

A Non-Bayesian Incentive
Mechanism Using Two-Part
Tariffs

by Ingo Vogelsang

Do not quote without the permission of the author.
©1987 Columbia Institute for Tele-Information

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

A NON-BAYESIAN INCENTIVE
MECHANISM USING
TWO-PART TARIFFS

BY
INGO VOGELSANG

Draft
1987

Research Working Paper Series.
Not for citation, quotation, reproduction, or distribution
without written permission.

Center for Telecommunication and Information Studies,
Graduate School of Business,
Columbia University,
New York, NY 10027.

A Non-Bayesian Incentive Mechanism Using Two-Part Tariffs

by Ingo Vogelsang, Boston University and RAND Corporation*)

Working Draft, September 1987

Paper Prepared for Conference on "Optimal Incentive Mechanisms for Cost-Efficiency and Risk-Sharing in Regulated Industries" at the Center for Telecommunications and Information Studies at Columbia University, New York, November 13, 1987.

*)This paper was partially funded through a grant from the John and Mary R. Markle Foundation.

I. Introduction

The regulatory incentive mechanism to be discussed in this paper may be seen as a contribution to the issue of the optimality of marginal cost pricing. The case for and against marginal cost pricing by public utilities has a somewhat dialectic history. Hotelling (1938) set the stage for the thesis by arguing that in decreasing cost industries buyers should only pay the marginal costs of serving them. The resulting deficit should simply burden the tax payers. Coase (1945, 1946) soon vehemently opposed this suggestion. He argued first that marginal cost pricing does not pass the test that consumers' total willingness to pay exceeds production costs of the good in question, second that subsidies jeopardize efficient operation of the monopoly supplier, and third that tax financing of subsidies results in an unjustified redistribution from general taxpayers to the consumers of goods produced under increasing returns. However, Coase's antithesis did not initially win the profession. This took much longer and resulted in Ramsey prices as the synthesis. Ramsey prices maximize total surplus under a balanced budget constraint for the public utility. Such a balanced budget constraint fulfills several functions. It neutralizes income distributional issues between shareholders of the firm and its customers. The shareholders exactly receive a competitive return, neither more nor less. Without any more specific information it further allows us to state that consumers in total value the output of the public utility at least at production cost. Third, it puts a (sometimes generous) cap on any inefficiencies in the production of the output. Last, it avoids subsidies and the accompanying distortions.

In spite of the virtues of Ramsey prices marginal cost prices have had a recent comeback. The reasons belong into two categories. First, Ramsey prices are hard to implement. They discriminate between low elasticity and high elasticity customer groups. This makes them politically unpopular. And they require the regulator to have substantial information about cost and demand functions. This makes them hard to calculate. Second, implementation problems for marginal cost prices have been reduced by now through various incentive schemes. Still, there seems to remain the problem that marginal cost prices may necessitate subsidies. In particular, subsidies may now become necessary as an incentive device. In this paper I instead suggest that the subsidy be paid by consumers in the form of the fixed part of a two-part tariff. The properties of these new tariffs are a further indication that the distinct superiority of two-part tariffs over Ramsey prices could form a new pricing synthesis for most regulated industries and public enterprises.

The economic literature on regulatory incentive mechanisms has grown substantially over the past ten years. Before this time economists were mostly concerned with the theoretical derivation of welfare optimality conditions for prices and costs of regulated monopoly firms. Practical approaches to regulatory pricing, such as rate-of-return regulation, were mostly criticized for failing to obey these conditions. They were not seen as necessarily imperfect attempts to implement these conditions. If anything has been brought out very clearly by the new incentive literature it is that the originally derived optimality conditions fail to be optimal

in an imperfect world with asymmetric information and risk aversion.¹ This message has come across most clearly through the Bayesian approach to regulatory incentive mechanisms (pioneered by Baron and Myerson, 1982).

In Bayesian mechanisms the regulator acts as a principal while the firm (or its management) is an agent whom the principal wants to induce to behave in the principal's interest. The principal wants to maximize a welfare function. He has certain (unbiased) a priori expectations, for instance on the firm's cost function, and can observe certain variables, for instance the firm's total expenses. The agent wants to maximize his own utility which is assumed to be a function of income (profit) or income and effort (or risk-taking). Maximizing welfare is usually not in the interest of the agent because that requires effort or a sacrifice of profits. The principal therefore has to compensate the agent for maximizing welfare. In the absence of lumpsum taxes such a compensation will itself influence welfare and hence affect the desired welfare optimum. The Bayesian literature has worked out this point very clearly. It has also provided a number of very basic insights which are discussed in other papers of this conference. I want to argue, however, that non-Bayesian approaches continue to be useful. In my view, the Bayesian approach has three major drawbacks.

