THE REGULATION OF MULTIPRODUCT FIRMS:
THEORY AND POLICY ANALYSIS

Jean-Jacques Laffont
Jean Tirole

No. 507 October 1988
The Regulation of Multiproduct Firms:
Theory and Policy Analysis*

by

Jean-Jacques Laffont**
and

Jean Tirole***

October 1988

* The authors are grateful to Frank Fisher, Paul Joskow and Dick Schmalensee for helpful comments, and to the Ford Foundation, the Pew Charitable Trust, the Guggenheim Foundation, the Center for Energy Policy Research at MIT and the National Science Foundation for financial support.

** Harvard University and Massachusetts Institute of Technology.

*** Massachusetts Institute of Technology.
ABSTRACT

The paper develops a normative model of the regulation of a multiproduct firm with private information about its technology and cost-reducing activity. It first studies the conditions under which only the cost-reimbursement rule (but not pricing) is used to provide incentives (the "pricing-incentives" dichotomy). Second, it shows how optimal pricing can be decentralized to the regulated firm in the absence of cost and demand information. Third, the model is extended to allow competition by regulated and unregulated firms, access pricing and second-degree price discrimination.

Key words: Regulation, multiproduct firm, competition, incentives, pricing.

JEL Numbers: 026, 511, 613.
1. Introduction.

Over the last twenty years, the policy discussion of several industries, most notably telecommunications, has emphasized multiproduct issues in the regulation of natural monopolies. The moves from cost-of-service regulation to fully-distributed-cost pricing, and, more recently, to price caps witnessed debates about whether these moves would promote incentives, reduce prices, remove "cross-subsidization" or affect the regulated firms' competitors. Concurrently, the theory of multiproduct cost functions and sustainability of natural monopolies developed, culminating with the book by Baumol, Panzar and Willig [1982]. By and large, this theoretical literature has ignored the regulators' informational environment.¹ The purpose of this paper is to develop a normative theory of rate-of-return and price control for multiproduct firms based on an explicit informational imperfection in the regulatory process. [A note on the terminology: we will use "cost-of-service (COS) regulation" to denote any scheme designed to make the firm break even.²]

The often used "rate-of-return regulation" terminology is misleading, as any regulatory scheme determines some economic rate-of-return, and this rate of return need not be zero.]

We consider a multiproduct firm with private information about its technology and engaging in a cost-reducing activity. The regulatory instruments are the firm's total cost and prices; in particular the regulator cannot fully distribute costs to particular products, except through an

¹ An exception is Sappington [1983], whose work we discuss in Section 3.

² In principle, cost-of-service regulation refers to the policy that consists of maximizing a social welfare function subject to the firm's making zero profit. The theory part was developed by Boiteux [1956] and Baumol-Bradford [1970], and later substantially extended to a competitive environment (see Faulhaber [1975], Baumol et al. [1982] and Sharkey [1982]). An excellent survey is Brown-Sibley [1986].
arbitrary accounting procedure. The regulator chooses an optimal regulatory policy subject to his informational gap. The paper starts with a brief taxonomy of the concepts of cross-subsidization (Section 2) and is then divided into three parts.

The first part (Section 3) considers a general model and derives the optimal rate-of-return and price regulation. It demonstrates that the regulated price of good k satisfies the equation

\[ L_k = R_k + I_k, \]

where \( L_k \) is the Lerner index (price-cost margin), \( R_k \) is a Ramsey index, and \( I_k \) is the "incentive correction for good k." In contrast with the Ramsey index obtained in the conventional cost-of-service regulation model (which imposes a somewhat ad hoc\(^3\) budget constraint for the regulated firm), our Ramsey index can be computed from publicly available data and requires no knowledge of the firm's cost function. The incentive correction, if any, reflects the regulator's desire to limit the firm's rent without destroying incentives, and, more specifically, depends on whether an increase in good k's price helps raise incentives. A fundamental theoretical question is whether the incentive and pricing issues can be disconnected \( (I_k = 0) \) or not \( (I_k > 0) \); that is, is the optimal rate-of-return regulation (which under some assumptions turns out to be linear in the firm's performance) a sufficient instrument to promote incentives? Using familiar techniques from the theory of aggregation, Section 3 gives necessary and sufficient conditions for the dichotomy between

---

\(^3\)The existence of a budget constraint and the prohibition of transfers from the government to the firm in the models of Ramsey and Boiteux are inconsistent with their assumption that the regulator is benevolent; for, with a benevolent regulator, there is no point imposing such welfare-reducing constraints. Such constraints must thus be derived from a "political economy" approach to regulation, which leads us to question the formalization of the regulator's objective function in these models.
incentives and pricing to hold. It first considers the case of a single
dimensional cost reducing activity. It then allows the firm to allocate its
cost-reducing activity among the diverse (unobserved) components of the cost
functions, assuming one of two functional forms: the shared fixed-cost
function, in which the n goods share a joint cost and otherwise have
individual subcost functions; and the shared variable-cost function, in which
a cost-reducing activity affects the marginal cost of producing each good (as
is the case in the peak-load pricing model).

In the second part (Section 4) we argue that the optimal incentive and
pricing regulatory scheme can be approximately implemented by a rule that has
some informational advantages over current regulatory rules. More
specifically, we offer to i) regulate the firm's overall profit, that
includes possibly the returns from unregulated product lines ii) adopt a
reward structure that is linear in realized aggregate profit and iii) give
the firm freedom to choose its products' prices subject to a "price tax." In
the benchmark case (in which, in particular, pricing and incentives can be
disconnected), the price tax is equal to \( \frac{\lambda \hat{q}_k}{(1+\lambda)} \) per unit of increase in
the price of good k where a) \( \lambda \) is the shadow cost of public funds, a datum
available from econometric studies (a reasonable estimate of \( \lambda \) for the U.S. is
.3)\(^4\) and b) \( \hat{q}_k \) is a rough estimate of the demand for product k at the optimal
price (we discuss its estimation in detail in Section 4).

Unlike COS regulation, this rule promotes incentives. And the price
setting requires less information: The abstract Ramsey number that emerges
from the maximization of welfare subject to the firm's budget constraint.

\(^4\) See Ballard-Shoven-Whalley [1985], Hausman-Poterba [1987], and the references
therein. For instance, Ballard et al. find \( \lambda \) in the range of .17 to .56
depending on the saving and labor supply elasticities, and the tax used to
raise public funds.
depends on a Lagrange multiplier that reflects the firm's technology and the demand functions. In contrast, once the budget constraint is endogenized through the introduction of a shadow cost of public funds, the appropriate Ramsey number turns out to be independent of technology and demand functions and the optimal pricing policy can be decentralized through the use of a price tax. Similarly our pricing scheme is also simpler to implement than price caps (PC), which also require substantial information about technology and demand. Furthermore, by letting the government share part of the firm's return, our scheme does a better job than caps at capturing monopoly rents.

The third part (Section 5) applies the general theoretical framework studied in the first part to specific pricing issues: third- and second-degree price discriminations; competition by regulated and unregulated firms; access pricing; taxation by regulation. We develop the appropriate Ramsey indices and we generalize the price tax scheme developed in Section 4 for each application. Section 6 summarizes our main conclusions.

2. Definitions of cross-subsidization.

Several definitions of cross-subsidization are explicit or implicit in the theoretical and applied literatures. They are not mutually inconsistent; rather they reflect different facets of the regulatory problem. We will restrict ourselves to three notions:

1) Cost-side cross-subsidization. Such cross-subsidization occurs when the firm's managers allocate investments and their time and energy inefficiently among the firm's product lines. That is, the marginal productivity of a cost reducing activity applied to some product exceeds that for another product; one can then say that the former product subsidizes the latter. Any regulation based on aggregate cost (COS regulation -- similar to a cost-plus
contract, PC regulation -- similar to a fixed-price contract, and our intermediate incentive regulation) avoids cost-side cross-subsidization, as it induces managers to allocate (given levels of) investment and effort so as to minimize cost. In contrast, fully distributed cost pricing, which allocates joint costs to services in an arbitrary accounting manner, is a natural breeding ground for such cross-subsidies (see Brennan [1987]).

ii) Demand-side cross-subsidization. This kind of cross-subsidization occurs when prices are distorted to favor a targeted class of consumers. For instance, telephone pricing in the US and the UK has reflected a strong concern about the level of access charges to residential customers, and less concern about business customers. Similarly, rural consumers have often been protected from cost-justified price discrimination. Formally, what Posner [1971] has dubbed "taxation by regulation," may correspond to a discrepancy between a good's Lerner index and the Lerner index that would emerge from a regulatory policy that maximizes the unweighted sum of consumers' surpluses. (A slight variant of) this definition can be stated in the context of our model: If \( L_k < R_k + I_k \), good \( k \) is demand-side cross-subsidized (cross-taxed if \( L_k > R_k + I_k \)).

iii) Contestable cross-subsidization. The recent theoretical literature has emphasized the quite distinct notion that a given production plan and given (inflexible) prices for a regulated firm may trigger (hit and run) entry by an unregulated competitor who faces the same technology (the contestable case).

---

5 Demand-side cross-subsidization may also stem from the different powers of the various interest groups, as emphasized by the political science literature. See Laffont-Tirole [1988a] for a political theory of cross-subsidization.
Faulhaber [1975]'s definition of cross-subsidization is that the regulated firm's cost of producing a subset of goods (the "stand-alone cost") is lower than the revenue associated with this subset of goods. Baumol et al. [1982] contains a thorough investigation of when Ramsey pricing is consistent with sustainability (the absence of contestable cross-subsidization).

Although we study (possibly inefficient) entry by competitors, we abstract from the concern of this literature by assuming that the market is not contestable. That is, even though other firms may produce similar or differentiated products, the technology facing the regulated firm is unique and cannot be costlessly duplicated.

3. The incentive-pricing dichotomy.

a) The model. Consider a regulated firm with aggregate cost function

\[ C = C(\beta, e, q), \]

where

- \( \beta \) is a technological parameter or firm's "type" (\( C_\beta > 0 \)).
- \( e \) is its managers' cost reducing effort (\( C_e < 0 \))

and

- \( q = (q_1, \ldots, q_n) \) is the firm's output vector (\( C_{q_k} > 0 \)).

The technological parameter \( \beta \) and the effort level \( e \) are single-dimensional.

In subsection 3d), we give conditions under which a multidimensional parameter space (with one parameter \( \delta_k \) per product line) can be reduced to our single-dimensional representation.

\[ \text{Subscripts other than those indexing goods (k,l) denote partial derivatives:} \quad C_{\beta} = \frac{\partial C}{\partial \beta}. \]
In subsections 3d) and 3e), we allow the effort to be allocated among product lines. We let \( E(\beta,C,q) \) denote the effort required for a firm of type \( \beta \) to produce \( q \) at cost \( C \):

\[
(3-2) \quad C = C(\beta,E(\beta,C,q),q).
\]

The partial derivatives of this effort function with respect to its arguments are denoted \( E_{\beta}, E_C \) and \( E_{q_k} \) (\( E_{\beta} > 0, E_C < 0, E_{q_k} > 0 \)).

The firm's objective function is

\[
(3-3) \quad U = t - \psi(e),
\]

where

- \( t \) is a monetary transfer from the regulator to the firm, and
- \( \psi(e) \) is the manager's disutility of effort (\( \psi' > 0, \psi'' > 0 \)).

By accounting convention, the government pays the firm's cost and receives the revenue from its sales (so \( t \) is a net transfer). The firm is willing to participate in the regulatory process if and only if \( U \geq 0 \).

Last the utilitarian social welfare function is the sum of consumer welfare and the firm's welfare:

\[
(3-4) \quad W = [V(q) - R(q) - (1+\lambda)(t+C(\beta,e,q)-R(q))] + U
\]

or, using (3-3),

\[
(3-5) \quad W = [V(q) - \lambda R(q)] - (1-\lambda)(\psi(e) - C(\beta,e,q)) - \lambda U,
\]

where

- \( R(q) \) is the revenue generated by the sale of the production vector \( q \) (\( R(\cdot) \) is assumed to be increasing and concave),
- \( V(q) \) is the consumers' gross surplus (assumed increasing and concave),
A is the shadow cost of public funds.

(3-5) shows that \( W \) can be decomposed into three terms: the generalized gross surplus \( V + \lambda R \), which accounts for the fact that government revenue is socially valuable when public funds are costly; the total cost \( \psi + C \) of operating the firm times the shadow price of this cost; and the social cost \( \lambda U \) of leaving a rent to the firm.

