Volume Title: Studies in Public Regulation

Volume Author/Editor: Gary Fromm, ed.

Volume Publisher: The MIT Press

Volume ISBN: 0-262-06074-4

Volume URL: http://www.nber.org/books/from81-1

Publication Date: 1981

Chapter Title: Open Entry and Cross-Subsidization in Regulated Markets

Chapter Author: Kenneth C. Baseman

Chapter URL: http://www.nber.org/chapters/c11435

Chapter pages in book: (p. 329 - 370)

7

Open Entry and Cross-Subsidization in Regulated Markets

Kenneth C. Baseman

Decisions by the Federal Communications Commission and other regulatory agencies to encourage competition in regulated markets have led to extensive consideration of the problem of cross-subsidization. In particular, the FCC has concerned itself (in docket 18128) with whether American Telephone and Telegraph could increase prices in its monopoly MTS-WATS markets in order to finance "unfair" competition against firms competing with AT&T in other markets.[1] The same issue arises in postal service. The U.S. Postal Service is a monopoly carrier for first class mail and competes with various firms in parcel delivery.

AT&T's position before the FCC and Congress is one which I suspect many economists would find appealing, at least at first blush:[2] that as long as revenues in competitive markets cover incremental costs, the pricing structure cannot be said to involve cross-subsidization even though the resulting prices exclude potential competitors from markets nominally open to entry. Revenues in excess of incremental cost in competitive markets allow those markets to make some contribution toward covering costs that are common to the provision of both competitive and monopoly services. Therefore, prices charged to consumers of monopoly services can be lower than if joint provision of the various services were not allowed. Further, consumers of competitive services are better off than if the monopolist's price were disallowed, since that price is lower than the price competing firms can offer. Therefore, correct application of the incremental-cost test ensures that consumers in all markets are better off with joint provision of the services.

My argument is that the above position, despite its intuitive appeal, is incorrect. I shall demonstrate that the combination of franchise monopoly markets and an incremental-cost pricing rule in competitive markets cannot be expected, in general, to lead the industry to the most efficient market structure. In particular, it will be shown that the above set of rules may allow the multiproduct firm to successfully monopolize both its own franchise monopoly markets and the competitive markets even when joint production is absolutely inefficient. In such situations, the price in the monopoly markets will be seen to lie above the cost of serving those markets alone. The results are then interpreted as statements about cross-subsidization by reference to Faulhaber's (1975a) game-theoretic approach to cross-subsidization.

A Simple Numerical Example

Suppose we have two products for which the demands are perfectly price-inelastic, and that the cost functions are

$$C(q_1,0) = 40,$$

$$C(0,q_2) = 40,$$

$$C^{m}(q_1,q_2) = 90.$$

Let S denote an input, used by the joint monopolist, that is necessary for the production of either service. Expenditures on S are common costs for the joint monopolist. Let A_i denote inputs that are used only in the production of service i. Expenditures on A_i are attributable to service i. Suppose all input prices equal unity.

Now let the two-product monopolist's most efficient input choice, with respect to his joint-production technology, be

$$S = 50,$$

$$A_1 = 20,$$

$$A_2 = 20.$$

Let service 2 be the joint monopolist's franchise monopoly market. The incremental cost of additionally serving market 1 is given by

$$C^{m}(q_1,q_2) - C^{m}(0,q_2) = 90 - 70 = 20. \tag{1}$$

The monopolist can clearly pick p_1^{m} to satisfy an incremental-cost rule, while beating the competition in market 1 by choosing p_1^{m} such that

$$20 < p_1^{m} < 40.$$

To satisfy a break-even constraint, the monopolist then chooses p_2^{m} to satisfy

$$90 - p_1^{m} = p_2^{m} > 50. \tag{2}$$

The final result is a joint monopoly over both markets, with total costs 10 above what they would be with separate production.

Now suppose that the monopolist's cost-minimizing input choice, using the joint production technology, is given by

$S = 8,$

$A_1 = 42,$

$A_2 = 42.$

Since the incremental cost of serving market 1 is now 42, it would appear that an inefficient joint monopoly could not pass an incremental-cost test. However, suppose that the elasticity of substitution between the inputs S and A_i is greater than zero (that is, suppose we are not dealing with a fixed-proportions technology). Then an inefficient joint monopoly may be feasible. Suppose that one of the points on the monopolist's isoquant is

$S = 20,$

$A_1 = 38,$

$A_2 = 38.$

Total costs are now 96, reflecting the fact that we have moved away from the cost-minimizing point for joint production. However, since A_1 is less than 40, joint monopoly can now be successful.

In remainder of the article I will distinguish, as was done above, between two types of inefficient joint monopoly: type 1, in which the monopolist's cost-minimizing input choice, using the joint-production technology, allows him to pass an incremental-cost test, and type 2, in which the joint monopolist cannot pass an incremental-cost test using the cost-minimizing input combination for joint production, but input substitutability allows the joint monopolist to serve both markets while satisfying an incremental-cost pricing rule.

More Examples of Inefficient Joint Monopoly

In this section, a two-product model will be used to further develop the notion of inefficient joint monopoly. The assumptions of the model, except for subadditivity and the specification of the monopolist's cost function, match the assumptions made in Faulhaber's (1975a) work on cross-subsidization.

Assumptions

A1. The demand functions are given by

$$q_i = q_i(p_1,p_2) \qquad (i = 1,2),$$

with

$$\frac{\partial q_i}{\partial p_i} < 0, \qquad \frac{\partial q_i}{\partial p_j} \geq 0.$$

A2. The cost functions for independent production of the two commodities are given by $C^1(q_1,0)$ and $C^2(0,q_2)$. The cost functions exhibit nonincreasing average cost:

$$\frac{d(C^i(q_i)/q_i)}{dq_i} \leq 0.$$

A3. The joint monopolist's cost function is given by

$$C^m(q_1,q_2) = \min_{S,A_1,A_2} C^m(S,A_1,A_2;q_1,q_2,p_{A_1},p_{A_2},p_S), \tag{3}$$

where S, A_1, and A_2 retain their definitions from the preceding section. The underlying production function is given by

$$q_1^m = q_1^m(S,A_1),$$

$$q_2^m = q_2^m(S,A_2).$$

The isoquants for the above production functions are convex:[3]

$$\left.\frac{dS}{dA_i}\right|_{dq_i=0} < 0, \qquad \left.\frac{d^2S}{dA_i^2}\right|_{dq_i=0} > 0.$$

A4. The overall cost function is superadditive:

$$C^m(q_1,q_2) > C^1(q_1,0) + C^2(0,q_2).$$

A5. The monopolist is confined to price pairs for which $\partial\pi^m/\partial p_i > 0$ $(i = 1,2)$. The monopolist's $\pi^m = 0$ locus might look something like that depicted in figure 7.1. Price pairs along the arc AB satisfy $\partial\pi^m/\partial p_i > 0$. The assumption is plausible in that, even if the firm is indifferent to whether or not its prices lie on an undominated portion of the $\pi^m = 0$ locus, the regulator will attempt to force the firm toward the "low" price pairs on AB. In addition, along any $\pi^m = k > 0$ locus, a profit-maximizing monopolist faced with either direct or indirect competition (in the form of a substitute product offered by a competing firm) has an incentive to choose the "low" price pairs for which $\partial\pi^m/\partial p_i > 0$.

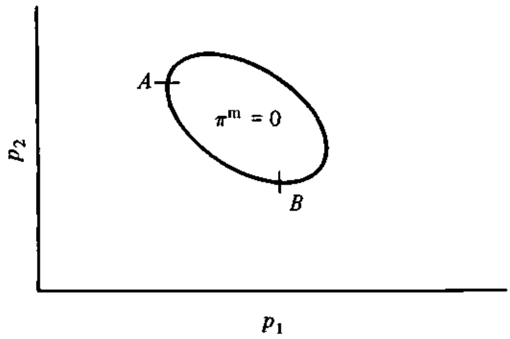


Figure 7.1 Region of non-negative profit for a two-product monopolist.

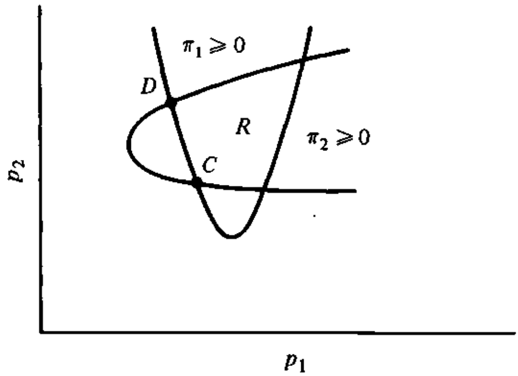


Figure 7.2 Regions of non-negative profit for two single-product firms.

We are also assuming that, at the noncooperative zero-profit equilibrium for separate production, $\partial\pi^i/\partial p_i > 0$ for $i = 1,2$. This would be the case if the profit functions $\pi^i(p_i;p_j)$ were concave in their own prices. Figure 7.2 shows regions of non-negative profits are shown for firms producing independently (with $\partial q^i/\partial p_j > 0$ and decreasing average costs). At C, a price increase by either firm would yield it positive profits. At any other zero-profit solution for the two firms, such as D, $\partial\pi^i/\partial p_i < 0$ for at least one of the firms (in this case it is firm 2). C dominates other solutions to $\pi^i(p_1,p_2) = 0$ for $i = 1,2$, since, at the other solutions, entry would force price reductions in at least one of the markets.

We know from assumption A4 that the joint monopolist will lose money at (p_1^c,p_2^c), since the two independent firms just break even at that point and the monopolist incurs larger costs to produce the same outputs. Denote by L the joint monopolist's losses at (p_1^c,p_2^c). R^m is the monopolist's revenue function. Thus,

$$\pi^{m}(p_1^c,p_2^c) = R^{m}(p_1^c,p_2^c) - C^{m}(q_1^c(p_1^c,p_2^c), q_2^c(p_1^c,p_2^c)) = L < 0. \tag{4}$$

Now assume that market 2 is a franchise monopoly market for our joint monopolist, but that entry is allowed in market 1. Can the joint monopolist, given a franchise monopoly in market 2 and an incremental-cost pricing rule in market 1, find a price pair that yields non-negative profits? That price pair (p_1^m,p_2^m) must satisfy the following conditions.

i. $p_1^m \leq p_1^*$, where p_1^* is the lowest price, given p_2^m, at which a firm serving only market 1 can break even.[4] A profit-maximizing monopolist will always choose to satisfy this requirement with equality. This can be most easily seen by referring back to figure 7.1. Positive profits occur for the monopolist at price pairs in the interior of $\pi^m = 0$. Thus, if our monopolist starts at some price pair on AB and finds his p_1^m more than sufficient to prevent entry in market 1, he can increase profits by increasing p_1^m. Thus, condition i is replaced by condition i′:

i′. $p_1^m = p_1^*$. This result, which means that for a given value of p_2^m profits are maximized at p_1^*, will simplify the analysis, since p_1^* can be derived implicitly from p_2^m and we need deal with only the single decision variable p_2^m.

ii. $p_1^m q_1^m \geq p_{A_1} A_1$. This is the incremental-cost pricing rule.

iii. $\pi^m(p_1^m,p_2^m) \geq 0$.

