Capacity Utilization under Alternative Regulatory Restraints: An Analysis of Taxi Markets
Author(s): Arthur S. De Vany
Source: *Journal of Political Economy*, Vol. 83, No. 1 (Feb., 1975), pp. 83-94
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/1833274

Capacity Utilization under Alternative Regulatory Restraints: An Analysis of Taxi Markets

Arthur S. De Vany

Texas A & M University

Both the Averch-Johnson (A-J) model and the Chamberlin model fail to consider the value of excess capacity to consumers. Service industries, whether they are regulated or not, will usually have excess capacity in the Chamberlinian sense because this capacity conserves time for consumers. This paper examines a model of the taxi industry where allowance is made for capacity to affect the value or quality of the service through its effect on waiting time. The central issue is to determine equilibrium output, capacity, and the utilization of capacity when the market is organized as a franchised monopoly, through a medallion system, and when there is free entry, and to exhibit the relationship among these variables and prices, cost of capacity and output, and policies of the regulator. It is found that many of the characteristics of taxi markets that would appear to confirm the monopolistic-competition thesis arise because of the nature of regulation of these markets.

I. Introduction

The effect of regulation on capacity utilization has been actively debated in recent years. Since the Averch-Johnson demonstration (1962), there has been some acceptance of the idea that regulation designed to secure some rate of return on capital of the firm will induce "excess" capacity. The point also has been made that free entry into markets having some degree of product differentiation results in "excess" capacity.[1] Casual examination of such markets as barbering, real estate sales, stock

I wish to acknowledge the valuable conversations I have had with Richmond Lloyd on the subject matter of this paper. Financial support was provided by the Texas A & M University Organized Research Fund.

[1] Chamberlin (1933) and Kaldor (1935) speak of excess capacity in terms of a rate of production below that which would yield minimum average cost. The A-J result would produce a similar situation with the overcapitalized firm operating where it experienced declining short-run average cost.

[*Journal of Political Economy*, 1975, vol. 83, no. 1]

brokerage, and taxicabs (in some cities) appears to support this proposition.[2] A recognized deficiency of these demonstrations is the failure to consider the value of capacity to consumers of the product.[3] That is, given any reasonable stochastic properties of demand, capacity serves to reduce waiting time and planning costs for consumers, and, therefore, it is not necessarily inefficient to have some unused capacity. This paper examines a model of an industry where allowance is made for capacity to affect the value or quality of the services received by consumers through its effect on waiting time. The vehicle for this examination is the taxicab industry, but the results obtained are generalizable to a number of service industries where waiting time is likely to be a function of the rate of utilization of each firm's capacity and the number of firms offering the service.

The central issue is to determine the industry response to regulation when the taxi market is organized as a franchised monopoly, when entry is limited through a medallion system, and when there is free entry. My interest is in finding equilibrium output, capacity, and utilization of capacity and in exhibiting the relationship between these variables and prices, the cost of capacity and output, and the policies of the regulator. It is found that a constant cost of quality gives rise to decreasing and then increasing cost of output. The monopolist produces where his marginal cost of output is equal to the regulated price, and this occurs on the nondecreasing portion of the average cost curve, even if profits are zero. Under free entry, each firm produces where marginal cost equals the regulated price at minimum average cost. Monopoly marginal cost is greater than marginal cost under free entry, and so there is less output, less capacity, and lower average cost of output under monopoly. Nonetheless, the monopoly outcome is generally inferior to the free-entry outcome, unless the regulated price is set well above the efficient level.

II. Demand Assumptions

Assume the industry demand function is

$$Q = Q(\pi, \pi_o, B),$$

where Q represents quantity of passenger trips by cab, π is the "full price" of a cab trip, π_o is an index of the full prices of all other goods,

[2] Samuelson (1973, p. 518), for example, suggests barbering is an industry that fulfills the predictions of the Chamberlin model. In this paper, I show that chronic excess capacity may be the result of regulation that prohibits price competition and sets price at an efficiently high level.

