Regulating a Monopolist with Unknown Demand

By TRACY R. LEWIS AND DAVID E. M. SAPPINGTON*

Optimal regulatory policy is derived in a setting where the firm has better knowledge of demand than the regulator. When marginal production costs increase with output, the regulator can induce the firm to use its private information entirely in the social interest. When marginal costs decline with output, however, the regulator is unable to derive any benefit from the firm's superior knowledge, and a single price is established that is invariant to demand.

We examine the design of regulatory policy when the regulated firm has better information about demand than the regulator from the outset of their relationship. Although this seems to be an important and relevant issue to explore, it has been largely overlooked in the literature. Most studies in the recent regulatory literature focus on the setting where a firm's private information concerns its *costs* rather than its demand (see, David Baron and David Besanko, 1984; Baron and Roger Myerson, 1982; Jean-Jacques Laffont and Jean Tirole, 1986; David Sappington, 1983; and Sappington and David Sibley, 1988).

There are a variety of reasons why a regulated firm is likely to have better information about the demand for its product than the regulator.[1] One reason is that the firm generally has superior knowledge of the quality and reliability of the product it is marketing. To the extent that demand is affected by these attributes, the firm is better able to predict the quantity demanded at any price than the regulator. Product quality and reliability are important determinants of demand in industries such as utilities, transportation, and communication, which have traditionally been regulated. Another reason regulated firms will commonly have superior knowledge of demand is because of the significant resources they devote to marketing studies. Although regulatory commissions may have access to the findings of these studies, they generally lack the manpower to fully analyze the findings.[2]

In light of the extensive research that has been devoted to the study of regulatory policy with asymmetric cost information, the corresponding study assuming asymmetric demand information would contribute relatively little if the same qualitative insights arose from the two lines of research. In fact, though, we find striking qualitative differences between the two cases. With asymmetric cost information, significant pricing

*University of California at Davis and Bellcore, 435 South Street, Morristown, NJ 07960-1961. We wish to thank two anonymous referees, Robert Russell, William Sharkey, David Sibley, and seminar participants at Boston, Brown, Duke, and Rutgers Universities, The Universities of Iowa and Southern California, and the Federal Trade Commission for helpful comments. The first author gratefully acknowledges financial support from the California Energy Research Group. The views expressed in this paper are not necessarily those of Bellcore.

[1]The fact that regulators commonly believe the firms they regulate to have superior knowledge of demand is reflected in the following excerpt from a report entitled, "Risk, Return, and Ratemaking," by the California Public Utilities Commission's Policy and Planning Division (October 1, 1986): [The Commission must evaluate... arguments with minimal information about a customer's economic situation. Utilities have greater manpower and closer contact with customers, and can probably make better judgments about the conditions customers face.]

[2]Of course, the extent of the information asymmetry in regulated industries is not entirely exogenous. The regulator can devote resources to improving information about demand, and can influence the resources the firm devotes to market studies. Thus, one may view the analysis in this paper as providing some indication of the value of information about product demand in a regulatory setting.

flexibility is generally delegated to the firm. Although the delegation generally results in rents for the firm, social welfare is enhanced when the firm employs its superior technological information to choose prices which track the movement of efficient prices. The regulator also finds it advantageous to induce prices in excess of realized marginal production costs in order to limit the firm's rents. In addition, the regulator will always choose to shut down the firm (i.e., terminate its operations) when realized costs are so high that the maximum net surplus from continued operation is negative. Shutdown may also be induced for particularly high cost realizations even though the potential social gains from production exceed the costs.

With asymmetric demand information, very different conclusions emerge. To illustrate, when marginal costs of production increase with output, the firm commands no rents from its superior demand information, and efficient prices are always implemented.[3] Although pricing authority is delegated to the firm in this setting, the firm can be induced (costlessly) to employ its superior demand information in the social interest. Furthermore, shutdown will occur when and only when production is not in the social interest. On the other hand, if marginal production costs decline with output, the regulator will generally not delegate any pricing authority to the firm.[4] Instead, the regulator will optimally use only his own imperfect knowledge of demand to set a single price that the firm must charge regardless of demand. This procedure results in a price that exceeds the marginal cost of production for small demand realizations and that falls short of marginal cost for large demand. The firm also receives rents under all but some intermediate levels of demand.

With declining marginal costs of production, the possibility of inefficient shutdown also arises. In contrast to the case with asymmetric cost information, if shutdown is induced, it will be implemented for intermediate rather than extreme realizations of the private demand information. Furthermore, the firm may be permitted to produce for some demand realizations even though the maximum net surplus from production is negative for those demand realizations.

The explanation and derivation of these findings proceeds as follows. In Section I, the basic model under consideration is described and a formal statement of the regulator's problem is presented. The first-best outcome is also introduced as a benchmark. The first-best outcome is the one the regulator would implement if he had perfect knowledge of demand. In Section II, it is shown that the regulator can ensure the first-best outcome despite his imperfect demand information when (and only when) marginal production costs rise with output. In Section III, the case of declining marginal costs is examined. It is shown that the regulator will restrict the firm to a single price in this setting. For simplicity, the possibility of shutdown is ignored in Sections I through III.

In Section IV, two extensions of the basic model are considered briefly. The shutdown decision is one of these extensions. The other is to allow for marginal production costs that are not everywhere declining or everywhere increasing with output. The optimal policy when the marginal cost curve is U-shaped is shown to be the natural "combination" of the policies described in Sections II and III.[5] Conclusions are drawn in Section V.

Before proceeding, we briefly distinguish our work from the few studies in the literature that do consider the possibility of

[3] Even if marginal costs are strictly increasing in output, average costs may still be falling in the relevant range. Thus, even a natural monopoly firm may have rising marginal costs of production.

[4] This conclusion requires that a common regularity condition be satisfied. The condition will be satisfied if, for example, the firm's private information concerns the position but not the shape of the demand curve.