The entry-deterring price[5] in market 1, p_1^*, is given implicitly by

$$p_1^* = p_1 | R^1(p_1,p_2^m) - C^1(q_1(p_1,p_2^m) = 0. \tag{5}$$

The derivative dp_1^*/dp_2^m is

$$\frac{dp_1^*}{dp_2^m} = -\frac{\partial R^1}{\partial p_2^m} - \frac{\partial C^1}{\partial q_1}\frac{\partial q_1}{\partial p_2}\Bigg/\frac{\partial R^1}{\partial p_1^*} - \frac{\partial C^1}{\partial q_1}\frac{\partial q_1}{\partial p_1^*}$$

$$= -\left(p_1^* - \frac{\partial C^1}{\partial q_1}\right)\frac{\partial q_1}{\partial p_2}\Bigg/\frac{\partial R^1}{\partial p_1^*} - \frac{\partial C^1}{\partial q_1}\frac{\partial q_1}{\partial p_1^*} \leq 0. \tag{6}$$

The denominator in the expression above is equal to $\partial\pi^1/\partial p_1^* > 0$ for $p_1^* \leq p_1^c$. The numerators are non-negative, being the product of two non-negative terms. Assumption A2 requires $\partial C^1/\partial q_1 \leq C^1(q_1)/q_1$, and equation (5) requires $p_1^* = C^1(q_1)/q_1$. Therefore, $p_1^* \geq \partial C^1/\partial q_1$. Also, $\partial q_1/\partial p_2 \geq 0$ by assumption A1.

We now examine a candidate price vector $(p_1^*,\bar{p}_2^m)$ to see if the joint monopolist can satisfy conditions i′ through iii. At $(p_1^*,\bar{p}_2^m)$ the joint monopolist's profits are

$$\pi^{m}(p_1^*,\bar{p}_2^m) = R^m(p_1^*,\bar{p}_2^m) - C^m(q_1(p_1^*,\bar{p}_2^m),q_2(p_1^*,\bar{p}_2^m)). \tag{7}$$

The change in the joint monopolist's profits when he moves from $(p_1^* = p_1^c, p_2^c)$ to $(p_1^*,\bar{p}_2^m)$ is given by

$$\int_{p_2^c}^{\bar{p}_2^m} \frac{d\pi^m}{dp_2^m} dw,$$

where w is a dummy variable of integration.
The total derivative $d\pi^m/dp_2^m$ is

$$\frac{d\pi^m}{dp_2^m} = \frac{\partial^+\pi^m}{\partial p_1^*}\frac{dp_1^*}{dp_2^m} + \frac{\partial^+\pi^m}{\partial p_2^m}. \tag{8}$$

A necessary condition for a successful joint monopoly is that equation (8) be positive. If the monopolist cannot obtain an increase in profits via an increase in p_2^m,[6] his initial losses (L) can never be overcome.

For given specifications of the demand and cost functions, equation (8) can be either positive or negative. Large unexploited monopoly profits in the franchise monopoly market will help the joint monopolist, since then $\partial\pi^m/\partial p_2^m$ will be large. Similarly, large losses in market 1, given dp_1^*, will tend to hurt the joint monopolist. Large substitution effects (in demand) and substantial scale economies for independent production of service 1 will hurt the joint monopolist, since these factors make $|dp_1^*/dp_2^m|$ large.

When demands are independent or when independent production is done at constant cost, then $dp_1^*/dp_2^m = 0$ and equation (8) reduces to

$$\frac{d\pi^m}{dp_2^m} = \frac{\partial\pi^m}{\partial p_2^m}. \tag{9}$$

This result is highly intuitive. Independent demands or constant costs imply that changes in p_2^m will not affect the entry-preventing price in market 1. Therefore, the only things that affect profits when p_2 changes are the changes in costs and revenues in market 2.

We are now ready to complete the analysis. Type 1 joint monopoly will obtain if

$$\int_{p_2^c}^{\bar{p}_2^m} \frac{d\pi^m}{dp_2^m} dw - L \geq 0 \tag{10}$$

and

$$P_{A_1}A_1 \leq p_1^*\bar{q}_1^m. \tag{11}$$

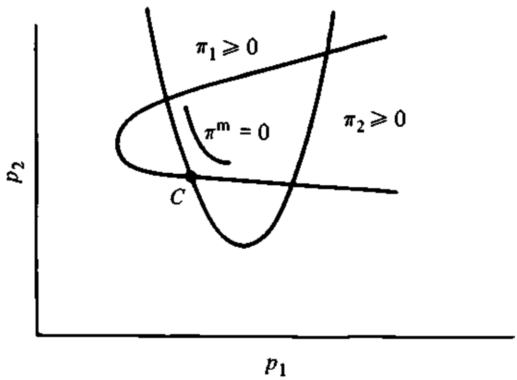


Figure 7.3 A case where joint monopoly cannot succeed.

When equation (11) is not satisfied by the joint monopolist's cost-minimizing input choices, given $\bar{q}_1^m$ and $\bar{q}_2^m$, the monopolist can turn to type 2 monopolization. Let the joint monopolist substitute away from A_1 and toward S until the point is reached (if one exists) where $p_{A_1}A_1 = p_1^*\bar{q}_1^m$. Denote by $\hat{A}_1$ the largest value of A_1 satisfying the incremental-cost constraint. Because the monopolist's isoquants are assumed to be convex, the total costs of supplying the two markets will increase with this substitution. Denote by ΔC^m the change in costs:

$$\Delta C^m = \min_{A_2,S} C^m(A_2,S;\hat{A}_1,\bar{q}_1^m,\bar{q}_2^m) - \min_{A_1,A_2,S} C^m(A_1,A_2,S;\bar{q}_1^m,\bar{q}_2^m) > 0. \quad (12)$$

Of course, the larger the elasticity of substitution, the smaller ΔC^m will have to be in order to satisfy the incremental-cost pricing rule.

The joint monopolist will now succeed if

$$\int_{p_2^c}^{\bar{p}_2^m} \frac{d\pi^m}{p_2^m} dw - L - \Delta C^m \geq 0. \quad (13)$$

The various possibilities for success or failure of inefficient joint monopoly can be diagramed. Demands are assumed to be substitutes in these figures. Figure 7.3 depicts a case where joint monopoly cannot succeed. The $\pi^m = 0$ locus lies "northeast" of the point C because of the superadditivity assumption. Joint monopoly cannot succeed, because the undominated portion of the $\pi^m = 0$ locus lies entirely within the $\pi_1 \geq 0$ region. At any price pair chosen by the monopolist, a firm offering only service 1 can successfully underprice p_1^m.

Figure 7.4 illustrates a situation where joint monopoly might succeed. Because a $\pi^m = k$ locus passes outside $\pi_1 \geq 0$, price pairs (such as point

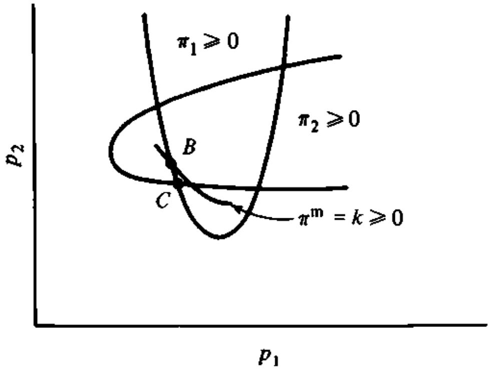


Figure 7.4 A case where joint monopoly might succeed.

B) can be chosen at which a firm serving only market 1 will not be viable. If at any such point revenues exceed incremental costs, type 1 joint monopoly is possible. If the incremental-cost rule is not satisfied at B, the monopolist can attempt to satisfy the rule by substituting away from A_1. If there exists an $\hat{A}_1$ that satisfies the rule, and if $k \geq \Delta C^m$, then a type 2 joint monopoly could survive.

Type 1 versus Type 2 Joint Monopolies

I have found it convenient for heuristic purposes to define two different types of joint monopolies. Both types (not just type 2) depend crucially on the assumption of variable proportions. If it is true that joint production is necessarily inefficient, then a cost-minimizing monopolist will forgo joint production and operate two subsidiaries that produce their services independently. Costs will be higher than under independent production by separate firms, owing to the presumably small costs of operating the holding company. To get from the holding company to a case of type 1 joint monopoly, the monopolist must substitute away from attributable inputs toward common inputs; otherwise the incremental-cost pricing constraint can never be satisfied. This is not to say that the distinction is purely heuristic. Type 1 joint monopolies could occur whenever a monopolist fails to minimize costs, so long as the failure is biased in the "right" direction. The failure could either be inadvertent (that is, a result of x-inefficiency) or deliberate. There are reasons why the monopolist might choose a technology other than the cost-minimizing one. For example, Averch-Johnson input distortion and rate-base padding are both due to decisions by the monopolist to choose input combinations

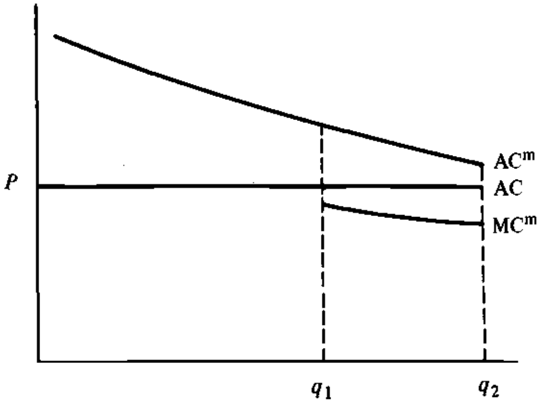


Figure 7.5 The case of an inefficient single-product monopoly.

that do not minimize costs. The profitability of such choices, of course, depends upon the monopolist's being subject to a rate-of-return constraint and not a break-even constraint.

