Monopoly, quality, and regulation

A. Michael Spence

Associate Professor of Economics
Stanford University

This paper deals with market problems that arise when a monopoly sets some aspect of product quality as well as price. It is argued that the market failure is associated with the inability of prices to convey information about the value attached to quality by inframarginal consumers. In the regulatory context, this market problem appears in the form of a difficult informational question for the regulator; what is the average valuation of quality over all the consumers in the market? The paper suggests that rate-of-return regulation may have attractive features when quality is a variable.

1. Introduction

■ The argument in this paper reduces to three basic points. Under regimes of monopoly (and monopolistic competition), product characteristics (which are often endogenous variables) are not usually optimally set under the pressure of market forces. Second, regulation is also beset with difficulties when price and quality are decision variables. These difficulties are informational, and are closely related to the sources of market failure in the unregulated market. Third, rate of return regulation may have attractive, second-best properties in such situations.

The source of the potential market failure is relatively easy to locate. Consider a firm which is contemplating a small increase in the quality of its product and assume for simplicity, though this is inessential, that each consumer buys one unit of the good. The increase in quality will increase costs, say by Δc. It will also increase revenues. The increase in quality increases the dollar benefits of the product *to the marginal consumer* (i.e., to the consumer who is just willing to pay the going price) by $\Delta p(x)$. The firm will increase revenues by $x \Delta p(x)$, where x is the number of purchasers. Thus, the increase is desirable for the firm if $x\Delta p(x) > \Delta c$. But $x\Delta p(x)$ is not necessarily an accurate measure of the social benefits of the increase in quality.

The quality increase is desirable if the average benefit $(1/x)\int_0^x \Delta p(v)dv$ exceeds the average cost, $\Delta c/x$. Equivalently, the total benefits $\int_0^x \Delta p(v)dv$ must exceed the cost of the increase

---

A. Michael Spence received the B.A. from Princeton University in 1966, the B.A. from Oxford University in 1968, and the Ph. D. from Harvard University in 1972. Currently his research centers on market structure and imperfect competition.

This work was supported by National Science Foundation Grant GS-39004 at the Institute for Mathematical Studies in the Social Sciences, Stanford University. I am grateful to Oliver Williamson and James Rosse for their extensive comments.

Δc. The social benefits correspond to the increase in the revenues of the firm only if the marginal consumer is average or representative, that is to say when $(1/x)\int_0^x \Delta p(v)dv = \Delta p(x)$. But there is nothing at all intrinsic to the market that guarantees that the marginal purchaser is representative. On the contrary, in many cases, the marginal consumer is quite unlikely to be representative in his marginal valuation of quality.[1]

The implications of this divergence between private and social benefits, for both regulated and unregulated monopolists, are developed below. The paper does not pretend to deal with the regulatory problem in operational terms. It nevertheless raises analytic issues that may be relevant to regulatory strategy. The paper begins with the unregulated monopolist.

2. Notation

■ Throughout, the following notation is employed. Price is p, quantity x, and product quality q. Quality is a scalar, though none of the results depend upon that simplifying assumption, which is adopted for ease of exposition. The demand for the product is $D(p,q)$ and the inverse demand is $P(x,q)$. Costs of producing x units of quality q are $c(x,q)$. Only one product is produced.[2]

Consumer surplus, denoted by S can be written in two ways:[3]

$$S = \int_0^x P(v,q)dv - xP(x,q) \tag{1}$$

or

$$S = \int_p^\infty D(v,q)dv. \tag{2}$$

Similarly, revenues can be written in two ways:

$$R = xP(x,q) \tag{3}$$

or

$$R = pD(p,q). \tag{4}$$

Profits, denoted by π are

$$\pi = R - c. \tag{5}$$

And finally, the total surplus is denoted by W:

$$W = S + \pi. \tag{6}$$

[1] These problems arise in much more general settings in which there is monopolistic competition and product differentiation. See Spence [5] and Dixit and Stiglitz [1].

[2] One might ask why a whole spectrum of products, differentiated in terms of both price and quality, is not produced. The assumption is that fixed costs or increasing returns and demand prevent the profitable production of more than one product of the type under consideration. More generally, fixed costs *limit* the number of products without reducing the feasible set to just one. This is the more difficult subject of monopolistic competition.

[3] Income effects that may make the consumer surplus inaccurate are ignored here. For a discussion of the accuracy of consumer surplus, see Willig [8].

■ The firm has three decision variables: price, quantity, and quality. Two of these are independent; the third is determined by the demand function. For the time being I shall take the decision variables to be x and q.

For a *given* q, the surplus is maximized with respect to x when $P = c_x$. Profits, however, are maximized when $P - c_x = -xP_x > 0$. This failure is the familiar one: the monopoly exploits its power over price. It raises price because marginal revenues fall short of marginal gross dollar benefits (in the absence of price discrimination).