In this section, we keep the \( V(\cdot) \) and \( R(\cdot) \) functions general. In a familiar case, \( R(q) = \sum p_k q_k - \sum S_k q_k \) (where \( p_k = S_k = \delta S / \delta q_k \) is the consumer price). But, as we show in subsection 5b), the model can be applied to other situations as well.

The regulator observes the firm's cost \( C \) and quantities \( q \) (or equivalently, prices \( p = (p_1, \ldots, p_n) \)). Unless otherwise stated, he regulates the firm's n outputs. The firm has private information about (knows) its technology parameter \( \beta \), which from the regulator's viewpoint is drawn from a cumulative distribution \( F(\cdot) \) on \( [\delta, \beta] \) with density \( f(\cdot) \). We make the standard assumption that the hazard rate \( f(\beta) / F(\beta) \) of the distribution is monotonic: \( \frac{d(F/F)}{d\beta} \geq 0 \). The cost-reducing effort \( e \) is also unobservable by the regulator. \( U(\beta) \) will be called the firm's (informational) rent.

b) The optimal regulatory allocation. The regulator maximizes the expectation over \( \beta \) of the social welfare function given by (3-5), over \((q(\beta), e(\beta), U(\beta)) \) (which amounts to maximizing with respect to \((q(\beta), e(\beta), \epsilon(\beta)) \)):

\[
(3-5) \quad E[\psi] = \int_{\beta} E[\psi(q)] = \int \left[ V(q) + \lambda R(q) - (1 + \lambda)(\psi(e) + C(\beta, e, q)) - \lambda U \right] f(\beta) d\beta,
\]

subject to the individual rationality and incentive constraints. Because the firm's rent is necessarily decreasing in \( \beta \), and because (from (3-5)), the
regulator does not wish to leave rents to the firm, the individual rationality constraint can be written:

\[(3-7) \quad U(\beta) = 0.\]

The derivation of the incentive constraint is standard (see, e.g., Guesnerie-Laffont [1984]). An intuitive argument is as follows: A firm with type \( \beta \) can produce the same output vector at the same cost as a firm with type \( \beta' \), and therefore obtains the same revenue and transfer, by reducing its effort relative to that of type \( \beta \) by \( \Delta e = E_\beta(\beta',C(\beta,e,q),q)d\beta \). This implies that the gradient of the firm's rent with respect to \( \beta \), \( \dot{U}(\beta) = dU/d\beta = \frac{\partial}{\partial \beta} (e)de \), is given by:

\[(3-8) \quad \dot{U}(\beta) = \frac{\partial}{\partial \beta} (e)E_\beta(\beta',C(\beta,e,q),q).\]

Taking \( e(\beta) \) and \( q(\beta) \) as control variables and \( U(\beta) \) as a state variable, the regulator maximizes (3-6) subject to (3-7) and (3-8). Appendix 1 shows that under standard technological assumptions, \( dC/d\beta \geq 0 \) and \( dq_k/d\beta \leq 0 \) for all \( k \) are sufficient conditions for the firm's second-order conditions to be satisfied. Whether those conditions hold for the solution to the first-order conditions must be checked in each application of our model (for instance, we will check them in Proposition 9).

The optimal regulatory policy is derived in Appendix 1. Here we content ourselves with its statement and intuition.

**Proposition 1:** The first-order conditions of the regulator's program with respect to effort \( e \) and output \( c_k \) are:

\[(3-9) \quad \psi'(e) = -c_e - \frac{\lambda}{1+\lambda} \frac{F(\beta)}{F'(\beta)} [\psi''(e)E_\beta + \psi'(e)E_\beta C_e] \]
To understand (3-9), note that in a symmetric information (first-best) world, the marginal disutility of effort $\psi'(e)$ should be equal to the marginal cost savings ($-C_e$). The last term in (3-9) is due to the regulator's desire to extract the firm's informational rent. The term in brackets

$$A = \psi''(e)E_{\beta}\psi'(e)E_{\beta}C_e$$

is the derivative with respect to $e$ of $|\hat{U}(\beta)|$. A sufficient condition for $A$ to be positive is that $C_{ee} = 0$ (there are decreasing returns in the cost reducing technology) and that $C_{\beta e} = 0$ (the marginal cost reduction is not higher for efficient types). From (3-6), $A$ is also the increase in the rent of all firms with type in $[\hat{r}, \beta]$ (which have probability $F(\beta)$), when the effort of type $\beta$ is increased by one. From (3-6), the social cost of the extra rent for the types in $[\hat{r}, \beta]$ is $\lambda F(\beta)A$. On the other hand, the distortion in effort for type $\beta$ relative to the first best level ($\psi'(e) + C_e$) has social cost $(1+\lambda)(\psi'(e) + C_e)$ from (3-5) and occurs with probability $f(\beta)$. The tradeoff between rent extraction and efficient effort thus yields (3-9). An implication of (3-9) that we later develop is that -- like in our single-product paper [1986] -- the regulator uses cost-reimbursement rules that are intermediate between the cost-plus contract (which induces $\psi'(e) = 0$) and the fixed-price contract (which induces $\psi'(e) = -C_e$).

Of particular interest for the subsequent analysis is the "modified Ramsey equation" (3-10). Under symmetric information, the marginal generalized gross surplus $d(V+AR)/dq_k$ is equal to the marginal social cost of production $d((1+\lambda)C)/dq_k$. Under asymmetric information, there may exist an incentive correction associated with the regulator's desire to extract the
firm's rent. From (3-8), a unit increase in output \( k \) affects the rent by

\[ \psi'(e) \frac{d}{dq_k} (E_\beta), \]

where \( d/dq_k \) (\( E_\beta \)) = \( E_\beta C_{q_k} + E_{q_k} \)

is a total derivative and is a measure of how output \( k \) affects the potential effort savings associated with an increase in efficiency. As in equation (3-9), the gain in reducing \( |\hat{U}| \) is proportional to the probability \( F(\beta) \) that the firm is more efficient than type \( \beta \), and the cost of the distortion relative to the first best is proportional to the probability \( f(\beta) \) that the firm has type \( \beta \), which explains equation (3-10).

Equation (3-10) yields a simple conclusion: incentives and pricing of good \( k \) are disconnected if and only if \( d(E_\beta)/dq_k = 0 \). In this case, the regulator uses only the cost reimbursement rule (implicit in (3-9)) to extract the firm's rent while preserving incentives. If \( d(E_\beta)/dq_k > 0 \) (respectively, < 0), the price of good \( k \) exceeds (respectively, is lower than) its symmetric information level, ceteris paribus. Proposition 2 below states that the incentive correction favors a high price (a low quantity) for good \( k \) if an increase in the output \( k \) increases the marginal rate of transformation between effort and efficiency in the cost function, that is, if this increase in output makes it easier for the firm to transform exogenous cost changes into rent.

**Proposition 2.** Ceteris paribus, the price of good \( k \) exceeds its symmetric information level if and only if

\[ \frac{\partial}{\partial c_{\beta k}} \left[ - \frac{c_{\beta}}{c_e} \right] > 0. \]
Proof of Proposition 2: Differentiating (3-2) with respect to $\beta$ yields

$$0 = C^B + C^E E^B.$$ 

Hence,

$$\frac{d}{dq_k} (E^B) = \frac{\partial}{\partial q_k} \left[ -\frac{C^B}{C^e} \right].$$

Substituting into (3-10) shows that $V + \lambda R - (1+\lambda)C^q_k$ is positive if and only if

$$\frac{\partial}{\partial q_k} \left[ -\frac{C^B}{C^e} \right] > 0.$$ 

Q.E.D.

c) The incentive-pricing dichotomy. As mentioned earlier, a central issue is whether pricing should be used to promote incentives (or to extract the firm's rent). Leontieff [1947]'s well-known aggregation theorem and Proposition 2 yield:

Proposition 3: The incentive-pricing dichotomy ($d(E^B)/dq_k = 0$ for all $k$)

holds if and only if there exists a function $\Phi$ such that

$$(3-11) \quad C = C(\Phi(\beta,e),q).$$

Proposition 3 is both important and straightforward. If efficiency and effort can be aggregated in the cost function, changing $q_k$ does not affect the extent to which the firm can convert exogenous productivity increases into rent, which is entirely determined by the $\Phi$ function.

At this stage it is worth comparing our approach with Sappington [1983]'s early work in incentives in multiproduct firms. Sappington, along the lines of Baron and Myerson [1982]'s single product contribution, shows that in general prices are distorted to limit the firm's informational rent. There are two related differences between his results and ours. First, Sappington's
(like Baron-Myerson's) symmetric information benchmark is marginal cost pricing and not Ramsey pricing because of the absence of a cost of public funds. Second, and mainly, Sappington finds that pricing cannot be disconnected from incentives as long as the uncertainty parameter affects marginal costs.

For instance, suppose that

\[ C = \Phi(\beta, e) \left( \sum_{k=1}^{n} q_k \right). \]

Equation (3-12) describes a technology in which efficiency and effort determine the constant marginal cost of a common output to be distributed in \( n \) different markets. From Proposition 2, the price of good \( k \) is set at the symmetric information level in our model, while it is set above the symmetric information level in Sappington's model. This second discrepancy is due to Sappington's assumption that the firm's cost (or a garbled version of it) is not observable. Thus the regulator cannot use cost reimbursement rules to extract rents and is forced to use price distortion as a substitute.\(^7\) While the condition for the dichotomy is strong in our model, it leads us to view the dichotomy as a benchmark from which a case must be built that price manipulations do promote incentives.

When the incentive-pricing dichotomy does not hold, it is interesting to know whether the incentive corrections have the same signs for all goods.

[The adjustment envisioned in subsection 4c, that consists of raising or lowering the cost of public funds to reflect the incentive correction, has then a uniform sign across all goods.] A sufficient condition for this is

\(^7\) Sappington's analysis, like most of those in this paper, assumes a single dimensional cost parameter. See Dana [1987] for an example of regulation (without cost observation) when the firm's subcosts are determined by different (although possibly correlated) parameters. See also Proposition 7.
Proposition 4. Suppose that there exists a function $\Gamma(q_1, \ldots, q_n)$ such that

$$(3-13) \quad C = C(\beta, e, \Gamma(q)).$$

Then the incentive correction has the same sign for all goods.

Proof of Proposition 4. For the functional form (3-13),

$$\frac{\partial}{\partial q_k} \left( - \frac{C_{\beta}}{C_e} \right) - \left[ \frac{\partial}{\partial \Gamma} \left( - \frac{C_{\beta}}{C_e} \right) \right] \Gamma_{q_k}.$$

Proposition 4 follows from the fact that all the $\Gamma_{q_k}$ have the same (positive) sign.

Q.E.D.

d) Effort allocation: the shared-fixed-cost model. In this subsection and in the next, we make more specific assumptions about the cost function. Our first paradigm is that of a joint fixed cost shared by several product lines, along with a variable cost for each product line (this paradigm might depict the case of switches used for several telecommunications services):

$$(3-14) \quad C = \psi(\beta, e_0, \ldots, e_n, q) = \sum_{k=1}^{n} C^k(\beta, e_k, q_k) + C^0(\beta, e_0).$$

In (3-14), $C^k$ is the subcost or variable cost of good $k$. $C^0$ is the fixed cost. The fixed cost and the subcosts cannot be disentangled from accounting data (an often reasonable assumption), so that only the total cost $C$ can be contracted on. The firm's managers allocate their cost reduction activity $e$ among the product lines and the joint activity:

$$(3-15) \quad e = \sum_{k=0}^{n} e_k,$$
so as to minimize $\mathcal{C}$ (this behavior is optimal as long as regulation is based on total cost):

$$\begin{align*}
(3-16) \quad C(\beta,e,q) &= \min_{\{e_0,\ldots,e_n\}} \mathcal{C}(\beta,e_0,\ldots,e_n,q) \\
\text{s.t. } (3-15).
\end{align*}$$

Let $\{e^*_\ell(\beta,e,q)\}_{\ell=0}^n$ denote the solution to (3.16).

The goal of this section is to obtain the pendants to Propositions 3 and 4 for this model. That is, we look for conditions on the subcost and joint cost functions so that the dichotomy holds or at least that all incentive corrections have the same sign when the firm allocates its cost reducing activity.