The incentive the monopolist will have to monopolize both markets (in either the type 1 or the type 2 case) also follows from the fact that the profit constraint is given in the form of a limit on the allowed rate of return. Averch and Johnson (1962) demonstrated that so long as the allowed rate of return exceeds the cost of capital and additional profits can be drawn from a monopoly market, a regulated monopolist will find it profitable to serve additional markets even if incremental costs are not covered in those markets. Serving such markets increases the size of the rate base, thereby allowing total profits to rise. The monopolist can cover any "losses" in competitive markets with additional revenues drawn from the monopoly market.

The Results in Terms of Cost Functions

The results in the preceding sections can be illustrated by looking directly at the cost functions. In figure 7.5 we assume that the monopolist has a franchise monopoly over the first q_1 units of output, but that for subsequent units $(q_2 - q_1)$ competition is allowed. The minimum-average-cost curve is shown as AC. The inefficient monopolist's average-cost curve is AC^m. The incremental cost for the monopolist of producing $q_2 - q_1$ (the area under MC^m) is less than the constant incremental cost of those units when produced by another firm. The inefficient monopolist can thus serve both markets. The market will look like a natural monopoly market to a regulator who does not have knowledge of the full technology set.

A similar result can be obtained in the multiproduct case. The assumption that the joint monopolist is inefficient is stated algebraically as

$$C^m(q_1,q_2) > C^1(q_1) + C^2(q_2),$$

$$C^m(q_1,0) > C^1(q_1),$$

$$C^m(0,q_2) > C^2(q_2).$$

If the joint monopolist passes an incremental-cost test in market 1 and the monopolist chooses p_1^m to "just" deter entry, we have

$$C^m(q_1,q_2) - C^m(0,q_2) \leq p_1^m q_1^m = C^1(q_1),$$

or

$$C^m(q_1,q_2) \leq C^m(0, q_2) + C^1(q_1)$$

or

$$C^m(q_1,q_2) < C^m(q_1,0) + C^m(0,q_2).$$

The monopolist's chosen technology will appear to be subadditive, although the cost-minimizing technology is not. A regulator without knowledge of the full technology set will think he is dealing with a two-product natural monopoly.

These results are not surprising. Under an incremental-cost test, incremental costs for each service will not sum up to total costs whenever the average cost function is declining or, with several products, when the cost function is subadditive. There will be a residual of common costs that cannot be attributed to any service and that reflect the fact that production of larger outputs allows the firm to realize lower average costs of production. The substitution away from attributable inputs toward common inputs thus requires the choice of a cost function that is "more subadditive" than the cost function associated with the cost-minimizing technology.

Inefficient Joint Monopoly and Cross-Subsidization

Readers familiar with Faulhaber's work on cross-subsidization have no doubt recognized that the prices charged by our joint monopolist are not subsidy-free. Faulhaber's basic insight is that cross-subsidization can be neatly and elegantly analyzed by employing the theory of cooperative games. The "players" of the game are the consumers in each market. Joint monopoly production is the outcome of that game if the monopolist chooses a price vector satisfying the core conditions. Among the core conditions are the requirements that $p_i^m \leq p_i^c$—that is, the monopolist must "beat" the prices available to each "player" if independent production is chosen. Our joint monopolist clearly does not satisfy the core conditions, since $p_2^m > p_2^c$. The price in market 2 is above "stand alone" cost, and service 2 users are subsidizing consumers of service 1. The franchise-monopoly rule prevents the consumers in market 2 from withdrawing from the joint monopolist's "coalition." Even though joint monopoly is inefficient and the prices are outside the core of the cooperative game, a "cooperative" solution—joint monopoly—is forced upon the users of service 2.

Open Entry

We have seen that a franchise monopolist can be inefficient and still compete successfully in other markets even when subject to an incremental-cost pricing rule in competitive markets. In addition, an incremental-cost pricing rule gives a franchise monopolist an incentive to choose inefficient technologies if his cost-minimizing joint technology does not allow him to satisfy an incremental-cost pricing rule in competitive markets. In either case, because resources are wasted in supplying the franchise-monopoly market, a natural policy implication is that the monopoly market ought to be opened up to entry. Work by Panzar and Willig (1977a) indicates that such a market test may backfire. A "true" natural monopoly may be unsustainable if entry restrictions are lifted.

Sustainability

Panzar and Willig (1977a) have demonstrated that the assumption that the cost function is subadditive (the "natural monopoly" assumption) is not sufficient to guarantee that a monopolist will be able to profitably maintain production of the full set of outputs if free entry is allowed. A natural monopoly, when faced with entry, may be forced to reduce its product offerings or face elimination from the market.

In the one-product case, if natural monopoly is unsustainable the effect of an open-entry rule will be a reduction in output. In the multiproduct case, when natural monopoly is unsustainable and entry is allowed, the

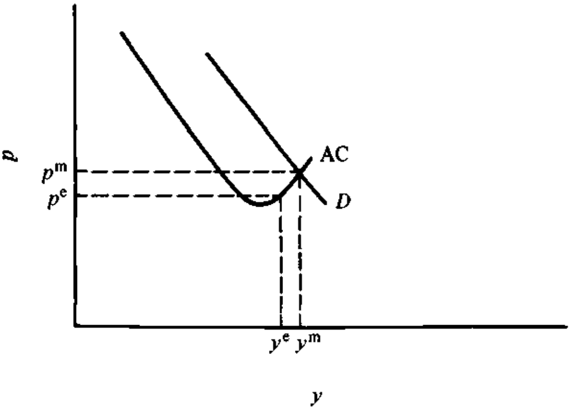


Figure 7.6 The case of an unsustainable single-product natural monopoly.

resulting equilibrium will feature fewer products. In either case, welfare (total surplus) could increase or decrease if entry is allowed and the natural monopoly is unsustainable. Thus, under certain circumstances, entry restrictions may be required in order to achieve the welfare maximum.

The sustainability models thus appear to raise troublesome policy questions concerning the role of entry restrictions in regulated markets. Many regulated markets exhibit declining average-cost curves and possible economies of joint production, which may give rise to sustainability problems. This is especially true where some sort of "networking" is present, as in telecommunications, mail service, power distribution, and transportation. In such instances there may be economies from the joint provision of several "links" in the network.

The question Panzar and Willig pose is: Under what conditions can a natural monopolist choose a price vector that guarantees that a potential entrant can anticipate only negative profits if entry is attempted? No consideration is given to the joint monopolist's reaction to attempted entry. The justification for this assumption, and the implications of relaxing the assumption, will be considered later in this article.

Sustainability with One Product

The "natural monopoly" assumption does not require the average cost curve to be monotonically decreasing. Figure 7.6 depicts a cost curve that satisfies the subadditivity assumption that the cheapest way to produce the output y^m is production by only one firm. In such a situation, where the average-cost curve turns up, the natural monopoly will be unsustainable. For example, an entrant, by choosing p^e, can supplant the original monopolist. Of course, if the average-cost curve is monotonically decreasing, the natural monopoly will be sustainable. This result was first pointed out by Faulhaber (1975b).

Sustainability with Two Products

(Note: Some cases considered by Panzar and Willig are not treated in this section. "Cooperative" entry and the case of independent demands will be discussed later.)

The monopolist's outputs are denoted by the vector $y^m = (y_1^m, y_2^m)$. The monopoly prices are $p^m = (p_1^m, p_2^m)$. The demand functions are given by

$$y_i^m = Q^i(p^m) \qquad (i = 1,2)$$

with

$$\frac{\partial Q^i}{\partial p_i} < 0, \qquad \frac{\partial Q^i}{\partial p_j} > 0.$$

The multiproduct minimum cost function is $C(y^m)$. This cost function is assumed to exhibit multiproduct scale economies characterized by least ray average cost; $(1/\lambda)C(\lambda y^m) > C(y^m)$ for $0 < \lambda < 1$). In addition, the average incremental cost of each output, with the amount of the other service held constant, is assumed to be declining.

The cost functions available to firms producing only one of the two products are $C(y_1,0)$ and $C(0,y_2)$. These cost functions are assumed to exhibit declining average costs. Costs are assumed to be subadditive:

$$C(y^m) = C(y_1^m, y_2^m) < C(y_1,0) + C(0,y_2). \tag{13}$$

The monopolist's chosen price vector is sustainable against an entrant offering only one of the products if and only if

$$p_i^e y_i^e - C(y_i^e) < 0 \qquad (i = 1,2),$$

where $p_i^e \leq p_i^m$ and $y_i^e \leq Q^i(p_i^e, p_j^m)$.

Panzar and Willig derived seven necessary conditions for the sustainability of the monopolist's price and output choice. Among those conditions, in addition to subadditivity and cost minimization by the monopolist, are the following:

• p^m is undominated; that is, there cannot exist a $(p_1^e, p_2^e) \leq (p_1^m, p_2^m)$ with $\pi(p^e) \geq 0$.

• $p_i^m y_i^m < C(y_i^m)$ for $i = 1,2$. If this condition did not hold, then an entrant could offer y_i^m at p_i^m and earn non-negative profits.

The second condition is related to Faulhaber's application of the theory of cooperative games to the problem of cross-subsidization. Multiproduct monopoly can be viewed as the core outcome of a cooperative cost game in which the "players" are the consumers of each service (one "player" is associated with each service). Denote by E_i the revenue allocation ("imputation") assigned to each customer group. The core conditions require that revenues cover costs, or

$$\sum_{i=1}^{2} E_i \geq C(y^m),$$

and that no customer group be charged more than its "stand-alone" costs, or

$$E_i < C(y_i^m) \qquad (i = 1,2).$$

The second condition requires that the imputation $E = (p_1^m y_1^m, p_2^m y_2^m)$ be in the core of the cost game. Clearly, it is first required that the cost game possess a nonempty core.[7] Panzar and Willig show that the assumption of weak cost complementarities is sufficient to ensure that the core of the cost game is nonempty. Weak cost complementarities require that the marginal cost of one service not increase with increased output of any other service.

Given a nonempty core to the cost game, the second condition plus $\pi^m \geq 0$ require that the market revenues lie in that core. This means that the demand functions must lie "out" far enough so that the monopolist can cover its costs (for example, if all demands are zero, market revenues are clearly outside the core). In addition, market demands allowing $\pi^m \geq 0$ must contain some points that allow the monopolist to price below stand-alone cost in both markets. Panzar and Willig provide some examples which indicate that, when the demands are insufficient to support independent production of the two commodities, the monopolist, in order to serve both products, may have to set price above stand-alone cost in one of the markets. In this section I shall be concerned only with sustainability problems that arise when the second condition is satisfied.[8] (Sustainability problems when that condition is not satisfied are discussed briefly below.)

