Controlling the Activities of Regulated Firms in Unregulated Markets

Tracy R. Lewis
David E. M. Sappington

Columbia Institute for Tele-Information
Graduate School of Business
809 Uris Hall
Columbia University
New York, New York 10027
(212) 854-4222

Abstract

We consider the role that entry into unregulated markets can play in controlling the incentive a regulated firm may have to exaggerate production costs, c, in the regulated market. We associate entry into unregulated markets with the acquisition of critical inputs for production in those markets. When the ultimate value of these inputs is positively correlated with c, an exaggeration of c amounts to a claim that profits in the unregulated markets are greater than they actually are. Hence, incentives to exaggerate c are mitigated. We also demonstrate the optimality of simple regulatory rules, wherein inflexible pricing rules are implemented and only limited discretion is afforded the firm.

1. Introduction

A topic of great current concern is whether firms that formerly operated solely in a single regulated market should be permitted to expand into other markets, possibly unregulated ones. In the telecommunications industry, for example, some argue that the local operating companies should be allowed to provide information processing services as well as information transmission services. Some also question whether these firms shouldn't be permitted to enter the market for cable television services.

Our paper is designed to examine the extent to which regulated firms are optimally permitted to participate in other markets, hereafter referred to as "unregulated" markets. Our particular focus is on how such participation can reduce the rents the firm would command from its private technological information if it operated solely in the regulated industry. There are studies in the literature that focus on how to best control the rents of a firm that operates in a single regulated market when the firm is endowed with better technological information than the regulator. (See, for example, Baron and Myerson [1982], Sappington [1983], and Baron and Besanko [1984].) Such studies tend to prescribe fairly complex regulatory rules. In particular, they generally recommend considerable delegation of pricing authority to the regulated firm. In practice, regulatory rules tend to be simpler, and the pricing authority delegated to the firm more restricted. Thus, a discrepancy generally emerges between theoretical prescription and the policy that is actually implemented.

In our model, however, prescription and practice appear more congruent. For a range of possible cost levels, the same price is set in the regulated

market. Only for particularly high or low realizations of cost will the regulator delegate pricing discretion to the firm.

The inflexible pricing rule is optimal in our model because of "countervailing incentives." When the firm operates only in the regulated market, it will have an incentive to exaggerate production costs in an attempt to secure greater compensation from the regulator. Participation in the unregulated market, on the other hand, will generally create an incentive for the firm to understate its earnings in that market. Hence, if entry is managed so that earnings in the unregulated market are positively correlated with costs in the regulated market, countervailing incentives are created. The firm's incentive to exaggerate production costs in the regulated market will be mitigated because such an exaggeration now implies a claim that participation in the unregulated market is more profitable than it really is. Therefore, the regulator's task of controlling cost exaggeration in the regulated market can be made less burdensome by allowing participation in unregulated markets. And the creation of incentives for the firm to exaggerate low cost realizations and to understate high realizations turns out to imply that a single price will always be enforced under the optimal regulatory policy for a range of intermediate cost realizations.

Before proceeding to develop these conclusions more carefully and more formally, we should point out that our model is designed to focus on a particular aspect of the complex decision about whether to allow regulated firms to participate in unregulated markets. Consequently, we abstract from other important considerations that warrant careful study in any complete analysis of the topic. For example, we do not consider the possibility that the regulated firm might have an unfair competitive advantage in the unregulated

market, or that it might permit the quality of its regulated product to deteriorate when its attention is focused on new markets.

In summary, there are two features of our model that are most important: the management of entry by the regulated firm into unregulated markets, and the optimality of simple, inflexible regulatory rules. We develop and explain these features as follows: In section 2, the model under consideration is described in detail. Our conclusions are reported in section 3, and an explanation of our findings is also offered there. Concluding remarks are contained in section 4.

2. The Model

There are two markets in the model we consider: a regulated and an unregulated market. There are also two primary actors: the regulator and the firm. One of the regulator's tasks is to dictate the extent to which the firm is allowed to participate in the unregulated market. The regulator also determines the price the firm can charge for its product in the regulated market. The regulator cannot dictate the price charged in the unregulated market.