The first drawback is that the a priori information of the regulator is non-verifiable. This would cause no problem if the regulator were the true

¹This is not the same as the second-best issue but related to it in spirit.

principal known from the simple principal-agent framework. However, the principal-agent framework is hardly a correct description of the regulatory process. The regulator is usually a public official or, more likely, a bureau. In both cases, unless being a dictator, the regulator should be acting on behalf of the polity. This in itself could constitute a principal-agent relationship if it were not for conceptual difficulties in formulating preference formation of the polity as the principal. Instead we see the regulator as being guided by legal rules and voting mechanisms. Then the problem of responsibility and third-party control arises, for instance, through courts or through auditors or via elections. How can judges, auditors or the electorate control regulatory decisions without any direct observation of the a priori beliefs of the regulator? There is a clear moral hazard problem in differentiating between subjective probabilities and political opinions. If the regulator has different welfare weights than the electorate, then he can implement these by misstating his true a priori probabilities. In addition to the moral hazard problem the different public officials in the regulatory bureau among themselves face an aggregation problem for their a priori probabilities. They have to find jointly held subjective probabilities before determining optimal regulation.

The second drawback is that the regulator's a priori information may be very poor and incomplete. Chances are high then that the resulting mechanism will provide the wrong incentives.

The third drawback is that optimal Bayesian incentive mechanisms are extremely hard to derive for all but the very simplest functional forms. This means that the problems treated by these mechanisms so far are quite remote from practical implementation. It does not mean that the general insights provided by Bayesian mechanisms are not empirically relevant, though.

In this paper we discuss an alternative to the Bayesian approach which is based on the tradition of adjustment processes known in the economic literature at least since Walras. The current process is a blend of two adjustment processes previously suggested by J. Finsinger and myself. The first of these processes (Vogelsang and Finsinger, 1979, in the following: V-F) is a regulatory adjustment process for private firms leading to Ramsey prices. The second one (Finsinger and Vogelsang, 1982, in the following: F-V) is a performance index for public enterprise managers leading to marginal cost prices. These approaches are quite different from the Bayesian approach in several respects.

First, uncertainty is not explicitly introduced. Rather, the firm (or its management) is assumed to know cost and demand functions for its outputs while the regulator is assumed to know only very general properties of these functions such as the sign of derivatives. The regulator, however, can ex post observe bookkeeping data on prices, quantities and total costs (expenses). Clearly, uncertainty could be introduced, but it is not an essential part of the framework.

Second, the approach is essentially dynamic, and it is based on a lagged adjustment. With the exception of mechanisms that converge in one period this means that the mechanisms will only develop their full properties in a stationary environment. Also, the mechanisms will deviate from the full information optimum in all periods before convergence to a steady state. Strategic behavior may occur that reduces the speed of convergence. Therefore, the discounted value of welfare levels provided by these mechanisms will differ from the present value of full information optima. I have not been able so far to derive any general formula for this deviation. On the other hand the mechanisms usually improve welfare in every period. They might therefore better be regarded as piecemeal approaches using a gradient method.

Third, the incentives or regulatory constraints used are simple approximations to the welfare change effected by the firm respectively its management. They are therefore simple formulas that can be easily understood by regulators and managers.

The first and the last of these properties clearly suggest that the approach is more readily implementable than the Bayesian approach. The second property is really what seems to count against it. However, this property is also a consequence of practicality. Lags are necessary for the observation of the cost and quantity data. If lags could be avoided altogether, then one could make the lag period of the mechanisms arbitrarily small and thereby achieve convergence in an essentially stationary environment.

In terms of philosophy the main difference between Bayesian mechanisms and the regulatory schemes discussed in this paper seems to lie in their view of the regulator. The Bayesian approach views the regulator as benevolent and well informed, the British ideal of a civil servant. Our approach views the regulator as a potentially imperfect executor of rules and laws, someone who has to be subject to third-party control.

In the next section we introduce a regulatory constraint as the fixed part of a two-part tariff. In order to bring home the main points it is first assumed that the regulatory constraint acts as a lumpsum tax on consumers. This assumption is then relaxed in Section III, and it is shown that the scheme is likely to work under more general conditions. Section IV contains some possible extensions that have not been fully worked out yet. The paper ends with short conclusions in Section V.

II. Two-part Tariffs with a Fixed Number of Customers

We use a discrete dynamic model in a stationary environment. Assume a regulated monopoly firm producing a single output in quantity q_t in period t . The firm faces a cost function $C(q_t)$, but it is not necessarily producing on it. The difference between its actual cost, C_t , and $C(q_t)$ could be any kind of inefficiency, but for simplicity here it is assumed to be pure waste, W_t . In the initial period 0 there is no regulatory constraint, although the firm knows that regulation will be installed in

period 1. Thus, the price of the product in period 0 is p_0 and the firm's profit is $\pi_0 = p_0 q_0 - C_0$. Starting in period 1 the regulatory constraint is introduced as the fixed portion F_t of a two-part tariff. The

general form of the constraint for period t is

$$F_t = -[\pi_{t-1} - (p_{t-1} - p_t)q_{t-1}]/N \quad (1),$$

where N is the number of consumers buying from the firm.* Equation (1) says that the firm must disburse its profits of the previous period either through a fixed fee (refund), F_t , or a price decrease, $p_{t-1} - p_t$, denominated at last period's quantity, q_{t-1} . Any combination, which makes the sum of the two changes equal to the previous profit, is feasible. Thus the firm may actually increase p as long as F is sufficiently decreased, and vice versa it may increase F as long as p is sufficiently decreased. Should the firm make a loss in a period it can similarly ask the customers to reimburse it for this loss in the next period.