Less familiar is the failure associated with the fact that quality is an additional decision variable. For a given quantity, x, the surplus is maximized with respect to quality when

$$\frac{\partial W}{\partial q} = \int_0^x P_q dv - c_q = 0. \tag{7}$$

On the other hand, profits are maximized with respect to quality when

$$\frac{\partial \pi}{\partial q} = xP_q - c_q = 0. \tag{8}$$

Since (7) and (8) differ, quality may not be optimally set. Is the monopolist above or below the socially optimal q for given x? The answer depends on the sign of P_{xq}. To establish the manner of the dependence, we note from (1) and (6) that

$$\frac{\partial W}{\partial q} = \frac{\partial S}{\partial q} + \frac{\partial \pi}{\partial q} = \int_0^x P_q(v,q)dv - xP_q(x,q) + \frac{\partial \pi(x,q)}{\partial q}. \tag{9}$$

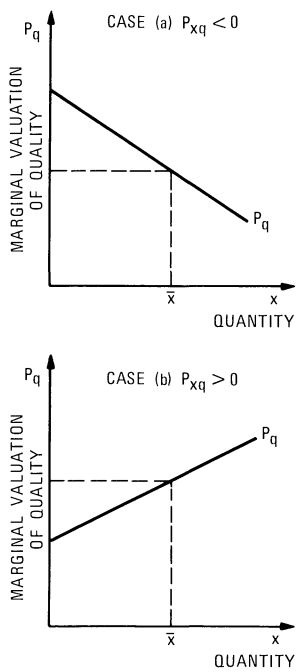
Therefore, when $\partial\pi/\partial q = 0$, the sign of $\partial W/\partial q$ depends on the relative magnitudes to $\int_0^x P_q dv$ and xP_q. The derivative, P_q, is the marginal valuation of quality by the marginal consumer. On the other hand, $(1/x)\int_0^x P_q dv$ is exactly the *average* valuation of quality (at the margin) over all the people in the market. The average is the relevant quantity for welfare, but the firm responds to the marginal individual. (For the firm to do otherwise would require the ability to price discriminate.) When the average valuation of quality increments exceeds the marginal valuation, the firm sets quality too low; it stops increasing quality too soon.

The quantity P_q is the marginal consumer's valuation of quality increments. Thus P_{xq} is the change in P_q as one moves down the spectrum of consumers ordered by their willingness to pay. If $P_{xq} < 0$, then the marginal value of quality falls as absolute willingness to pay falls. When this is true, the average value attached to quality exceeds the marginal consumer's valuation. (See Figure 1.) The firm sets q too low for given x. We can summarize these partial effects in a proposition.

Proposition 1: (A) For given x, the firm undersupplies quality relative to the optimum when $(1/x)\int_0^x P_q dv > P_q$ and conversely. (B) It is sufficient for $(1/x)\int_0^x P_q dv > P_q$ that $P_{xq} < 0$, and conversely.

FIGURE 1

THE MARGINAL VALUATION OF QUALITY SCHEDULES

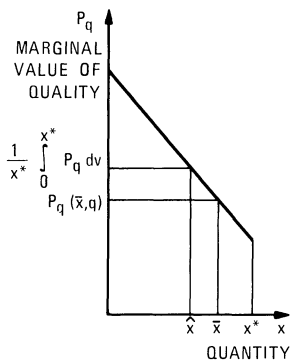


This aspect of market failure has very little to do with monopoly. It is, rather, a result of the fact that price signals carry marginal information, while averages or totals are required in locating the optimum. Any profit-oriented supply side runs into similar difficulties.[4]

The firm with market power deviates from the optimum in two respects. It maintains a markup above marginal cost, or sets quantity too low, and it sets quality high or low depending on the relative valuations of quality of the marginal and average consumer. The overall performance of the monopolist is the conjunction of these two forces or partial effects.[5]

FIGURE 2

OPTIMAL vs. PROFIT MAXIMIZING QUALITY WITH $P_{xq}<0$



In the general monopoly case, the relationship between optimal and profit maximizing quality levels is determined by the interaction of the average marginal effect (i.e., the sign of P_{xq}) and the extent to which the monopolist restricts output. Consider Figure 2 in which P_q is shown as a declining function of x. Assume quality is at the profit maximizing level, and let $\bar{x}$ be the profit maximizing quantity. Let x^* be the socially optimal quantity given q. Assume costs are $c(q)x$.