The crux of the sustainability problem lies in the twin assumptions

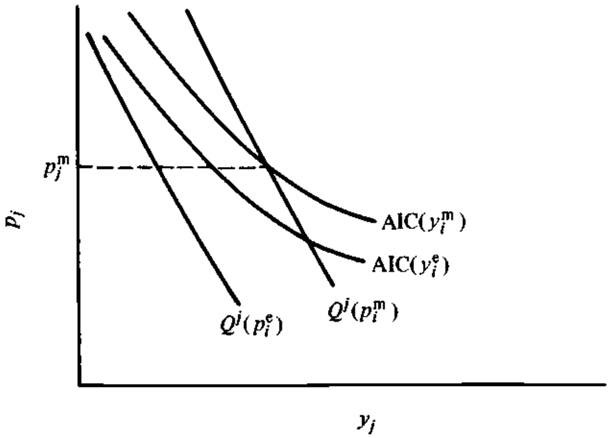


Figure 7.7 The case of an unsustainable natural monopoly that cannot recover the incremental costs of providing a substitute service.

of single-product economies of scale (for both entrant and monopolist) and demand substitutability. The first of these effects allows the potential entrant, facing p^m, to contemplate a substantial price cut in one of the markets. The potential entrant, looking only at the demand for one of the products, enters if $\pi^e \geq 0$ can be achieved (the market demand curve lies above his average cost curve). This gives a new $p_i = p_i^e$. The question, for the sustainability calculation, is whether the monopolist could have chosen a vector $p^m = (p_i^e, p_j^m)$ that would have yielded non-negative profits. Such a choice may not have been feasible. The lower price for service i implies that the demand curve for substitute service j will have shifted "backward." At lower outputs, the joint monopolist's average incremental cost (AIC) of providing service j will increase. If the average incremental cost of providing service j increases "fast" enough, and the demand curve shifts back "far" enough, there may be no points in common that allow the monopolist to recover the incremental costs of providing service j. Such a situation is depicted in figure 7.7.

The move from (p_i^m, y_i^m) to (p_i^e, y_i^e) has two effects. AIC_j shifts down, because production of service j is now complemented by larger production of service i.[9] The demand curve for service j shifts back because the products are substitutes. As drawn, the new demand curve lies nowhere above the new AIC curve, so the natural monopoly is unsustainable. Moreover, by subadditivity, the average-cost curve for a firm supplying only service j lies above AIC_j. Thus, service j will not be provided at all. The total surplus forgone in market j is the area above p_j^m and below $Q^j(p_i^e)$. If this forgone surplus exceeds the gains in surplus in market i,

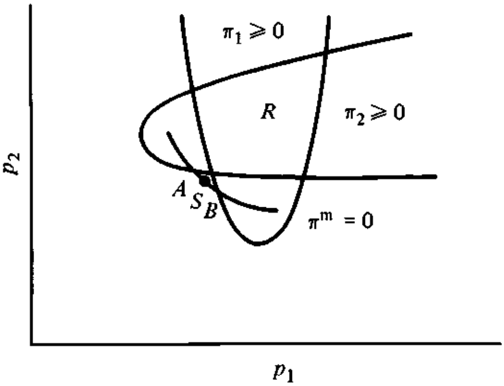


Figure 7.8 A sustainable natural monopoly.

then the decision to allow entry will have resulted in a decline in welfare.

In the sustainability calculation, then, the presence of strong product-specific scale economies for both the entrant and the joint monopolist, and a large cross-price elasticity of demand tend to favor the entrant over the monopolist. On the other side, of course, large economies of joint production tend to favor the joint monopolist.

The sustainability problem is further illustrated in figures 7.8 and 7.9. Figure 7.8 depicts a sustainable natural monopoly. The region of prices that could simultaneously support independent production of both services is denoted R. By subadditivity, the $\pi^m = 0$ locus must lie "south-west" of R. The sustainable range of prices lies on the arc AB. From any point on that arc (such as the point S), price reductions for either service cannot move an entrant into a region of non-negative profits.

Figure 7.9 shows an unsustainable natural monopoly. An entrant in market 2 can, by a large price reduction, come to rest in the $\pi_2 \geq 0$ region, thus successfully supplanting the original monopolist. The point M satisfies all the necessary conditions for sustainability. However, M is unsustainable against, for example, the point E.

Sustainability with Monopoly Reaction

That a monopolist does not react when entry occurs is a very tenuous assumption. Panzar and Willig justify the use of the Cournot-Nash assumption with regard to prices as follows:

...the Cournot-Nash expectation need not be self-fulfilling.
Competitive entry would plunge the previously breaking-even monopoly

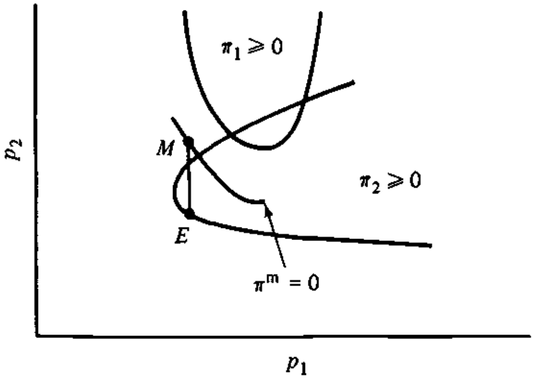


Figure 7.9 An unsustainable natural monopoly.

into the red and would necessitate changes in its prices. However, if the entrant is "small" relative to the industry, the magnitude of these required adjustments will also be "small" and may hence be justifiably ignored by the entrant. Also, the entrant may believe that the regulator would insure that the price adjustments would not be retaliatory with respect to the entrant's product lines, and that instead the monopolist's losses would be recovered from changes in prices of services only tangentially related to the entrant's offerings. (Panzar and Willig 1977a, p. 4.)

The second of these reasons is plausible enough. Regulators have the power to forbid price reactions, and occasionally have been known to exercise that power. However, in two of the three versions of the sustainability models presented by Panzar and Willig the entrants are quite large. With several products, scale economies, and declining average incremental cost, the entrant must supply more of the product than the original monopolist had chosen to provide. In the single-product case, with a U-shaped average cost curve, the subadditivity assumption requires that the entrant be at least half the size of the original firm. Successful entry at a smaller scale would violate the subadditivity assumption, implying that we were not dealing initially with a natural monopoly. Only in the case where the monopolist cannot find (or is not allowed by the regulator to choose) a price vector satisfying the core conditions can a small-scale entrant be successful. Discussion of this last case is deferred until later in the article.

With a large-scale entrant it is reasonable to ask whether the regulator would or should disallow a price reaction by the monopolist when entry occurs. The costs of disallowing monopoly reaction can be enormous, quite apart from sustainability issues. Suppose an entrant can produce the monopolist's output at a cost only slightly less than that of the monopolist. If we allow entry and disallow a monopoly price reaction, the total cost to society of obtaining (approximately) the monopolist's original output will be the entrant's total costs plus the monopolist's fixed costs. Also, given the general dissatisfaction with industry performance in markets where regulators have restricted price reductions by firms of roughly equal size, such as airlines, it is at least worth taking a look at what happens in the sustainability problem when the monopolist is allowed to react to a large-scale entrant.

With regard to the first reason given in the quote above, there is something inherently incongruous about an entrant small enough to be ignored and large enough to drive the monopolist from one of his markets. Admittedly, the two-product case exaggerates the effect of entry, since in that case successful entry means that the monopoly disappears. However, even in the n-product case, where the effect of entry in any one market may only be to force the monopolist to offer a smaller product set, the assumption that the monopolist will not react is noticeably at variance with the behavior of real-world regulated monopolies.

What I propose to examine is the short-run profit-maximizing response of the monopolist to attempted entry. It will turn out, not surprisingly, that the range of situations under which monopoly production is sustainable is enlarged substantially. As in the Panzar-Willig model, sustainable natural monopoly production can be viewed as a Nash equilibrium, although in this case it is the monopolist's reaction function and not his price against which the entrant holds the Cournot-Nash expectation. The monopolist's reaction function will be derived by solving for his profit-maximizing (loss-minimizing) price and quantity choices, given the price and quantity choices of the entrant.

To examine the short run aspects of the problem we need to specify in more detail the monopolist's cost function. Let the cost function be given by

$$C(y_1^{\mathrm{m}},y_2^{\mathrm{m}}) = \min_{F,F_1,F_2,V,V_1,V_2} C(F,F_1,F_2,V,V_1,V_2;y_1^{\mathrm{m}},y_2^{\mathrm{m}}, \text{input prices}), \tag{14}$$

where F and V are the fixed and variable inputs that are common costs for both services and F_i and V_i are the fixed and variable inputs used specifically in providing service i. The incremental, or attributable, cost of providing service i is denoted by IC^i.