[5] Another straightforward extension of our findings is to the case where the regulator and firm share the same imperfect knowledge of the cost function. Thus, our conclusions do not hinge on the assumption (maintained for expositional convenience) that costs are known perfectly to both parties.

asymmetric knowledge of demand. A key distinction is that our analysis is a Bayesian analysis in that the regulator is assumed to have well-defined prior beliefs about the relevant demand parameter. In contrast, a number of analyses (such as Ingo Vogelsang and Jörg Finsinger, 1979; and Mo-Yin Tam, 1985, for example) assume the regulator knows *nothing* about demand (or cost) functions. In this regard, our work is more closely related to the Bayesian analyses of Michael Riordan, 1984, and Daniel Spulber 1988. Spulber considers a bargaining problem in which consumers have private demand information and the firm has private cost information. Riordan's model is more similar to ours in that the firm (alone) has private information only about demand. However, he restricts attention to the case where marginal production costs are everywhere constant in order to focus on the issue of motivating the optimal choice of capacity.

A third Bayesian model that deserves mention is the general analysis of Roger Guesnerie and Laffont (GL, 1984), which admits the interpretation of private demand information for a regulated firm. Despite a key difference involving a monotonicity condition on the firm's profit function (which (GL) impose but which is not satisfied in our model), many of the proof techniques in (GL) are readily adapted for our purposes. Therefore, we will refer the interested reader at points to (GL) for the proofs of some of our conclusions.

I. Statement of the Model

The regulatory environment under consideration is the following. The cost of producing output q is known to be given by the function $C(q)$. It is also common knowledge that demand for the firm's product is given by $q = Q(p,\theta)$, where p is the unit price for the good. θ captures the firm's private information about consumer demand. Formally, θ is the realization of a random variable described by the density function $g(\theta)$. $g(\theta)$ is assumed to have strictly positive support on the interval $[\underline{\theta},\bar{\theta}]$. $g(\cdot)$ is common knowledge, but the realization of θ is observed only by the firm. The firm observes θ before any interaction with the regulator begins.[6]

In Section III, it will be assumed that the higher the realization of θ, the larger the quantity of output consumers demand at any nonnegative price, p, that is, $Q_\theta(p,\theta) > 0$, where subscripts denote partial derivatives (here and throughout). In this interpretation, θ is a measure of the intensity of demand, or the willingness of consumers to pay for the firm's product. Throughout the analysis, it will be assumed that for all realizations of θ, more output is demanded the lower the regulated price; that is, $Q_p(p,\theta) < 0$. Further, to avoid characterization of local minima rather than maxima, we assume $|C''(Q(p,\theta))| < |Q_p(p,\theta)|^{-1} \quad \forall p,\theta$, when $C''(\cdot) < 0$, where primes denote derivatives. In words, demand curves are assumed to be more steeply sloped than the firm's marginal cost curve when the latter is downward sloping.[7] Finally, for simplicity we abstract from shutdown considerations until Section IV. Thus, we temporarily assume that demand is sufficiently great relative to costs for all realizations of θ that it is preferable to have the firm operate rather than shut down.

The regulator is endowed with the power to establish a unit price for the firm's output and to specify a transfer, T, from consumers to the firm. The funds for the transfer payment might be generated by the fixed component (or "access charge") of a two-part

[6]The firm need not know the demand curve it faces exactly for our basic conclusions to hold. The key point is that the firm's knowledge of demand is better than the regulator's knowledge. Exact knowledge of demand does, however, ensure the socially efficient shutdown decision when marginal costs increase with output. (See Section IV.)

[7]This is a standard type of assumption which ensures the regulator's problem is concave. If the relevant demand curve were flatter than the downward-sloping marginal cost curve over some range of output, the possibility arises that in that range, incremental production costs would exceed the incremental benefits to consumers. In particular, production at a level given by the equality between price and marginal cost could provide a minimum of social surplus rather than a maximum.

tariff, for example.[8] The regulator can observe (and therefore enforce) the regulated price. The quantity sold at the regulated price is presumed prohibitively costly for the regulator to monitor directly. However, the regulator can still be sure that the firm fulfills its mandate to serve all demand at the regulated price. To do so, the regulator need only invite consumers to report any incident in which they were (either) refused service at the established price (or charged a higher price), and penalize the firm for such rationing.

The assumption that the quantity of output sold by the firm is unobservable is an important one, and so warrants explanation.[9] One simple interpretation of the assumption stems from the observation that there are a variety of attributes of a "unit of output" that are not readily monitored by the regulator. For example, telephone calls of a given duration differ markedly in quality depending upon how quickly the connection is made, the amount of static on the line, etc. Thus, our assumption may be interpreted as a statement that it is prohibitively costly for the regulator to monitor precisely the number of units of "quality-adjusted" output that are sold. Alternatively, even if quality considerations are unimportant, it will generally be very costly for the regulator to collect his own data on sales. And if the regulator relies solely on the firm for information, the firm may find it advantageous to engage in "creative bookkeeping" to misrepresent sales data. In what follows, we abstract from these issues by assuming the regulator cannot observe the firm's sales at all. It is worth noting that when marginal costs rise with output, the inability of the regulator to monitor output is of no consequence. (See Corollary 2 below.)

The regulator's objective is to maximize a weighted average of expected consumers' surplus (S) and profit (Π). $\alpha \in (\frac{1}{2},1]$ will represent the weight on consumers' surplus, and $1-\alpha$ the weight on profit.[10] Formally, the regulator's problem [RP] is the following:

$$\underset{p(\theta),T(\theta)}{\text{Maximize}} \int_{\underline{\theta}}^{\bar{\theta}} \{\alpha[S(Q(p(\theta),\theta))-T(\theta)] + [1-\alpha]\Pi(\theta|\theta)\}g(\theta)\,d\theta$$

subject to:

$$\text{(IR)} \quad \Pi(\theta|\theta) \geq 0 \ \forall\theta \in [\underline{\theta},\bar{\theta}],$$

$$\text{(IC)} \quad \Pi(\theta|\theta) \geq \Pi(\hat{\theta}|\theta) \ \forall\theta,\hat{\theta} \in [\underline{\theta},\bar{\theta}],$$

where

$$\Pi(\hat{\theta}|\theta) \equiv p(\hat{\theta})Q(p(\hat{\theta}),\theta) - C(Q(p(\hat{\theta}),\theta)) + T(\hat{\theta}).^{11}$$

In the statement of [RP], $p(\theta)$ and $T(\theta)$ are the price and transfer payment that will be implemented when the firm's private de-

[8] Under this interpretation, the fixed component of the two-part tariff (that consumers must pay regardless of the number of units of the product they ultimately purchase) is assumed not to affect demand.