**Proposition 5.** Assume an interior solution to program (3.16). The incentive-pricing dichotomy holds in the shared-fixed-cost model if and only if for all $k$

$$\begin{align*}
(3-17) \quad \sum_{\ell=0}^n \frac{\partial}{\partial e^*_\ell} \left( \frac{C^\ell}{C^e} \right) \frac{\partial e^*_\ell}{\partial q^k} + \frac{\partial}{\partial q^k} \left( \frac{C^k}{C^e} \right) &= 0. \tag{8}
\end{align*}$$

**Proof of Proposition 5.** From Proposition 3, a necessary and sufficient condition for the dichotomy to hold is that $\frac{\partial (C^e/C^e)}{\partial q^k} = 0$ for all $k$. Let

$$\frac{\partial e^*_k}{\partial q}$$

the change in effort $e^*_k(\beta,e,q)$ resulting from a unit increase in output $k$, is obtained by totally differentiating the $(n+1)$-equation system

$$\begin{align*}
(C^\ell \quad e^*_\ell = C^k_{e^*_k}) \quad \text{for all } \ell \text{ in } \{0,\ldots,k-1,k-1,\ldots,n\} \quad \text{and } \sum_{\ell=0}^n e^*_\ell = e \quad \text{w.r.t. } q^k.$$
\( e_k^* \) denote the optimal effort levels (\( \sum_{k=1}^{n} e_k^* = e \)). One has:

\[
C_{\beta} = \sum_{k=0}^{\infty} \left[ C_{\beta}^{k} + \frac{\partial e_k^*}{\partial \beta} C_{\beta}^{k} \right].
\]

Furthermore, program (3.16) implies that

\[
C_e = C_{e_k}^k \quad \text{for all } k.
\]

Hence

\[
C_{\beta} = \sum_{k=0}^{\infty} C_{e_k}^k,
\]

using \( \sum_{k} \frac{\partial e_k^*}{\partial \beta} = 0 \). This implies that

\[
\frac{C_{\beta}}{C_e} = \sum_{k=0}^{\infty} \left[ \frac{C_{\beta}^k}{C_{e_k}^k} \right].
\]

Hence, \( C_{\beta}/C_e \) is independent of outputs if and only if (3.17) holds. 

Q.E.D.

Application: Adopting the convention that \( q_0 = 1 \), the following functional form

(3-18) \( C_k^k(\beta, e_k, q_k) = (\exp(-\mu_k e_k)) z_k(\beta) z_k^k(q_k) \quad \text{for } k = 0, 1, \ldots, n \)

satisfies (3.17), because \( C_{\beta}/C_{e_k}^k \) depends only on \( \beta \). Hence for the cost function described by (3.18), the incentive-pricing dichotomy holds. Another example of a functional form for which \( C_{\beta}/C_{e_k}^k \) depends only on \( \beta \), and thus for which the dichotomy holds, is

\( C_k^k(\beta, e_k, q_k) = (\beta - a_k e_k)^{b_k} z_k^k(q_k) \quad \text{for } k = 0, 1, \ldots, n. \)
Remark: A difference between the fixed effort allocation model of subsection 3c (which can be thought of as allocating effort in fixed proportions among the product lines and joint cost) and this flexible effort allocation model is that the regulator may lose some power to extract rent through price distortions when the firm can allocate effort among product lines. For example, assume that

\[(3-19) \quad C = (\exp(-a_1e))r^1(\beta)z^1(q_1) + (\exp(-a_0e))r^0(\beta),\]

with \(a_0 + a_1 = 1\) (so \(e_0 = a_0e\) and \(e_1 = a_1e\), where \(a_0\) and \(a_1\) are positive and predetermined). [This technology can be viewed as a two-product cost function, in which one of the "products" -- with associated subcost the fixed cost -- is produced in quantity one.] From Proposition 3, the dichotomy always fails for the cost function described in (3-19). However, from Proposition 5, it holds for the analogous cost function with flexible effort allocation:

\[(3-18') \quad C = (\exp(-e_1))r^1(\beta)z^1(q_1) + (\exp(-e_0))r^0(\beta).\]

To understand why this is so, suppose for instance that \(r^0(\beta) = r^0(\beta)\) might then be the quality or price of an input and \(\exp(-e_1)\) a measure of the efficiency of the transformation of the input into the output). The incentive correction for (3-19) (whose sign is given in Proposition 1) consists in lowering \(q_1\) relative to the symmetric information case. This reduces \(\frac{\partial e}{\partial \beta}\) and \(C\) constant and thus the firm's rent. In contrast, for the technology described by (3-18'), the firm can lower \(e_1\) without lowering \(e_0\) when \(\beta\) decreases (and is indifferent to doing so at the margin), so that marginal changes in \(q_1\) do not affect the slope of the rent function.

More generally, when do all incentive corrections have the same sign in the shared-fixed-cost model? From Proposition 4, a sufficient condition for
this is that there exist a function $\Gamma$ such that the solution to (3-16) can be written $C = C(\beta, e, \Gamma(q))$.

Formally, the existence of such a $\Gamma$ is equivalent to the possibility of partially aggregating capital goods in the production function of a firm that allocates labor optimally to capitals of different vintages. [That is, the outputs $q_k$ correspond to the vintages of aggregated capital $K_{1,k}$; the efforts $e_k$ correspond to the labor inputs $L_k$ allocated to machines of vintage $k$; $\beta_k = \beta$ is the analog of a vintage specific capital $K_{2,k}$ that is not aggregated.] The necessary and sufficient condition for the partial aggregation of capital goods was developed by Fisher [1965] (see also Gorman [1968]). Applied to our context, Theorem 5.1 in Fisher [1965] yields:

**Proposition 6:** A necessary and sufficient condition for the indirect cost function obtained in (3-16) to have the form $C(\beta, e, \Gamma(q))$ is that

$$\frac{\partial^2 C}{\partial q_k \partial \beta} - \frac{\partial^2 C}{\partial q_k \partial e} \frac{\partial^2 C}{\partial e \partial \beta} = 0 \quad \text{for all } k \in \{1, \ldots, n\}. \tag{3-20}$$

and

$$\frac{\partial^2 C}{\partial e_k \partial q_k} = g \left( \begin{array}{c} \frac{\partial C}{\partial q_k} \\ \frac{\partial C}{\partial e_k} \end{array} \right). \tag{3-21}$$

Hence, if (3-20) and (3-21) hold, all incentive corrections have the same sign.
Application. The conditions stated in Proposition 6 are obviously very strong, but they allow us to, for instance, find classes of cost functions such that all incentive corrections have the same sign. Suppose that the subcost functions have the following multiplicative form:

\[(3-22) \quad c^k - r^k(\beta)s^k(e_k, q_k) \quad \text{for all } k \in (1, \ldots , n).\]

The reader can check that conditions (3-20) and (3-21) hold if and only if

\[\frac{\partial}{\partial e_k} \left( \frac{\partial s^k}{\partial q_k} \right) = 0 \quad \text{for all } k \in (1, \ldots , n),\]

(as is the case for instance if \(s^k(e_k, q_k) = z^k(q_k) - \mu^k e^k\)).

The same techniques of partial aggregation of capital goods can be applied to the question of when a vector of cost parameters can be reduced to a single-dimensional cost parameter. To this purpose, it suffices to reverse the roles of the exogenous cost parameters and quantities in Proposition 6. More specifically, suppose that each product line is affected by a specific cost parameter:

\[(3-23) \quad C = C(\beta_0, \ldots , \beta_n, e, q) = \min_{(e_0, \ldots , e_n)} \left[ \sum_{k=1}^{n} C^k(\beta_k, e_k, q_k) + C^0(\beta_0, e_0) \right] \]

\[\text{s.t. } \sum_{k=0}^{n} e_k \leq e.\]

The set of technological parameters \((\beta_0, \ldots , \beta_n)\) is now \((n-1)\)-dimensional. We allow the joint distribution over \((\beta_0, \ldots , \beta_n)\) to be arbitrary. The manager knows all parameters, while the regulator knows none.

The validity of the single-parameter representation in this case depends on the possibility of partially aggregating technological parameters in the
cost function. For, suppose that there is a function \( \Lambda(\cdot) \) satisfying:

\[
(3-24) \quad C(\beta_0, \ldots, \beta_n, e, q) = C(\Lambda(\beta_0, \ldots, \beta_n), e, q).
\]

One then defines \( \beta = \Lambda(\beta_0, \ldots, \beta_n) \). The distribution over \( \beta \) results from the joint distribution over \( (\beta_0, \ldots, \beta_n) \). All the results obtained so far in the one-dimensional case can then be applied.

Thus, Fisher (1965)'s Theorem 5.1 also yields:

**Proposition 7:** Assuming an interior solution to (3-23), a necessary condition for the indicated cost function obtained in (3-23) to have the form \( C(\Lambda(\beta_0, \ldots, \beta_n), e, q) \) is that

\[
(3-25) \quad \frac{\delta^2 C^k}{\delta q_k \delta e_k} = \frac{\delta^2 C^k}{\delta e_k \delta e_k} - \frac{\delta^2 C^k}{\delta e_k \delta e_k} \frac{\delta^2 C^k}{\delta e_k \delta e_k}
\]

and

\[
(3-26) \quad \text{there exists a function } h(\cdot) \text{ such that for all } k \text{ in } \{0, \ldots, n\},
\]

\[
\frac{\delta^2 C^k}{\delta e_k^2} = h\left(\frac{\delta^2 C^k}{\delta e_k^2}\right).
\]

**Application:** The subcost functions

\[
(3-27) \quad C^k(\xi_k, e_k, q_k) = (\xi_k - e_k)^2 q_k^2
\]

satisfy the conditions of Proposition 7 (if an interior solution to (3-23) prevails) so that the technological parameters can be aggregated into a single-dimensional parameter.
e) **Effort allocation: the shared-marginal-cost model.** In several regulated industries, the same physical good is sold at different points of time (or in different geographical areas). A reduction in the marginal cost of producing one good (where a good is the physical good indexed by time, say) cannot be separated from a reduction in the marginal cost of producing other goods, in contrast with the technology considered in 3d.

Consider an electric utility facing time-varying demand for electricity. A fraction $x_k \in (0,1)$ of the time, the demand for electricity is $q_k$ (which depends on the prices charged at different points of time). One has $\sum_{k=1}^{n} x_k = 1$. Without loss of generality, let us assume that $q_1 < \ldots < q_n$ (so $n$ denotes the peak period). The firm installs capacity for each incremental output $(q_k - q_{k-1})$: $q_1$ units of base capacity, $(q_2 - q_1)$ units to supplement the base capacity in period 2, etc. [By convention, $q_0 = 0$.] Let $X_k = \sum_{t=k}^{n} x_t$ denote the load factor of capacity of type $k$. The optimal exploitation policy implies that the (constant) marginal production costs associated with all types of capacity satisfy: $c_1 \leq c_2 \leq \ldots \leq c_n$. Let $G(q_1, q_2, q_1, \ldots, q_n - q_{n-1})$ denote the total cost of installing the different types of capacities; we will let $C_k = \frac{\partial G}{\partial (q_k - q_{k-1})}$, and $\tilde{c}_k = c_k X_k$. The firm's cost function can be written:

\[(3.28) \quad C = \varnothing(\beta, e_1, \ldots, e_n, q) \]

\[= c_1(\beta, e_1)q_1X_1 + c_2(\beta, e_2)(q_2 - q_1)X_2 + \ldots + c_n(\beta, e_n)(q_n - q_{n-1})X_n \]

\[+ G(q_1, q_2 - q_1, \ldots, q_n - q_{n-1}) \]

\[= \sum_{k=1}^{n} \tilde{c}_k(\beta, e_k)(q_k - q_{k-1}) + G(q_1, q_2 - q_1, \ldots, q_n - q_{n-1}). \]

To give an example with $n = 2$, $e_1$ and $e_2$ can be thought of as the managers' efforts to supervise or improve the technology of the nuclear power division.
(which supplies both peak and off-peak demands) and the coal division (which supplies peak demand only). As before, we define

(3-29) \[ C(\beta, e, q) = \min_{\beta_1, \ldots, \beta_n, q} \\] 
\[ \text{s.t. } \sum_{k=1}^{n} e_k = e. \]

**Proposition 8.** Assume an interior solution to program (3.29). The incentive-pricing dichotomy holds for the shared-marginal-cost model if and only if, for all \( k \),

\[
(3-30) \quad \sum_{\ell=1}^{n} \frac{\partial}{\partial e_\ell} \left( \frac{\partial^2 c}{\partial \beta \partial e_\ell} \right) \frac{\delta e_\ell^*}{\delta q_k} = 0,
\]

where the functions \( \{e_\ell^*\}_{\ell=1}^{n} \) are the solutions of (3.29).