In the short run, the monopolist finds himself already having chosen F, F_1, F_2, and V.[10] We retain the assumption that $p_2^{\mathrm{m}}y_2^{\mathrm{m}} < C(y_2^{\mathrm{m}})$. However, given the single-product scale economies, the entrant can contemplate offering service 2 at a price lower than the monopolist's current price. Since the entrant must exploit economies of scale in order to enter, he is also offering to supply a larger quantity than that provided by the monopolist. Thus, if the monopolist is to supply any of service 2 at all, he must at least match the entrant's price.[11]

The monopolist's profit-maximization problem[12] is given by

$$\underset{p_1^m, p_2^m, y_1^m, y_2^m}{\text{maximize}} \quad \pi^m = \sum_{i=1}^{2} [p_i^m y_i^m - IC^i(y_i^m)] - \text{other costs}$$

$$(p_2^e - p_2^m \geq 0)$$

and

$$Q^i(p_1^m, p_2^m) - y_i^m \geq 0 \qquad (i = 1,2).$$

The second of the above constraints indicates that the monopolist's output choice is limited to outputs that can be sold. The Lagrangian for the optimization problem is

$$\mathscr{L} = \pi^m + \sum_{i=1}^{2} \lambda_i [Q^i(p_1^m, p_2^m) - y_i^m] + \mu(p_2^e - p_2^m). \tag{15}$$

The Kuhn-Tucker necessary conditions for an interior solution are

$$\frac{\partial \mathscr{L}}{\partial p_2^m} = y_2^m + \lambda_1 \frac{\partial Q^1}{\partial p_2^m} + \lambda_2 \frac{\partial Q^2}{\partial p_2^m} - \mu = 0, \tag{16}$$

$$\frac{\partial \mathscr{L}}{\partial p_1^m} = y_1^m + \lambda_1 \frac{\partial Q^1}{\partial p_1^m} + \lambda_2 \frac{\partial Q^2}{\partial p_1^m} = 0, \tag{17}$$

$$\frac{\partial \mathscr{L}}{\partial y_1^m} = p_1^m - \frac{\partial IC^1}{\partial y_1^m} - \lambda_1 = 0, \tag{18}$$

$$\frac{\partial \mathscr{L}}{\partial y_2^m} = p_2^m - \frac{\partial IC^2}{\partial y_2^m} - \lambda_2 = 0, \tag{19}$$

$$\frac{\partial \mathscr{L}}{\partial \lambda_i} = Q^i - y_i^m \geq 0, \qquad \lambda_i \geq 0, \qquad \lambda_i(Q^i - y_i^m) = 0, \tag{20}$$

$$\frac{\partial \mathscr{L}}{\partial \mu} = p_2^e - p_2^m \geq 0, \qquad \mu \geq 0, \qquad \mu_i(p_2^e - p_2^m) = 0. \tag{21}$$

Successful entry requires $\lambda_2 = 0$; otherwise $Q^2 - y_2^m = 0$ (the monopolist serves the entire market). We then know immediately from equation (19) that the monopolist will set price equal to marginal cost in market 2.

With $\lambda_2 = 0$, equation (17) becomes

$$y_1^m + \lambda_1 \frac{\partial Q^1}{\partial p_1^m} = 0. \tag{22}$$

Since $\partial Q^1/\partial p_1^m < 0$ and $y_1^m > 0$ at an interior solution, equation (22) requires $\lambda_1 > 0$, which carries the unsurprising implication that in a market unthreatened by entry the monopolist supplies the entire available demand.

Equation (13) can be rewritten as

$$\mu = y_2^m + \lambda_1 \frac{\partial Q^1}{\partial p_2^m} > 0, \tag{23}$$

which implies, from (21), that the monopolist sets $p_2^m = p_2^e$.

Equations (19) and (21) taken together then yield

$$p_2^m = \frac{\partial \mathrm{IC}^2}{\partial y_2^m} = p_2^e. \tag{24}$$

The monopolist's short-run profit-maximizing response to entry is thus to match the entrant's price and choose the output level at which short-run marginal cost equals that price.

Of course, in addition to the Kuhn-Tucker necessary conditions, the monopolist's price and output choices must satisfy the standard loss-minimization conditions for continued production:

$$p_i^m y_i^m \geq p_{V_i} V_i \qquad (i = 1,2)$$

and

$$\sum_{i=1}^{2} p_i^m y_i^m \geq p_V V + p_{V_1} V_1 + p_{V_2} V_2,$$

where the p_Vs are the variable input prices.

The effect of the monopolist's short-run response to entry is illustrated in figure 7.10. We begin with the monopolist choosing an initial vector of outputs and prices. The point M, with $\hat{p}_2^m$ and $\hat{y}_2^m$, is assumed to represent the solution in market 2 to that problem. The choice of $\hat{y}_2^m$ and the inputs necessary to produce it also allow us to specify the short-run average-incremental-cost ($\mathrm{AIC_{sr}}$), average-variable-incremental-cost ($\mathrm{AVIC_{sr}}$), and marginal-cost ($\mathrm{MC_{sr}}$) curves shown in the figure.

To position the demand curve for service 2 requires some statement

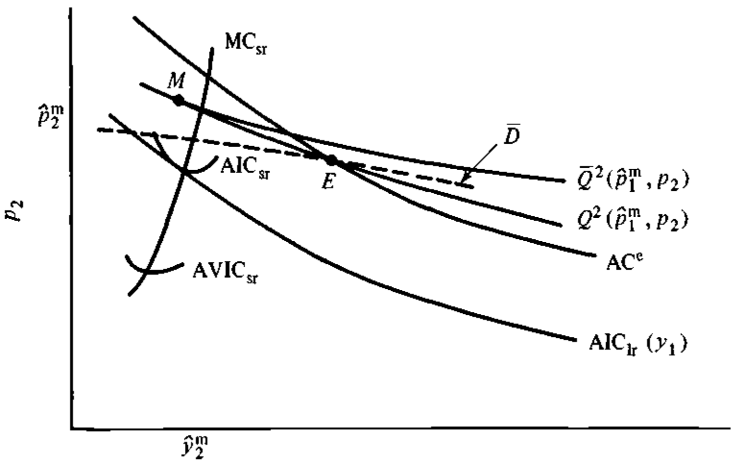


Figure 7.10 Monopolist's short-run response to entry.

about the monopolist's postentry choice of p_1^m. In general, the sign of dp_1^m/dp_2 is indeterminate.[13] The monopolist, when faced with announced entry at p_2^e, could choose to raise, lower, or hold constant p_1^m. To make things as simple as possible, I will assume that p_1^m is held constant at $\hat{p}_1^m$. Therefore, the demand curve for service 2 will not shift as a result of entry at p_2^e.[14] By the Panzar-Willig criteria, successful entry would be observed whenever the market demand curve lay above the entrant's average-cost curve. For example, at E in figure 7.10 successful entry would be feasible.

Successful entry, given the monopolist's short-run reaction, requires that the residual demand available to the entrant lie someplace above his average-cost curve. The residual demand is given by the horizontal distance between the market demand curve and the monopolist's short-run marginal-cost curve.[15] The $\bar{Q}^2$ market demand curve allows successful entry at E. The residual demand available to the entrant is shown as the dashed line $\bar{D}$.

Given successful entry in market 2, one can easily verify that, even with reaction by the monopolist, sustainability is not ensured and entry may reduce welfare. All one has to do when constructing the demand and cost curves for market 1 is to make sure the long-run AIC^1 curve lies well out from the origin and the postentry demand curve is tucked just inside it, thereby ensuring the long-run infeasibility of continued service in market 1 and the loss of large amounts of surplus.

Thus, even though the monopolist's short-run profit-maximizing response to entry increases the likelihood of sustainability (because larger unserved market demands are required for successful entry), that response by no means ensures sustainability. The troublesome policy issues concerning open entry remain unanswered, at least on an *a priori* basis.

Sustainability with Monopoly Reaction and One Product

The monopolist's reaction to entry in the one-product case is less obvious. Recall that for a single-product natural monopoly to be unsustainable the average-cost curve must turn up. This implies that the entrant will be offering to supply an output less than the monopolist's original output (Panzar and Willig refer to this as "cooperative" entry). Unfortunately, in this case we will not be able to derive the monopolist's reaction function from the new firm's entry "offer." We were able to do so in the preceding section because the entrant had to offer to supply an output larger than the joint monopolist's original output. The monopolist then took the entrant's price as given and chose that quantity which equated his marginal cost to the entrant's announced price. With cooperative entry, we (and the entrant) must make an assumption about the monopolist's reaction.

Suppose the monopolist chooses to react along his short-run marginal cost curve by, again, accepting the entrant's price and setting quantity so as to equate his marginal cost to the entrant's price. This reaction will

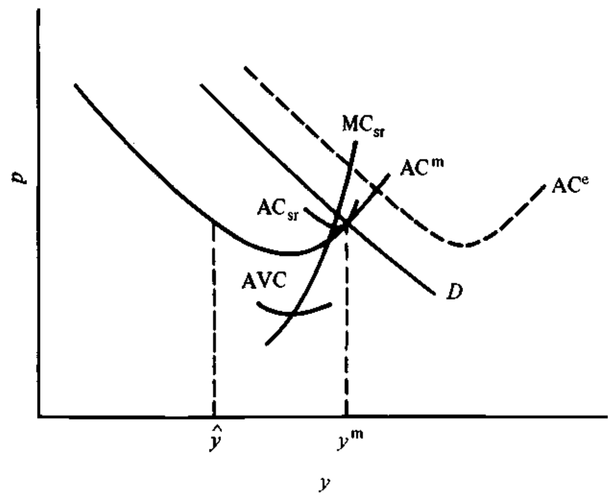


Figure 7.11 Monopolist's reaction ensures sustainability of a single-product natural monopoly.

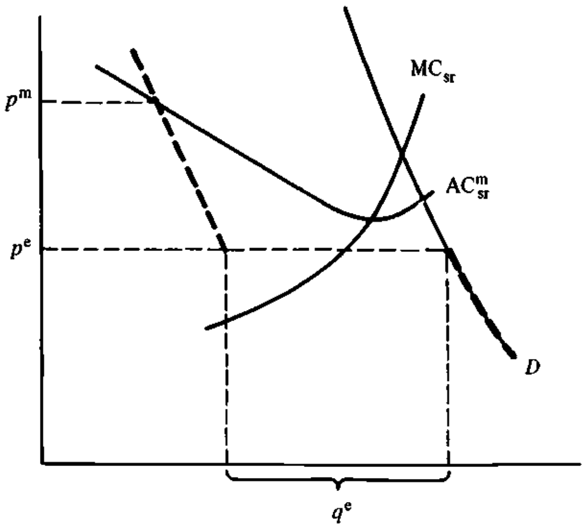


Figure 7.12 Monopoly reaction in the one-product case.

deter entry if costs are subadditive over the market range (and not just up to the monopolist's output) and if the monopolist's minimum average variable cost is less than his minimum average cost. Figure 7.11 shows such a situation. We can check for subadditivity by constructing a new average-cost curve, denoted AC^e, from a vertical axis at $\hat{y}$, the output at which average cost is equal to the average cost at y^m. As long as the demand curve lies below AC^e, costs are subadditive over the market range. Monopoly reaction along the short-run marginal-cost curve then deters entry by leaving the entrant a residual demand (D − MC$_{sr}$) that is less than $D - \hat{y}$.

Reaction along the short-run marginal-cost curve is not necessarily the monopolist's profit-maximizing response to entry. To see this examine figure 7.12, which shows the demand curve (the discontinuous dashed line) facing the monopolist when an entrant offers to sell q^e units at price p^e. If the monopolist matches or beats p^e, the market demand is his. His profit-maximizing choice is, then, to match p^e and choose quantity as described above. However, an alternative strategy that might yield higher profits (or lower losses) would be to set p^m above p^e and allow the entrant to sell q^e units. This price increase, coupled with a substantial reduction in output, could allow the original firm to break even. The entrant would of course raise his price to match p^m; but if he cannot increase output much past q^e in the short run, this strategy might work for the monopolist.