[3] Alchian and Allen (1969, pp. 152–57) offer an extensive discussion of the value of excess capacity in the form of empty barber chairs, hotel rooms, retailer inventories, and so on, to consumers having less-than-perfect information and scarce time.

and B is a demand-shift parameter. This is the Becker "full price" demand function[4] with the property that $\partial Q/\partial \pi < 0$. Let $\pi = p + vt(H)$, where p is the fare per cab hour, v is the value of time of passengers, and $t(H)$ is waiting time for a cab. The H represents cabs supplied simultaneously and I assume that $t' = \partial t/\partial H < 0$.[5] Differentiating the demand function gives the properties

$$\frac{\partial Q}{\partial p} = \frac{\partial Q}{\partial \pi} < 0, \qquad \frac{\partial Q}{\partial H} = \frac{\partial Q}{\partial \pi} vt' > 0.$$

Demand is assumed to be uniformly distributed across the day.

III. Monopoly Solution

Assume a single firm has a charter from a city council or some such agency granting it monopoly rights to pick up passengers within a market area, and that the fare is also set by this agency.[6] It is assumed that cabs can be added to operation at constant cost, so the cost function is

$$C = cH,$$

with c a parameter.[7]

The firm's program is to maximize

$$pQ[p + vt(H)] - cH$$

by choice of H subject to

$$p = p^o,$$
$$Q \leq H,$$

where p^o is the regulated fare. Equilibrium must occur in the interior of the capacity constraint, or else waiting time would be indefinitely long, implying an indefinitely high full price so long as time has a positive value. The solution H^* of

$$p \frac{\partial Q}{\partial H} - c = p \frac{\partial Q}{\partial \pi} vt' - c = 0 \tag{1}$$

[4] Becker (1965).

[5] Orr (1969) and Douglas (1972) adopt similar assumptions on the waiting-time distribution, but do not use the full price-demand function.

[6] This description closely characterizes the Los Angeles taxi market (see Eckert 1973).

[7] The cab's cost-minimizing operating period is assumed to span the demand period, and so variable cab hours is strictly equivalent to the number of cabs operated. No distinction is made between cruising or carrying a passenger in the cost function, since it is assumed that the operating cost of a cruising cab does not differ from the cost of a cab engaged in carrying a passenger.

will be a maximum-profit point provided that at that point

$$p \frac{\partial^2 Q}{\partial H^2} = p \left[\frac{\partial Q}{\partial \pi} vt'' + \frac{\partial^2 Q}{\partial \pi^2} (vt')^2 \right] < 0. \tag{2}$$

The monopolist operates where marginal revenue of cabs at the regulated price equals marginal cost of cabs, and marginal revenue is declining. By equation (2), these conditions will be fulfilled if the marginal reduction in waiting time due to cabs is declining ($t'' > 0$), and $\partial^2 Q/\partial \pi^2 \leq 0$, or when

$$\frac{\partial^2 Q}{\partial \pi^2} > 0, \qquad \text{and} \qquad \frac{\partial^2 Q}{\partial \pi^2} (vt')^2 < \left| \frac{\partial Q}{\partial \pi} vt'' \right| .$$

Suppose the demand function is of constant elasticity with respect to full price. Now, the price and cab elasticities of demand can be written

$$\beta = \frac{\partial Q}{\partial p} \frac{p}{Q} = \varepsilon \frac{p}{\pi}, \qquad e = \frac{\partial Q}{\partial H} \frac{H}{Q} = \varepsilon v t' \frac{H}{\pi},$$

where ε is the full price elasticity of demand. The consequences of this assumption are that as cab hours are increased, the elasticity of demand with respect to cabs falls, and the price elasticity of demand rises (absolutely). In the neighborhood of equilibrium, quality improvements increase price elasticity of demand and decrease quality elasticity of demand. I shall assume a constant full price elasticity of demand.