The environment is characterized by imperfect and asymmetric information about the firm's operating costs. We let c represent the firm's realized constant marginal cost of producing the regulated product. Initially, neither the firm nor the regulator know the exact realization of c. Both parties share the same beliefs about this cost parameter. Their beliefs are represented by the density function f(c) that has strictly positive support on the interval $[\underline{c}, \overline{c}]$. F(c) is the corresponding distribution function for c. To simplify the analysis, the following regularity condition is imposed:

(RC) $\frac{d}{dc}\left\{\frac{F(c)}{f(c)}\right\} > 0$ and $\frac{d}{dc}\left\{\frac{1-F(c)}{f(c)}\right\} < 0$ $\forall\, c\in[\underline{c},\bar{c}]$.

This condition rules out beliefs that are not well behaved and thus would unduly complicate the analysis.[1] Notice that the uniform distribution, $f(c) = \frac{1}{\bar{c}-\underline{c}}\ \forall\, c\in[\underline{c},\bar{c}]$, satisfies (RC).

At this initial stage when costs are uncertain, the terms of the regulatory policy are announced. The policy includes a menu of possible unit prices, p, for the regulated product and associated taxes, T, on the firm.[2] The firm is ultimately permitted to select the {p,T} pair it desires after it learns its operating costs. Since the demand function Q(p), for the regulated product is known to both the regulator and the firm, the selected price will determine a known level of output that the firm must produce to satisfy all demand in the regulated market.

When he specifies the menu of prices and taxes in the regulated market, the regulator also specifies and installs a level of capacity, K, that dictates the extent to which the firm can participate in the unregulated market. Installing capacity is assumed to be a time consuming process; so K must be chosen before all cost uncertainty is resolved. For simplicity, we assume exactly one unit of capacity is required to produce one unit of the product sold in the unregulated market. The regulator "buys" capacity at unit cost c_e, which is the expected unit value of employing the capacity in the unregulated market.[3] Given this "fair" valuation of the capacity by the market place, the regulator views this venture into the unregulated market as a "breakeven" operation... he does not anticipate strictly positive profits for the firm in the unregulated sector that can be used to subsidize

operations in the regulated sector. Therefore, if the regulator does choose to allow participation in the unregulated market by choosing $K > 0$, he does so solely to limit the rents the firm could otherwise secure in the regulated market.

The realized unit return from employing the capacity in the unregulated market is positively correlated with the costs of producing in the regulated market. For simplicity, we assume perfect correlation: Thus, c is both the realized unit cost of production in the regulated market and the unit profit from employing the capacity in the unregulated market. This direct relationship might arise, for example, if the firm purchases its productive inputs on the spot market and the inputs employed in the regulated sector are substitutes for the capacity used to operate in the unregulated market.

After capacity is installed and a start-up phase is completed in the regulated market, the firm learns how costly it will be to operate in the regulated market, i.e., the firm learns the realization of c. At the same time, the firm also learns the unit profit (c) that can be earned by employing its capacity (K) in the unregulated market. Then, the firm chooses from the menu presented by the regulator the price - tax pair that maximizes its profit. All demand, Q, that is forthcoming at the selected price in the regulated market is then produced by the firm. The firm also chooses its level of participation in the unregulated market at this time. Since $c > 0 \; \forall \; c \in [\underline{c}, \overline{c}]$, the firm will always employ all of its capacity in the unregulated market. This timing in the model is summarized in Figure 1.

The regulator designs the regulatory policy to maximize the expected level of consumers' surplus, $S(\cdot)$, in the regulated sector. The regulator knows that to induce the firm to produce the regulated product, it must be

guaranteed the level of profit it could earn by operating elsewhere. This level of profit is normalized at zero. Notice that the regulator can influence the firm's realized profit through his choice of capacity. A large capacity increases the chances that the firm will be able to generate some profit in the unregulated sector to complement its earnings in the regulated sector.