I suspect that this result is somewhat unlikely and, in any event, clearly undesirable. The regulator could effectively block it by not allowing the monopolist to decrease output and increase price in response to entry. Such a policy would force the monopolist to react along his short-run marginal-cost function. The sustainability of socially desirable natural monopolies would then be ensured as long as p^e remained greater than the monopolist's minimum average variable cost.

Sustainability and Entry Restrictions

We have seen that the Panzar-Willig approach to sustainability problems can be extended to allow consideration of a monopolist's short-run profit-maximizing response to entry. Those authors mention that further insight into sustainability could, and should, be obtained through a dynamic version of the model and through consideration of nonlinear prices.[16]

The range of policy options when sustainability is a problem can also be expanded. The sustainability problem is "merely" another case where market demand is insufficient to support continued production.[17] One policy solution to that problem is to impose entry restrictions or minimum price regulation in other markets, so that sufficient excess revenues can be generated to pay for production in the unsustainable market.

Another policy option is the imposition of a set of taxes and subsidies. A subsidy, defined either as a certain amount per unit of output or as a percentage of revenues, could be offered to any firm willing to serve the otherwise unsustainable market. Presumably, given the economies of joint production, the market will be served by our joint monopolist. The revenues to finance this subsidy would come from a set of taxes, defined in a similar manner, on the remaining products in the industry. Competition for these remaining markets is thereby feasible, with the lowest-cost producer serving each market. If the remaining markets do in fact make up a natural monopoly, then it can be preserved without resort to entry restrictions. Of course, running such a program would have its costs. However, there are also social costs—which could be very large—to protecting an inefficient joint monopolist behind a wall of entry restrictions.

Sustainability in Intercity Telecommunications: A Short Case Study

The particularly vexatious aspect of the sustainability problem is the fact that optimal resolution of the public-policy issues may require knowledge of the behavior of demand and cost functions outside the range of observed price and output choices. This would certainly be true if one tried to assess whether large-scale entry were feasible in any market, since the answer to that question depends on the behavior of demand and cost cruves beyond the range of current observations.

Fortunately, we may often be able to say something about what happens in markets where, if entry restrictions were to be lifted, production might be infeasible, We will often have observations on the long-run average-cost cruve in such markets. If forgone scale economies are "small" as output is reduced, then sustainability problems cannot be expected to arise. Either production remains feasible over widely varying outputs (in which case a natural monopoly will be sustainable) or, if natural monopoly is unsustainable, the forgone surplus in the "lost" market will tend to be "small." Thus, the welfare calculation, in which these losses must be balanced against the welfare gains in the market where entry occurs, will tend to favor the postentry world.

An argument along these lines can be put forth in the area of intercity telecommunications. Waverman (1975) argued convincingly that economies of scale for microwave transmission are exhausted at output levels well below the observed outputs of many telecommunications links.[18] In addition, our observation of the market responses of many commercial telecommunications users to the FCC's "above 890" decision[19] verify that microwave scale economies are small enough so that large individual commercial users, given appropriate prices, can "go it alone" with private microwave.[20] Finally, the demand for private line service is a derived demand for telecommunications as an input. Since telecommunications expenditures are typically a small fraction of total cost for a firm, one would expect the demand for private line service to be quite price-inelastic. Combined, these factors suggest that any forgone surplus would be small if open entry in public message-toll service rendered unsustainable the continued provision of private line service.

Even without monopoly reaction, the increased surplus accompanying successful entry in message-toll service would be quite large. For example, if the scale elasticity for independent production were 1.3, the initial MTS price advantage over independent production were 10 percent, and the demand function were linear, increased surplus in MTS would be half the product of the 10 percent price reduction and an output increase of one-third. With monopoly reaction, the surplus would be larger, since the market demand would have to be larger in order to accommodate the entrant. If the monopolist's short-run marginal-cost curve were roughly vertical, an output increase of 133 percent would be required to overcome the initial 10 percent monopoly advantage.

To carry this multiproduct interpretation of each telecommunications link one step further, one can ask why the cost function might be subadditive. The most plausible reason is economies of scale in transmission. The various services share the same transmission facilities.[21] For a given level of transmission capacity, the joint monopolist, especially at the *ex ante* planning stage,[22] can freely alter the mix between private line and message toll service without forgoing transmission scale economies. Thus, as the product mix is varied, the average incremental cost of transmission capacity for each service will not change. Therefore, the joint monopolist will not, for reasons arising from the transmission technology, run into a declining average-incremental-cost curve when performing the *ex ante* adjustment in output mix necessary to deter entry. This implies that sustainability problems should never arise for private line service, since the joint monopolist is offering message toll service in addition to private line service and thus will always be "farther down" the transmission average-cost curve than any potential entrant offering only private line service.

Sustainability problems could conceivably arise in message toll service, since, in addition to transmission, a switching technology must be employed in order to route each call to the desired destination. The threat of entry in private line service may force the joint monopolist to reduce private-line prices. Some users will then shift at least part of their demands from MTS to private-line service. If there are economies of scale in switching, this demand shift away from MTS will force the joint monopolist "backward and up" along a declining average-incremental-cost curve for message toll service. In a qualitative sense, at least, the stage has been set for the possible infeasibility of continued provision of message toll service.

Such an outcome is not to be expected. A very substantial body of message-toll users would not contemplate using private lines at any conceivable price for private-line service as long as message-toll service was available at current prices. The move to private line requires a large enough demand along any one route to cover the costs of a dedicated circuit. Residential toll users, with small individual demands along any one route, will prefer MTS to private-line service, even given drastic price reductions for private line, if a switching technology is available which allows the pooling of their individual demands at prices close to current MTS prices. We can infer that such a technology is available from the variety of switched private-line services currently offered, at small scales of production, by the specialized common carriers.

Of particular relevance is MCI's "Execunet" service, which is a very close substitute for AT&T's message-toll service. Execunet is a switched service, and an Execunet subscriber can place a call to any phone on the public network in any city served by MCI. Execunet is offered at a scale that is minuscule in comparison with the scale realized by AT&T for message-toll service on the same routes.[23] I do not have market-share figures for each firm on a route-by-route basis, but the national totals give some idea of relative magnitudes. In the fiscal year ending March 31, 1977, MCI's Execunet sales were $28 million (*Telecommunications Reports* 1977). AT&T's message-toll revenues in 1975 were $12.9 billion, of which WATS accounted for $1.4 billion (FCC 1975). Admittedly, these figures overstate the relative size of AT&T, since MCI does not offer service on most routes served by AT&T. However, MCI does serve most high-density routes, and even if one assumes that AT&T derives only 10 percent of its revenues from routes also served by MCI, MCI's market share is only 2.1 percent. In addition, at this point the example is biased against the probability that message-toll service can be provided at a small scale by a joint monopolist also offering private-line service, since MCI's private-line volume is but a tiny fraction of AT&T's. Thus, MCI offers Execunet even though, by hypothesis,[24] it must forgo scale economies in both transmission and switching. The joint monopolist facing entry in private-line service would be forgoing only whatever scale economies there are in switching as he attempted to market both services.

An alternative view of the intercity telecommunications market definition is that only one product—transmission capacity—is being provided. The observed rate differentials for a circuit between A and B are then viewed as evidence of price discrimination. In this one-product world, nonincreasing average cost is then sufficient for the sustainability of each link. Scale elasticities greater than unity have been found in virtually every empirical investigation of telecommunications, so the assumption is certainly defensible. However, even if subadditivity holds but nonincreasing average cost does not, natural monopoly will generally be sustainable if the monopolist reacts along his short-run marginal-cost curve.

If each link is sustainable, can sustainability problems arise across links? The answer would appear to be no as long as the costs of serving a collection of links are subadditive, as long as weak cost complementarities hold across links (that is, the marginal cost of serving link i does not increase with increased output on link j), and as long as market revenues are in the core. The demands for service along different links are arguably independent. The demand is for service between A and B; the consumer is indifferent to the prices charged on other routes, since a call to C cannot be substituted for the desired call to B.

Under these assumptions, proposition 3 of Panzar and Willig 1977a applies and the natural monopoly network will be sustainable. Since each link considered alone is sustainable, successful entry on any one link would require that the lower post entry price pull in consumers from other links. Independent demand implies that consumers will not move across links. Thus, the natural monopoly network will be sustainable.

We are left with what could be a source of sustainability problems in telecommunications. AT&T currently averages its message-toll rates, with prices on high-density routes set above "cost" and prices on low-density routes set below "cost." It may be that AT&T, in attempting to restructure its rates when faced with an open-entry rule, would be unable to find an initial price vector satisfying the core conditions. In other words, consumers on low-density routes might be unwilling to pay prices high enough to cover the average incremental cost of being served. Whether this is a realistic possibility I cannot say, since the size of the subsidies arising from nationwide rate averaging is unknown. The size of those subsidies will depend on the "true" scale elasticity for telecommunications and on the demand mix of toll users along low-density routes. The latter factor is important because a toll user along a low-density route could experience lower toll prices under an open-entry rule if a large enough fraction of his toll calls were to be carried over high-density routes.

If toll service in some rural areas proves to be either infeasible or much more expensive under an open-entry rule, then some source of subsidy for those consumers may be desirable. As I indicated earlier, there are ways of providing that subsidy while still allowing entry.

The views expressed are those of the author and should not be taken as representing official views or positions of the U.S. Department of Justice. John Hoven, Richard Ippolito, Bruce Owen, Robert Reynolds, James Rosse, and Lee Sparling provided helpful comments at various stages in my research on these issues. The author is solely responsible for any errors.

Notes

1. MTS: message-toll service. WATS: wide-area telephone service. AT&T no longer has a *de jure* monopoly on MTS. The D.C. Court of Appeals overturned the FCC's Execunet

decision. Execunet is a close substitute for MTS offered by MCI. The FCC had disallowed MCI's Execunet tariff on the grounds that MCI had not been authorized to offer public switched service.

2. Before the FCC, in docket 18128, AT&T proposed as a pricing standard the "Baumol burden test," which is an extension of the simple incremental-cost test. The "burden test" shares the deficiencies of the incremental-cost test.

3. The convexity of the isoquants ensures that the second-order necessary and sufficient conditions for the monopolists's cost-minimization problem are satisfied.