Using the definition of elasticity of demand with respect to cab hours, the monopolist's marginal revenue of cabs is

$$MR_H = p \frac{\partial Q}{\partial H} = pe \frac{Q}{H}, \tag{3}$$

and so $MR_H = AR_H$ when $e = 1$, $MR_H > AR_H$ when $e > 1$, and $MR_H < AR_H$ when $e < 1$. The firm operates where $e < 1$ if it earns positive profits; $e = 1$ is a zero-profit point, and $e > 1$ is a negative-profit point. The monopolist operates where the occupancy rate of its cabs declines with further increases in its supply of cabs.

Equilibrium is depicted diagrammatically in figure 1. The long-run average and marginal cost curve of cab hours is $AC_H = MC_H$, and marginal and average revenue of cab hours are MR_H and AR_H at the regulated price. Equilibrium hours are H^m.

These relationships are particularly interesting when translated from cab hours into passenger trips. What is revealing about this translation is that the supply of cab hours H may now be regarded as a dimension of quality of output. Consider first the relationship between H and the demand curve. An increase in H reduces time cost and increases the quantity demanded at each price—the extent of this increase depends on

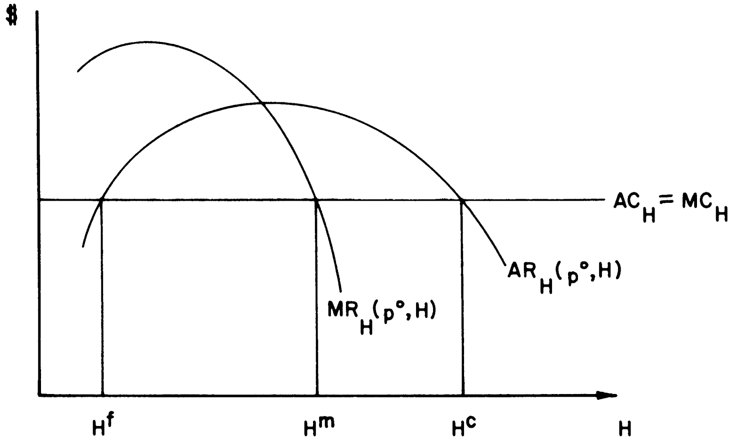

FIG. 1.—Equilibrium cab hours at p^{o}

the value of time, the reduction in waiting time achieved, and the full price elasticity of demand. Under the assumptions behind the AR_H curve, the demand curve shifts more than proportionately with increases in H initially, and then less than proportionately as e falls below unity, with the firm settling into the latter region.

Now consider how costs expressed in terms of passenger output behave as H is varied. Define the output-cost function $C(Q)$ by

$$C(Q) - cH \equiv 0. \tag{4}$$

Using the implicit-function rule gives

$$\frac{\partial C}{\partial Q} = \frac{c}{\partial Q/\partial H} \tag{5}$$

as marginal cost of output MC_Q. Marginal cost is positive since $\partial Q/\partial H > 0$, and rising at equilibrium, since $\partial^2 Q/\partial H^2 < 0$ by equation (2). Average cost of output is

$$AC_Q = \frac{C(Q)}{Q} = \frac{cH}{Q},$$

and, using the definition of e in (5),

$$MC_Q = \frac{cH}{Qe}. \tag{6}$$

So $MC_Q = AC_Q$ where $e = 1$, and at this point AC_Q is at a minimum, while AR_H is at a maximum. Thus, the output-cost curves are U-shaped, even though cost per cab hour is constant. These properties are reflected in the cost curves drawn in figure 2. Furthermore, applying equation (3)