[9] With the cost function known, if output were observable the firm could simply be compensated for its realized costs. Under this policy, the firm would have no incentive to produce a level of output other than the one most preferred by the regulator. Thus, the problem we analyze is only interesting if the firm's sales cannot be observed perfectly by the regulator. For simplicity, we assume they are completely unobservable to the regulator.

[10] Thus, we focus on the case where the regulator "cares more" ($\alpha > \frac{1}{2}$) about consumers than the firm. The regulator's problem is an uninteresting one when $\alpha < \frac{1}{2}$, since unbounded transfers from consumers to the firm are optimal in this case.

[11] Notice that $\pi_\theta(\hat{\theta}|\theta)$ is positive (negative) when price exceeds (falls short of) marginal cost. This partial derivative will generally change sign in the solution to [RP]. This contrasts with Guesnerie and Laffont (1984, p. 359) where the sign is restricted to be everywhere positive or everywhere negative.

mand information is θ. $Q(p(\theta),\theta)$ is the corresponding production level. The optimal regulatory policy can be interpreted as follows. The regulator allows the firm to select a price-transfer pair from a carefully designed menu of such pairs. The incentive compatibility constraints (IC) ensure that the firm prefers the $\{p(\theta),T(\theta)\}$ pair to any other when its private information is θ. The individual rationality constraints (IR) guarantee the firm nonnegative (extranormal) profit from the chosen price-transfer pair. Thus, the firm's voluntary participation in the regulated industry is ensured.

Of central concern is how the optimal regulatory policy is affected by the regulator's imperfect knowledge of demand. As a benchmark, we record the *first-best* policy, which is the one the regulator would implement if the realization of θ were observed publicly.[12]

Definition 1: The first-best policy consists of prices $p^*(\theta)$ and transfer payments $T^*(\theta)$ that satisfy the following properties $\forall\theta\in[\underline{\theta},\bar{\theta}]$:

$$\text{(a)}\quad p^*(\theta)=C'(Q(p^*(\theta),\theta)),\quad\text{and}$$

$$\text{(b)}\quad T^*(\theta)=C(Q(p^*(\theta),\theta)) - p^*(\theta)Q(p^*(\theta),\theta).$$

Thus, the ideal outcome from the regulator's perspective (for any welfare weights $\alpha\in(\frac{1}{2},1]$ and $1-\alpha$) holds the firm to zero profit and implements marginal cost prices. In the next section, we demonstrate that under cost conditions that are commonly assumed in formal analyses, the regulator can implement the first-best policy even if he has no information about demand.

[12] In subsequent discussion, we will refer to the outcome that results under the first-best policy as the first-best outcome. Similarly, we will refer to the prices that are charged in the first-best outcome as first-best prices. Note that in the definition presented here, strictly positive levels of output are assumed to be in the social interest for all demand (θ) realizations. In Section IV, we will consider the possibility that for some sufficiently small realizations of θ, the firm will not produce at all under the first-best policy.

II. Nondecreasing Marginal Costs

In this section, we examine the solution to [RP] when marginal production costs are everywhere nondecreasing. No restrictions are placed on average costs of production. Thus, in particular, average costs of production may be everywhere declining in the relevant range, so the firm may be a natural monopoly. No restrictions are placed on the demand uncertainty either. In particular, $Q_\theta(p,\theta)$ can be both positive and negative in different regions, so demand curves can "cross." The solution to the regulator's problem in this setting is recorded in Proposition 1.

PROPOSITION 1: *If* $C''(q)\geq 0\ \forall q\geq 0$, *then the solution to* [RP] *is the first-best policy.*

PROOF:

It need only be shown that the first-best policy is a feasible solution to [RP] when $C''(q)\geq 0\ \forall q\geq 0$. To do so, consider any two possible distinct realizations of $\theta:\theta_1,\theta_2\in[\underline{\theta},\bar{\theta}]$ with $\theta_2>\theta_1$. Now, for the purposes of this proof only, define $p_i\equiv p^*(\theta_i)$ and $T_i\equiv T^*(\theta_i)$ for $i=1,2$. It remains to show that when $\theta=\theta_i$, the firm will prefer to implement the first-best price and transfer payment $\{p_i,T_i\}$ rather than any other $\{p,T\}$ pair. To do so, it suffices to show that the following two inequalities hold for any $\theta_1,\theta_2\in[\underline{\theta},\bar{\theta}]$:

$$(1)\quad p_2Q(p_2,\theta_1)-C(Q(p_2,\theta_1))+T_2\leq 0,$$

$$(2)\quad p_1Q(p_1,\theta_2)-C(Q(p_1,\theta_2))+T_1\leq 0.$$

There are three cases to consider.

Case 1. $Q(p_1,\theta_1)=Q(p_1,\theta_2)$.

In this case, $p_1=p_2$, and $T_1=T_2$; so (1) and (2) are obviously satisfied.

Case 2. $Q(p_1,\theta_1)<Q(p_1,\theta_2)$.

In this case, $p_1\leq p_2$ since $Q_p(p,\theta_2)<0$, with strict inequality if $C''(Q(p_1,\theta_1))>0$. From the definition of first-best prices, whenever $p_2>p_1$, we know $Q(p_2,\theta_2)>Q(p_1,\theta_1)$. Therefore, since $Q_p(p,\theta_1)<0$, we have $Q(p_2,\theta_1)<Q(p_1,\theta_1)<Q(p_2,\theta_2)$.