**Proof of Proposition 8.** The proof is essentially the same as that of Proposition 5. From (3.29), we know that \( \frac{\partial}{\partial e_k} C_k(\beta, e_k)(q_k - q_{k-1}) = C_e \) for all \( k \). Therefore:

\[
\frac{\partial}{\partial q_k} \left( - \frac{C_\beta}{C_e} \right) = - \sum_k \frac{\partial}{\partial q_k} \left( \frac{\partial^2 c}{\partial \beta \partial e_k} \right) \left( \frac{\partial c_k}{\partial e_k} / \partial q_k \right).
\]

Because of constant returns to scale, \( C_k \) and its partial derivatives depend on \( q_k \) only through the effect of \( q_k \) on \( e_k^* \). This, together with Proposition 2, yields (3-30).

Q.E.D.

22
Examples: (3-30) and therefore the incentive-pricing dichotomy holds for the marginal cost functions

\[ (3-31) \quad \bar{c}_k(\beta, e_k) = (\beta - e_k)^{b_k} \]

and

\[ (3-32) \quad \bar{c}_k(\beta, e_k) = \beta \log e_k . \]

[For marginal cost function (3-31), \( \frac{\partial}{\partial e_k} \left( \frac{\partial \bar{c}_k}{\partial \beta} \right) = 0 \), for all \( t \).

For marginal cost function (3-32), the same term is equal to -1 for all \( t \); the equality \( \sum_{l=1}^{n} \frac{\partial e_l^*}{\partial q_k} = 0 \) then yields (3-30).]

f) Linearity of cost reimbursement rules. Linear cost (or profit) sharing rules have the attractive property that they are still optimal under an additive cost disturbance or accounting error \( (C = C(\beta, e, q) + \epsilon) \) as long as the parties are risk-neutral. One may wonder whether the optimal regulatory policy can be implemented through a menu of linear cost reimbursement rules. Strong assumptions are needed to get exact linearity. Finding the minimal such assumptions is a difficult task, and we content ourselves with stating a sufficient condition for a menu of linear contracts to be optimal.

**Proposition 9:** Assume that

(i) \( C(\varepsilon, e, q) = G(\beta, \varepsilon)H(q) \), where \( G' > 0, G'' > 0 \), and the curvature of \( G \) is "not too high":

\[ \max(G''/G') \leq \min(\varepsilon(F/F')(F/F')). \]

(ii) \( \psi'' \geq 0 \).

(iii) Letting \( (e(\beta), q(\beta)) \) denote the solution to (3-9) and
\[
\frac{d}{d\beta} G(\beta-e(\beta)) \geq 0 \quad \text{and} \quad \frac{d}{d\beta} (\mathbf{q}_k(\beta)) \leq 0 \quad \text{for all} \ \beta.
\]

Then the optimal regulatory policy can be implemented through a menu of linear contracts.

Assumption (iii) in Proposition 9 is reasonable. For the optimal policy, a firm with a higher \( \beta \) is likely to produce less \( (d\mathbf{q}_k/d\beta \leq 0) \) and to have a higher \( G(\beta-e(\beta)) \). [It is easy to find sufficient conditions for these to hold. Note that from Proposition 4, the dichotomy holds, so that the outputs are the solutions to \( V + AR = (1+\lambda)C \) for all \( k \).] Assumption (ii) is technical and is much stronger than necessary. It ensures that the optimal regulatory scheme is not stochastic. Assumption (i) is strong. The proof of Proposition 9 can be found in Appendix 2.

4. Policy implications.

The purpose of a normative approach to regulation is to specify optimal rate-of-return and pricing decisions as functions of the cost and demand environment, including the distribution of uncertainty and the informational constraints faced by the regulator. In contrast, considerable weight is given in the policy arena to rules that are "simple." We define a rule as being simple if it can be prescribed by an economist lacking a detailed knowledge of cost and demand functions, and if it is nevertheless comprehensive enough to be readily applicable by -- i.e., to leave little discretion to -- regulators. This concern for simplicity stems from the possibility that the regulatory hierarchy (administrative agency and its oversight structure) does not use its information efficiently. On the one hand, a regulator may be put off by -- or may not have the incentive to devote time to -- the application of complex
regulatory rules. On the other hand, while a regulatory agency is unlikely to be as well informed as the regulated firm about the technology (and possibly demand), it still has a substantial informational edge over its oversight structure (Congress, courts,...). This second asymmetry of information raises the possibility of collusion between the agency and the regulated firm. Simple rules (in the above terminology), by removing agency discretion, give less scope to collusion than complex rules relying on an intimate knowledge of the likely distribution of the cost and demand parameters that the oversight structure is unlikely to possess; that is, simple rules are immune to the threat of regulatory capture by the industry.

Thus, under an inefficient regulatory hierarchy (that is, a hierarchy which, for either of the above two reasons, is unable to implement the optimal incentive scheme conditional on its finest information structure), a normative theory that ignores this imperfection (like the one presented in this paper) is ultimately judged by whether its optimal allocation can be approximately implemented through simple rules.

The purpose of this section is to briefly discuss the informational requirements of the two existing policy mechanisms: cost-of-service (COS) and price capping (PC); and then to propose an alternative mechanism based on our normative analysis.

A) COS and PC regulations. The economist's benchmark for "efficient pricing" is marginal cost. For instance, the 1967 White Paper recommended that the prices of UK public enterprises reflect long-run marginal costs (with potential adjustments towards short-run marginal costs for low capacity levels). Marginal cost pricing is not only theoretically unsatisfactory (see equation (3-10)); it is also impractical and incomplete: impractical because regulators have little knowledge of marginal costs; incomplete because it does
not specify the cost reimbursement or rate-of-return rule that settles the tradeoff between incentives and rent extraction.

A much more influential paradigm is that of cost-of-service rate making. A naive description of this rule (corresponding to the theoretical Ramsey-Boiteux model) is that the regulator sets prices so that the revenue meets the firm's cost. To be applied *stricto sensu*, COS ratemaking would require a formidable knowledge of the cost and demand functions from regulators. In practice, cost estimates are based on the firm's previous performance, creating notorious incentive problems: the firm has little incentive for cost control and innovation as such policies feed back into lower allowed prices and revenue in the future. The approximate cost-plus character of COS rate making has been rightly decried by many academic economists lately.⁹

In principle, PC regulation attempts to regulate prices, but not cost or rate-of-return (in this latter respect, it resembles a "fixed-price contract"). It imposes ceilings on the firm's rates, either on individual goods or on a weighted average of a basket of goods. For instance, the 1984 regulatory reform in Great Britain required that British Telecom's weighted average rate of change of prices of certain services do not increase at a rate of more than the inflation rate minus three percent ("RPI-3"), where the three-percent term reflects an estimated annual average productivity growth. Because it in principle offers much more scope for incentives,¹⁰ PC regulation

---

⁹ For sharp criticisms of this scheme and its fully distributed cost variant, see e.g., the comments (in particular Schmalensee's) on the matter of Policy and Rules Concerning Rates for Dominant Carriers before the U.S. Federal Communications Commission (CC Docket 87-313), as well as Joskow-Schmalensee [1986].

¹⁰ However, it is recognized that at the regulatory review (in 1988 for British Telecom) the regulators are likely to adjust prices to reflect past performance (cost, profit), so that PC regulation may be closer to COS regulation than it might appear.
has been accepted as an improvement over COS regulation.

PC regulation also has serious drawbacks (Joskow-Schmalensee [1986]). First, its fixed-price nature makes no attempt at capturing the regulated firm's rent. 11 This neglect is socially very costly when the shadow cost of public funds is high. Second, a "fundamental shortcoming of price-cap regulation is the danger that the values of the price caps will be set at inappropriate levels" (Baumol-Willig [1987, pp. 11-12]). On the one hand, the appropriate levels must be related to some formal approach. On the other hand, like COS rate making, PC regulation requires information about cost and demand that regulators are unlikely to possess. In the presence of considerable uncertainty about its correct levels (assuming the latter is determined by a theory), a price cap may be chosen so high that it is not binding for the firm; in this case, the monopoly price falls below the cap, a clear perversion of the regulatory process. Or the cap may be binding (lower than the monopoly level), in which case it can still be chosen too high or too low. Last, there is no guarantee that the firm is viable at the caps chosen by the regulator.

To summarize, COS and (the more attractive) PC regulations are informationally demanding with respect to both price and rate of return setting; yet they have normative drawbacks.

3) Implementation of the optimal regulatory scheme.

We now offer an alternative rule based on our analysis of section 3 in the context of third-degree price discrimination. This rule is simple with

11 At least in the absence of regulatory review: see previous footnote. Regulatory reviews lead to some capture of rent, but at the expense of discouraging cost control and innovation.
respect to price setting and informationally demanding with respect to rate-of-return determination.

a) Pricing and competition.

Let us assume that the gross surplus function is \( V(q) = S(q) \) and that the revenue function is \( R(q) = \sum_k q_k \mu_k = \sum_k q_k \mu_k \). The demand function can be written \( q_k = q_k(p) = D_k(p) \). Let \( \eta_{kl} = \frac{\partial q_k}{\partial p_l} q_k \) and \( \eta_k = -\frac{\partial q_k}{\partial p_k} q_k \). \( \eta_{kk} \) denote the elasticities of demand.

For this case, equation (3-10) can be written:

\[
(4-1) \quad p_k + \lambda \left[ p_k + \sum_\ell \frac{\partial p_\ell}{\partial q_k} q_\ell \right] - (1+\lambda) C_k - \lambda F(\beta) \psi'(e) \frac{d}{dq_k} (\psi(\beta)) = 0.
\]

or

\[
(4-2) \quad L_k = R_k + I_k.
\]

where

\[
(4-3) \quad L_k = \frac{(p_k - C_k q_k)}{p_k} \quad \text{is the Lerner index,}
\]

\[
(4-4) \quad R_k = -\frac{\lambda}{1+\lambda} \int_0^\infty \frac{\partial q_k}{\partial p_k} q_k \quad \text{is the Ramsey index,}
\]

and

\[
(4-5) \quad L_k = \frac{\lambda F(\beta) \psi'(e)}{\left(1-\lambda\right) F(\beta) p_k} \frac{d}{dq_k} (\psi(\beta))
\]

is (a multiple of) the incentive correction studied in Section 3. Possibly up to the incentive correction, the pricing structure is determined by a familiar Ramsey formula. For instance, for independent demands,
As mentioned earlier, a difference with the traditional Ramsey formula is that the "Ramsey number" \( \frac{\lambda}{1+\lambda} \) is determined economy-wide and in particular does not depend on the unknown technology, in contrast with the traditional cost-of-service theory of Ramsey, Boiteux and Baumol-Bradford.\(^\text{12}\)

There are other familiar expressions for the Ramsey index. We can for instance rewrite (3-10) taking the derivatives of the social welfare function with respect to prices instead of quantities. This yields

\[
(4-6) \quad (1+\lambda) \sum_{\ell=1}^{n} (p_{\ell} - c_{\ell} \frac{\delta q_{\ell}}{\delta p_{k}}) + \lambda q_{k} \cdot F(\beta) \psi' (e) \left[ \sum_{\ell=1}^{n} \left\{ \frac{d}{dq_{\ell}} (E_{\beta}) \frac{\delta q_{\ell}}{\delta p_{k}} \right\} \right] = 0.
\]

For instance, if we assume that \( C = C(\phi(\beta, e), q) \) so that \( I_{k} = 0 \) for all \( k \), and that \( n = 2 \), it can be shown that

\[
R_{k} = \frac{\lambda}{1+\lambda} \frac{1}{\eta_{k}}, \quad k = 1, 2,
\]

where

\[
\hat{\eta}_{1} \equiv \eta_{1} \left[ \frac{1 - \frac{\eta_{1} \eta_{2}}{\eta_{1} \eta_{2}}}{1 + \frac{p_{2} q_{2}}{p_{1} q_{1}} \frac{\eta_{2}}{\eta_{2}}} \right]
\]

and symmetrically for \( \hat{\eta}_{2} \). The "superelasticities" \( \hat{\eta}_{k} \) were first derived by Boiteux [1956] (see Brown-Sibley [1985] for a general exposition). Thus, with

\(\text{12}\) This approach assumes that the regulator has full information and maximizes \( S(q) - C(q) \) subject to the constraint \( R(q) - C(q) \geq 0 \). Letting \( \mu \) denote the Lagrange multiplier of the constraint, the optimization yields \( L_{k} = R_{k} \), where \( R_{k} \) is the same as \( R_{k} \) except that \( \lambda \) is replaced by the (cost and demand contingent) \( \mu \).
substitutes ($\eta_{k\ell} > 0$ for $k \neq \ell$), the price of good $k$ should be adjusted upwards to account for the fact that such an increase raises the revenue received on good $\ell \neq k$.