The revelation principle (e.g., Myerson [1979]) ensures there is no loss of generality in representing the equilibrium price and tax as functions of the firm's truthful report of c. In other words, the firm can be thought of as announcing the cost realization, $\hat{c}$, and the regulator as implementing the price-tax pair $\{p(\hat{c}), T(\hat{c})\}$ that he initially promised to implement if he received the report of $\hat{c}$. Formally, the regulator's problem [RP] is the following:

$$\underset{p(c),\, T(c),\, K}{\text{Maximize}} \quad \int_{\underline{c}}^{\overline{c}} [S(p(c)) + T(c)]\, f(c)\, dc - c_e K$$

subject to

$$\int_{\underline{c}}^{\overline{c}} \pi(c)\, f(c)\, dc \geq 0 , \qquad \text{(IR-1)}$$

$$\pi(c) \geq 0 \qquad \forall\, c \in [\underline{c},\overline{c}], \qquad \text{(IR-2)}$$

$$\pi(c/c) \geq \pi(\hat{c}/c) \qquad \forall\, c,\, \hat{c} \in [\underline{c},\overline{c}], \qquad \text{(TR)}$$

where

$$\pi(\hat{c}/c) \equiv [p(\hat{c}) - c]\, Q(p(\hat{c})) - F_0 - T(\hat{c}) + c\, K,$$

$\pi(c) \equiv \pi(c/c)$, and

F_0 = the fixed costs of producing in the regulated market.[4]

The ex ante individual rationality constraint (IR-1) ensures that the firm will initially agree to the terms of the regulatory policy announced by the regulator. The firm will do so provided its expected profit under the policy is nonnegative. The ex post individual rationality constraint (IR-2) reflects the fact that after learning its operating costs, the firm can "withdraw" from the regulatory charter without penalty, i.e., it can produce no output in the regulated sector and in return receive zero compensation and forfeit its right to employ the installed capacity in the unregulated market. Notice that this restriction rules out the possibility that the firm might be required to post a large financial bond with the regulator initially which would be forfeited if the firm ever decided not to produce in the regulated market.[5]

The truthful revelation constraints (TR) define $\{p(c), T(c)\}$ as the price-tax pair that will be implemented when c is the realized value of operating costs in the regulated market. The constraints ensure that when c is realized, the firm will prefer to truthfully reveal that fact rather than claim that some other cost level, $\hat{c}$, was realized.

3. The Optimal Regulatory Policy

In this section we characterize the solution to [RP]. The general properties of the optimal regulatory policy are described in Proposition 1 and illustrated in Figures 2 and 3. We also illustrate the specific properties of the optimal policy with an example.

Before proceeding to Proposition 1, consider the strategy the regulator would follow if he could observe the realization of c along with the firm. In this case, the regulator would establish the regulated price at the level of realized marginal cost (i.e., $p(c) = c \quad \forall\, c \in [\underline{c},\overline{c}]$. He would also set $T(\cdot)$ to ensure exactly zero profits for the firm. Furthermore, the regulator would have no strict preference for permitting operation in the unregulated market. Such operation offers no direct gain in expected operating profits, nor could it serve to reduce the profits earned by the firm in the regulated market since these profits are already minimized in this hypothetical case.

More generally, when the firm acquires superior knowledge of production costs, it will receive rents from its private information. To limit these rents, it is advantageous for the regulator to permit entry into unregulated markets. The regulator will also adopt a rather "inflexible" pricing policy in the regulated market, as noted in Proposition 1.

Proposition 1. The solution to [RP] has the following features:

(i) Inflexible pricing rules, i.e.,

$$p(c) = p_K \quad \forall\, c \in [c_1,c_2], \text{ where } \underline{c} < c_1 < c_2 < \overline{c};$$

(ii) Limited pricing efficiency, i.e.,

$$p_K = \int_{c_1}^{c_2} c\, f(c)\, dc\ ,$$

$$p(\underline{c}) = \underline{c},\ p(\overline{c}) = \overline{c}, \text{ and } p(c_K) = c_K \text{ for some } c_K \in (c_1,c_2);$$

(iii) Rent limitation, i.e., $\pi(c) = 0 \quad \forall\, c \in [c_1,c_2]$, although $\pi(c) > 0$ for $c \in [\underline{c},c_1)$ and $c \in (c_2,\overline{c}]$;

(iv) Moderate participation in unregulated markets, i.e.,

$$K = Q(p_K) \quad \varepsilon \quad (Q(\underline{c}), Q(\overline{c})).$$

(v) Prices above marginal cost, i.e.,

$$p(c) > c \qquad \forall\, c \,\varepsilon\, (\underline{c}, c_K); \text{ and}$$

(vi) Prices below marginal cost, i.e.,

$$p(c) < c \qquad \forall\, c \,\varepsilon\, (c_K, \overline{c}).$$

The proof of Proposition 1 can be found in Lewis and Sappington (1987b). Rather than repeat the formal proof here, we concentrate on the intuition that underlies the conclusions in the Proposition. If the firm were prohibited from operating in the unregulated market, it would always be tempted to exaggerate its true costs of operating in the regulated market. By doing so, it would hope to convince the regulator that it needs more favorable compensation to break even in the regulated market.