Now, substituting for p_i and T_i, $i=1,2$, from the definition of the first-best policy, inequalities (1) and (2) can be rewritten, respectively, as

$$(3)\quad C'(Q(p_2,\theta_2)) \geq \frac{C(Q(p_2,\theta_2))-C(Q(p_2,\theta_1))}{Q(p_2,\theta_2)-Q(p_2,\theta_1)},$$

$$(4)\quad C'(Q(p_1,\theta_1)) \leq \frac{C(Q(p_1,\theta_2))-C(Q(p_1,\theta_1))}{Q(p_1,\theta_2)-Q(p_1,\theta_1)}.$$

It is apparent that (3) and (4) will hold $\forall\theta_1,\theta_2\in[\underline{\theta},\bar{\theta}]$ with $\theta_2>\theta_1$ if (and only if) $C''(q)\geq 0\ \forall q\geq 0$.

The proof for Case 3, where $Q(p_1,\theta_1)>Q(p_1,\theta_2)$ follows in similar fashion.

To achieve the first-best outcome without knowledge of demand in this case, the regulator need only ask the firm to report the actual demand curve, and then implement the first-best price and transfer payment conditional on the firm's report. The firm will have no incentive to misrepresent its private demand information under this arrangement when marginal costs of production increase with output.[13] Demand exaggeration results in authorization to charge a higher unit price for output; but any revenue enhancement is more than offset by the associated reduction in transfer payment. And should the firm understate demand, the loss in revenue that results from the lower regulated price more than offsets the increment in allowed transfer payment.

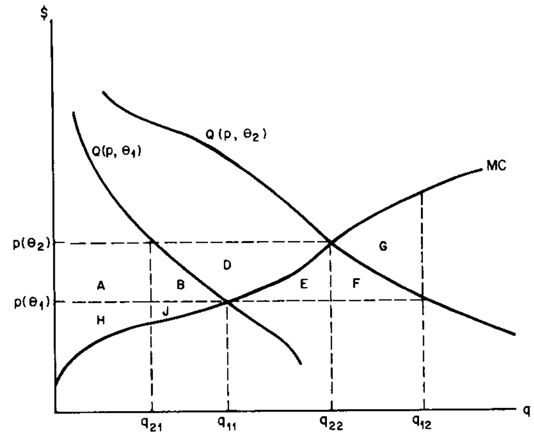

FIGURE 1. INCENTIVES WITH INCREASING MARGINAL COSTS

To examine these effects in more detail, consider Figure 1, which is drawn for the case where $Q_\theta(p,\theta)>0\ \forall p,\theta$. Two representative demand curves are depicted, one for "low" demand (θ_1) and one for "high" demand (θ_2). Quantities are represented on the horizontal axis, with $q_{ij}=Q(p(\theta_i),\theta_j)$, $i,j=1,2$. The first-best policy will implement marginal cost prices, $p^*(\theta_i)$, and award transfer payments, $T^*(\theta_i)(i=1,2)$, which are calculated to provide exactly zero profit to the firm when θ_i is the realized demand parameter. Thus, $T^*(\theta_1)$ will be equal to the fixed costs of production less areas H and J in Figure 1. $T^*(\theta_1)$ will exceed $T^*(\theta_2)$ by the sum of areas A, B, and D.

The conclusion in Proposition 1 can be explained by demonstrating that the firm will strictly prefer to set price $p^*(\theta_i)$ and receive transfer $T^*(\theta_i)$ when $\theta=\theta_i$ rather than select the alternative pair $\{p^*(\theta_j), T^*(\theta_j)\}$, $j\neq i$, $i,j=1,2$. First, consider the change in profit for the firm when $\theta=\theta_1$, but price $p^*(\theta_2)$ is charged. The higher price provides greater revenues on q_{21} units of output. This gain is given by area A in Figure 1. The firm also suffers a direct loss given by area J, since the higher price selected leads to a reduction in demand. In addition, the firm's transfer payment is re-

[13] The ensuing discussion is focused on the case where marginal production costs increase strictly with output. When marginal costs are constant, the first-best policy is readily implemented with an allowed price set equal to the constant marginal cost of production and a transfer payment equal to the fixed costs of production. (See Riordan, 1984.)

duced by the sum of areas A, B, and D when the higher price is selected. Hence, the net loss to the firm from choosing the $\{p^*(\theta_2), T^*(\theta_2)\}$ pair rather than $\{p^*(\theta_1), T^*(\theta_1)\}$ when $\theta = \theta_1$ is the sum of areas B, D, and J.

Now, suppose $\theta = \theta_2$ and the firm chooses to set the lower price $p^*(\theta_1)$ rather than $p^*(\theta_2)$. A direct loss in revenues given by the sum of areas A, B, D, and E results, which more than offsets the increment in transfer payments (areas A, B, and D) that is provided when the lower price is charged. An additional loss given by areas F and G results because the cost of producing the incremental demand $(q_{12} - q_{22})$ exceeds the allowed revenues.

It is also straightforward to verify that the conclusion of Proposition 1 is not altered if the firm cannot be prevented from rationing customers. In Figure 1, when θ_2 is the true demand parameter and the firm chooses to set unit price $p(\theta_1)$, the profit-maximizing output level is q_{11}. But at this level of production, the loss in net revenues exactly offsets the increment in transfer payments that is generated by setting the lower price.

COROLLARY 1: *If $C''(q) \geq 0\ \forall q \geq 0$, then the regulator can implement the first-best policy even if he cannot detect rationing of customers.*[14]

Since the first-best outcome is attained in this setting, no strict gains would be realized if it were possible to verify the firm's sales exactly.

COROLLARY 2: *Provided $C''(q) \geq 0\ \forall q \geq 0$, any information about realized demand in the industry has no value to the regulator.*

It remains to determine whether the first-best policy might be feasible even when marginal production costs decline with output over some range. As noted in the proof of Proposition 1, this is not the case.