1. Price setting: The regulator forms a rough estimate of the demand $q_k$ for each of the firm's products at the optimal prices. $q_k$ can result from survey data. Or, as will be shown in 4e, it can be obtained from historical data, a policy more in line with the simplicity -- i.e., absence of discretion -- requirement. The regulator imposes a price tax on good $k$ equal to $p_k\tilde{q}_k$ (in one variant, see 4e for another variant), where $\lambda$ is a data obtained in the public finance literature (see footnote 4). He otherwise gives entire price freedom to the firm. The firm maximizes its profit $\pi$ like a monopolist subject to the price tax. Thus

$$\pi = \max_{\mathbf{p}} \left\{ \sum_{k=1}^{n} p_k D_k(p) \cdot C(\beta, e, D(p)) - \sum_{k=1}^{n} \frac{p_k \tilde{q}_k}{1+\lambda} \right\},$$

where $\mathbf{p} = (p_1, \ldots, p_n)$ and $D(p) = (D_1(p), \ldots, D_n(p))$.

2. Rate-of-return setting: The regulator uses linear rate of return rules. The firm's rate of return is thus

$$t = D + K\pi,$$

where $D$ is a fixed fee, and $K$ belongs to $[0,1]$ ($K = 0$ corresponds to a cost-plus contract and $K = 1$ to a fixed-price contract). Optimally the regulator offers a menu of such contracts from which the firm chooses one. Alternatively, the firm can be asked to announce its expected profit $\pi^e$ and is then paid
where \( \pi - \pi^e \) is the excess profit (or loss). Incentive compatibility requires that both \( \bar{D} \) and \( \bar{R} \) increase with \( \pi^e \) (see Laffont-Tirole [1986]).

3. Scope of regulation. The profit of the entire firm is subject to the rate-of-return setting. Thus, if the firm produces unregulated goods \( k = n+1, \ldots, m \), these goods should not be treated separately: Their costs and revenues should be included in the definition of \( \pi \).

To investigate the properties of the price-setting rule, let us consider the "benchmark case," in which the cost technology satisfies the separability condition of Proposition 2 (so there is no incentive correction), and there is no pricing correction due to the existence of an unregulated competitor with market power or subsidized, or to the existence of a regulated competitor (see Section 5). Assuming also that all the firm's products are regulated, the maximization of (4-7) with respect to \( p_k \) yields

\[
(4-10) \quad \sum_{\ell} \left( p_\ell - C_{Q_\ell} \right) \frac{\partial D_\ell}{\partial p_k} + \left[ D_k - \frac{\hat{q}_k}{1 + \lambda} \right] = 0.
\]

If \( \hat{q}_k = D_k \), then (4-10) coincides with equation (4.6), and the pricing structure is socially optimal.

In practice, the regulator is likely to make substantial errors in estimating demand, and thus the optimal \( \hat{q}_k \) (see 4e for the use of historical data to reduce errors). Yet the errors in pricing\(^{13}\) are likely to be lower

\[\text{For instance, if demands are independent, if good } k \text{'s elasticity of demand is equal to } 2 \text{ and if } \lambda = 1/2, \text{ the optimal price of good } k \text{ (given by (4-6)) is } 20\% \text{ above its marginal cost. However, equation (4-10) yields a price } 11\% \text{ (respectively } 30\%) \text{ above marginal cost if the regulator overestimates}\]

\[31\]

\(^{13}\)For instance, if demands are independent, if good k's elasticity of demand is equal to 2 and if \( \lambda = 1/2 \), the optimal price of good k (given by (4-6)) is 20% above its marginal cost. However, equation (4-10) yields a price 11% (respectively 30%) above marginal cost if the regulator overestimates.
than those resulting from COS or PC regulations, which, besides relying on a theoretically less satisfactory approach, compound errors in cost and demand estimations.

Next the rate of return setting is inspired by Proposition 9, which establishes that under some circumstances, optimal cost reimbursement or profit-sharing rules are linear in the firm's performance (cost or profit). Now optimal rules are certainly not always linear, but from a policy viewpoint, Proposition 9 tells us that the simple linear rules may not be too off-track as approximations to the optimal scheme. The coefficient \( K \) in \([0,1]\) reflects the tradeoff between rent extraction (which requires \( K = 0 \)) and incentives (which require \( K = 1 \)). Ideally, the regulator ought to offer a menu of linear contracts, including one with \( K = 1 \) and several with \( 0 < K < 1 \). While it is possible to further characterize the optimal menu of contracts without knowledge of the exact distribution of uncertainty (see above), we realize that our rate-of-return prescription is informationally demanding, as its details depend on the regulator's formulating a subjective distribution about the technological parameter. So is any prescription aiming at promoting incentives (as it seems that only a cost-plus contract can abstract from the question of whether the firm will remain viable or conversely enjoy an excessive rent if the regulator errs in estimating the distribution of uncertainty). It is also important to note that the firm's pricing decisions are insensitive to errors in the choice of the rate-of-return rule.

Last, along with other authors, we propose to regulate the firm's overall profit rate to avoid cross subsidization through arbitrary cost allocations. For instance, an electric utility would gain \((1-K) > 0\) on every dollar of profit shifted back to an unregulated oil, coal or power plant subsidiary. (respectively underestimates) \( \hat{c}_K \) by 20%.
Or, even if the subcosts of regulated and unregulated products can be separated, the managers will devote more effort to unregulated products (cost-side cross subsidization). 14

c) Corrections. Subsection 4b shows that, in the benchmark case, the price tax scheme yields approximately the socially optimal prices. In other cases, the price structure may have to be adjusted to reflect non separabilities in the cost function (incentive correction of section 3), or, as we will see, to internalize externalities on other regulated firms (see 5b), or to correct distorted pricing by unregulated competitors (see 5c). For instance, the price of a product should be corrected upwards when the product competes with those of other regulated firms. In contrast, the regulated firm facing an unregulated competitor who has market power or is subsidized may have to adjust its price downwards. We view the unit price tax $q_k/(1+\lambda)$ as a benchmark from which adjustments must be made. The regulator can raise $\lambda$ (or equivalently, reduce $q_k$) when an upward adjustment is needed, and conversely for a downward adjustment.

d) Quality issues. A concern with PC regulation is that the firm might evade a binding price cap by an unverifiable cut in product quality. Indeed,

14 Because of rapid technological change (in particular in the telecommunications industry) or of legal limitations on the scope of regulation, regulatory agencies continually face a dilemma between allowing and prohibiting entry by regulated firms into new markets. Prohibition has obvious drawbacks in terms of product diversity offered to consumers. Assuming that immediate regulation of new products is infeasible, it still may be preferable to let the regulated firm enter and charge a monopoly price on a new market, which it would do under our regulatory scheme. Of course, the theory of industrial organization tells us that a monopoly may introduce too many or too few products. Furthermore, it might be the case that expansion into new markets leads to a loss of control by the regulator concerning the optimal cost reimbursement rule. These points are well worth further investigation.
Vickers and Yarrow [1988, p. 228] using Oftel's quality-of-service indicators note that "British Telecom's quality of service has not deteriorated since privatization, but that it had not improved much either. Given the rate of advance of telecommunications technology, this record is poor." A similar problem may arise for our price tax regulation. To simplify notation, consider the shared-fixed-cost model and independent demands. Under a price tax regulation, the regulated firm maximizes, for each k:

\[
\text{Max} \left\{ \frac{p_k \hat{D}_k(p_k, s_k)}{1+\lambda} \cdot \mathcal{C}^k(\beta, e_k, \hat{D}_k(p_k, s_k)) \right\}
\]

over the price \( p_k \) and the quality of service \( s_k \). Like a pure monopoly (Spence [1975]), the firm chooses a quality that is higher or lower than the socially optimal level, depending on whether the marginal consumer values quality more or less than the average consumer.\(^{15}\) So, in general, the regulated firm can under- or oversupply quality.

**Remark:** There is a case in which the firm clearly undersupplies quality under our scheme (and under price caps). Suppose that price reductions and quality increases are perfect substitutes for both the consumers and the firm

\[
(D_k(p_k, s_k) = \tilde{D}_k(p_k - s_k) \quad \text{and} \quad \mathcal{C}^k(\beta, e_k, \hat{D}_k(p_k, s_k)) =
\]

\[
\tilde{C}^k(\beta, e_k, \tilde{D}_k(p_k - s_k)) + s_k \tilde{D}_k(p_k - s_k). \]

The program (4-11) then yields the monopoly outcome. The firm has an incentive to drive \( p_k \) to 0 to escape the price tax, and to choose a quality level equal to minus the monopoly price\(^{16}\) if such a low level of quality is feasible and unverifiable by the regulatory agency.

\(^{15}\) Note that, unlike Spence, we do not have to add the qualifier "for a given price," as the regulated firm's price is optimal given the quality.

\(^{16}\) which maximizes \( (p_k \tilde{D}_k(p_k) - \tilde{C}_k(\beta, e_k, \tilde{D}_k(p_k))) \).
Furthermore the allocation (including the firm's profit) is the same as under a price cap. All this is not surprising. If the regulator cannot monitor quality and if quality is a perfect substitute for price, everything is as if the price itself were not regulated. Such goods naturally fall under the heading of unregulated goods. Any attempt at regulating pricing seems to require quality control or at least an imperfect substitutability between price and quality.

e) Using historical data to construct the price tax. An alternative to basing the price tax on good \( k \) on a rough estimate of \( \hat{q}_k \) is to use past quantity observations, if such exist. But the regulator must be careful not to offer perverse incentives to a foresighted firm. To simplify notation, let us assume in the whole subsection that there is a single product, and that the cost function satisfies the separability condition of Proposition 3. One scheme based on historical data consists in taxing price changes. More precisely, let

\[
T(p_t, p_{t-1}, q_{t-1}) = \frac{(p_t - p_{t-1})q_{t-1}}{(1+\lambda)(1-\delta)}
\]

be the tax levied on the firm at date \( t \), where \( q_{t-1} = D(p_{t-1}) \) is the date \(-(t-1)\) sale. Faced with this price tax, the firm maximizes

\[
\sum_{t=0}^{\infty} \delta^t [K_t[p_t D(p_t) - T(p_t, p_{t-1}, D(p_{t-1})) - C(T, e_t, D(p_t))] - \psi(e_t)],
\]

Under a price cap \( \hat{p}_k \), the profit attributable to good \( k \) is obtained by maximizing

\[
\pi_k = p_k D_k(p_k, s_k) - c_k(T, e_k, D(p_k, s_k)) \text{ subject to } p_k \leq \hat{p}_k.
\]

The two cases yield a generalized price \( p_k - s_k \) equal to the monopoly generalized price and a profit \( \pi_k \) equal to the monopoly profit.
assuming that the cost function has the separable form so as to be able to abstract from the incentive correction. \( \delta \) denotes the discount factor. \( K_t \) is the slope of the firm's incentive scheme at date \( t \).

The normative analysis implies that the incentive scheme should be time invariant (as is easily seen). So let us complete our tax-on-price-changes scheme by a time invariant rate of return rule; in particular let \( K_t = K \) for all \( t \). The first-order condition with respect to \( p_t \) in (4.13) yields:

\[
(4.14) \quad (p_t - C)^\prime D'(p_t) + D(p_t) - \frac{\delta((p_{t+1} - p_t)D'(p_t) - D(p_t))}{(1+\lambda)(1-\delta)} = 0.
\]

In steady state, \( p_{t-1} = p_t = p_{t+1} \), and hence

\[
(4.15) \quad (p - C)^\prime D + \frac{AD}{1+\lambda} = 0,
\]

which is the socially optimal pricing policy.