When participation in the unregulated market is permitted, however, a countervailing incentive is introduced. Now if the firm exaggerates the realization of c, it also exaggerates the profits it can earn from employing its capacity ($K > 0$) in the unregulated market. This is because production costs in the regulated sector are high precisely when the capacity on hand is particularly valuable. With the implicit claim of higher profits in the unregulated sector, the firm provides an opportunity for the regulator to reduce compensation for the firm in the regulated market. Consequently, the firm's initial incentive to exaggerate the costs of operating in the regulated sector is mitigated. This is precisely why the regulator will finance entry into the unregulated market in our model: participation in the unregulated sector helps limit the firm's incentive to exaggerate production costs in the regulated sector.

Notice from property (iv) of Proposition 1 that participation in the unregulated market is limited, however. If K were chosen to be too large, the firm would always find it advantageous to understate c. In doing so, it would claim that greater compensation is required in the regulated market because the returns in the unregulated sector are too meagre. The key point is that the regulator will always find it advantageous to maintain the presence of countervailing rather than unilateral incentives. In other words, participation in the unregulated market will be structured so that the binding constraint is to prevent the firm from exaggerating low realizations of c and from understating high realizations of c.

The incentive to exaggerate low cost realizations is optimally mitigated by reducing the output demanded when the firm reports higher values of c. (See Figure 2.) The smaller output reduces the difference in total cost when c is low rather than high. Consequently, the gain in profit from exaggerating c is reduced relative to the case when efficient quantity levels are implemented. The optimal reduction in output is effected by setting the price charged in the regulated market above realized marginal cost. As indicated in property (v) of Proposition 1, this will be done for all cost realizations in the region $(\underline{c}, c_K)$. This is the region of low values of c where the dominant incentive for the firm is to exaggerate c.[6]

The complementary region of high cost realizations $(c_K, \bar{c})$ is the region where the dominant incentive for the firm is to understate the true value of c. To prevent understatement of c, output is optimally expanded beyond efficient levels when reports of low c are made, as illustrated in Figure 2. The additional output is more costly to produce when costs are truly high; hence, the attraction of understating c is reduced.[7]

The cost realization c_K defines the "boundary" between the two regions of countervailing incentives. To best control the firm's incentive to exaggerate costs below c_K and to understate cost realizations above c_K, the regulator would like to set prices above cost for all $c < c_K$ and below cost for all $c > c_K$. To do so, however, would require that for some range of cost realizations around c_K, lower prices and associated larger output levels would be associated with higher costs. To induce greater output from the firm when its costs are truly high requires sufficiently generous compensation that the firm will find it attractive to exaggerate its costs to some extent. Doing so will secure the generous compensation for expanded output even when costs are low. Formally, output in the regulated sector must fall as costs rise if the truthful revelation (TR) constraints in [RP] are to be satisfied.

Therefore, prices must not decrease as costs increase. As a result, in the region $[c_1,c_2]$ around c_K, the regulator can do no better than to set a single price, p_K, that will be implemented for all cost realizations in this region. (See Figure 2.) Thus simple, inflexible pricing rules are optimal, as reported in property (i) of Proposition 1. p_k will be set equal to expected marginal cost in the interval $[c_1,c_2]$, as reported in property (ii). Thus, given the restriction to a single price in this region, the "best" price in terms of expected total welfare is selected.[8]