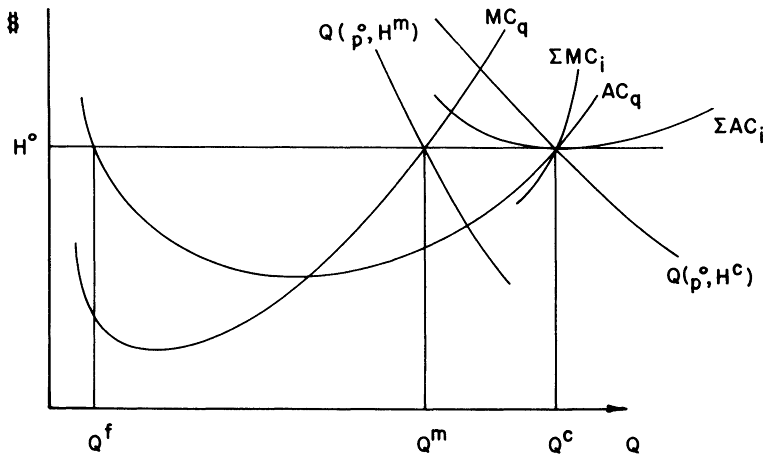

FIG. 2.—Equilibrium output at p^o

to equation (1) gives

$$pe\frac{Q}{H} - c = 0,$$

so

$$p = \frac{cH}{Qe}, \tag{7}$$

from which it follows that, in equilibrium, marginal cost of output is equal to the regulated price.[8] It should be noted that, even though the firm produces with excess capacity, it produces to the right of the minimum point of the average cost curve as long as it is earning profits. A zero-profit equilibrium occurs at minimum average cost and not at a point of tangency of the average cost-and-demand curves.

IV. Competitive Solution

Assume the industry is comprised of owner-operated cabs, one cab per owner or firm. The price to be charged is set by the regulator, but entry is unrestricted. If there are any economies associated with maintenance of large fleets and in the dispatching of cabs, these services can be provided by an independent maintenance contractor and by an impartial dispatcher maintained by an association of the cab owners. Assume operating cost per cab hour is U-shaped, declining initially because of some fixed costs, and eventually rising as the cab is operated ever more hours per

[8] Since it is fixed, price is marginal revenue, so this is not an unexpected result. The novelty is in the fact that, at the going price, output may only be varied by varying product quality, and it is the cost of changing output in this manner that counts.

day because of mounting maintenance costs and because costs include the imputed value of the owner's time, which rises as he drives more hours per day, leaving less time for other uses. Assume cabs are operated at the rate that minimizes cost per cab hour and that this cost is c. Assume also that this operating period spans the demand period. Then new cabs may be added at constant cost, and industry cost will be cH, where $H = \sum h_i$.

Given a reasonably large number of cabs, there will be no differentiation of cabs: the information structure assumed is that each potential cab rider knows that there is some expected interval between wanting a cab and getting one and uses this knowledge to form expected price. Aggregating individual demand functions, since price is fixed and equal for all cabs, gives the industry demand function

$$Q = Q[p + vt(H)].$$

Industry equilibrium occurs where average revenue per cab hour equals average cost; this is point H^c in figure 1. Under the cost assumptions, both the franchised monopoly and the competitive industry will have identical, and constant, costs with respect to cab hours.

Assuming market share equals share of industry cab hours, the profit function of the individual firm is

$$pQ[p + vt(H)]\,\frac{h_i}{H} - C(h). \tag{8}$$

The ith firm's program is to maximize profits by choice of h_i subject to the regulated fare. The maximum-profit point satisfies

$$p\,\frac{\partial q_i}{\partial h_i} - \frac{\partial C}{\partial h_i} = 0, \tag{9}$$

which, using the definition $e_i = \partial q_i/\partial h_i(h_i/q_i)$, may be written

$$pe_i - \frac{\partial C}{\partial h}\,\frac{h_i}{q_i} = 0. \tag{10}$$

The elasticity of demand with respect to cab hours experienced by the firm may be written

$$e_i = 1 + (e - 1)\,\frac{h_i}{H} + (e - 1)\,\frac{h_i}{H}\sum_{j=1}^{n-1}\frac{\partial h_j}{\partial h_i},$$

where $\partial h_j/\partial h_i$ is the change in rival's cab hours induced by i's change in cab hours, and e is the industry elasticity. If it is assumed that there are a large number of competing cab companies, each operating independently of the others, then $h_i/H \to 0$, so that $e_i \to 1$. This means that the firm will feel that it can sell all the cab hours it wishes to supply at the industry equilibrium-utilization rate. When $e_i = 1$, then from equation (10)

$$p - \frac{\partial C}{\partial h}\frac{h_i}{q_i} = 0, \tag{11}$$

and price is equal to the marginal cost of capacity (cab hours) times capacity per unit of output. In terms of output, rather than cab hours, we have, as in the monopoly case

$$p = \frac{\partial C}{\partial h}\frac{h_i}{q_i e_i}. \tag{12}$$

But, since $e_i = 1$,

$$p = \frac{\partial C}{\partial h}\frac{h_i}{q_i} \tag{13}$$

for each firm, and price is equal to marginal cost of output. Since price also equals average cost, marginal and average cost will be equal, implying that output is produced at minimum average cost for the equilibrium level of capacity utilization.