Actually the price tax schemes based on survey data and on historical data may be used jointly. For new products in particular, the survey data may be used to initiate the price tax (i.e., to yield a \( q_0 \)), to then let the price tax to be later determined by (4.12). This procedure may yield faster convergence than a price tax initiated by the monopoly output.

The dynamic behavior of price \( p_t \) is given by (4.14), the initial price \( p_{-1} \) and demand \( q_{-1} = D(p_{-1}) \) at date \( -1 \) and the transversality condition. Let us make the following assumption:

\[
A: \quad D(p) = a - bp; \quad C = (\delta - e)q; \quad \psi(e) = e^2/2.
\]

Given a slope \( K \) of the incentive scheme, the date-\( t \) effort is given by \( \psi'(e_t) = Kq_t + e_t = K(a - bp_t) \). Substituting \( e_t \) into (4.14) yields a second-order linear difference equation in the price.
Proposition 10: Under assumption A, the optimal price is given by

\[ p_t = p^* + (p^* - p^*) r^{t-1}, \]

where \( p^* \) is the steady state solution given by (4-15) (i.e., the socially optimal price) and

\[ r = \frac{Kb}{2(1+\lambda)(1-\delta)} \left( 1 + \frac{1+\lambda(1-\delta)}{2(1+\lambda)(1-\delta)} - \frac{Kb}{2(1+\lambda)(1-\delta)} \right)^2 - \delta. \]

(A sufficient condition for \( r \) to be less than 1 is that \( (1-\delta)(1+\lambda) < 1 \). So for \( \lambda = .3, \delta > .23 \) is sufficient for \( r < 1 \).)

The proof of Proposition 10 is available upon request from the authors.

We conclude that the delegated price converges in a non-oscillatory way toward the socially optimal price.

5. Applications.

This section applies the general results of Section 3, in particular equation (3-10), to specific pricing problems.

a) Pricing and competition.

Because the (cost-based) incentive correction has been discussed in detail in Section 3, we devote this subsection to a discussion of the (demand determined) Ramsey index for some competitive variants of the third degree price discrimination model. The following analysis is simple and often will only state the results; yet it contains some pitfalls. Its general goal is to explain how to compute elasticities correctly. To simplify, we will assume that the demands for the regulated firm's \( n \) products are independent, and that a competitor produces a good (\( n+1 \)) that is an imperfect substitute for good \( n \). Thus, the consumer gross surplus can be written:
\[ S(q, q_{n+1}) = \sum_{k=1}^{n-1} S^k(q_k) + S(q_n, q_{n+1}). \]

The competitor produces with cost function \( C^{n+1}(q_{n+1}) \). [We abstract from possible incentive issues for the competition. The absence of uncertainty about the competitor also implies that the regulator cannot learn from the quantities traded information he does not infer from the regulated firm's choice of prices, which he could do if the two firms' technologies were correlated.] We will also assume to simplify that \( C = C(\Phi(\beta, e), q) \) so that \( I_k = 0 \) for all \( k \in (1, \ldots, n) \). Incentive corrections arising for alternative functional form would be additive and would not affect the Ramsey terms developed below.

1) Regulated competition. Suppose that good \((n+1)\) is produced by another regulated firm. For instance, electricity and local gas distribution are both regulated at the state level in the U.S. and are substitutes for heating and cooking. The competitor's revenue value has then social value \( \lambda p_{n+1} q_{n+1} - \lambda q_{n+1} (\delta S / \delta q_{n+1}) \). Then the \( n \)th Ramsey index for the multiproduct firm is given by a superelasticity formula:

\[
(5-1) \quad P_n = \frac{\lambda}{(1+\lambda)\eta_n} \left[ 1 + \frac{p_{n+1} q_{n+1} \eta_{n+1}}{p_n \eta_n} \right] \left[ 1 - \frac{\eta_{n,n+1}\eta_{n+1}}{\eta_n \eta_{n+1}} \right].
\]

That is, aside from the incentive corrections (which are here equal to zero from our technological assumptions), the price of good \( n \) should be regulated as if the two firms were merged. A regulated firm must be induced to internalize the loss in another regulated firm's revenue due to its pricing behavior.
This conclusion is very natural, yet it does not hold in the traditional cost-of-service model.\textsuperscript{18,19} The conclusion is more subtle than it appears as it hinges on the separability between incentives and Ramsey correction and, especially, on the shadow price of public funds being firm independent.

We now suppose that the competitor is unregulated. We maintain the assumption that good \(n\) and \(n+1\) are substitutes (one may think of the competition between AT&T and unregulated long-distance carriers or between the railway and road freight. The case of complements -- e.g., public transportation and private accommodation -- can be treated analogously). We assume that the regulated and the unregulated firms compete in prices and we consider both the sequential and simultaneous timing. In the sequential timing, the regulator is able to choose a regulatory scheme and induce price setting by the regulated firm before the unregulated competitor chooses its price. In the other timing, the two decisions are simultaneous, so the

\textsuperscript{18}In the COS model, even if the definition of social surplus is modified to include the externality on the competitor (by adding the term \(\mu_{n+1}p_{n+1}q_{n+1}\) where \(\mu_{n+1}\) is the shadow cost of the competitor's budget constraint), there is no reason why a formula of the type (5-1) should hold (as the shadow price of the multiproduct firm's budget constraint, \(\mu\), in general differs from \(\mu_{n+1}\)). See Brauetigam [1984]. The correct superelasticity is obtained in the COS model only in Boiteux [1956]'s second case of a single budget constraint for the public sector (as opposed to the individual budget constraints of the traditional COS model) or in Brauetigam [1984]'s "viable industry Ramsey optimum," in which the regulator can impose lump-sum transfers among the regulated firms.

\textsuperscript{19}Brauetigam [1979] considers a model of an increasing returns-to-scale firm facing constant-returns-to-scale competitors. In the "totally regulated second best," the increasing-returns firm is regulated in a Ramsey-Boiteux fashion, and the competitors are taxed. However, the taxes do not enter the social welfare function, as the shadow cost of public funds is assumed equal to 0. The conclusion thus differs from (5-1). The Ramsey-Boiteux firm in Brauetigam is instructed to behave as if there were no competition (in our notation: \(q_n^\ast = \lambda/(1+\lambda)q_n\)). See also our previous footnote.
regulator loses his commitment power. Needless to say, conclusions that are insensitive to the timing are more attractive.

In the social welfare function, the competitor's revenue now has no weight, as it cancels with the consumers' expenditure on good n+1: The generalized consumer surplus minus production costs is equal to

$$\sum_{k=1}^{n-1} S_k(q_k) + S(q_n, q_{n+1}) + \alpha p_n q_n - (1+\lambda)C\left[\phi(\beta, \epsilon), q\right] - C^{n+1}(q_{n+1}).$$

[To obtain the social welfare function, one must further subtract $(1+\lambda)\psi(\epsilon) + \lambda U].20$

a2) Unregulated competitive fringe. Suppose that the fringe is a price taker $(p_{n+1} = \delta S/\delta q_{n+1} = dC^{n+1}/dq_{n+1})$, and consider first the simultaneous timing. The optimal regulatory policy then yields:

$$L_n = \frac{\lambda}{1+\lambda} \frac{1}{\eta_n}.$$  

That is, the regulated firm should "ignore" the competitive fringe (i.e., not internalize the effect of its pricing, unlike in (5-1)).21

When the regulator is a Stackelberg leader (sequential timing), $p_n$ has an effect on $p_{n+1}$. We let

$$\epsilon = \frac{dp_{n+1}}{dp_n} = \frac{p_{n+1}}{p_n},$$

20 We are here considering the polar case in which none of the unregulated fringe's profit goes to the government. If a fraction $r$ of profits were captured through a profit tax, the fringe's profit would have social value $(1+\lambda r)$ instead of 1. In particular, if $r = 1$, the analysis is the same as when the fringe is regulated.

21 This result was first obtained by Braeutigam [1979] in a Ramsey-Boiteux framework.
denote the elasticity of the fringe's price reaction. Let us define the "generalized or net elasticity of demand for good n" (which takes into account the fringe’s reaction):

\[ \bar{\eta}_n = -\frac{\partial q_n}{\partial p_n} + \frac{\partial q_n}{\partial p_{n+1}} \frac{dp_{n+1}}{dp_n} \left/ \frac{q_n}{p_n} \right. - \eta_n - \eta_{n+1}. \]

For strategic complements ($\alpha > 0$) and demand substitutes ($\eta_{n,n+1} > 0$), $\bar{\eta}_n < \eta_n$. The optimal regulatory policy for the sequential timing yields:

\[ (5.3) \quad I_n = \frac{\lambda}{1+\lambda} \frac{1}{\bar{\eta}_n}. \]

That is, the price of good n is higher under sequential timing, because a high $p_n$ induces the fringe to raise its price, which increases the regulated firm's revenue.

A useful conclusion is that both cases can be summarized in a single recommendation. The relevant elasticity of demand is the net elasticity of demand, which may or may not coincide with the ordinary elasticity of demand. This result is important because it, together with the price tax analysis of Section 4, implies that the optimal price decision can be delegated to the firm. That is, the regulator need not know the exact game firms are playing (sequential or simultaneous), as long as the regulated firm knows. The regulated firm will automatically adopt the net elasticity as the correct measure of a change in price on its demand.

32) Unregulated competition with distorted pricing. Subsection 32 assumed that the producer of good (n+1) charges the social marginal cost. In practice, his price may be distorted because of either market power or subsidies.

\[ \text{Competitor with market power.} \] Suppose that the competitor equates
marginal revenue and marginal cost:

\[ q_{n+1} + p_{n+1} \frac{\delta q_{n+1}}{\delta p_{n+1}} = \frac{\delta c_{n+1}}{\delta q_{n+1}} \frac{\delta q_{n+1}}{\delta p_{n+1}}. \]

Under simultaneous competition, the optimal Lerner index for good n is:

\[ L_n = \frac{\lambda}{(1+\lambda)\eta_n} + \frac{(p_{n+1}q_{n+1})\eta_{n+1,n,n}^n}{(p_nq_n)(1+\lambda)\eta_n\eta_{n+1}}. \]

Equation (5-4) implies that the price of good n should exceed the level corresponding to the ordinary elasticity of demand. This conclusion is not surprising. A standard result is that the demand for a monopolized product should be encouraged by a commodity subsidy. Here, reducing the demand for good n has a similar effect as a (missing) commodity subsidy on good (n+1), in that both encourage consumption of good (n+1).  

Under sequential competition, the optimal Lerner index for good n is:

\[ L_n = \frac{\lambda}{(1+\lambda)\eta_n} + \frac{(p_{n+1}q_{n+1})\eta_{n+1,n,n}^n}{(p_nq_n)(1+\lambda)\eta_n\eta_{n+1}}. \]

Equation (5-5) unveils two new and straightforward effects. By raising \( p_n \), the regulated firm induces the competitor to raise \( p_{n+1} \), which induces a further distortion in the consumption of good (n+1); this effect shows up in the numerator of (5-5). But the increase in \( p_{n+1} \) also raises the regulated firm’s revenue, which has a social value;  

The new term in the denominator explains the new term in the denominator.

---

22 Note that our framework implies that commodity subsidies are an imperfect instrument because of the existence of a shadow cost of public goods.

23 As long as the previous effect is not so powerful that it induces the regulated firm to charge below marginal cost (an unlikely occurrence).
Subsidized competition. It may happen that the unregulated competitor's pricing is distorted by subsidies. For instance, it is often asserted that railroads face "unfair competition" from subsidized competitors (barges which may not pay for the use of facilities and trucks which may not pay for congestion costs or highway maintenance); and that subway tokens should be subsidized to account for the fact that charging congestion and pollution costs on city drives is too costly.

Let us assume that the competitor is a price taker and that the subsidy in question is the absence of payment of some social cost (pollution, congestion cost). The consumer gross surplus function is thus

\[ S(q, \tau) = q - \frac{1}{2} \tau q^2 + \frac{1}{3} \tau q^3, \]

where \( \tau \) is the implicit subsidy on good \( q \). The competitor produces until \( p_n \) equals his private marginal cost \( (dC/dq_n) \). The optimal Lerner index for good \( n \) in the simultaneous mode is given by:

\[ L_n = \frac{1}{(1+\lambda)\eta_n} - \frac{(\tau_{n+1} q_{n+1})\eta_{n+1}}{(p_n q_n)(1+\lambda)\eta_n}. \]

Thus the price of good \( n \) is lowered to reduce the demand for, and hence the negative externality created by good \( n+1 \).