Of course, the reason the regulator chooses to introduce price distortions and permit entry into unregulated markets is to limit the rents the firm commands from its private information. The pattern of rents that does accrue to the firm in the solution to [RP] is illustrated in Figure 3. In the region of rules rather than discretion, $[c_1,c_2]$, where a single price, p_K, is charged, the firm earns no rents. Positive profits are garnered for

smaller and larger cost realizations, however.[9] The firm can secure profits when small values of c are realized because of the inherent advantage in producing in the regulated sector when costs are low. Profits also arise for the larger realizations of c because participation in the unregulated sector is particularly lucrative when c is large. Again, the regulator could reduce the rents associated high cost realizations by denying the firm the right to operate in the unregulated market (i.e., by setting $K = 0$). Entry into unregulated markets, however, introduces countervailing incentives, which limit the rents associated with smaller cost realizations. On balance, some entry into unregulated markets is always optimal.

An important element of the optimal regulatory policy concerns the sharing of risk between consumers and the firm. The risk under consideration here arises from the initial uncertainty about production costs, c. In our model, consumers are risk-neutral; so in a "first-best" world where the ultimate realization of c is observed publicly, consumers would bear all the inherent risk. The regulated price would optimally be set at the level of realized marginal cost, and the firm's profits would remain constant at zero. Because of the inherent incentive problems, however, prices are less sensitive to realized cost. The optimal regulated price exceeds cost for small realizations of c and falls short of cost for large values of c. Thus, prices rise more slowly than costs, thereby "insulating" consumers from the variance that would arise absent incentive concerns. The insulation is most extreme for intermediate cost realizations, where regulated prices are entirely insensitive to cost variations. Note that in this region, though, the stability for consumers is not achieved by imposing greater risk on the firm. As illustrated in Figures 2 and 3, both the level of realized profit and the

level of realized consumers' surplus is constant in this region. Thus, the absence of "risk" is achieved at the expense of foregone surplus in this region. But overall, the "stickiness" of prices results in a higher level of expected consumers' surplus, even when consumers are risk neutral.

What remains is to explore the magnitude of the effects identified in Proposition 1. To do so, consider the case of uniform beliefs about c. Suppose that initially, all realizations of marginal production cost in the unit interval are thought to be equally likely, i.e., $f(c) = 1 \quad \forall c \in [0,1]$. In this case, the same price, $p_K = 1/2$, will be established for fifty percent of the possible cost realizations, i.e., $\forall\, c \in [1/4, 3/4]$, as illustrated in Figure 4. For the smallest cost realizations $c \in [0, 1/4]$, the regulated price will be twice realized cost, i.e., $p(c) = 2c$. For the highest cost realizations $c \in [3/4, 1]$, the regulated price will be one less than twice the realized cost, $p(c) = 2c - 1$. The level of participation in the unregulated market that will be permitted is given by the level of demand for the regulated product at the price of $p_K = 1/2$, which is the expected level of marginal cost. Thus, in this example, considerable entry into the unregulated market is facilitated, and the inflexibility in prices in the regulated market is quite pronounced.

4. Conclusions

Our purpose in undertaking this investigation was two-fold. First, we sought to examine the divergence between regulatory policy in practice and the prescriptions from economic theory. Second, we wished to determine whether by facilitating entry into unregulated markets, a regulator could better limit the rents of a firm operating in a regulated market. We found a close

connection between these two apparently disparate issues.

Our conclusion is that a regulator may be able to enhance the level of expected consumers' surplus generated in a regulated market by allowing the regulated firm to enter unregulated markets. We have associated "entry" into unregulated markets with the possession of "capital" that constitutes a critical factor of production in the unregulated market. The key feature of this capital is that its value is positively correlated with production costs in the regulated industry. Consequently, if the firm attempts to exaggerate its costs of producing in the regulated market, it simultaneously exaggerates the profits it will earn in the unregulated sector, enabling the regulator to reduce the allowed compensation in the regulated sector. In effect, allowing entry into unregulated markets introduces a countervailing incentive for the firm which limits its tendency to exaggerate production costs. This enables the regulator to better control the profits earned by the regulated firm.

The presence of countervailing incentives gives rise to "stickiness" in optimal regulated prices. To limit the incentive to exaggerate low cost realizations, prices are set in excess of realized marginal cost. To mitigate the tendency to understate high realizations of c (and thereby understate earnings in the unregulated sector), prices are set below realized marginal cost. For a wide range of intermediate cost realizations, these countervailing incentives result in a single regulated price being optimal. Thus, the optimal policy here is more congruent with regulatory policy in practice: a simple pricing rule is instituted rather than a complex pricing formula that affords considerable discretion to the firm.