At $\bar{x}$, $P_q(\bar{x},q) = c'(q)$ because of profit maximization. On the other hand

$$\frac{dW}{dq} = \int_0^{x^*} P_q(v,q)dv - c'(q)x^*. \tag{10}$$

[4] To illustrate, let me consider the case in which the firm was constrained to set price equal to marginal cost: $P(x,q) = c_x(x,q)$. It is still true that quality would be set incorrectly except in special cases. To illustrate this point suppose that $P_{xq} < 0$. On the line $P = c_x$,

$$\frac{dx}{dq} = -\frac{p_q - c_{xq}}{P_x - c_{xx}}. \tag{a}$$

Let us assume further that $c_{xq} \leq c_q/x$. That implies that average costs go up faster than marginal costs as quality is increased. With $P = c_x$ as a constraint, welfare is maximized when

$$\frac{dW}{dq} = (P - c_x)\frac{dx}{dq} + \int_0^x P_q dv - c_q = \int_0^x P_q dv - c_q = 0. \tag{b}$$

At that point

$$\frac{d\pi}{dq} = (P - c_x + xP_x)\frac{dx}{dq} + xP_q - c_q$$

$$= xP_q - c_q - xP_x\frac{P_q - c_{xq}}{P_x - c_{xx}} < xP_q - xc_{xq} - xP_x\frac{P_q - c_{xq}}{P_x - c_{xx}} = -xc_{xx}\frac{P_q - c_{xq}}{P_x - c_{xx}}$$

However, since $P_{xq} < 0$, $P_q - c_{xq} < P_q - c_q/x < \int_0^x (P_q/x)dv - c_q = 0$, from (b). Thus $d\pi/dq < 0$ at the optimal point, and the firm would reduce quality. When constrained to set price equal to marginal cost, the profit maximizing firm does not set quality at the optimal level.

[5] Of somewhat less interest is the case where price is fixed or taken as given. In that case, the firm always sets quality too low. To see this note that

$$\frac{\partial W}{\partial q}(p,q) = \int_p^{\infty} D_q(v,q)dv + \frac{\partial \pi}{\partial q} > \frac{\partial \pi}{\partial q}.$$

Thus, when $\partial\pi/\partial q = 0$, $\partial W/\partial q > 0$. It is this tendency that has led regulatory authorities to specify minimum quality standards in the context of price regulation. The argument above implies that for any fixed price, the firm sets quality too low. It does *not* imply that if one compares the global optimum in price and quality and the profit maximizing point, that one will always find quality lower in the latter.

Thus, $dW/dq > 0$, if $(1/x^*)\int_0^{x^*} P_q dv > c'(q) = P_q(\bar{x},q)$. The crucial question then is whether the average $(1/x^*)\int_0^{x^*} P_q dv$ is greater than or less than $P_q(\bar{x},q)$. The answer is determined by the size of $\bar{x}$. If $\bar{x}$ is near x^*, then $(1/x^*)\int_0^{x^*} P_q dv > P_q(\bar{x},q)$, $dW/dq > 0$ and optimal quality is above the profit maximizing level. If $\bar{x}$ is small, the reverse holds. (See Figure 2.) Both of these conclusions are reversed, however, if $P_{xq} > 0$, as can be seen from Figure 3. Table 1 summarizes the results in a qualitative way. To recapitulate, the two determining factors are (i) the relation between the marginal and the average consumer in terms of valuation of quality, and (ii) the extent to which the firm restricts quantity. This second factor is determined by the shape of the demand curve, or roughly by the elasticity of demand.

Implicit in the foregoing is the influence of demand elasticity on quality distortion, and in particular the effect of quality increases on the elasticity of demand. To investigate this, let

$$\bar{W}(q) = \max_x W(x,q), \tag{11}$$

and let

$$\bar{\pi}(q) = \max_x \pi(x,q). \tag{12}$$

We define the ratio of maximized profits to maximized surplus to be

$$\beta(q) = \frac{\bar{\pi}(q)}{\bar{W}(q)}. \tag{13}$$

The sign of the slope of $\beta(q)$ determines whether quality is over or undersupplied by the monopolist. Taking logs and differentiating, we have

$$\frac{\beta'}{\beta} = \frac{\bar{\pi}'}{\bar{\pi}} - \frac{\bar{W}'}{\bar{W}}. \tag{14}$$

Thus when $\bar{\pi}' = 0$

$$\frac{\beta'}{\beta} = -\frac{\bar{W}'}{\bar{W}}. \tag{15}$$

Proposition 2: If $\beta' < 0$, quality is undersupplied by the monopolist and conversely.