In the sequential mode, the formula is the same as (5-6), except that the elasticities must be replaced by the net elasticities (which includes the elasticity of reaction of the competitive fringe):

\[ \tilde{\tau}_n = \tau_n - (n_{n+1} + 1 + p_n) \]

and

\[ \tilde{\eta}_{n+1} = \eta_n - (n_{n+1} + 1 + p_n) \]

Assuming that (5-6) yields pricing above marginal cost (which is not
guaranteed), the fringe's reaction lowers the net elasticity of demand for good n, and thus calls for a higher price $p_n$. Furthermore, the fringe's reaction to an increase in $p_n$ reduces demand for the subsidized good (n+1), a socially useful effect. Hence, the two effects have the same sign, so that the sequential timing calls for a higher price $p_n$ than the simultaneous timing.

a4) Access pricing. We now analyze a variant of the unregulated competitive fringe of subsection a2) to study the pricing of an input by a regulated multiproduct firm to a competitor producing a good that competes with another product of the regulated firm. Such "access pricing" is a common issue and has been the subject of much legal and regulatory activity lately. For instance, before the divestiture, AT&T, a regulated firm, had monopoly over the local exchange and competed with a couple of unregulated companies on the long-distance market; the rivals needed access to AT&T's local network to supply toll calls. A railroad may have a monopoly position from A to B, and compete with trucks or another railroad from B to C; to carry freight from A to C, the competitors need access to the segment AB. In the energy field, British Gas' rivals need access to British Gas' distribution network to compete for gas customers. Similarly, the U.S. Public Utility Regulatory Policy Act of 1978 requires electric utilities to purchase power from "qualified" independent suppliers.

The main theoretical question is whether the regulated firm has too much incentive to foreclose the competitive segment of the market by charging a high transfer price. To analyze this question, suppose that one unit of good 1, an intermediate good produced by the regulated firm, is consumed to produce each unit of good n (produced by the regulated firm) or of good (n+1) (produced by the fringe) total production of good 1 is thus $q_n + q_{n+1}$. Letting
p_1 denote the transfer price, the fringe's cost is \[ c_{n+1}(q_{n+1}) + p_1 q_{n+1} \], while the regulated firm's cost is \[ C(\Phi(\beta, e), q_n + q_{n+1}, q_2, \ldots, q_n) \]. Price taking behavior by the fringe implies that \( p_{n+1} = \frac{dC_{n+1}}{dq_{n+1}} + p_1 \). The social welfare function can then be written:

\[
W = S_2(q_2) + \ldots + S^{n-1}(q_{n-1}) + \bar{S}(q_n, q_{n+1}) + \lambda \left[ (p_{n+1} - \frac{dC_{n+1}}{dq_{n+1}}) q_{n+1} + \sum_{k=2}^{n} p_k q_k \right] - (1+\lambda)(\psi(e)+C) - \lambda U \cdot C^{n+1}(q_{n+1}).
\]

Let \( \eta_1 = -\frac{\delta q_{n+1}}{\delta p_1} \frac{p_1}{q_{n+1}} = -\frac{p_1}{q_{n+1}} \left[ \frac{\delta p_{n+1}}{\delta q_{n+1}} - \frac{d^2C_{n+1}}{dq_{n+1}^2} \right] \) denote the fringe's elasticity of demand for good 1, and \( R_1 = p_1 q_{n+1}, R_n = p_n q_n, R_{n+1} = p_{n+1} q_{n+1} \) denote the revenues on goods 1, n and \((n+1)\). The notation for the other elasticities follows our previous convention. Optimal pricing requires that:

\[
(5-7) \quad L_k = \frac{\lambda}{1+\lambda} \frac{1}{\eta_k} \quad k = 2, \ldots, n-1
\]

\[
(5-8) \quad L_n = \frac{p_n - \frac{\delta C}{\delta q_n} - \frac{\delta C}{\delta q_1}}{p_n} = \frac{\lambda}{1+\lambda} \frac{1}{\eta_n}
\]

where \( \hat{\eta}_n = \eta_n \left[ \frac{1 - \frac{\eta_{n+1} q_{n+1}}{\eta_n q_n}}{\frac{R_{n+1}}{R_n} \frac{\eta_{n+1} q_{n+1}}{\eta_{n} q_n}} \right] < \eta_n \)

is good n's superelasticity (see the introduction to this subsection).

The Lerner index for the access good is given by
The key to understanding equations (5-8) and (5-9) is to view goods \( n \) and \( n+1 \) as substitutes. An increase in \( p_n \) raises the demand for good \( (n+1) \), which in turn raises the demand for the access good by an equal amount. It is therefore not surprising that the Lerner index for good \( n \) is given by a familiar superelasticity formula. Similarly, an increase in \( p_1 \) raises the price of good \( (n+1) \), which raises demand for good \( n \). Hence, the externality on good \( n \) requires that the price of good \( 1 \) exceeds the price \((\lambda/(1+\lambda)\eta_1)\) given by a "myopic" inverse elasticity rule.

The second insight of this subsection is that the optimal pricing policy can be decentralized by a price tax of the form discussed in Section 4. More precisely, form estimates \( \hat{q}_n \) and \( \hat{q}_{n+1} \) of the demands for goods \( n \) and \( 1 \) at the optimal prices, and let the regulated firm maximize its profit subject to a price tax \((p_1 \hat{q}_{n+1} + p_n \hat{q}_n)/(1+\lambda)\). [That is, the firm chooses \((p_1, \ldots, p_n)\) so as to maximize \((p_1 \hat{q}_{n+1} + \sum_{k=2}^{n} p_k \hat{q}_k + C - (p_1 \hat{q}_{n+1} + p_n \hat{q}_n)/(1+\lambda))\).] The reader will easily check that the resulting pricing structure converges to the one given by equations (5-7) through (5-9) when the estimates \( \hat{q}_n \) and \( \hat{q}_{n+1} \) become more and more accurate. The possibility of decentralized price making is natural in view of the similar result obtained for a competitive fringe in subsection 4.2. We conclude that, in the context of our model, the price-tax scheme allows the
regulator to free even the regulated firm's access and competitive prices without fear of excessive market foreclosure or cross-subsidization.

b) *Second-degree price discrimination.* Let us now assume that the regulated firm produces a single physical good and practices non-linear pricing. To see why we can apply the general multiproduct framework to this situation, consider two types of consumers, \( k = 1,2 \). Consumers of type \( k \), who are in proportion \( a_k \) (such that \( a_1 + a_2 = 1 \)), have gross surplus function \( S_k(q_k) \), with \( S'_2 > S'_1 \). Consumers of type 2 are thus the "high-demand consumers." The consumers' gross surplus function is then

\[
V(q) = a_1 S_1(q_1) + a_2 S_2(q_2).
\]

The cost function can be written \( C(\beta, e, Q) \), where \( Q = a_1 q_1 + a_2 q_2 \) is total output. Note that Proposition 4 implies that all incentive corrections have the same sign (indeed, the right-hand side in (3-10) is the same for all \( k \)).

We analyze fully non-linear price (subsection b1) and two-part pricing (subsection b2).

b1) *Fully non-linear pricing.*

Let \((T_1, q_1)\) and \((T_2, q_2)\) denote the payment-quantity pairs for the two types of consumers. Incentive compatibility requires that high demand consumers do not pretend they have low demand:

\[
(5-10) \quad S_2(q_2) - T_2 \geq S_2(q_1) - T_1.
\]

Furthermore, we assume that even the low-demand consumers consume the good.\(^{24}\)

\(^{24}\) By a familiar reasoning, the other incentive compatibility and individual rationality constraints are not binding.
(5-11) \[ S_1(q_1) - T_1 \geq 0. \]

Because public funds are costly, (5-10) and (5-11) are binding. The revenue function can then be written:

(5-12) \[ R(q) = \alpha_1 T_1 + \alpha_2 T_2 - S_1(q_1) + \alpha_2 (S_2(q_2) - S_2(q_1)). \]

Letting \( p_k = S_k'(q_k), \) we can now apply equation (3.10):

(5-13) \[ \alpha_1 p_1 + \lambda [p_1 - \alpha_2 S_2'(q_1)] = \alpha_1 (1+\lambda) C_1 Q + I \]

and

(5-14) \[ \alpha_2 p_2 + \lambda \alpha_2 p_2 = \alpha_2 (1+\lambda) C_2 Q + I, \]

where \( I = \frac{\lambda F(\beta) e'(\beta)}{f(\beta)} \frac{d}{dQ} (E_{\beta}) \) is the common incentive correction. For instance, if \( C = C(\beta, \varepsilon, Q), \) so that \( I = 0, \) we get:

(5-15) \[ L_1 = \frac{\alpha_2}{\alpha_1} \frac{\lambda}{1+\lambda} \left[ \frac{S_2'(q_1) - S_1'(q_1)}{S_1'(q_1)} \right] > 0 \]

(5-16) \[ L_2 = 0. \]

That is, we obtain the familiar result that the high consumption should not be distorted, while the low consumption is lower than the one when the firm has perfect information about preferences. The shadow cost of public funds induces the regulated firm to behave somewhat like a monopolist and to distort the consumption vector.

There are several interesting extensions of the general approach. For instance, one could allow for outside competition in this second-degree price discrimination model. The possibility of bypass (see, e.g., Einhorn [1987]) in general puts other constraints on the above maximization problem (for

48
instance, high-demand consumers of long-distance services may build their own access to the long-distance carriers and bypass the local exchanges. We will not pursue this issue here (see Laffont-Tirole (1988b)).

b2) Two-part tariffs. We now restrict attention to combinations of a fixed fee and a marginal price: \( T(q) = \lambda + pq \). This restriction is interesting in its own right, and will offer us some insight on how non-linear pricing by a regulated firm can be decentralized, that is, on how the price tax can be generalized.

Let \( S_i(p) \) and \( S^n_i(p) \) denote the gross and net surpluses of consumers of type \( i \) for marginal price \( p \) (\( dS_i/dp = pD'_i(p) \), \( dS^n_i/dp = -D_i(p) \)). For a given marginal price \( p \), the fixed fee is optimally chosen equal to the low demand consumers' net surplus: \( \lambda = S^n_1(p) \). Thus, the firm's revenue is:

\[
R(p) = S^n_1(p) + pD(p),
\]

where \( D(p) = \alpha_1 D_1(p) + \alpha_2 D_2(p) \) is total demand at price \( p \). It is easily checked that the optimal marginal price is given by

\[
(5-17) \quad L = \frac{pC_0}{p} = \left[ \frac{\alpha_2 D_2}{\alpha_1 D_1 + \alpha_2 D_2} \right] \left[ \frac{\lambda}{1 + \lambda \eta} \right],
\]

where \( \eta \) is the elasticity of total demand (\( \eta = \frac{dD}{dp}/D(p) \)). The marginal price should thus lie between marginal cost and the optimal linear price (which satisfies \( L = \frac{\lambda}{1 + \lambda \eta} \)). Its location between these two values depends on the relative demands of the two types of consumers.

An issue we did not address in the fully non-linear pricing analysis is whether the optimal tariff can be delegated to the firm. It turns out that there is a simple way to do so. Form an estimate of total demand \( q \) at the optimal tariff, and impose the tax.
(5-18) \[ T(A, p) = \frac{1}{1+\lambda} |A+pq| \].