In closing, we wish to emphasize that our model is designed to examine only certain elements of the decision to permit regulated firms to enter

unregulated markets. In practice, there are many other elements that warrant careful consideration. For example, there may be concern that the regulated firm will have an "unfair" competitive advantage in unregulated markets. Alternatively, entry into unregulated markets may complicate cost accounting procedures and thereby facilitate undesired cross subsidies. Furthermore, the possibility exists that a regulated firm might allow the quality of the regulated service to deteriorate when its attention is focused in other markets. These issues await additional research.

Footnotes

1. This condition is commonly imposed in the literature. In most agency models, the condition ensures a completely separating equlibrium, i.e., it guarantees that the optimal regulated price differs strictly with the realized production cost. (RC) does not guarantee such an equilibrium in our model.

2. We admit the possibility of negative taxes, which imply transfer payments from consumers to the firm. Such transfers might be effected as the fixed (access) charge in a two-part tariff, for example.

3. The regulator "buys" capacity in the sense that the cost is borne by the ratepayers rather than directly by the firm. Intuitively, the cost of the capacity is allowed in the rate base as soon as it is installed.

4. The level of fixed cost, F_o, is assumed to be known to both the regulator and the firm.

5. For simplicity, we adopt the assumption that the firm can't post <u>any</u> bond with the regulator. (Alternatively, the assumption is that <u>no</u> penalty can be imposed on the firm if it decides not to produce in the regulated industry.) The qualitative conclusions we report below continue to hold provided the bond the firm can post (or, alternatively, the penalty that can be imposed on the firm) is sufficiently small relative to the maximum expected total surplus that could be generated in the regulated sector.

6. For $c \in [\underline{c}, c_1]$, $p(c) = c + \frac{F(c)}{f(c)}$ in the solution to [RP].

7. For $c \in [c_2, \bar{c}]$, $p(c) = c - \frac{1-F(c)}{f(c)}$ in the solution to [RP].

8. Notice also from property (ii) of Proposition 1 that price will be set at the level of marginal cost for the smallest and largest realizations of cost. Deviations of price, $p(\hat{c})$, from cost, $\hat{c}$, are instituted to reduce the attraction of falsely reporting $\hat{c}$. But there are no cost realizations below $\underline{c}$ for which the firm will exaggerate and report $\underline{c}$, and there are no realizations above $\bar{c}$ the firm will understate and report $\bar{c}$. Hence, no pricing distortions will be implemented for these extreme cost realizations.

9. The rate at which equilibrium profits vary with realized costs, c, is given by $\pi'(c) = K - Q(p(c))$. The fact that the firm earns zero profits for an entire range of cost realizations $[c_1, c_2]$ follows from the simplifying assumption that profits in the unregulated market are a linear function of c. If these profits were an increasing, strictly concave function of c, the (IR) constraint in [RP] would bind for only a single cost realization. However, the solution to [RP] would still involve a nondegenerate region of cost realizations for which the same price is implemented. (For details, see Lewis and Sappington [1987a].)

References

Baron, David and David Besanko. "Regulation and Information in a Continuing Relationship," Information Economics and Policy. Vol. 1 (1984), pp. 267-302.

Baron, David and Roger Myerson. "Regulating a Monopolist with Unknown Costs," Econometrica. Vol. 50 (1982), pp. 911-930.

Lewis, Tracy and David Sappington. "Countervailing Incentives in Agency Problems," University of California Discussion Paper, September 1987a.

Lewis, Tracy and David Sappington. "Rules Rather Than Discretion in Agency Problems," University of California Discussion Paper, October 1987b.

Sappington, David. "Optimal Regulation of a Multiproduct Firm with Unknown Technological Capabilities," Bell Journal of Economics. Vol. 14 (1983), pp. 453-463.