Thus, the question of whether the firm over or undersupplies quality relative to the optimum translates into the question of how the fraction of the total potential surplus that is capturable by the firm varies with quality. This question can be asked of any demand function $D(p,q)$, but the case in which the elasticity is independent of price (or quantity) is illuminating. Thus, let us suppose that the demand functions are of the constant elasticity type for any given level of quality, so that $P(x,q) = A(q)x^{-n(q)}$. In this case, it can be verified that

$$\beta(q) = [1 - n(q)]^{\frac{1}{n(q)}}. \tag{16}$$

FIGURE 3

OPTIMAL vs. PROFIT MAXIMIZING QUALITY WITH $P_{xq}>0$

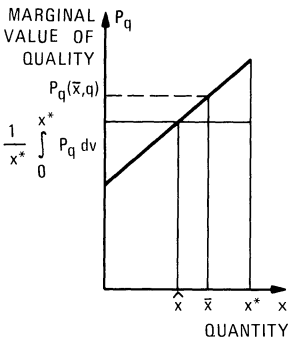


TABLE 1

CONDITIONS UNDER WHICH THE OPTIMUM QUALITY IS ABOVE AND BELOW THE PROFIT MAXIMIZING QUALITY

	$P_{xq} < 0$	$P_{xq} > 0$
QUANTITY RESTRICTION MODERATE	ABOVE	BELOW
QUANTITY RESTRICTION SEVERE	BELOW	ABOVE

The function, $(1 - n)^{1/n}$ is declining in n.[6] Moreover, n is the inverse of the price elasticity of demand. The fraction of the surplus captured by the firm is an increasing function of price elasticity of demand. If price elasticity declines with q, then $n'(q) > 0$, and $d\beta/dq < 0$. In that case, when $\pi' = 0$, $W' > 0$ from (17), and quality is set too low by the monopolist.[7]

Proposition 3: If the price elasticity is not a function of the price and if the elasticity declines with quality, then quality is undersupplied and conversely. If the elasticity does not vary with quality, quality is set at the optimal level by the monopolist. More generally, the market mechanism is biased (in welfare terms) against low elasticity products.[8]

4. Regulatory issues—general

■ The preceding problems in the unregulated market raise analogous problems in the context of regulation. I want to make three points in this section. The first is that when price and quality are variable, the interaction between the firm and the regulatory authority can usefully be characterized as a duopoly problem in which part of the task of the firm and the regulatory authority is to avoid inefficient outcomes, that is, outcomes in which both profits and surplus are unnecessarily sacrificed. In oligopolistic interaction, collectively rational outcomes are achieved through the adoption of reasonable reaction patterns by the players in the game.[9] The same is probably true in the regulatory context. The second point is that the regulatory authority's optimal reaction function is a line or schedule in price-quality space, along which consumer surplus is constant. And

---

[6] Taking logs and differentiating, we have

$$\frac{\beta'}{\beta} = -\frac{1}{n^2}\left[\frac{1}{1-n} + \log(1-n)\right] = -\frac{1}{n^2}\left[\frac{n}{1-n} - \int_{1-n}^{1}\frac{dx}{x}\right] < 0.$$

[7] One can verify that for demands that are linear in x, $\beta(q) \equiv 1/2$. As a result $W'(q) = 0$ if and only if $\bar{\pi}'(q) = 0$ and quality is optimally set by the monopoly. Distortions do not arise with linear demands.

[8] Durability can be thought of as product quality. It will generally be correctly supplied by the monopolist. Let q = durability, x = quantity of the good, s = services delivered by the good, p = price of the good and r = price of the services. Then by definition $s = qx$ and $p = rq$. The inverse demand for the good is $qT(xq)$. Assuming constant marginal costs for each level of q,

$$\pi = xqT(xq) - c(q)x$$

$$W = q\int_0^x T(vq)dv - c(q)x = \int_0^{xq} T(v)dv - c(q)x.$$

Let $y = xq$. Profits and surplus become

$$\pi = yT(y) - \frac{c(q)}{q}y.$$

$$W = \int_0^y T(v)dv - \frac{c(q)}{q}y.$$

Thus both are maximized with respect to q when $c(q)/q$ is minimized with respect to q, i.e., the unit cost of the services the goods deliver are minimized. Generalizing slightly, any situation which has the property that inverse demand can be written $\alpha(q)T(x\alpha(q))$ where $\alpha(q)$ is any increasing function of q, will cause no quality problems.

[9] Reaction function equilibria and tacit coordination are discussed in Spence and Untiet [6].