Noting the definition of π_t we may rewrite the constraint as

$$F_t = C_{t-1}/N - p_t q_{t-1}/N = \underline{c}_{t-1} - \underline{q}_{t-1} p_t \quad (2),$$

where \underline{c}_{t-1} and \underline{q}_{t-1} are average cost respectively average quantity per customer in period $t-1$. Equation (2) clearly shows the tradeoff between F_t and p_t . Here F_t can be interpreted as the difference between average cost and average variable revenue per customer.

*For simplicity in the formal arguments we assume this constraint always to be binding. We will argue below, however, that an inequality constraint (\leq) might do better.

In a sense, the constraint turns customers into shareholders of the firm, but as in a cooperative they receive their dividend on a per capita basis. Here the fixed fee can be seen as a membership contribution which entitles the member to a price discount. Similarly, a negative fixed fee would appear as a form of profit distribution to the members. So we could interpret the situation as one which turns the public utility into a cooperative. Note that cooperatives tend to distribute profits, both, as per capita dividends and as price discounts. Henceforth we therefore call F either the 'consumer dividend' or the 'membership fee' depending on whether it is negative or positive; and p will simply be the 'price'.³

In the current section we take N , the number and identity of members of the cooperative, to be fixed and, in particular, to be independent of the firm's pricing policy. Hence, F_t is a lumpsum tax respectively lumpsum subsidy for the consumers. Now, post-dividend profits of the firm in period t are

$$\begin{aligned}\hat{\pi}_t &= p_t q_t - C_t + F_t N \\ &= p_t (q_t - q_{t-1}) - (C_t - C_{t-1}) \\ &= \pi_t - \pi_{t-1} + q_{t-1} (p_{t-1} - p_t),\end{aligned}\tag{3}$$

where π_t is defined as in π_t above. In the following, $\hat{\pi}_t$ is referred to as total profit. All demand has to be served at prices p_t .

³Also note that in connection with cost overruns for nuclear power stations there are strong tendencies in the U.S. to engage customers in public utility financing and thus de facto to turn public utilities into cooperatives.

Assume that the firm maximizes the discounted stream of future profits

$$W_t^{\max} \hat{\pi} = \sum_{t=0}^{\infty} [p_t q_t - C(q_t) - W_t + F_t N] \beta^t \quad \text{s.t. (1) and } F_0 = 0$$

or

$$W_t^{\max} L^{\hat{\pi}} = \sum_{t=0}^{\infty} [p_t q_t - C(q_t) - W_t + F_t N - \mu_t (p_t q_{t-1} - C_{t-1} + F_t N)] \beta^t \quad (4).$$

Here $\beta = 1/(1+i)$ is the firm's discount factor, and μ_t is the Lagrangean multiplier associated with the constraint in period t .

The first order conditions for this problem are⁴

$$\frac{\delta L}{\delta p_t} \hat{\pi} = q_t + \frac{\delta q_t}{\delta p_t} (p_t - \frac{\delta C_t}{\delta q_t}) - \mu_t q_{t-1} - \beta \mu_{t+1} \frac{\delta q_t}{\delta p_t} (p_{t+1} - \frac{\delta C_t}{\delta q_t}) = 0 \quad (5)$$

$$\frac{\delta L}{\delta F_t} \hat{\pi} = N - \mu_t N = 0 \quad \text{or } \mu_t = 1 \quad (6)$$

$$\frac{\delta L}{\delta W_t} \hat{\pi} = -1 + \mu_{t+1} \beta = -1 + \beta \leq 0 \quad \text{and } W_t (1 - \beta) = 0 \quad (7)$$

$$\frac{\delta L}{\delta \mu_t} \hat{\pi} = p_t q_{t-1} - C_{t-1} + F_t N = 0. \quad (8)$$

Accounting for (6) the steady state solution to (5) is

$$\frac{\delta L}{\delta p} \hat{\pi} = \frac{\delta q}{\delta p} (p - \frac{\delta C}{\delta q}) (1 - \beta) = 0, \quad (9)$$

where dropping of the time subscript indicates the steady state.

Thus, provided $\beta < 1$, price will equal marginal cost in the steady state.

The question then is whether the sequence of prices converges. Here it helps to see that the current two-part pricing scheme is closely related to F-V (1982) and V-F (1979). F-V (1982) suggest a performance index for the

⁴Note that there is no period -1, and we have assumed $F_0 = 0$. Thus, the firm can freely maximize its profits in the first period. As it turns out, the first order conditions for period 0 are such that the firm will not waste, but will in general use some strategic pricing (marginal revenue \neq marginal cost).

managers of a public enterprise of the form $I_t = \pi_t - \pi_{t-1} + q_{t-1}(p_{t-1} - p_t)$. Thus, the problem set up in (4) is exactly the same as in F-V (1982). On the other hand in their Ramsey pricing problem V-F (1979) suggest that the regulator impose a regulatory constraint of the form $p_t q_{t-1} - C(q_{t-1}) \leq 0$ on the regulated firm for each successive period $t = 1, \dots, \infty$. The firm then is constrained on average to reduce its price(s) by the previous period's profit margin. To make V-F and F-V comparable I_t can easily be translated into an equivalent tax/subsidy TS_t for a private enterprise by setting $TS_t = -\pi_{t-1} + q_{t-1}(p_{t-1} - p_t)$. For the owner-manager of the private firm this yields the same incentive as I_t does for the manager of a public enterprise. Thus, two-part tariffs using the formula of the V-F constraint in the fixed part exactly mimic the subsidy-based F-V performance index.

