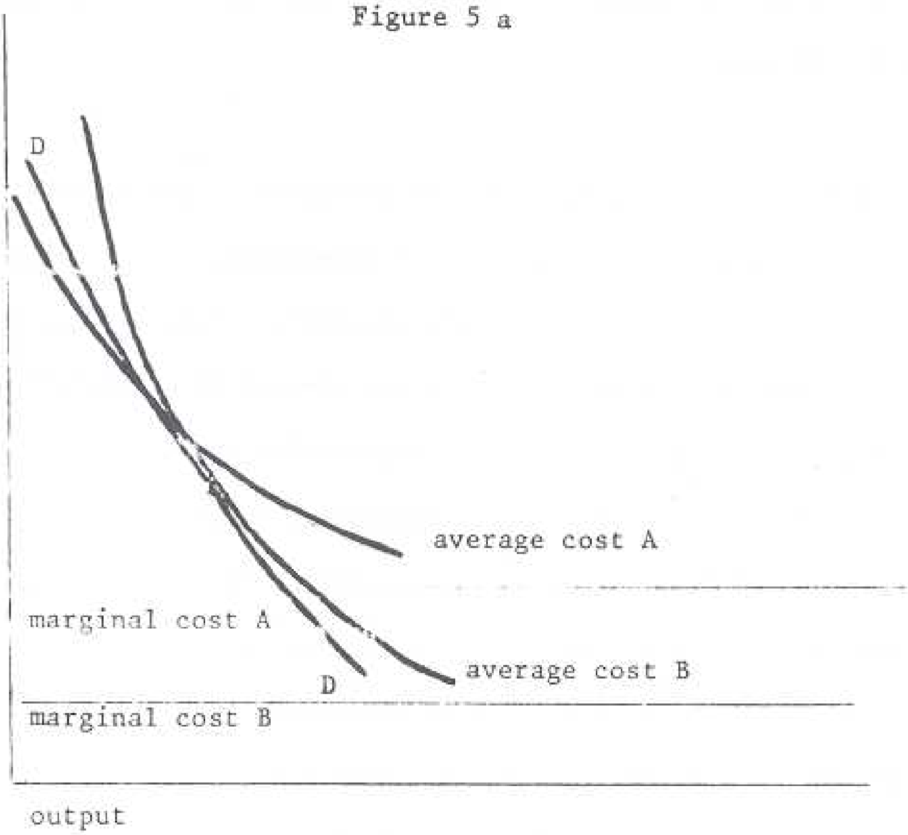

Figure 5 b

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.

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 5b. 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 surplus; indeed, since were it to be produced, B would produce at a much higher level than A, there is a much larger consumer 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.

On the side of production, there is one related problem that we should mention.[13/] We have assumed that the fixed cost for any firm is independent of the number of firms in existence. However, it is often thought that economies of scale in a primary production or servicing industry, or results of standardisation, will mean that the cost $A(n)$ of setting up n firms is not proportional to n. If this is so, we must specifiy the manner in which this is allocated between firms. If each firm is charged the average set-up cost $A(n)/n$, this introduces an externality among firms: setting up a new firm affects the cost charged to existing firms. Competitive equilibrium can exist even if $A(n)/n$ is declining, in the standard manner of Marshallian parametric external economies, but there is now one more reason for it to be inefficient. On the other hand, if each firm is charged the marginal set-up cost, i.e. the i^{th} firm pays $A(i)-A(i-1)$, then economies of scale will mean non-existence of competitive equilibrium, since the infra-marginal firms will be paying higher fixed costs and thus making losses when the marginal one is just breaking even. Also, the optimum will involve a complicated asymmetry that is not easy to handle.

7. Concluding remarks

We have constructed in this paper a series of models to study various aspects of the relationship between market and optimal resource allocation in the presence of some non-convexities. 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 characterised by too few firms in the monopolistically competitive sector.

When demand curves were independent, we also observed a bias against products with low price elasticities or high fixed costs. With interdependent demands, the failure of each firm to take account of the effects on other firms would presumably lead to the possibility of further biases. The polar case examined here confirmed that hypothesis and indicated some particular outcomes of wrong product choice.

A more general theory would attempt to incorporate these various effects into a single model. The problem is one of suitable parametrization to yield interesting results. Michael Spence has considered one such model in a partial equilibrium context. The Lancaster approach of relating interdependence in demand to product attributes is another possibility. Such general models are a subject for further research.

FOOTNOTES

1. Chamberlin (1950), p.89.

2. Kaldor (1935), p.50.

3. Hotelling (1929). However the article by Nicholas Stern casts doubt on that presumption even in the context of location.

4. Stiglitz (1973).

5. An earlier version of this article considered the aesthetically more pleasing case of a continuum of products. However, it was discovered that technical difficulties of that case led to unnecessary confusion.

6. Michael Spence (1974) focuses on this issue in greater detail.

7. See e.g. John Green (1964), p.21.

8. See e.g. Avinash Dixit (1975), Theorem 1.

9. Chamberlin (1933), p.94.

10. Chamberlin appears to have confused the issue by saying that "monopoly is necessarily a part of the welfare ideal", see his article (1950), p.86. As far as the first best is concerned, that is not so. See also Bishop (1967) and Starrett (1974) for analyses of the first best.

11. We are indebted to Michael Spence for pointing out to us the strong implications of assuming constant elasticity functions.

12. For a more exhaustive treatment of these questions, see Michael Spence (1974).

13. Some aspects of this generalization are discussed by Joseph Stiglitz (1973).

REFERENCES

Bishop, Robert L.	"Monopolistic competition and welfare economics", in R. Kuenne (ed). Monopolistic Competition Theory, New York: Wiley, 1967.
Chamberlin, E.	The Theory of Monopolistic Competition, Cambridge, Mass: Harvard University Press, first edition, 1934.
- - - - -	"Product heterogeneity and public policy", American Economic Review, Papers and Proceedings, May 1950, pp.85-92.
Dixit, Avinash	"Welfare effects of tax and price changes", Journal of Public Economics, 4(2), May 1975, pp.
Green, H.A.John	Aggregation in Economic Analysis, Princeton, University Press, 1964.
Hotelling, Harold	"Stability in competition", Economic Journal, 39(1), March, 1929, pp.41-57, reprinted in Stigler and Boulding (eds), op.cit.
Kaldor, Nicholas	"Market imperfection and excess capacity", Economica, N.S.II(1), February 1934, pp.33-50, reprinted in Stigler and Boulding (eds), op.cit.
Spence, Michael	"Product selection, fixed costs and monopolistic competition", 1974, manuscript.
Starrett, David	"Principles of optimal location in a large homogeneous area", Journal of Economic Theory, 9(4), December 1974, pp.418-48.
Stern, Nicholas	"The optimal size of market areas", Journal of Economic Theory, 4(2), April 1972, pp.159-73.
Stiglitz, Joseph E.	"Monopolistic competition and optimal product differentiation : an example", manuscript, March 1973.
Stigler, George and Boulding Kenneth (eds)	Reading in Price Theory, Homewood, Ill.: Irwin, 1952.