The industry marginal- and average-cost curves are U-shaped and identical to the franchised-monopoly curves. As entry initially occurs, waiting time falls more than proportionately, so that demand increases by more than the supply of cab hours, the utilization rate increases, and average cost per unit of Q falls. In this initial stage, entry of new firms increases profits of existing firms, since the average utilization rate realized by existing firms rises when new firms enter. This is not a stable situation, and it is clear that industry equilibrium may only occur in the nondecreasing range of the industry cost curve, where further entry decreases profits of existing firms. The competitive industry equilibrium occurs at Q^c in figure 2. The competitive output is greater than monopoly output, the competitive industry demand curve lies above the monopoly demand curve—because of the higher quality or lower waiting time—and average cost at the competitive output is higher than at the monopoly output, though marginal cost is equal for both.

The industry marginal-cost-of-output curve is also the monopolist's marginal cost curve, but it is not the sum of the MC curves of the individual firms. The reason for this is that the industry, on the whole, experiences decreasing returns to cab hours, but each individual firm realizes constant returns at any point of industry equilibrium. The sums of the marginal- and average-cost curves of all firms are shown as $\sum MC$ and $\sum AC$ at the industry equilibrium point. The average cost curve shows social marginal cost, but only at the regulated price, it is not the long-run industry supply curve.

V. The Medallion Solution

Assume one medallion per cab, and, as before, that the optimal operating period of the cab spans the demand period. Referring to figure 1, the feasible region begins at H^f, where revenue covers the cost of providing the service. If the number of medallions issued were less than H^f, none would be taken and no service would be provided. Any number of medallions in excess of H^c would not be taken if there were free entry, and no more than H^m would be taken by the monopoly. The rent per medallion is the difference between AC and AR associated with the number in existence. The medallion solution ranges between H^f and H^c, with zero rents and no service at H^f, and zero rents and maximum service at H^c. The effect of increasing the number of medallions on rents depends upon the initial conditions: if initially the number of medallions is less than H^m, then rents will increase as more are issued, but will fall with more medallions to the right of H^m.

VI. Limited Entry with Unrestrained Price

Suppose now the regulator simply sets the number of cabs allowed to operate, but leaves price free to find its own equilibrium. There is insufficient space in this paper to consider this problem in a thorough manner; I confine myself to a description of equilibrium when it is assumed that there are a large number of individually operated cabs competing for sufficiently well-informed passengers such that price equals average cost. Then price is a well-defined function of the number of cabs and the waiting-time function.

From the assumption of constant industry cost in cab hours, it follows that the long-run average cost of output is a U-shaped curve, because of the increasing and then decreasing productivity of cab hours in reducing delay. The *LRAC* curve is drawn in figure 3. The open-entry competitive equilibrium will be on the nondecreasing portion of this curve; assume it is point Q^o. If the regulator limits cabs to less than the quantity associated with Q^o, there will be longer waiting time, lower effective demand, less output, and a lower price. It can be shown that Q^o is the efficient level of output for the industry if it is the open competitive equilibrium.[9] If the regulator imposed Q' as a limit (through imposition of a limit on cabs), average cost and price would be lower than at Q^o, but the demand curve would also be shifted back from D^o to D', thus reducing consumer surplus. The result of this limitation would be a reduction in welfare, because the reduction in price achieved is too little to offset the reduction in service quality—that is, the increase in waiting time.[10] On the other hand, if the number of cabs were severely enough limited

[9] This is proved in De Vany (1973).
[10] This would follow if Q^o were efficient as asserted.