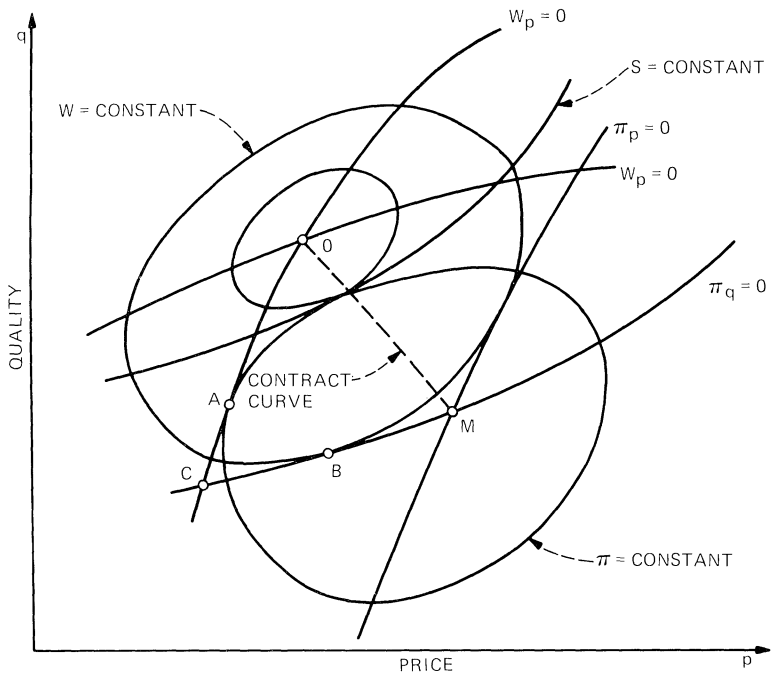
thirdly, the locating of this schedule presents difficult informational problems for the regulators. Simple market information is insufficient even for evaluating small, local changes in price and product (or service) quality. Having developed these points, I shall then argue in the following section that rate of return regulation may be, under certain circumstances, an attractive second-best strategy.

There are a variety of possible modes of regulatory behavior, some preferable to others. Price is taken to be the regulatory authority's decision variable. Quality is set by the firm. The assignment of decision variables is somewhat arbitrary. The important questions, however, concern *how the regulatory authority (and the firm) respond or react to the moves of the other party,* and not the particular decision variables they manipulate.

It will facilitate the analysis to depict graphically the contours along which total surplus and profits do not vary and to introduce other features of the incentive structure of the game. A typical configuration, in price-quality space, is illustrated in Figure 4. There are two sets of closed contours, one surround-

FIGURE 4

THE INCENTIVES IN THE REGULATORY GAME



ing the point O, the other around M. The former are iso-total surplus contours and the latter are iso-profit contours. Total surplus is maximized at O, profits are maximal at M. The points of tangency of these two sets of contours are on the line OM. This line is the contract curve: at each point on it, the surplus is as large as possible given profits and conversely. In addition there are four more lines; two through O, and two through M. For example, the line $W_p = 0$ is the line along which surplus is maximized with respect to p for each level of q. Similar state-

ments hold for the other curves, which are labeled. Finally, there is a line of constant consumer surplus, S = constant. It is tangent to both an iso-total surplus line and an iso-profit line, the tangency occurring on the contract curve. The double tangency follows from the fact that $W = S + \pi$.

The incentive structure of this game is closely akin to that of an ordinary duopoly. What is somewhat distinctive are the decision variables. The game is nonzero sum. Both parties have a common interest in avoiding outcomes for which there are alternatives that dominate in both profits and surplus. Like a duopoly game, this problem has a variety of equilibria. For example, C is the Cournot or Nash equilibrium, while A and B are the von Stackelberg equilibria. Similarly, there is a plethora of outcomes that correspond to the adoption of different pairs of reaction functions by the firm and the regulatory authority. Points on the contract curve can be sustained as reaction function equilibria. Naive behavior is not likely. For example regulators commonly supplement price regulation with minimum quality standards, to avoid outcomes like C in Figure 4.

Efficiency (in the sense of the contract curve) is achieved and sustained through the adoption of suitable reaction functions that implicitly police the desired outcome. To focus upon the regulators, their appropriate strategy is to confront the firm with a schedule of prices and qualities that correspond to an iso-consumer surplus line. If such a reaction function is adopted, the firm will maximize against it by selecting a point on the contract curve, because the contract curve is the locus of tangencies of profit and consumer surplus contours. Notice that this implies that the regulatory authority can, *in principle,* confine itself to representing consumer preferences in the market.

What distinguishes the regulatory duopoly, in degree if not in kind, is the severe informational problem facing the regulatory authority. In addition to the familiar difficulty of knowing costs (especially marginal costs), there are two additional informational problems.

One is simply measuring quality. This problem has often been noted, to which I have nothing to add. But there is a second and, I think, equally serious problem relating to demand.