4. I assume that if $p_1^m = p_1^*$ the joint monopolist will satisfy the entire demand. This assumption is made for convenience only. Alternatively, I could require $p_1^m < p_1^* - \varepsilon$ and carry the ε throughout the analysis.

5. p_1^* is the entry-deterring price in a "friction-free" world without entry barriers. With entry barriers, the entry-deterring price will exceed p_1^*.

6. When equation (8) is negative the opposite strategy of increasing p_2^m will not aid the monopolist, since then p_1^m must increase in the face of competition in market 1.

7. The assumption of subadditivity is sufficient to guarantee a nonempty core in the two-player game. With more than two players, weak cost complementarities are required.

8. The distinction between the two sources of unsustainability becomes blurred when consideration is given to the monopolist's reaction to entry. In the Panzar-Willig world, if the second condition is not satisfied natural monopoly can be displaced by local price changes. With monopoly reaction a large price cut is needed for successful entry whether or not the initial prices are in the core.

9. Alternatively, AIC_j need not shift down if $C_{ij} = 0$ over the range of output changes being considered. In that case, the source of the economics of joint production is a savings in the inframarginal fixed or startup costs when two services are produced jointly rather than separately. Thus, although the increasing output of service i may dictate larger fixed expenditures by the monopolist, the increased use of fixed inputs generates no additional cost complementarities.

10. To make things simple I am assuming that F and V come in fixed proportions. This is the case if a fixed maintenance program must be followed to keep a unit of F operating efficiently.

11. Again, I am assuming that at $p_2^m = p_2^e$ the monopolist can supply as much of good 2 as he chooses. The alternative, requiring $p_2^m \leq p_2^e - \varepsilon$, yields equivalent results but is less convenient.

12. The monopolist's problem and solution are precisely the same as the Panzar-Willig "cooperative" entrant's solution to the problem of what prices and outputs to choose if entry is considered at less than the monopolist's existing output and the monopolist's average-cost curve is declining.

13. dp_1^m/dp_2 is obtained by totally differentiating equation (17) and solving for dp_1^m/dp_2.

14. $p_1^m \gtrless p_1^m$ would imply $Q^2(p_1^m, p_2^e) \gtrless Q^2(\hat{p}_1^m, p_2^e)$.

15. For prices less than the monopolist's minimum AVIC, the entire market demand is available to the entrant.

16. I have looked at the sustainability problem with nonlinear prices, but have been unable to generate any clear-cut results. It is clear that some price discrimination scheme would, *ceteris paribus*, increase the likelihood of markets, unthreatened by entry, being served by the monopolist. However, the effects of nonlinear pricing strategies are unclear in markets where entry is threatened. The monopolist, by appropriate use of declining

block-rate tariffs, can inch down the demand curve, leaving a smaller residual demand for the entrant. However, once the monopolist has positioned himself, the entrant now has the more powerful tools of price discrimination to get at the available residual demand. The net effects on the likelihood of entry (and the probability of sustainability problems) are ambiguous.

17. Regulators now deal routinely with such situations. An example is subsidized air service on low-density routes.

18. Waverman's work indicates that microwave scale economies are exhausted somewhere between 1,000 and 4,000 voice circuits, with a cost premium of only a few percent at around 600 circuits. A 100-circuit system could be constructed at a cost premium of about 60 percent. By way of comparison, AT&T's L4 cable system has a capacity of 32,400 circuits.

19. The "above 890" decision allowed telecommunications users, for the first time, to build and operate their own microwave communications systems.

20. AT&T's response to the "above 890" decision was a substantial price reduction (the "Telpak" tariffs) to large-volume private-line customers. As a result, there has been very little investment in private microwave systems beyond that observed before the Telpak tariffs became effective.

21. If the various services do not share transmission facilities, as would be the case with physically separate analog and digital transmission systems, then one must search elsewhere for a source of subadditivity.

22. The substitution can also occur after the system is in place. AT&T currently transfers existing circuits back and forth between message-toll and private-line service.

23. MCI's operation at a small scale is not inconsistent with scale economies in telecommunications. AT&T's rates may be set above "cost" on the high-density routes over which the firms compete.

24. That is, here we are assuming, for the sake of argument, that the "cream skimming" explanation of MCI's entry is valid. An alternative view, held by MCI and other specialized common carriers, is that their entry can be defended because their costs, even at very low scale, are lower than AT&T's.

References

Areeda, Phillip, and Turner, Donald F. 1975. "Predatory Pricing and Related Practices under Section 2 of the Sherman Act." *Harvard Law Review* 88: 697–733.

Averch, Harvey, and Johnson, L. L. 1962. "Behavior of the Firm under Regulatory Constraint." *American Economic Review* 52: 1053–1069.

Baumol, William, Bailey, Elizabeth, and Willig, Robert. 1977. "Weak Invisible Hand Theorems on the Sustainability of Multiproduct Monopoly." *American Economic Review* 67: 809–822.

Faulhaber, Gerald. 1975a. "Cross-Subsidization: Pricing in Public Enterprise." *American Economic Review* 65: 966–977.

———. 1975b. Increasing Returns to Scale: Optimality and Equilibrium. Ph.D. diss., Princeton University.

Federal Communications Commission. 1976. Recommended decision, docket 18128.

———. 1975. *Common Carrier Statistics.*

———. 1976. Report of domestic telecommunications policies, docket 20003.

Kahn, Alfred E. 1970. *The Economics of Regulation*, vol. 1. New York: Wiley.

Panzar, John C., and Willig, Robert D. 1977a. "Free Entry and the Sustainability of Natural Monopoly." *Bell Journal of Economics* 8: 1–22.

———. 1977b. "Economies of Scale in Multi-Output Production." *Quarterly Journal of Economics* 91: 481–494.

Scherer, F. M. 1977. "Predatory Pricing and the Sherman Act: A Comment." *Harvard Law Review* 89: 869–890.

Seneca, Rosalyn. 1973. "Inherent Advantage, Costs and Resource Allocation in the Transportation Industry." *American Economic Review* 63: 945–956.

Spence, Michael. 1977. "Entry, Capacity, Investment and Oligopolistic Pricing. *Bell Journal of Economics* 8: 534–544.

Telecommunications Reports 43 (no. 25): 13.

Waverman, Leonard. 1975. "The Regulation of Intercity Telecommunications." In Almarin Phillips (ed.), *Promoting Competition in Regulated Markets*. Washington, D.C.: Brookings Institution.

Comment

William J. Baumol

My comments are devoted primarily to the first portion of Kenneth Baseman's illuminating article—the portion dealing with appropriate tests of cross-subsidization rather than that devoted to the issue of sustainability. This part of his article is entirely critical in character, seeking to demonstrate weaknesses in the incremental-cost and burden-test approaches to the construction of criteria of cross-subsidization. As a general comment, one might have wished that the author had gone beyond mere criticism, and that he had offered some counterproposals—tests that in his view are more effective than those to which he objects. Such an attempt to approach issues more constructively is, of course, always to be encouraged.

In my view, however, an alternative criterion of cross-subsidization may not be necessitated by the paper's criticisms. This is not meant to deny that the analysis has brought to our attention a substantial policy issue. Indeed, the issue implicitly raised has probably been important throughout economic history. That issue is governmental grant of monopoly in cases where it cannot be justified by superior performance—a topic that has loomed large in the economic literature since the publication of *The Wealth of Nations.* What I will undertake to show is that it is this issue, and *not* the adequacy of the proposed tests of cross-subsidization, with which Baseman's analysis deals. Rather, I will show that in the circumstances envisioned in the paper the marginal-cost test of cross-subsidization performs exactly as one would wish it to, and that the problem the examples raise is that of a monopoly franchise unjustifiably granted.

While the paper presents a number of variants of its argument on cross-subsidization, I will deal only with the simplest example, since it brings out the issues so clearly.

In the example in question, Baseman examines a case in which there are diseconomies of scope—it is more expensive to have a single firm produce both of two commodities than to have the same quantities produced separately by two specialized firms. The monopolist's total cost of producing the two items together is 90, and the incremental cost of each item is only 20. Each item can be produced by a specialized competitor at a total cost of 40. The unspecialized firm has a monopoly franchise in commodity 2 and therefore fears no competition in this market. It sets a price of, say, 35 for the competitive product, good 1, thoroughly exceeding its incremental cost, and covers the remainder of its 90 cost by charging a price of 55 for its monopoly product, good 2, thereby depriving the consumers of the opportunity of buying the good from a specialized firm at 40. Clearly, the consumers of good 2 are burdened by the arrangement.

But the story, as told so far, obscures the true source of the burden. To make the issue clear, instead of considering only the case in which there is a monopoly franchise, let us examine two possibilities: that in which entry is prevented by law (the monopoly franchise) and that in which entry is free. In each case we contrast the results of an incremental-cost floor for competitive product 1 with those of a full-cost floor for that product, where for illustration it is assumed that consumers of each good must then be charged half their 90 total cost (each must pay 45). The results are summarized in the following table.

	Incremental-Cost Pricing		Full-Cost Pricing	
	P_1	P_2	P_1	P_2
Legal monopoly in market 2	35	55	45	70
Free entry in market 2	35	?	45	?

Consider first what happens if there is a legal monopoly franchise in good 2. Under the incremental-pricing regime, as before, the consumers of good 1 pay 35 and the consumers of good 2 pay the residue, 55. But with full-cost pricing the firm loses its market for good 1, and customers of the monopolized product will then have to pay both the fixed cost, 50, and the incremental cost, 20. That is, the monopoly product's cost will rise from 55 to 70, because of the loss of the contribution of 15 in excess of incremental cost which product 1 would otherwise have provided.

Now consider the case where entry is free. Then, if in order to meet competition in product 1 the unspecialized firm charges 35 for that product under an incremental-cost floor, it will simply be unable to maintain its market for product 2, for either it will charge 55 and lose that market to an entrant or it will charge less than 40 and go broke. Ultimately, the unspecialized firm will be driven from both markets, for it cannot survive serving only market 1 at a price of 35.

With entry, the same will happen under full-cost pricing. Indeed, what must ultimately happen with free entry is that the unspecialized firm will be unable to remain in either market, and that is just as the public welfare requires.