COROLLARY 3: *The first-best policy is a feasible solution to* [RP] *only if $C''(q) \geq 0\ \forall q \geq 0$.*

Hence, it remains to characterize the optimal regulatory policy when marginal production costs decline with output. The analysis begins in Section III.

III. Decreasing Marginal Cost and Single Crossing

The main conclusion to be drawn from Section II is that with rising marginal costs of production, the firm's incentives can be made to coincide with those of society, so the information asymmetry between regulator and firm regarding demand is largely inconsequential. Notice that this conclusion was shown to hold even though virtually no restrictions were placed on the nature of the information asymmetry.

Matters are somewhat more complicated with declining marginal costs. Here, additional structure is required for a complete characterization of the solution to [RP]. To begin, we suppose higher realizations of θ correspond to greater demand at all prices; that is, $Q_\theta(p, \theta) > 0\ \forall p, \theta$. Furthermore, we impose the single-crossing property (SCP), which is commonly assumed in self-selection problems of the type considered here. (See Russell Cooper, 1984, for a survey and general treatment.) Intuitively, the SCP guarantees the firm is systematically more willing (or systematically less willing) to forego transfer payments to obtain a higher unit price the greater is demand. As reported below, the SCP will hold, for example, when increases in θ correspond to parallel outward shifts in the demand curve.

When the SCP holds, it turns out that the firm must be induced to select a higher price the greater is demand under the optimal regulatory policy. (See Lemma 2 below). Thus, an obvious conflict arises between

[14] The firm will actually be indifferent between implementing the marginal cost price while serving all demand and charging a lower price while limiting output to the profit-maximizing level. Thus, the "loss" that arises when rationing cannot be prevented is that the firm may only weakly prefer to implement the first-best policy, whereas the preference will be strict when rationing can be detected and punished.

feasible prices and first-best prices (which fall with demand when marginal costs are declining). As is reported in Proposition 2, this conflict is always resolved by instructing the firm to set the same price for its product, independent of its private demand information. In effect, the firm's superior knowledge of demand must be ignored when marginal costs decline with output because it is too costly to make the firm's interests coincide with those of society. Rather than delegate any pricing authority to the firm, the regulator uses his own imperfect demand information to establish the regulated price.

To state these conclusions formally, two definitions are helpful.

Definition 2: A price schedule, $p(\theta)$, is *incentive compatible* if it (along with the associated transfer schedule, $T(\theta)$) satisfies the (IR) and (IC) constraints in [RP].

Definition 3: The *single-crossing property* (SCP) holds if the firm's marginal rate of substitution of price for transfer payment ($\Pi_p(p,\theta)$) is monotonic in θ $\forall\theta\in[\underline{\theta},\bar{\theta}]$, and strictly monotonic for some $\theta\in[\underline{\theta},\bar{\theta}]$, where $\Pi(p,\theta)\equiv pQ(p,\theta)-C(Q(p,\theta))+T(\theta)$.[15]

As noted above, the SCP ensures the firm's preferences vary smoothly and consistently with its private information. Lemma 1 reports that if the firm's private information pertains only to the position of the demand curve and not to its shape, then the SCP is automatically satisfied. (The proof of the Lemma follows from straightforward differentiation of $\Pi(p,\theta)$.)

LEMMA 1: *If* $Q_{p\theta}(p,\theta)=0$ $\forall\theta\in[\underline{\theta},\bar{\theta}]$ *and* $\forall p\geq 0$, *then* $\Pi_{p\theta}(p,\theta)\geq 0$ $\forall p,\theta$, *so the* SCP *holds.*

Thus, any setting in which higher realizations of θ correspond to parallel outward shifts in demand constitutes a setting where the results cited in this section are valid.[16] Of course, the single-crossing property will also be satisfied more generally.

An implication of the SCP is the following.

LEMMA 2: *If the* SCP *holds, then* $p'(\theta)\geq 0$ $\forall\theta\in[\underline{\theta},\bar{\theta}]$ *in any incentive-compatible price policy.*

The intuition that underlies Lemma 2 is fairly straightforward. (The proof is omitted. See Guesnerie and Laffont, 1984, for the essential details.) To illustrate, take the case where price exceeds marginal production cost. If the firm is to be induced to set a lower price ($p_2<p_1$) in this case, it must receive a higher transfer payment ($T_2>T_1$). The increase in T required to offset the decrease in revenues will have to be relatively large when demand is high ($\theta=\theta_2$). However, for smaller realizations of demand ($\theta_1<\theta_2$), the same drop in price results in a smaller decrease in revenues. Therefore, if the firm (weakly) prefers $\{p_2,T_2\}$ to $\{p_1,T_1\}$ when $\theta=\theta_2$, the preference will be strict when $\theta=\theta_1$. Consequently, the price charged by the firm in any incentive-compatible policy cannot decline with demand.

But when the regulated price must rise as demand increases to ensure incentive compatibility, price is proceeding in a direction opposite to that of the first-best price. Thus, allowing the firm to use its private knowledge of demand to select the market price becomes very costly...so costly, in fact, that no delegation of pricing authority is allowed.

[15]As its name implies, the SCP guarantees that any two iso-profit curves for the firm in $p-T$ space, each corresponding to a different realization of θ, intersect at most once. The additional structure implied by the SCP facilitates the characterization of incentive-compatible price policies: provided the (IC) constraints in [RP] are satisfied locally, they will also be satisfied globally when SCP holds.

[16]One setting in which $Q_{p\theta}(p,\theta)=0$ $\forall p,\theta$ is the following. Suppose the utility function $H(\cdot)$ of every consumer is separable in income (y) and the quantity (q) of the regulated commodity consumed, and is of the form: $H(q,y,\theta)=h(y)+z(q+\theta)$, where $h(\cdot)$ and $z(\cdot)$ are increasing functions of their arguments. θ might, for example, represent an endowment of the regulated commodity. It is straightforward to verify that the stated property of aggregate demand for the regulated commodity is satisfied in this setting.