A consumer surplus line is defined by the equation

$$\int_p^\infty D(v,q)dv = \text{constant}. \tag{17}$$

The implied schedule of prices and qualities is difficult to compute and this is the source of the informational problem. The problem is best illustrated by considering small changes in price and quality. Suppose that the firm proposes changes in price and quality (dp, dq) that its marketing people believe, on the basis of some local experiments with price and quality, will be profitable: i.e.,

$$d\pi = \pi_p dp + \pi_q dq > 0. \tag{18}$$

The regulators must assess whether consumer surplus also rises, i.e., whether

$$dS = -Ddp + [\int_p^\infty D_q(v,q)dv]dq > 0. \tag{19}$$

It is not difficult to observe the current quantity, D. The problem is the integral term. No small experiments with price and quality of the type that permits estimates of π_p and π_q will generate the required information. The reason is that quality changes affect the welfare of the entire set of inframarginal consumers. These effects must be estimated to evaluate a change in price and quality. But the data generated by local changes in the parameters will not yield such estimates.

Expression (19) can be rewritten in the following form:

$$\left[\frac{\int_p^\infty D_q(v,q)dv}{D}\right]dq > dp. \tag{20}$$

In this form, it says that the average valuation of the quality change (the left-hand side) must exceed the price increase, dp, if the change is to be accepted by the regulator. Thus the regulatory authority has the problem of estimating the average valuation of the quality shift for all consumers in the market. This information is not conveyed by prices, or local experiments with price and quality.

The implication of this line of reasoning is that regulatory agencies facing firms with discretionary control over aspects of product quality require nonmarket information to evaluate changes in prices and quality. There may be ways to generate this type of information: consumer surveys suggest themselves as a starting point. But additional information is needed. Shifts in product or service quality affect inframarginal consumers. These people must be consulted if a correct evaluation of a quality change is to be made. The argument can be reversed. Surveys of consumers designed to establish their willingness to pay for increases in service quality will *not* necessarily provide information about the *profitability* of quality improvements.

A random survey of current users will yield accurate information about the average willingness to pay for service quality. But the fact that average willingness to pay exceeds the cost does not imply that an increase in service quality will be profitable.

That the regulatory problem should be informationally complicated in this particular way is simply a reflection of the fact that the price signals are misleading in the unconstrained market, as we saw earlier. The sources of the two problems are the same. From a welfare point of view the inframarginal consumers matter.

Because of the informational difficulty of accurately setting an iso-consumer surplus schedule, it seems desirable to inquire into reasonable second-best alternatives. The regulatory authority needs to confront the firm with a price-quality schedule in which prices rise quickly enough with quality to provide the firm with an incentive to raise quality. But it needs to set the schedule in the absence of detailed knowledge of the cost function, $c(x,q)$.

In the absence of quality variation, the difficulty of observing costs has pushed regulatory practice in the direction of controlling through constraining the rate of return on capital. While this mode of behavior leads, in principle, to input distortions, it may

nevertheless be a reasonable response to limited information. I want to argue now that with quality variation, rate of return regulation may continue to be a reasonable constrained second-best mode of behavior, because rate of return regulation implicitly constrains the firm with a positively sloped schedule of prices and quantities.

5. Rate of return regulation

■ Consider first the simple case where capital is the only input.[10] Let $k(x,q)$ be the amount of capital required to produce x units of output of quality q. Let r be the cost of capital and s be the "fair" rate of return. Costs are $c(x,q) = rk(x,q)$. The rate of return constraint takes the form

$$\frac{pD}{k} \leq s, \tag{21}$$

or

$$p \leq \frac{s}{r}\frac{c(D,q)}{D}. \tag{22}$$

Let $c(D,q)/D = a(D,q)$, the average cost and assume $a_D \geq 0$ and $a_q > 0$. It follows that the rate of return constraint is a positively sloped function in price-quality space. By differentiating the equality version of (22), we have

$$\frac{dp}{dq} = \frac{a_D D_q + a_q}{\frac{r}{s} - D_p a_D} > 0. \tag{23}$$

Thus a rate of return constraint confronts the regulated firm with a positively sloped constraint, which, while it may not be the optimal one, is preferable to a simple fixed price. The comparison is depicted in Figure 5. As depicted, the ROR outcome is at E. It is not on the contract curve, but the best fixed outcome is F, where the locus $\pi_q = 0$, which is the firm's reaction curve, is tangent to the iso-total surplus contour. F is likely to be less desirable than the ROR constrained outcome because of the absence of responsiveness of price to quality increases.

To illustrate, suppose that $c(D,q) = \phi(q)D$, so that costs are linear in quantity. The rate of return constraint (23) becomes $p = (s/r)\phi(q)$. If s is near r, the rate of return constraint is near $W_p = 0$, or $p = c(q)$. Thus one could approach the point A in Figure 4. In general, the position of the rate of return constraint will depend on the allowed rate of return, s. As s increases, p rises faster with q and greater investment in quality is induced. The dashed lines in Figure 5 illustrate a family of ROR constraints.