Arithmetic arguments can be confusing, yet it is easy to draw a number of general conclusions from this discussion:

- The consumer is obviously poorly served in the case illustrated in the Baseman paper, but this is because the firm has been granted a monopoly in a market where there is no cost justification for such a grant. Thus, under an incremental-cost regime customers are not as well off as they might otherwise be. But they are worse off still under full-cost pricing, which forces *both* customer groups to pay more than under incremental-cost pricing.
- If despite the absence of any cost justification the government is determined to impose a monopoly in market 2, then the issue is what pricing scheme minimizes the resulting damage. It is easy to see that it is the incremental pricing scheme.
- Under these circumstances it is simply wrong to say that incremental pricing introduces cross-subsidization by the customers of the monopoly product. Given the idiotic decision to impose the monopoly in market 2, under a ceiling on total company profits the customers of the monopoly product are better off when the competitive product 1 is sold at any price above incremental cost that is low enough to keep the competitive market for the unspecialized firm, for then the competitive-market sales make a net financial contribution to the company and consumer payments for the monopoly product must be reduced accordingly.
- Most important for the current discussion, pricing based on an incremental-cost test achieves the right results in either of the two cases, the case in which the monopoly is imposed in market 2 and the case where it is not. In the first case it saves money for both customer classes; in the second case it drives the unspecialized firm out of business, as it deserves because of its relative inefficiency.

I conclude from all this that tests of cross-subsidization can and must be incremental in character. To say that the sale of a product benefits from a cross-subsidy means that its sale is at the expense of the customers of other products. But if the product in question contributes net revenues in excess of its net incremental costs, then surely its market is not being served at the expense of the customers of other products, whether or not those other products have a protected monopoly.

Finally, I should like to emphasize the desirability of further work on both subjects raised by the discussion: criteria that can be used to identify which products do merit a monopoloy franchise, and further sharpening of the tests of cross-subsidization. With regard to the latter, while I believe firmly that appropriate tests will have to be based on a comparison of incremental revenues and costs, there is, as always, room for further improvement in the testing procedures that have so far been designed. I hope that the author of the article will turn his considerable abilities to these issues.

I thank the National Science Foundation for its support of the writing of these comments.

Comment

John C. Panzar

My comments will deal with the second half of Mr. Baseman's article, in which he discusses some alternative approaches to the theory of monopoly sustainability and some of the policy issues involved.

The author's exposition of the essence of the sustainability problem for the two-product case is intuitive and concise. I was particularly intrigued with his diagrammatic use of the average-incremental-cost curve to show why the monopolist may not be able to match the price cut of a single-product entrant and still break even (figure 7.7). This device is a useful complement to the Panzar-Willig price-space diagram of figure 7.9. The use of such a price-quantity diagram also makes clearer the structural similarity of the sustainability problem to the literature on entry deterrence. Since the bulk of the paper's analytics are devoted to Baseman's attempt to develop an alternative to the Nash behavioral assumption underlying the Panzar-Willig analysis of sustainability, it may be worthwhile to develop a little perspective on this issue.

Anytime one attempts to model a potential entrant's decision problem, it is necessary to make some assumption about what the entrant perceives will be the response of the monopolist to his entry. The limit-pricing theory of Bain and Sylos, for example, was based upon the entrant's assuming that the monopolist would keep his output constant in the face of entry. The theoretical results of such models are, of course, quite sensitive to the behavioral assumptions employed, and debates in the literature tend to focus upon whether one assumption is more or less "reasonable" than another.

When discussing a regulated industry, it would seem that such debates are beside the point, since postentry price responses can be controlled by the regulator. In principle, allowable price responses can therefore be known objectively by all potential entrants, and the nature of such allowable responses becomes an important policy issue. In this light, it becomes evident that the policy tradeoff between the salutary stimulus that free entry may provide for efficiency and innovation and the possibility of wasteful, duplicating entry cannot be analyzed independently of regulatory price-response policy. Two extreme examples should help clarify this point:

- "Predatory pricing," if allowed as a response to entry, may deter many entrants who in fact possess superior technology or products.
- Welcoming entrants into the regulatory "fold" through favorable price adjustments will certainly lead to wasteful entry. (In Baseman's figure 7.8, even the point S may not be sustainable if the entrant perceives that the regulator will allow it and the original monopolist to "come to rest" at the southwest corner of R.)

The sustainability discussion of Panzar and Willig, then, is best interpreted as an analysis of the potential problems with a policy of free entry *given* that the regulator does not allow a price response—the simplest such regulatory policy to analyze. It demonstrated that there may indeed be serious problems with a policy of open entry under that condition. Therefore, an important direction for additional research is to examine open entry and sustainability under meaningful alternative assumptions about the responses to entry the regulatory will allow the monopolist to make. The goal of such analyses would, of course, be to determine whether or not a proposed policy tends to make sustainability problems more or less severe. While I believe Baseman's model to be a step in this right direction, I do have a few comments to make about his analysis and choice of regulatory scenario.

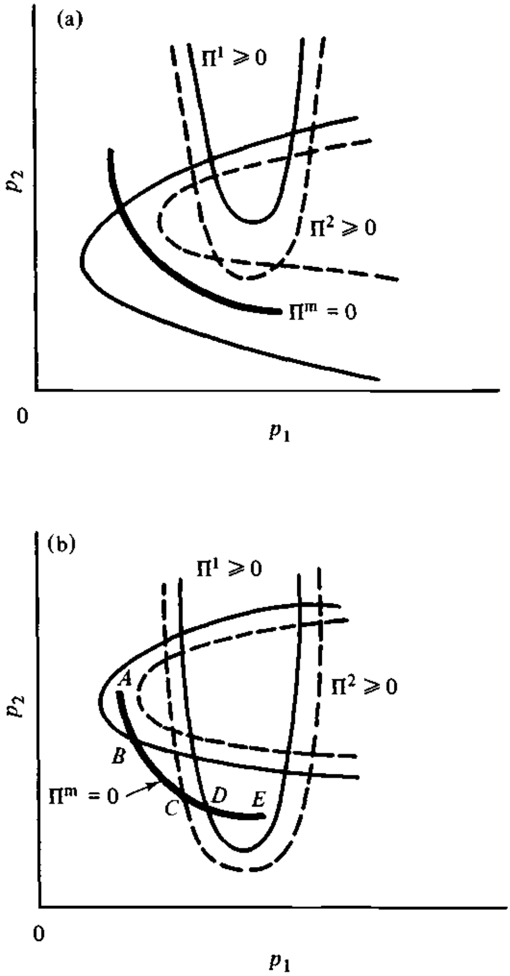
Baseman assumes that the entrant expects the regulator to allow the monopolist to respond to entry in a manner that maximizes his profits in the short run. In the light of the above discussion, is this an interesting policy alternative to analyze? In the first place, it is difficult to reconcile the short-run behavior of the monopolist with the (presumably) long-run concerns of the entrant. Perhaps the entrant is only interested in earning some short-run profit, but unless for some unexplained reason the technology he uses is extremely fungible he must also consider the possibility of losing some of his investment when the monopolist makes his long-run response. Put another way, what is the regulator's policy to be toward the (initial) monopolist in the long run? Will he prevent the monopolist from adjusting his input choices indefinitely? Baseman is not at all clear on these points.

Second, would it ever make sense for the regulator to adopt such a policy? Baseman raises this question in connection with the no-response scenario, but offers little in the way of justification for his own model in this regard. Presumably, the fact that regulation is effective means that there are unexploited profits to be had, even in the short run. This might induce some form of collusion between the monopolist and potential entrants, since tolerating some entry in market 2 may provide the opportunity to reap large profits in market 1.

Finally, Baseman's basic conclusion that "the monopolist's short-run profit-maximizing response to entry increases the likelihood of sustainability (because larger unserved market demands are required for successful entry)" depends upon the assumption that the monopolist does not choose to raise the price in the unthreatened market. If he did, the market demand for service 2 might shift out far enough so that even the entrant's residual demand would exceed the original (constant p_1) market demand. In view of this difficulty and the above discussion about unexploited profits, it may be more interesting to analyze the model by explicity assuming that the regulator allows the monopolist to change only the price in the service threatened by entry. Unfortunately, if the monopoly *is* unsustainable, it is easy to show that the monopolist cannot lose part of market 2 and break even in the long run without raising the price in market 1. While hardly conclusive, Baseman's discussion has indicated the usefulness of modeling alternative price responses to entry in a sustainability framework.

Another potential policy option for dealing with the sustainability problem that Baseman mentions only briefly involves taxes and subsidies. I find this concept quite intriguing analytically, so let me briefly elaborate. Consider the unsustainable situation depicted in figure 7.9. If it is required that any firm serving market 2 pay a (lump sum) penalty and that an equal bonus be paid to any firm offering to serve market 1, the $\Pi^2 \geq 0$ set will "shrink" and the $\Pi^1 \geq 0$ set will expand, without affecting the $\Pi^m = 0$ locus. (Baseman discusses per-unit or *ad valorum* taxes and subsidies, which, while they may be easier to administer, introduce an additional price distortion into the analysis.) It is at least possible that the situation of figure 7.9 may thereby be "transformed" into that of figure 7.8, which would lead to the emergence of a region of sustainable prices. The same scheme could also be used, in principle, to allow movement from one sustainable price vector to a socially preferred, but previously unsustainable, set of prices (for example, in figure 7.8, from *S* to a point to the left of *A*). These possibilities are illustrated in the accompanying diagrams. In (a), there are initially no sustainable price vectors. However, after the imposition of an appropriate tax-and-subsidy scheme has shifted the $\Pi^1 = 0$ and $\Pi^2 = 0$ loci to the positions indicated by the dashed lines, all prices on the $\Pi^m = 0$ locus are sustainable. In (b), a tax-and-subsidy policy has shifted the sustainable region from BCD to ABC, allowing the regulator to choose a price vector between *A* and *B* if he finds it to be socially desirable.

I shall conclude with a few remarks about the policy implications of



Movement to previously unsustainable prices.

Baseman's analysis. Baseman's heuristic empirical argument suggests that sustainability is not likely to be a problem in intercity telecommunications. Even if his perceptions are correct, there is a great need for additional empirical work in this area on which to base intelligent policy analysis. The knowledge that there exist one or more sustainable price vectors is, in and of itself, of little use to the policymaker. The relevant question is whether or not there are any socially desirable prices which are sustainable. If not, the decisionmaker faces a tradeoff between the benefits resulting from a policy of open entry and the concomitant loss of pricing flexibility. The important insights to be gained from Baseman's paper are that the terms of this policy tradeoff may be improved through the implementation of appropriate tax-and-subsidy schemes and through alternative regulatory policies toward allowable postentry responses by the multiproduct monopolist.

The views expressed in this comment are solely those of the author and do not necessarily reflect those of Bell Laboratories or the Bell System.