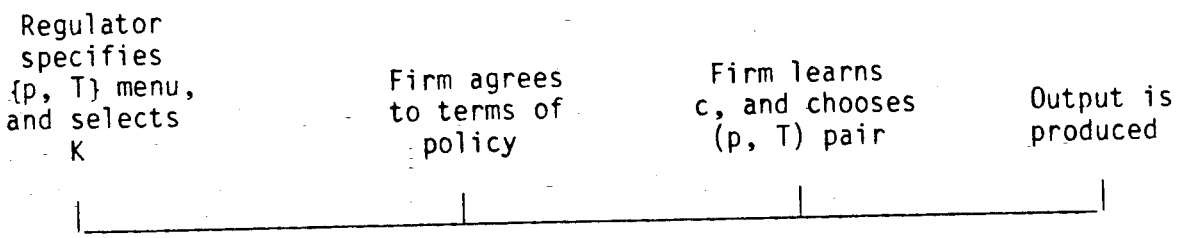

Figure 1. Timing in the Model.

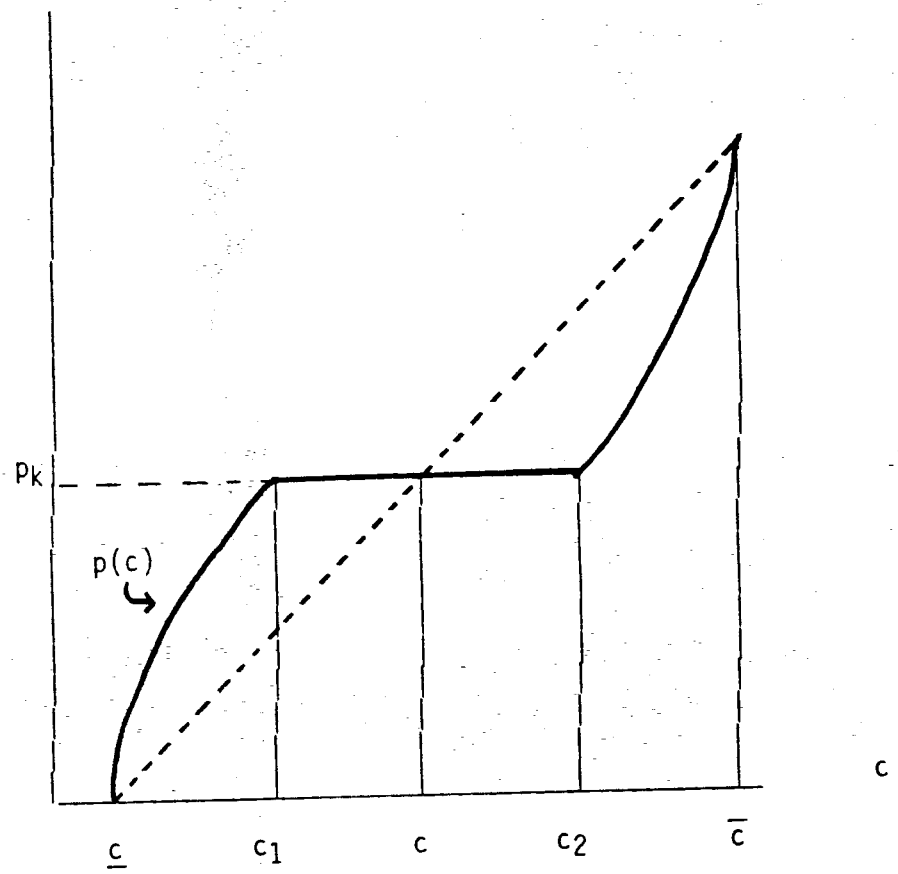

Figure 2. The Optimal Price Schedule.