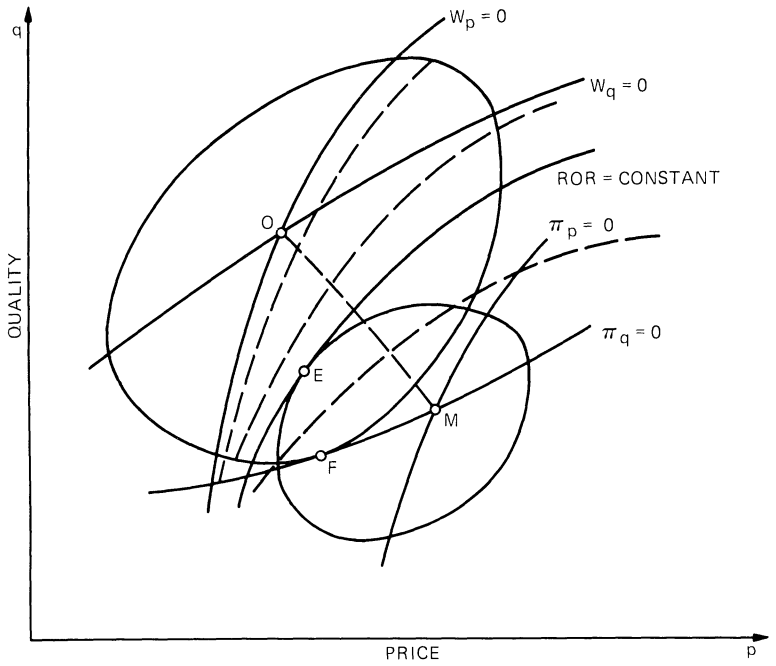
The problem becomes more complicated when quality interacts with input choices. It seems intuitively clear, however, that rate of return regulation will be attractive when quality is a capital-using attribute—where by capital-using I mean specifically that $(\partial/\partial q)(f_l/f_k) < 0$, so that the marginal rate of substitution of capital for labor rises with quality.[11] The reason that rate of

[10] Alternatively, one can assume that capital and labor must be combined in fixed proportions.

[11] This would be true for example, if higher quality required more capital equipment, but no more labor.

FIGURE 5

RATE OF RETURN CONSTRAINTS IN THE REGULATORY GAME



return regulation is attractive under these conditions is that the tendency of an ROR constraint to cause the substitution of capital for other inputs may partially compensate for the firm's tendency to undersupply quality. The converse is also true. If quality is labor-using, as it might be for certain kinds of services, rate of return regulation is likely to exacerbate the quality problem.

To explore these points, recall first that a rate of return constraint is equivalent to a lower bound on the capital stock. Therefore, the imposition of an upper bound on the rate of return is equivalent to imposing a lower bound on capital.

Profits can be written $\pi(x,q,k) = xP(x,q) - c(x,q,k)$, where $c(x,q,k)$ is defined by $c = wl + rk$ and $x = f(k,l,q)$: f is the production function.

Let

$$\hat{\pi}(k) = \max_{x,q} \pi(x,q,k). \tag{24}$$

Similarly, let

$$\hat{T}(k) = T(x(k),q(k),k) = \int_0^{x(k)} P(s,q(k))ds - c(x(k),q(k),k), \tag{25}$$

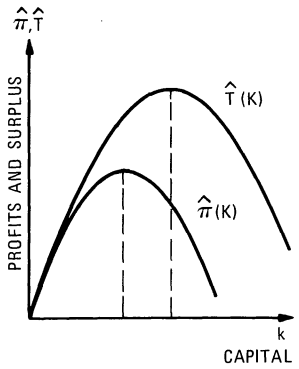
where $x(k)$ and $q(k)$ maximize profits for a given k, i.e.,

$$\pi_x = 0 \quad \text{and} \quad \pi_q = 0. \tag{26}$$

The question to be asked, now, is as follows: given that total surplus is constrained to the second-best level $\hat{T}(k)$, will forcing the capital stock above the profit maximizing level increase $\hat{T}(k)$ or not? In terms of Figure 6, the question is whether the peak of $\hat{T}(k)$

FIGURE 6

PROFITS AND SURPLUS AS A FUNCTION OF CAPITAL



is to the right of the peak of $\hat{\pi}(k)$ or to the left. This can be ascertained by obtaining the expression for $d\hat{T}/dk$, namely,

$$\frac{\partial \hat{T}}{\partial k} = (P - c_x)\frac{\partial x}{\partial k} + \left[\int_0^x P_q ds - c_q\right]\frac{\partial q}{\partial k} - c_k, \tag{27}$$

where

$$\frac{\partial x}{\partial k} = \frac{\pi_{qq}c_{xk} - \pi_{xq}c_{qk}}{\Delta}, \tag{28}$$

$$\frac{\partial q}{\partial k} = \frac{\pi_{xx}c_{qk} - \pi_{xq}c_{xk}}{\Delta},$$

and $\Delta = \pi_{xx}\pi_{qq} - \pi_{xq}^2 > 0$.