PROPOSITION 2: *Suppose* $C''(q) < 0 \; \forall q \geq 0$ *and the* SCP *holds. Then in the solution to* [RP], *the same price (and transfer payment) will be implemented for all realizations of demand.*

The stark contrast between Propositions 1 and 2 warrants emphasis. (The proof of Proposition 2 is in the Appendix.) When marginal production costs rise with output, the firm is readily motivated to use its private knowledge of demand to implement the outcome that is socially most preferred. On the other hand, when marginal costs decline with output in the setting under consideration, it is so costly to induce the firm to use its superior knowledge in the social interest that any attempt to do so is abandoned: the regulator relies solely on his imperfect demand information to structure regulatory policy.[17]

The single price that is implemented by the regulator depends upon his prior beliefs, upon the firm's cost structure, and upon α. In order to limit the firm's rents, the price generally differs from the price that maximizes expected total surplus. The extent and nature of the difference can be shown to depend upon the third derivative of the firm's cost function.[18]

IV. Extensions of the Model

To facilitate an intuitive understanding of the essential elements of the optimal regulatory policy, a number of simplifying assumptions were maintained in the preceding sections. In particular, marginal costs were assumed to be everywhere declining or everywhere increasing. Also, the possibility of shutting the firm down altogether was ignored. The purpose of this section is to extend the analysis to consider the possibilities of U-shaped marginal cost curves and shutdown.[19]

We begin by examining the optimal regulatory policy when marginal costs of production initially decline over some range of output ($q \in [0, q^L)$) and then rise for higher output levels ($q \in (q^L, \infty)$). We continue to assume the firm's cost structure is common knowledge, and the single-crossing property (SCP) holds. For simplicity, the possibility of shutdown is not considered yet.

The solution to [RP] in this case is illustrated in Figure 2.[20] As is apparent, the optimal price schedule $p^R(\theta)$ is a "combination" of the schedules that are optimal when marginal costs are everywhere declining and when they are everywhere increasing. A constant price, $\bar{p}$, will be charged whenever the realized demand curve intersects the marginal cost curve in its declining range (i.e., for $\theta \in [\underline{\theta}, \theta^L]$, where $p^{*\prime}(\theta) \lesseqgtr 0$ as $\theta \lesseqgtr \theta^L$). This same price will also be implemented when the point of intersection lies in the region where marginal costs begin to increase (i.e., where $p^{*\prime}(\theta) > 0$). However, after some point ($\tilde{\theta}$) in the region where marginal costs (and thus first-best prices) are rising, the first-best prices will be charged.[21]

[17]A similar conclusion is reported in Guesnerie and Laffont (1984). The authors present conditions under which a social planner will "ignore" any productivity information possessed by workers in a labor-managed public firm. When these conditions are satisfied, the social planner will simply mandate a fixed level of labor supply for all possible realizations of productivity.

[18]Suppose expected total surplus is a single-peaked function of price; and suppose $Q_{p\theta}(p, \theta) = 0 \; \forall p, \theta$. Then when $C'''(q) > 0 \; \forall q$, the regulator will set price above the level (p^w) that maximizes expected total surplus. The optimal regulated price will be below p^w when $C'''(q) < 0$. And if production costs decline linearly with output, the regulator will implement price p^w. These conclusions hold regardless of the regulator's preferences concerning the distribution of surplus between consumers and the firm (i.e., they hold $\forall \alpha \in (\frac{1}{2}, 1]$).

[19]As noted in fn. 5, the conclusions derived above also carry over in straightforward fashion to the case where the cost function is not known with certainty, but where the regulator and firm share the same imperfect knowledge of costs.

[20]A formal proof of the results that underlie Figure 2 is omitted, as the techniques employed are analogous to those explored by Guesnerie and Laffont (1984).

[21]Note that it is possible to have $\tilde{\theta} = \bar{\theta}$. This is most likely to be the case when $\bar{\theta} - \theta^L$ is small (so that marginal costs decline with output over most of the relevant region) and when $p^*(\bar{\theta}) - p^*(\theta^L)$ is small (so that the ascent of marginal costs is not steep in the region where they are rising). When $\tilde{\theta} = \bar{\theta}$, $p^R(\theta) = \bar{p} \; \forall \theta \in [\underline{\theta}, \bar{\theta}]$, as in the case where marginal costs are everywhere declining.

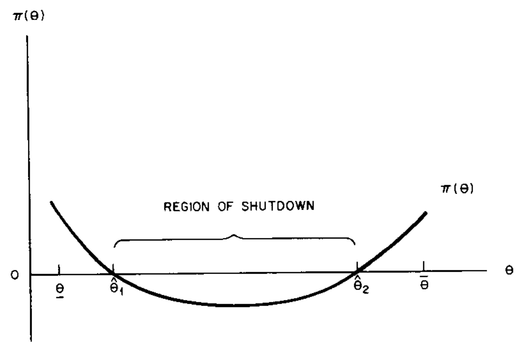

FIGURE 2. U-SHAPED MARGINAL COSTS

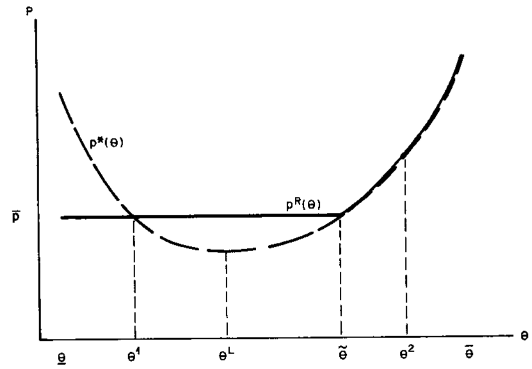

FIGURE 3. DECLINING MARGINAL COSTS AND SHUTDOWN

In the region $[\tilde{\theta}, \bar{\theta}]$ where first-best prices are charged, the firm's profit will be invariant to the realization of θ, as was the case in Section II. However, to induce the firm to set the higher first-best price rather than $\bar{p}$ when $\theta \in [\tilde{\theta}, \bar{\theta}]$, the firm must be afforded the rents it could secure if $\tilde{\theta}$ were the realized demand parameter. Hence, the transfer payment to the firm for setting a price above $\bar{p}$ includes $\Pi(\tilde{\theta} \mid \tilde{\theta})$.