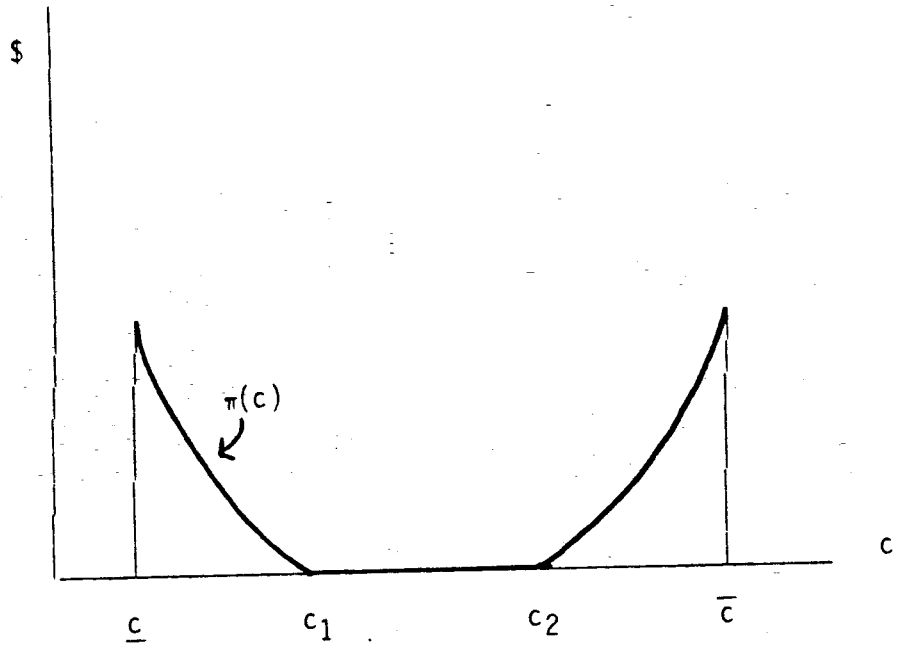

Figure 3. Profits in the Solution to [RP].

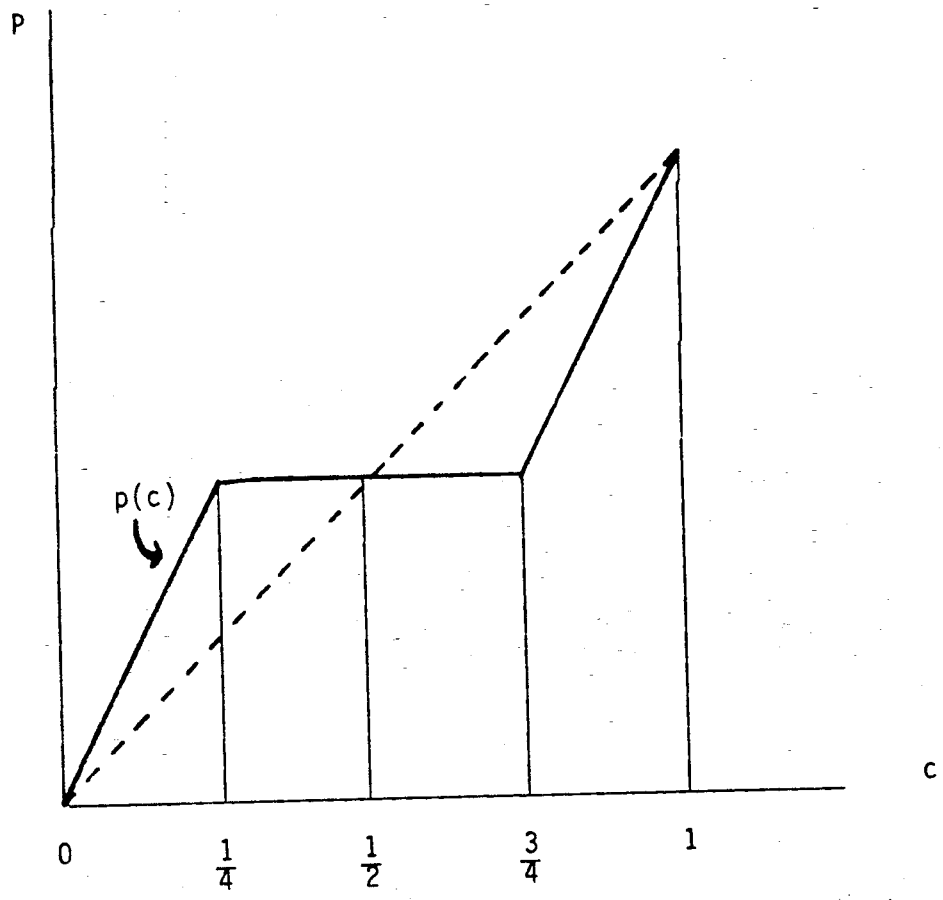

Figure 4. Optimal Prices with Uniform Beliefs.