At the profit maximizing k, $c_k = 0$, $P > c_x$, and if $P_{xq} < 0$, $\int_0^x P_q ds - c_q > 0$. Therefore $\partial\hat{T}/\partial k$ will certainly be positive if $\partial x/\partial k$ and $\partial q/\partial k$ are positive.

From the second-order conditions for profit maximization $\pi_{xx} < 0$, $\pi_{qq} < 0$, and $\Delta > 0$. Moreover, let me assume that π_{xq} is small, so that the second terms in the numerators of $\partial x/\partial k$ and $\partial q/\partial k$ are not large. It is easily verified that $c_{xk} < 0$, i.e., that marginal variable costs fall as the capital increases. That leaves c_{qk} as the crucial quantity. From the fact that $c = wl + rk$ and $x = f(k,l,q)$, we have

$$c_{qk} = w\frac{\partial}{\partial_q}\left(-\frac{f_k}{f_l}\right). \tag{29}$$

But the quantity, $-f_k/f_l$ is the marginal rate of substitution of labor for capital. Thus, if raising q reduces the substitutability of labor for capital (i.e., quality is capital using), then $c_{qk} < 0$. And if π_{xq} is small, both x and q will increase with k from (25), and the surplus will necessarily increase. On the other hand, if $c_{qk} > 0$ (i.e., quality is labor-using), then quality may fall. That is desirable if $P_{xq} > 0$. But if $P_{xq} < 0$, total surplus may fall as a result of the rate of return constraint.

The results are not free from ambiguity, but the forces at work are clear. Rate of return constraints force the capital stock up. That will improve quality if quality is capital-using and conversely. And the desirability of raising quality depends on the sign of P_{xq}, as before.

6. Concluding remarks

■ The conclusions of this analysis are somewhat discouraging. The unregulated monopolist's selection of product characteristics is likely to be biased away from the social optimum because of possible differences between marginal and average valuations of quality by consumers. Regulation is, in part, a two-party nonzero sum game with incomplete information. Ideally the regulatory authority would manage price-quality tradeoffs by confronting the firm, on behalf of consumers, with a reaction function that reflects rates of substitution between price and quality on the demand side of the market. But these rates of substitution are difficult to determine, because computing them requires knowledge of the value attached to quality by the full

range of inframarginal consumers. Rate of return regulation has some merit in these circumstances as a second-best strategy, especially when (1) profit maximizing quality tends to be too low and (2) quality is a capital-using attribute. If, however, some other aspect of quality is labor-using, it will likely be diminished under rate of return regulation.

Finally, let me comment on past attention to aspects of the quality of service in regulated industries. It has certainly been appreciated that simple price constraints may cause quality to be set well below optimal levels. The Cournot equilibrium is usually avoided. Second, in public transportation especially, considerable attention has been devoted to attributes of the service like comfort, speed, and predictability. In particular, there have been attempts to estimate the benefits of these attributes. But the reason for doing the estimates is that decisions on attributes frequently involve irreversible capital investments, so that the profitability of the investments needs to be established in advance. That, of course, is a good reason for surveying potential customers by itself. I hope, however, that the preceding analysis has succeeded in demonstrating that profitability is not a sufficient criterion for deciding on the social value of service quality, and that this, too, constitutes a reason for being interested in nonmarket, survey-type information.

References

1. DIXIT, A. K. AND STIGLITZ, J. E. "Monopolistic Competition and Optimum Product Diversity." Technical Report No. 153, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1974.
2. LEVHARI, D. AND PELES, Y. "Market Structure, Quality, and Durability." *The Bell Journal of Economics and Management Science,* Vol. 4, No. 1 (Spring 1973).
3. PANZAR, J. "A Model of Regulated Oligopoly with Product Quality Variation: The Case of Passenger Airlines." Memorandum # 157, Center for Research in Economic Growth, Stanford, November 1973.
4. ROSSE, J. N. "Product Quality and Regulatory Constraints." Memorandum # 137, Center for Research in Economic Growth, Stanford University, November 1972.
5. SPENCE, A. M. "Product Selection, Fixed Costs, and Monopolistic Competition." Technical Report No. 157, Institute for Mathematical Studies in the Social Sciences, Stanford University, February 1975. Forthcoming in the *Review of Economic Studies.*
6. ——— AND UNTIET, C. "Tacit Coordination of Industry Activity." Draft, Stanford University, March 1974.
7. SWAN, P. L. "Durability of Consumer Goods." *The American Economic Review,* Vol. 60, No. 5 (December 1970).
8. WILLIG, R. D. "Welfare Analysis of Policies Affecting Prices and Products." Memorandum #153, Center for Research in Economic Growth, Stanford University, 1973.