That is, price increases and fixed fee increases are taxed analogously. Decentralize the choice of the tariff \((A, p)\) to the firm, which then maximizes total revenue minus cost minus the price tax. It is easily seen that, when \(q\) becomes more and more accurate, the firm's choice of tariff converges to \((S^R, p)\) where \(p\) is given by (5.17). Thus, it is straightforward to generalize the price tax scheme to two-part pricing.

c) **Taxation by regulation.** In the analysis of subsections 4a and 4b, the gross consumer surplus was taken to be the unweighted sum of individual gross consumer surpluses. In practice the distribution of income is not socially optimal, and one may wonder whether regulation should be used as a substitute for a (non-existent) perfect taxation system. In more technical terms, should the Lerner index differ from the sum of the Ramsey index and the incentive correction for specific goods? That is, in the terminology of Section 2, should there be demand-side cross subsidization? Posner [1971] has argued that "existing views of regulation do not explain well the important phenomenon of internal subsidization" (page 28) and offers to "modify existing views by admitting that one of the functions of regulation is to perform distributive and allocative chores usually associated with the taxing or financial branch of government" (page 23). Examples of regulated pricing that seem to be motivated by redistributive concerns are the use of uniform rates for mail and telephone, avoiding cost-based price discrimination between rural and urban areas, or British Telecom's license agreement preventing low-usage residential customers' line rental charges from growing faster than the retail price index plus two percent per year.
The issue of demand-side cross-subsidization closely relates to the debate on direct vs. indirect taxation in public finance about whether the responsibility for redistribution should be assumed by income taxation only, or by differential excise taxes as well. One of the main findings of this literature (Atkinson-Stiglitz [1976], Mirrlees [1976]) is that using indirect taxation to redistribute income is a less obvious policy than one might have thought. Indeed, if i) consumers differ in their ability (or wage $w$ earned per unit of labor), ii) their utility function exhibits weak separability between consumption and labor ($U = U(A(q_1, \ldots, q_n), L)$, where $L$ is labor), and iii) the government uses optimal excise taxes on goods and optimal non-linear income taxation (so taxes paid by a consumer of "type" $w$ have the form $\sum_{k=1}^{n} \tau_k q_k + T(wL)$), the government need not use indirect taxation. [That is, it can pick $\tau_k = \tau$ for all $k$; and, without loss of generality, $\tau$ can be taken to be zero from the homogeneity of the consumers' budget constraints.]

This result does not imply that taxation by regulation is socially inefficient, simply that it is not a foregone conclusion. There are several ways to reintroduce motives for demand-side cross-subsidization. First, the strong assumption of weak separability may be relaxed; however, as Atkinson and Stiglitz [1980, p. 437] note, nearly all empirical studies of demand and labor supply functions have made this assumption and are therefore of little guidance for our purpose. Second, one may assume that income is imperfectly observable because of domestic production or of tax evasion (Deaton [1977] for instance takes this route). Third, consumers may differ with respect to other attributes than ability. To these three reasons listed by Atkinson-Stiglitz [1980, p. 440] can be added a fourth. There may exist situations in which commodity taxation can be based on information not available for income taxation, because of different arbitrage possibilities. For instance, it may be costly to make income taxation contingent on the rural or urban character.
of the consumer's place of residence, because of potential manipulation of fiscal address. However, the consumption of electricity or telephone in urban and rural areas are less subject to arbitrage. So, for instance, if the distribution of the unit wage \( (w) \) profile in urban areas first-order stochastically dominates that for rural areas, one can show that, under some conditions on preferences, the prices of electricity and telephone in rural areas should be subsidized relative to those in the urban areas.

Now, one may take the political science view that a number of instances of demand-side cross-subsidization result from interest group politics. While this view is quite relevant, we feel that taxation by regulation is an instrument not to be neglected even by benevolent regulators. The public finance literature simply shows that this is a delicate matter that warrants further attention.

Last, in the regulatory policy debate (for instance, in the telecommunications industry and the airline industry before deregulation), it is often suggested that entry in some regulated markets should be prohibited to prevent cream-skimming of profitable products from jeopardizing the supply (understand: "the subsidization") of essential services by regulated firms. While we agree that entry may in some circumstances thwart the taxation of goods mainly consumed by the rich, the logic of our model (in particular, the absence of an ad-hoc firm-level budget constraint) implies that there is no necessary consequence for the supply of essential goods (unless there are economies of scope between taxed and subsidized goods).

6. Conclusion.

This paper developed a normative model of the regulation of a multiproduced firm based on a full description of the informational constraints faced by the regulator. Let us summarize the main findings:
a) Optimal pricing requires that each product's Lerner index be equal to the sum of a Ramsey term and an incentive correction.

b) The Ramsey term is entirely determined by the industry demand data and the economy-wide shadow cost of public funds. In contrast to the Ramsey-Boiteux model, the Ramsey term is independent of the firm's cost structure.

c) The possibility of using pricing to extract the firm's informational rent may give rise to an incentive correction. The incentive correction, if any, is entirely cost-determined (is independent of the demand function) and can be analyzed by using aggregation techniques. But prices should not necessarily be used to promote incentives (in which case the incentive correction is equal to zero), because cost reimbursement (or profit sharing) rules may be a sufficient instrument to do so. Indeed, we view the dichotomy between the cost reimbursement rule as an incentive device and pricing as an allocation device as a good benchmark. A strong case must be built that the cost technology is conducive to incentive corrections.

d) In the absence of incentive correction, a regulator who lacks demand and cost information may implement optimal (Ramsey) pricing by delegating the pricing decision to the firm using the "price-tax" methods described in sections 4 and 5. This possibility of implementing optimal pricing despite limited information is in stark contrast with the implications of the Ramsey-Boiteux precepts. 25

e) The use of historical data reduces the impact of likely errors in estimating the quantities that form the basis for the price tax. A tax based on price changes and past quantities yields convergence toward the optimal

25 As Brown and Sibley [1985, p. 60] note: "Because of the difficulty in estimating [Ramsey prices] precisely, practitioners of [Fully Distributed Cost] pricing have often derided Ramsey pricing as impractical." Braustigan [1979, pp. 41,46] also discusses the informational requirements of Ramsey-Boiteux-type models.
prices.
f) The price-tax method yields optimal pricing when the regulated firm faces either no competition or an unregulated competitive fringe (see the subsections on intermodal competition and on access pricing). When competition exhibits distorted pricing due to market power or untaxed externalities, or else is regulated separately, the price tax must be adjusted upwards or downwards as described in Section 5.
g) Supply-side-cross subsidization (the non-cost-minimizing allocation of effort or investment among product lines) does not occur under an optimal regulation. The normative analysis of demand-side cross-subsidization (pricing for redistributive purposes) follows the precepts of public finance.

---

26 See Laffont and Tirole [1988] for a political theory of cross subsidization and more generally for a theory of regulation that relaxes the assumption of a benevolent regulator.
REFERENCES


Appendix 1: Proof of Proposition 1.

Assuming differentiability for convenience (differentiability almost everywhere can be proved using standard techniques) we first derive the first-order condition of incentive compatibility for the firm:

\[
\max_{\hat{\beta}} \left[ t(\hat{\beta}) + R(q(\hat{\beta})) - \psi(E(\beta,C(\hat{\beta}),q(\hat{\beta}))) \right]
\]

yields the F.O.C. (at \( \hat{\beta} = \beta \)):

\[
\frac{d}{d\beta}(t(\beta) + R(q(\beta))) - \psi' \left[ E_C \frac{dC}{d\beta} + \sum_k q_k \frac{dq_k}{d\beta} \right] = 0
\]

The sufficient second-order condition is (see Guesnerie-Laffont [1984]):

\[
\psi'' E \left[ E_C \frac{dC}{d\beta} + \sum_k q_k \frac{dq_k}{d\beta} \right] - \psi' \left[ E_C \frac{dC}{d\beta} + \sum_k q_k \beta \frac{dq_k}{d\beta} \right] = 0
\]

If \( E_\beta > 0 \), \( E_C < 0 \), \( E_{q_k} > 0 \) for all \( k \)

then \( \frac{dC}{d\beta} \geq 0 \), \( \frac{dq_k}{d\beta} \leq 0 \) for all \( k \) ensure that second-order conditions are satisfied. As usual, we do not impose these inequality constraints, but check that they are satisfied by our solutions.

Letting \( U(\beta) \) denote the rent of the firm, we can rewrite the first-order condition.

(1) \( \dot{U}(\beta) = -\psi'(e)E(\beta,C(e,q),q) < 0 \).

Because \( U(\beta) \) is decreasing, the individual rationality constraint of the firm is achieved by imposing

(2) \( U(\hat{\beta}) \geq 0 \).

The regulation maximizes (3.6) under (1) and (2) with respect to \( e(\hat{\beta}), q_k(\hat{\beta}) \), for all \( k \) treating \( U \) as the state variable. The Hamiltonian is:
\[ H - \left[ V(q) + \lambda R(q) - (1 + \lambda)(\psi(e) + C(\beta, e, q)) - \lambda U \right] f(\beta) \]
\[ - \mu(\beta) \psi'(e) E(\beta, C(\beta, e, q), q) \]

We have: \( \dot{\mu}(\beta) = -\frac{\partial H}{\partial U} = \lambda f(\beta) \). Since \( \mu(\beta) = 0, \mu(\beta) = \lambda F(\beta) \). Substituting \( \mu(\beta) \) into \( H \), the first-order conditions \( \frac{\partial H}{\partial e} = 0 \) and \( \frac{\partial H}{\partial q_k} = 0 \) for all \( k \) yield (3-9) and (3-10).

Q.E.D.

Appendix 2: Proof of Proposition 9.

Let \( \tilde{C} = C/H(q) \) and \( I = G^{-1} \) (with \( I' = 1/G' > 0 \) and \( I'' = -G''/G'^3 < 0 \)). Thus,

\[ \beta - e = I(\tilde{C}) \]

Because \( \tilde{C} \) is a "sufficient statistic" for \( e \) given \( \beta \), the transfer function can be made contingent on \( \tilde{C} \) only: \( t(\tilde{C}) \). Linearity with respect to \( C \) is equivalent to linearity with respect to \( \tilde{C} \). The firm chooses \( \tilde{C} \) so as to maximize:

\[ t(\tilde{C}) - \psi(\beta - I(\tilde{C})) \]

yielding the first-order condition:

\[ \frac{dt}{d\tilde{C}} = -\psi' I' \]

and the second-order condition (after using the first-order condition as an identity):

\[ I' \frac{d^2}{d\beta^2} (\tilde{C}) \geq 0 \]

or
\[
\frac{d}{d\beta} \left[ \frac{C(\beta)}{H(q(\beta))} \right] \geq 0.
\]

Condition (i) in Proposition 9 thus guarantees that the firm's second-order condition is satisfied.

The regulator can use a menu of linear cost reimbursement rules if and only if the transfer function \( t(C) \) is convex (the linear cost reimbursement rules are then given by the tangents to the convex curve). But

\[
\frac{d^2 t}{dC^2} \geq \psi'' \left[ \frac{d\beta}{dC} - 1' \right] I' - \psi' I''.
\]

Using the definition of \( C = G(\beta - e) \) and \( I' = 1/G' \), one has:

\[
\frac{d\beta}{dC} - I' = \frac{\dot{\psi}}{(1 - e)G'},
\]

where \( \dot{\psi} = \frac{d\psi}{d\beta} \).

Differentiating (3.9):

\[
\dot{\psi} = \frac{HG'' - \lambda \frac{d}{d\beta} \left[ \psi'' + (\Sigma H_{k} \dot{q}_{k}) G' \right]}{\psi'' + HG'' + \frac{\lambda}{1 + \lambda} \frac{d}{d\beta} \psi''}.
\]

Now, if \( \psi'' \geq 0 \) and \( \dot{q}_{k} < 0 \) for all \( k \),

\[
\frac{\dot{\psi}}{1 - \dot{e}} \leq \frac{HG'' - \lambda \frac{d}{d\beta} \left[ \psi'' + (\Sigma H_{k} \dot{q}_{k}) G' \right]}{\psi''},
\]

so that, using \( \frac{d(F/E)}{d\beta} \geq 0 \),

\[
\frac{d^2 t}{dC^2} \geq \frac{\psi' G''}{G' - \psi''} \geq \frac{HG'' - \lambda \frac{d}{d\beta} \left[ \psi'' \right]}{G' - \psi''}.
\]

Replacing \( \psi' \) using (3.9) yields, after some simplifications:
\[
\frac{d^2 t}{dc^2} \geq \frac{\lambda}{1+\lambda} \left[ \frac{\psi'}{G'F} \right] \left[ G \cdot \frac{d\{F\}}{d\beta} - G^r F \right].
\]

Hence, if the condition (i) in Proposition 7 holds, \( \frac{d^2 t}{dc^2} \geq 0 \) (recall that the hazard rate is monotonic). This implies that the regulator can use a menu of linear contracts. The slope of each incentive contract is given by the firm's first-order condition:

\[
\frac{dt}{dc} + \psi' (\beta - I(\tilde{C})) I'(\tilde{C}) = 0;
\]

it suffices to invert the equilibrium function \( \tilde{C} = \tilde{C}(\beta) \) into \( \beta = \beta(\tilde{C}) \) and eliminate \( \beta \) in the previous equation.

\[Q.E.D.\]