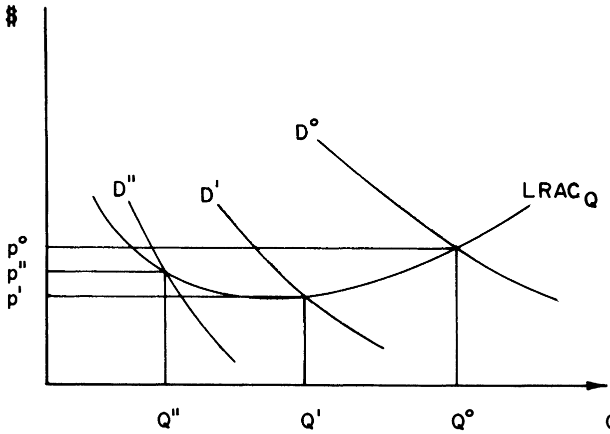

FIG. 3.—Equilibrium with entry limitations and unregulated price

so as to result in, say, output Q'', the demand curve would be shifted back to D'', and price and average cost would be increased. In this region it would appear that the industry experiences decreasing costs.

VII. Efficient Pricing

Assuming equal value of time for all consumers, the efficient levels of output and cab hours (or waiting time) occur where

$$-vt' = c,$$
$$-vt' = p\frac{Q}{H}, \tag{14}$$

or, equivalently,

$$p = c\frac{H}{Q}. \tag{15}$$

The first condition requires that the marginal value of capacity equal its marginal cost, and the second that average and marginal value of capacity be equal. Together, these conditions give equation (15), which indicates that price is equal to marginal cost of output, or, what is the same thing, price equals marginal cost of capacity times the number of units of capacity per unit of output.[11]

These conditions are equivalent to those that ensure that Q is max-

[11] Since it is assumed that the marginal cost of carrying a passenger on a cab that is already cruising is zero, to set price equal to this marginal cost would create losses for the industry. Efficient pricing demands that price equal marginal cost of output at the efficient level of capacity. True marginal cost is the cost of adding a passenger without reducing quality, i.e., holding H/Q constant. This is given by cH/Q, which equals price in competitive equilibrium. Cost is rising at equilibrium, since $c > 0$; and industry marginal cost cH/Qe exceeds average cost, since $e < 1$. Monopoly marginal cost cH/Qe is greater than true marginal cost, so monopoly is inefficient, even though price equals marginal cost at monopoly equilibrium.

imized subject to a zero-profit constraint, and so the efficient price maximizes output.[12] The effect of price variation on monopoly supply of cab hours is:

$$\frac{\partial H}{\partial p} = -\frac{\beta + 1 + \varepsilon_p^e}{p\,\partial^2 Q/\partial H^2}, \tag{16}$$

where $\varepsilon_p^e \leq 0$ is the elasticity of the capacity elasticity of demand with respect to price.[13] A price increase will increase monopoly capacity only if demand is quite price inelastic. In the case of open entry, the change in industry capacity is

$$\frac{\partial H}{\partial p} = -\frac{Q(\beta + 1)}{pQe - cH}, \tag{17}$$

which, since $(pQe - cH) < 0$, depends upon the price elasticity of demand as well. The competitive derivative lacks ε_p^e, and so a comparable increase in the regulated price will be more likely to expand capacity under competition than under monopoly. For either case, the change in output due to a change in the regulated price is

$$\frac{dQ}{dp} = \frac{\partial Q}{\partial \pi} + \frac{\partial Q}{\partial \pi} v \frac{\partial t}{\partial H}\frac{\partial H}{\partial p}, \tag{18}$$

and the full change in output depends upon a (negative) direct price effect and an indirect induced quality effect which may be positive or negative.