Therefore the results from F-V (1982) carry through and also extend to the multiproduct case with differentiated fixed fees. In particular, it is shown there that the firm will never use pure waste and that prices will eventually converge to marginal cost prices. Besides $N = \text{constant}$ and the usual differentiability assumptions the further assumptions that have to be imposed on demand and cost functions are surprisingly weak: Demand has to be such that consumers' surplus is convex in prices, p . Furthermore, in each period the regulator has to be able to observe quantities, prices and total costs for the last period.⁵

⁵Following an argument made by Sappington and Sibley (1985) it can also be shown that (at given prices) the firm will invest in a socially optimal manner as long as i is also the social discount rate. In this case C_t would be the expenses of the firm in period t rather than the cost.

The reason why the mechanism will converge to marginal cost prices is that under the scheme the firm's total profit $\hat{\pi}_t$ is an approximation to the change in social surplus. In Figure 1 the shaded area gives the firm's

Insert Figure 1

profit, $\hat{\pi}_t$, while $F_t N$ is given by the rectangle EFGH. As can be seen from this and from the last formulation in equation (3) the change in producer surplus is exact, while consumers' surplus is quite crudely approximated by $Q_{t-1}(p_{t-1} - p_t)$.

By convexity of consumers' surplus, however, this approximation is always weakly smaller than the true change in consumers' surplus. The difference is the nonappropriated triangle GKL. Thus, the firm will in any period receive a total profit $\hat{\pi}_t$ that is no greater than the change in social surplus. If there is no improvement in surplus then profit will be zero or negative. In terms of figure 1, if prices are increased above the previous level, then there is an additional triangle outside the demand curve that is lost by the firm. The firm can in any period t assure itself at least of zero profits by setting $p_t = p_{t-1}$. Due to the differentiability assumption the firm can reap any increase in surplus through infinitely small price changes. Thus, if surplus can be increased through price changes the firm will do just that. The firm, however, will usually refrain from infinitely small steps, because profits incurred later have to be

discounted. Hence, convergence in prices will pursue at a reasonable pace. Because the producer surplus change in equation (3) is exact, the firm will not want to waste, and it will invest optimally. Hence, while optimal pricing takes time, cost minimization occurs instantaneously.⁶

III. Two-part Tariffs with a Variable Number of Customers

By making a strong assumption on consumer behavior ($N = \text{constant}$) we were able to show the equivalence between a subsidy scheme and a two-part tariff. This kind of equivalence has potential empirical importance.⁷ The case of constant N probably covers the vast majority of cases for two-part tariffs by public utilities, such as electricity, gas, water and even telephone services, in highly industrialized countries. However, in developing countries and for the poor in developed countries, such as the U.S., participation in these services may still depend on pricing. We therefore have to address the following question: What happens if for each consumer i demand depends on both, the price p and the membership fee (consumer dividend) F , in such a way that the number of consumers varies as F and p are changed? In terms of the cooperative interpretation membership now becomes a variable. Thus, $q_t^i = q_t^i(p_t, F_t)$ and $N_t = N_t(p_t, F_t)$.

⁶For a complete proof see the appendix to F-V (1982). The average cost curve in figure 1 is drawn as upward sloping in order to suggest the generality of the approach. Under natural monopoly conditions we would usually expect decreasing average cost.

⁷It has been noted before, for instance, by David Sibley in an oral statement to the author.

Before describing the consequences of this generalization we make the following additional assumptions:

1. We have to redefine the regulatory constraint (1). This now becomes

$$F_t = (C_{t-1} - p_t q_{t-1}) / N_{t-1} \quad (1^0).$$

2. $C(q)$ exhibits (weakly) decreasing average costs.
3. $\hat{\pi}_t > 0$.
4. Consumers' surplus $V(p_t, F_t)$ is convex in both p_t and F_t .

$V(p_t, F_t)$ is twice differentiable with $\delta V / \delta p_t = -q_t$ and $\delta V / \delta F_t = -N_t$.

5. The inverse demand functions $p(q, N)$ and $F(q, N)$ shall be continuous and nonnegative for all $q \in \mathbb{R}_+$, $N \in \mathbb{R}_+$.

$\lim_{r \rightarrow \infty} (p^0 q^0 + F^0 N^0) \rightarrow \infty$, where $q^0 = r q$, $N^0 = r N$, $r \in \mathbb{R}_+$. p^0 and F^0 are the price respectively fixed fee that fulfill $q(p^0, F^0) = q^0$ and $N(p^0, F^0) = N^0$.

These assumptions require some explanation.