The second extension of the basic model we consider is the shutdown decision. Consider, first, the case in which marginal production costs are known to be everywhere nondecreasing with output. In this case, recall, the firm earns no profit if it operates. Hence, if it is provided with exactly zero profit when it shuts down, the firm will have no incentive to exaggerate demand to avoid being shut down; nor will it gain by terminating operations when production is in the social interest. Thus, when marginal productions costs increase with output, the first-best outcome can be ensured while delegating all pricing and operating decisions to the firm.[22]

Now consider the case where marginal production costs decline with output and the single-crossing property is satisfied. It is not difficult to show that in this case, the regulator will continue to offer a single price and transfer payment for all realizations of demand, even when shutdown may result. Therefore, since the firm's profit increases with θ at a rate directly proportional to the difference between price and marginal cost (see equation (A1) in the Appendix), profit will decline with θ for small demand realizations, and increase with θ for large demand realizations. Consequently, the firm's profit will be minimized, and hence the (IR) constraints will bind, at *intermediate* realizations of θ. Therefore, whenever the solution to [RP] imposes shutdown, production will be absent for some intermediate realizations of demand, as illustrated in Figure 3.[23]

There are a number of interesting implications of this fact. First, the regulator generally cannot induce the firm to terminate operations only when it is socially desirable to do so (i.e., when θ is sufficiently small that the net surplus from operation is negative); additional shutdown may be unavoidable for realizations under which production is so-

[22] We make the common assumption here that when it is indifferent among actions, the firm chooses in the social interest.

[23] The fact that shutdown occurs for *intermediate* rather than extreme realizations of θ warrants emphasis, in part because it is in contrast to the findings in related studies in the regulated literature. (See, for example, Baron and Myerson (1982) and Guesnerie and Laffont (1984).) The difference stems from the aforementioned monotonicity of the firm's profit function in θ that is imposed in other studies, but is not imposed here. (Recall fn. 11.)

cially desirable. Consequently, because this undesired but unavoidable shutdown can be very costly, the fact that net social losses will result from operation for small realizations of θ is not sufficient to ensure shutdown will occur for these realizations. On the other hand, circumstances may arise in which shutdown is induced even though the firm generates positive net social surplus for all demand realizations.[24] Although shutdown does impose some direct losses in this setting, it also reduces the rents the firm can command when it does operate. This latter effect guarantees that if the expected losses from terminating operations are sufficiently small (as when $g(\theta)$ is small for intermediate values of θ), then some degree of shutdown will be optimal. In summary, we have:

PROPOSITION 3: *The fact that the net surplus from efficient operation is strictly negative for some realizations of θ is necessary and sufficient for shutdown to be implemented for those realizations when $C''(q) \geq 0$ $\forall q \geq 0$. It is neither necessary nor sufficient for shutdown when $C''(q) < 0$ $\forall q \geq 0$. In this latter case, if shutdown does occur, it occurs for intermediate demand realizations $[\hat{\theta}_1, \hat{\theta}_2] \subset [\underline{\theta}, \bar{\theta}]$.*[25]

V. Conclusions

We have analyzed the optimal regulatory policy when the firm is endowed with superior information about demand, but the regulator and firm share the same technological information. The optimal policy was shown to be sensitive to the firm's technology and to the nature of the demand uncertainty. When marginal costs of production increase with output, the information asymmetry about demand is inconsequential for the regulator. He can (costlessly) induce the firm to employ its superior knowledge in the social interest, thereby implementing the first-best outcome.

On the other hand, when marginal costs decline with output under the demand conditions we have posited, no pricing authority is delegated to the firm.[26] The regulator uses his own imperfect information exclusively to establish the regulated price. In the case of declining marginal costs, the fact that the maximum net surplus from production under the realized demand curve is negative is neither necessary nor sufficient for shutdown of the regulated firm. And when shutdown is optimal, it will generally occur for intermediate rather than extreme realizations of demand.

An important implication of these findings is that the optimal regulatory policy will differ in important qualitative dimensions when the firm is privately informed about demand rather than about costs. Thus, before attempting to draw any general conclusions about appropriate regulatory policy in the presence of asymmetric information, it is imperative that the nature of the information asymmetry be carefully identified.

An important possibility in practice is that the firm will have better information about both its demand and cost functions. It can be shown that if the firm has better information about both demand and the fixed costs

[24] Recall that similar predictions are derived from regulatory models with cost uncertainty. See, for example, Baron and Myerson (1982).

[25] It is also possible to show that any region of shutdown $[\hat{\theta}_1, \hat{\theta}_2]$ will be a nondegenerate interval. To see this, note from equation (A1) in the Appendix that if shutdown occurred for a single realization of θ, say $\hat{\theta}^m$, then $p(\hat{\theta}^m) = C'(Q(p(\hat{\theta}^m), \hat{\theta}^m)$, $\Pi(\hat{\theta}^m) = 0$, and $\Pi'(\hat{\theta}^m) = 0$ would all be true. Thus, since shutdown first occurs at a point of efficient operation, social losses from shutdown initially rise at an infinite rate. And since $\Pi'(\theta) = 0$ at $\theta = \hat{\theta}^m$, there is initially no reduction in the firm's expected rents from shutdown. Therefore, since the optimal extent of shutdown is that at which the expected marginal gains and losses are equated, shutdown must occur over a nontrivial interval if it occurs at all.