The efficient, or output-maximizing, price occurs where (18) is zero. Since $\partial H/\partial \pi$ and $\partial t/\partial H$ are negative, $\partial H/\partial p$ must be positive at maximum output, implying that output is maximized at a lower price than capacity. The price which maximizes monopoly output is less than the efficient price which maximizes competitive output, and monopoly output and capacity are always less than competitive output and capacity. These relationships are shown in figure 4. A price higher than the efficient one (p^*) will reduce competitive output and expand capacity beyond the efficient level, as long as price is not set higher than the level that maximizes capacity (p'). A higher price than the capacity-maximizing one reduces both competitive capacity and output. The competitive outcome is superior to the monopoly outcome at every price up to and including the efficient level. At a price above p^*, competitive capacity will be expanded relative to output, and the marginal cost of capacity will exceed its marginal value. There will exist a price above the efficient level at which marginal cost of capacity under competition will be so far

[12] Douglas (1972) shows this to be true. However, the present model makes no allowance for such possible externalities as street congestion caused by taxis. Nor does it allow for differences in the value of time of consumers, in which case a uniform price would prevent product differentiation.

[13] Implicit differentiation of eq. (1) and some manipulation yields this result.

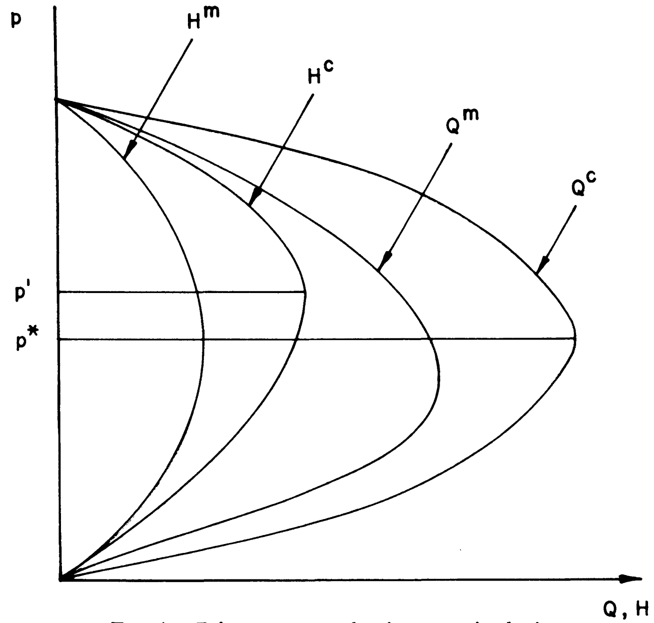

FIG. 4.—Price-output and price-capacity loci

in excess of marginal value that the lesser capacity offering of the monopolist will be more efficient than the competitively supplied capacity. This does not come about because competition is wasteful; rather, it is caused by the regulator's setting price above the efficient level.

References

Alchian, A. A., and Allen, W. R. *Exchange and Production Theory in Use*. Belmont, Calif.: Wadsworth, 1969.

Averch, H., and Johnson, C. C. "Behavior of the Firm under Regulatory Constraint." *A.E.R.* 52 (December 1962): 1053–69.

Becker, Gary S. "A Theory of the Allocation of Time." *Econ. J.*, vol. 75 (September 1965).

Chamberlin, E. H. *The Theory of Monopolistic Competition*. Cambridge, Mass.: Harvard Univ. Press, 1933.

De Vany, A. "The Allocation of Time in Competitive Markets." Mimeographed. Department of Economics, Texas A & M Univ., 1973.

Douglas, G. W. "Price Regulation and Optimal Service Standards: The Taxicab Industry." *J. Transport Econ. and Policy* 20 (May 1972): 116–127.

Eckert, R. D. "On the Incentives of Regulators: The Case of Taxicabs." *Public Choice* 14 (Spring 1973): 83–101.

Friedman, M. *Price Theory*. Chicago: Aldine, 1962.

Kaldor, N. "Market Imperfection and Excess Capacity." *Economica* 11, n.s. (1935): 33–50.

Orr, D. "The Taxicab Problem: A Proposed Solution." *J.P.E.* 77, no. 1 (January/February 1969): 141–47.

Samuelson, P. A. *Economics*. 9th ed. New York: McGraw-Hill, 1973.