In assumption 1 the reason for choosing N_{t-1} instead of N_t in the denominator on the right hand side of (1⁰) is that otherwise the constraint would be on the revenues generated from the fixed fee rather than on the fixed fee itself. This would allow the firm to choose a higher fixed fee and a lower number of customers than otherwise.

Assumptions 2, 3 and 5 are needed to assure feasibility of the process in the sense that the firm can always survive without making losses.

The nonnegativity of the fixed fee in assumption 5 is not necessarily compatible with the regulatory constraint (1⁰). We do not want to restrict the constraint, though. Instead, we show below, that the firm can always make profits and thus avoid a negative fixed fee.

Assumption 4 is necessary to assure convergence and optimality. It imposes some restrictions. First, the number of customers (respectively members), N_t , always has to be so large that it can be treated as a continuous variable. Also, the exit or entry of a customer has such a small effect that it does not hurt the differentiability of $V(p_t, F_t)$. Second, the assumed convexity comes out as a standard result of consumption theory if we postulate the strong axiom of revealed preference to hold and income effects to be absent. In that case V would be linear (and hence weakly convex) with respect to F , and it would be convex with respect to p (V-F, 1979, Appendix). But note that even with declining marginal utility of income convexity of $V(p_t, F_t)$ could still hold in the aggregate. Each consumer j would have a reservation price p^j for given F and a reservation membership fee F^j for given p such that the demanded quantity at any higher price or fee is zero. Hence, individual welfare is not affected by price changes occurring in the range above these prices. Now, if the distribution of consumer demands is sufficiently spread out, then any increase in p or F will result in the exit of customers which is essentially a convexifying phenomenon.

The interpretation of (1⁰) is similar to that of (1). In the cooperative interpretation the difference is that the consumer dividend

(membership fee) now is paid to (by) the current members but is based on membership in the previous period. This is not out of line with cooperative practice. Total profit of the firm now becomes

$$\hat{R}_t = p_t q_t - C_t + F_t N_t = p_t q_t - C_t + (C_{t-1} - p^* q_{t-1}) / N_{t-1} \quad (11).$$

Whenever we have $N_t \neq N_{t-1}$, will (11) differ from the F-V performance index. Hence, one wonders in what way the optimization problem is changed.

The firm now wants to maximize

$$\max_{W_t, p_t} \hat{R}^{\infty} = \sum_{t=0}^{\infty} [p_t q_t - C(q_t) - W_t + F_t N_t] \beta^t \text{ s.t. } (1^0) \text{ and } F_0 = 0 \quad (12).$$

Although, compared to the case where N is constant, the only change is in pricing and in the shape of the demand function, the possibility for waste now arises. The first order condition for (12) with respect to waste is analogous to (7). However, in general $\mu_t \neq \mu_{t+1} \neq 1$. It can be seen that the absence of waste is guaranteed as long as $\mu_{t+1} < 1+i$. The intuition behind this is that current profit reductions due to waste only pay if they are at least offset by an equivalent discounted relaxation in the constraint next period. In the following it is assumed that waste does not pose a problem.* Then (4) becomes

$$\max_{p_t} L^{\hat{R}} = \sum_{t=0}^{\infty} [p_t q_t - C(q_t) + F_t N_t - \mu_t (p_t q_{t-1} - C(q_{t-1}) + F_t N_{t-1})] \beta^t \quad (13),$$

where $q_t = q_t(p_t, F_t)$ and $N_t = N_t(p_t, F_t)$.

*The condition $\mu_{t+1} < 1+i$ can, by use of the first-order conditions of the problem, be translated into a condition containing only prices, quantities and first derivatives of demand. However, I have so far been unable to sign the expressions and thereby derive a meaningful condition on the absence of waste. Waste may, therefore, occur, but will never do so in the steady state.

The first order conditions for this problem are:

$$\frac{\delta L}{\delta p_t} \hat{\pi} = q_t + \frac{\delta q_t}{\delta p_t} (p_t - \frac{\delta C_t}{\delta q_t}) + F_t \frac{\delta N_t}{\delta p_t} - \mu_t q_{t-1} - \beta \mu_{t+1} [\frac{\delta q_t}{\delta p_t} (p_{t+1} - \frac{\delta C_t}{\delta q_t}) + F_{t+1} \frac{\delta N_t}{\delta p_t}] \quad (15)$$

$$= 0,$$

$$\frac{\delta L}{\delta F_t} \hat{\pi} = N_t + \frac{\delta q_t}{\delta F_t} (p_t - \frac{\delta C_t}{\delta q_t}) + F_t \frac{\delta N_t}{\delta F_t} - \mu_t N_{t-1} - \beta \mu_{t+1} [\frac{\delta q_t}{\delta F_t} (p_{t+1} - \frac{\delta C_t}{\delta q_t}) + F_{t+1} \frac{\delta N_t}{\delta F_t}] \quad (16)$$

$$= 0$$

and

$$-\frac{L}{\mu_t} \hat{\pi} = p_t q_{t-1} - C(q_{t-1}) + F_t N_{t-1} = 0 \quad (17).$$