[26] If the firm's private information concerns both the position and shape of the demand curve in such a way that the SCP is not satisfied, then the optimal regulatory policy may entail prices which decline with demand when marginal production costs decline with output. Consequently, it may be possible to induce the firm to use its private information to set prices which vary with θ in the direction that first-best prices vary with θ. Thus, some delegation of pricing authority may be optimal. A complete characterization of the optimal regulatory policy in this case is problematic because the (IC) constraints in [RP] will generally not hold globally even though they hold locally.

of production, but the constant marginal cost of production is common knowledge, the optimal regulatory policy will be as described by Baron and Myerson (1982): price will be set equal to marginal cost, so the only inefficiency that will arise comes in the form of shutting down the firm in cases where total surplus is strictly positive. Thus, the presence of asymmetric information about demand does not alter or seriously complicate optimal regulatory policy in this case because the demand and cost uncertainty are effectively "separable." Absent this separability, however (as when there is asymmetric information both about demand and about marginal production costs), interesting interactions between the two types of uncertainty arise. (For details, see Lewis and Sappington (1988).)

APPENDIX

PROOF OF PROPOSITION 2.

From Lemma 2, $p'(\theta) \geq 0\ \forall\theta \in [\underline{\theta},\bar{\theta}]$ in any feasible solution to [RP]. Let $p(\theta)$ represent the proposed solution to [RP], and let $p^*(\theta)$ represent the first-best price schedule. Also let $\theta^0 \in [\underline{\theta},\bar{\theta}]$ represent the value of θ at which the two schedules, $p(\theta)$ and $p^*(\theta)$, intersect. Since $C''(q) < 0\ \forall q \geq 0$, the schedules intersect at most once on the domain $[\underline{\theta},\bar{\theta}]$. We will return to consider the possibility that they do not intersect at all. The proof is completed for the case where the schedules intersect by showing that the regulatory policy $p^R(\theta) = p^*(\theta^0) = p(\theta^0)$ and $T^R(\theta) = -p(\theta^0)Q(p(\theta^0),\theta^0) + C(Q(p(\theta^0),\theta^0))\ \forall\theta \in [\underline{\theta},\bar{\theta}]$ is a feasible schedule that is strictly preferred by the regulator to $p(\theta)$.

Since $\{p^R(\theta), T^R(\theta)\}$ consists of a single price and transfer payment, it is obviously incentive compatible. Now, using techniques that are standard in the literature (see, for example, Baron and Myerson (1982), Cooper (1984), and Guesnerie and Laffont (1984)), it is readily shown that in any feasible solution to [RP], $\{p(\theta), T(\theta)\}$, that satisfies the (IC) constraints,

$$\text{(A1)}\quad \Pi'(\theta) = [p(\theta) - C'(Q(p(\theta),\theta))] \times Q_\theta(p(\theta),\theta)\quad \forall\theta \in [\underline{\theta},\bar{\theta}].$$

Therefore, the proposed regulatory policy also satisfies (IR) because $\Pi(\theta^0) \equiv \Pi(\theta^0|\theta^0) = 0$ and $p^R(\theta) < C'(Q(p^R(\theta),\theta))$ for $\theta \in [\underline{\theta},\theta^0)$ and $p^R(\theta) > C'(Q(p^R(\theta),\theta))$ for $\theta \in (\theta^0,\bar{\theta}]$.[27]

Next note that by construction, the gap between price and marginal cost is (weakly) reduced under the proposed $p^R(\theta)$ schedule. This has two effects. First, expected total surplus is increased. Second, from (A1), the rate at which the firm's realized profit increases with reductions in θ below θ^0 and with increases in θ above θ^0 is everywhere smaller under $p^R(\theta)$ than under $p(\theta)$. Hence, with expected total surplus increased and the firm's share of the total reduced under the $p^R(\theta)$ schedule, it will provide a larger value for the objective function in [RP], contradicting the presumption that $p(\theta)$ is a solution to [RP].

Finally, notice that if the proposed solution $p(\theta)$ lies everywhere above (below) the first-best price schedule, $p^*(\theta)$, the preceding arguments carry through in analogous fashion with $p^R(\theta) = p^*(\underline{\theta})(p^*(\bar{\theta}))\ \forall\theta \in [\underline{\theta},\bar{\theta}]$. For an alternative proof of a similar result, see Guesnerie and Laffont (1984).

[27]Standard analyses also reveal that if the SCP holds, any regulatory policy that satisfies (A1) also satisfies the (IC) constraints in [RP].

REFERENCES

Baron, David and Besanko, David, "Regulation, Asymmetric Information, and Auditing," *Rand Journal of Economics,* Winter 1984, *15*, 447–70.

______, **and Myerson, Roger,** "Regulating a Monopolist with Unknown Costs," *Econometrica,* July 1982, *50*, 911–30.

Cooper, Russell, "On Allocative Distortions in Problems of Self-Selection," *Rand Journal of Economics,* Winter 1984, *15*, 568–77.

Guesnerie, Roger and Laffont, Jean-Jacques, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics,* July 1984, *25*, 329–69.

Laffont, Jean-Jacques and Tirole, Jean, "Using Cost Observation to Regulate Firms," *Journal of Political Economy,* June 1986, *94*, 614–41.

Lewis, Tracy and Sappington, David, "Regulating with Unknown Demand and Cost Functions," *Rand Journal of Economics*, Winter 1988, *19*, forthcoming.

Riordan, Michael, "On Delegating Price Authority to a Regulated Firm," *Rand Journal of Economics,* Spring 1984, *15*, 108–15.

Sappington, David, "Optimal Regulation of a Multiproduct Monopoly with Unknown Technological Capabilities," *Bell Journal of Economics,* Autumn 1983, *14*, 453–63.

______, **and Sibley, David,** "Regulating Without Cost Information: The Incremental Surplus Subsidy Scheme," *International*

Economic Review, May 1988, *29*, 297–306.

Spulber, Daniel, "Bargaining and Regulation with Asymmetric Information about Demand and Supply," *Journal of Economic Theory,* April 1988, *44*, 251–68.

Tam, Mo-Yin, "Reward Structures in a Planned Economy: Some Further Thoughts," *Quarterly Journal of Economics,* February 1985, *50*, 279–90.

Vogelsang, Ingo and Finsinger, Jörg, "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," *Bell Journal of Economics,* Spring 1979, *10*, 157–71.