In the steady state (15) and (16) become

$$\frac{\delta L}{\delta p} \hat{\pi} = \{ \frac{\delta q}{\delta p} (p - \frac{\delta C}{\delta q}) + F \frac{\delta N}{\delta p} \} (1 - \mu\beta) + q(1-\mu) = 0 \quad (18),$$

$$\frac{\delta L}{\delta F} \hat{\pi} = \{ \frac{\delta q}{\delta F} (p - \frac{\delta C}{\delta q}) + F \frac{\delta N}{\delta F} \} (1 - \mu\beta) + N(1-\mu) = 0 \quad (19),$$

$$\frac{\delta L}{\delta \mu} \hat{\pi} = pq - C + FN = 0 \quad (20).$$

Let us compare these conditions to the first order conditions of the corresponding welfare maximization problem. Since the process does not depend on subsidies and since we know by construction of F_t that profits vanish in the steady state, the relevant welfare maximum is the constrained two-part tariff optimum defined by

$$S^* = \max \{ S(p, F) : \hat{\pi}(p, F) + V(p, F) \text{ s.t. } \hat{\pi} = 0 \} \quad (21),$$

where $V(p, F)$ is aggregate consumer surplus.

The first order conditions to this problem are

$$\frac{\delta L^S}{\delta p} = \left\{ \frac{\delta q}{\delta p} \left(p - \frac{\delta C}{\delta q} \right) + F \frac{\delta N}{\delta p} \right\} (1 - \tau) - \tau q = 0 \quad (22),$$

$$\frac{\delta L^S}{\delta F} = \left\{ \frac{\delta q}{\delta F} \left(p - \frac{\delta C}{\delta q} \right) + F \frac{\delta N}{\delta F} \right\} (1 - \tau) - \tau N = 0 \quad (23),$$

$$\frac{\delta L^S}{\delta \tau} = pq - C + FN = 0 \quad (24).$$

By dividing equation (22) through equation (23) and (18) through (19) we get for both cases

$$\frac{\frac{\delta q}{\delta p} \left(p - \frac{\delta C}{\delta q} \right) + F \frac{\delta N}{\delta p}}{\frac{\delta q}{\delta F} \left(p - \frac{\delta C}{\delta q} \right) + F \frac{\delta N}{\delta F}} = \frac{q}{N} \quad (25).$$

Since the constraint on profit is the same in both problems in the steady state, the first order conditions must be the same in both problems. That means, we must have $\mu = [1 - \tau(1 - \beta)]^{-1}$.

We have chosen a framework here that differs somewhat from that used in the standard literature on this subject (e.g. Ng and Weisser, 1974, Schmalensee, 1981). In the literature, consumers are usually ordered by some criterion y such as income. There is a density function of consumers $f(y)$ defined on this criterion. The marginal consumer is defined by the condition that her surplus from consuming the good at price p and fixed fee F just vanishes: $V^0(p, F, y^0) = 0$. Our approach is fully compatible with this characterization. Just note that $\delta N / \delta y = -f(y)$. Also, by assuming that the set of marginal consumers is of measure 0 we get $\delta V / \delta p = -q$ and $\delta V / \delta F = -N$ as in assumption 4 above.

Note that marginal cost pricing is not the usual outcome of the welfare maximizing problem (21). This would be true even for an inequality constraint that is not binding ($\tau = 0$). We are here in an essentially second-best world where nondistortionary head taxes are not possible. Hence, marginal cost pricing could be suboptimal. Also note that in a situation of decreasing returns to scale the optimum may involve a negative F and/or p . The Ramsey pricing equivalent to the first-order condition (22) is

$$\frac{\delta L^S}{\delta p} = \frac{\delta q}{\delta p} \left(p - \frac{\delta C}{\delta q} \right) (1 - \theta) - \theta q = 0 \quad (26).$$

It is not clear in what way the Lagrange multipliers μ and θ differ. Assuming that they are the same we can see by comparing (22) and (26) that under two-part tariffs the (variable) price is going to be closer to marginal cost than under Ramsey prices. This is because $\delta N/\delta p$ is going to be negative, while F is going to be positive or negative depending on whether average cost is decreasing or increasing at the optimum. More intuitively, for the relevant case of decreasing average cost the fixed fee F allows one to raise revenues. These can be used to bring price p closer to marginal cost and still fulfill the balanced-budget constraint.

PROPOSITION: Under assumptions 1 through 4 the process described by problem (13) will converge to the constrained welfare optimum described by equations (21) through (24).

The proof to this proposition contains four steps.⁹ In step 1 it is shown that profitability of the firm under the process is assured in all periods $1, \dots, \infty$ given that $\hat{\pi}_1 > 0$. The reasoning is by induction: Given that $\hat{\pi}_{t-1} > 0$ the constraint in period t is more stringent than in period $t-1$. The firm therefore has to offer the output at a price combination that allows the consumers to continue to buy the quantity they bought in the previous period and pay less. By convexity of consumer surplus they will buy more. This larger output can be sold at a profit because average cost is falling. Therefore $\hat{\pi}_{t,1} \geq 0$.

Hence the firm can always find a sequence of prices which lead to nonnegative profit in every single period. Should the firm decide to make losses in a period it will do so for strategic reasons, making up for these losses later. Hence, the regulator in this model should not take losses as a malfunctioning of the process and therefore should let the process continue after loss-making periods.

In step 2 it is shown that social surplus increases monotonically under the process. The reason is that for each period the change in social surplus is always larger than or equal to the profit of the firm in this period.

The change in social surplus between two periods t and $t+1$ is

⁹For the formal proof see Vogelsang (1987).

$$\begin{aligned}
 \Delta S &= \Delta \pi + \Delta V \\
 &= \hat{\pi}_{t+1} - \hat{\pi}_t + V(p_{t+1}, F_{t+1}) - V(p_t, F_t) \\
 &\geq \hat{\pi}_{t+1} - \hat{\pi}_t + (p_{t+1} - p_t) \delta V / \delta p + (F_{t+1} - F_t) \delta V / \delta F \\
 &= p_{t+1} q_{t+1} - C(q_{t+1}) + N_{t+1} F_{t+1} - p_t q_t + C(q_t) - N_t F_t - q_t (p_{t+1} - p_t) \\
 &\quad - N_t (F_{t+1} - F_t) \\
 &= \hat{\pi}_{t+1} \tag{26}.
 \end{aligned}$$

The inequality in (26) holds by convexity of $V(p, F)$ or, alternatively, by an aggregate revealed-preference argument. To see this we only consider the N_t customers purchasing in period t , since others, by definition, can only increase their purchases. Now, if the old customers in period $t+1$ want to purchase their previous quantities they will have to pay a total of $p_{t+1} q_t + F_{t+1} N_t = p_{t+1} q_t + ([C(q_t) - p_{t+1} q_t] / N_t) N_t = C(q_t)$. But this is less than they paid in period t , since then the firm was making a profit. This revealed preference argument follows directly from assumption 4: If $p_{t+1} q_t + F_{t+1} N_t \leq p_t q_t + F_t N_t$ and $q_t = q(p_t, F_t)$, then $V(p_{t+1}, F_{t+1}) \geq V(p_t, F_t)$.

Also, for the last equality in (26), note that $C(q_t) - p_{t+1} q_t - F_{t+1} N_t = 0$. The last equality in (26) then defines total profit π_t as the performance index I_t applied to two-part tariffs rather than linear prices.

We have thus established that every profitable period for the firm results in a welfare increase (weakly) greater than the firm's current

profits. Should the firm make a loss during a period, then the reduction in welfare is less than the loss to the firm. Hence, any sequence of profits and losses to the firm between two periods m and n will lead to a net welfare gain as long as the sum of profits minus losses from m to n is positive. We know that the firm can always make a profit in the period following a profitable period. Therefore it must be true that loss periods only occur in order to have more profitable periods later. The firm will not want to move in cycles because summed up that means a net loss. This tendency is enhanced by the fact that future profits are discounted. Hence, if a loss period follows a profitable period, then this implies the existence of a finite sequence with a positive sum. Such a sequence can always be ended at a profitable period. Combining such finite sequences into megaperiods of unequal length we get a new sequence of welfare levels that is monotonically increasing.

Step 3 establishes that this sequence is bounded by the constrained welfare maximum and therefore converges.

In step 4 we show that welfare converges to this welfare maximum. We have already seen in equation (25) that the necessary condition for welfare maximization holds in the steady state. Here, we want to make this fact more plausible with an economic argument. Assume the process converged to $\underline{S} < S^*$. Then welfare could still be increased at the point of convergence. Now, by definition of F in the steady state $\hat{\pi} = 0$. On the other hand the potential for a welfare increase indicated by $\underline{S} < S^*$ can be translated into profits by the firm, as long as price changes are sufficiently small.

An illustrative way to see this is through Figure 2. Since the zero-profit constraint is assumed to be binding at the optimum we can forget about profit in the maximand. Thus, the welfare problem becomes $\max V(p, F)$

Insert Figure 2

s.t. $\hat{\pi} = 0$. Then in the (p, F) -space pictured in figure 2 the welfare maximum is characterized by the tangency between the zero-profit contour $\hat{\pi} = 0$ and the maximal iso-welfare line V^* . Note that welfare increases toward the origin. Also note that the gradient to any iso-welfare line is given by $\nabla V = -(p, F)$. Further, the constraint can be rewritten as $0 = C(q_{t-1}) - p_t q_{t-1} - F_t N_{t-1}$. Hence, the constraint is parallel to the tangent through (p_{t-1}, F_{t-1}) . The constraint simply moves the pricing decision of the firm in the direction of the steepest welfare increase. Now, assume that the process has converged and the point of convergence is not the constrained welfare optimum. At the point of convergence we have $\hat{\pi} = 0$ by the definition of the constraint. Also, the constraint has to go through the stationary point (p^*, F^*) . At the same time it is tangent to an iso-welfare line. Hence the first-order condition has to hold.

Theoretically it is possible that the second order conditions for a maximum do not hold at this point, that is, the iso-welfare line could have a stronger curvature than the zero-profit contour. This would be a case where the firm could lower both p and F and still make a profit. The regulator should be aware of this possibility and agree to such a price

reduction if requested by the firm. Then the inequality constraint (that is, $V-F$), as noted above, might do better than the equality constraint.

IV. Extensions

The validity of assumption 1 (decreasing average cost) is hard to prove for a regulator. This is amply demonstrated by the inconclusive results in the literature on economies of scale in public utility industries like electricity or telephone services. What happens if $C(q)$ is no longer restricted to decreasing average costs? The process would still converge to optimal prices if it can be shown that the sum of the firm's profits from any finite time period to infinity is positive. Then the rest of the above proof would go through, because only the first step makes use of the assumption of decreasing average cost. For the case of constant N losses always can be avoided. However, with variable N and increasing costs positive profits cannot be guaranteed for every period t , in particular not if $\hat{r}_{t-1} > 0$. But any loss by the firm via the constraint (1⁰) leads to the permission of a price increase in the following period. This may or may not be enough to compensate for the loss. Discounting poses an additional problem if losses occur earlier. In this case the feasibility of the process hinges on the size of the discount rate. The firm could avoid all this by charging welfare optimal prices in period 0 and stay there. It is not clear, though, that this strategy maximizes the discounted stream of all profits.

In addition, with increasing average costs F will eventually become negative turning from a fee into a dividend. This could raise moral hazard problems for the consumers leading to a discontinuity in demand $N(F, p)$. For instance, it would become beneficial for households to split up their accounts; or new customers could subscribe to the services of the regulated firm without consuming anything. This problem would, however, vanish in an expanding system in which consumers finance investment through their membership fee.

Moral hazard caused by the availability of dividends could also be avoided through discriminatory two-part tariffs. These would lead to an extension of the above mechanism that would also mitigate any existing income distributional concerns over high fixed fees. Such tariffs could be linked to household incomes. Instead, we here would like to link them to last period's consumption, assuming that income is harder to observe. Then the aggregate constraint that the firm faces on its total revenues from the membership fee would formally remain as before. The membership fee (dividend) for each customer (or customer group) j , however, would become $F_t^j = -[\pi_{t-1} - (p_{t-1} - p_t)q_{t-1}^j]/q_{t-1}^j$. There could also be a differentiation in the variable price p . This could make the approach one of optional tariffs where the customers self select into groups. The firm could be allowed to attract new customers with any combination (p, F) it wants to, but subsequently the above formula holds. This formulation makes it necessary to adapt our assumptions on consumers' surplus and demand. In particular, consumers could now act strategically since the current fixed

fee would depend on last period's individual quantity of consumption. However, if individual consumer demands do not cross, then the firm can always find an optional two-part tariff that is incentive compatible. A simpler type of discriminatory tariffs feasible under our approach would be for the regulator to define lifeline rates or low-price blocks for certain large customer groups. Such prices would then be fixed in advance and be outside the optimization of the regulated firm.

Related to discriminatory two-part tariffs is to generalize our framework to a multiproduct monopolist. This has already been achieved for the case of a constant number of customers N . Also, as long as the multiproduct analogue to assumptions 1 through 4 holds, the process with variable N_t should converge in a similar way as in the singleproduct case. In particular, the condition on decreasing average cost becomes one on decreasing ray average cost. The framework also allows the firm to introduce new products. Since a product introduced in period t would have no sales and no purchasers in period $t-1$, its price and fixed fee would be unregulated in period t . This is analogous to period 0 for the already existing products. Quantity sold and number of buyers in period t would then form the basis for the constraint in period $t+1$.

A disturbing feature of our process is the possibility of non cost-minimizing behavior by the firm at the beginning of the process. This has been noticed as a major drawback of other regulatory schemes as well and has recently led to suggestions for predetermined price caps. The main

incentive property of such price caps is that they do not depend on cost factors that the firm can influence. Thus, C_{t-1} in formula (10) would have to be replaced by a target cost figure that comes out of negotiations between the firm and the regulator before the process starts. Initial calculations show that a process defined on the resulting price caps could also converge to welfare optimal prices.

A last extension is to ask what happens if the firm management does not know its own cost and demand functions. This case is of obvious importance. Can the firm still find profitable prices satisfying the regulatory constraint? Starting with $\hat{\pi}_t > 0$ we know that setting $p_{t+1} = p_t$ will lead to $\hat{\pi}_{t+1} = 0$ provided that there are no income effects and the number of customers, N_t , stays constant. In case of decreasing marginal cost the firm could also set $F_{t+1} = F_t$ and reduce p_{t+1} to a level that obeys the constraint. Then $\hat{\pi}_{t+1} > 0$. In both cases the process would generally stop short of welfare maximizing prices.

V. Conclusions

In this paper we have discussed a regulatory two-part pricing mechanism with a number of desirable properties. It is anonymous in the sense that it does not require regulators to have detailed prior information on the regulated firm and its environment. Also, subsequent observations of the regulator can be restricted to verifiable bookkeeping data. At the same time the pricing formula is simple and easily interpreted. The firm is free to choose the variable price of the two-part tariff as long as the fixed

part obeys the formula (1) respectively (1⁰). The firm will then in every period receive a profit that approximates, but is smaller than the welfare change caused by its price changes over the last period. The approach taken in this paper is more realistic than most other suggestions for incentive pricing because it requires no government subsidies for the regulated firm.

REFERENCES

- Baron, D., and R. Myerson, 1982, "Regulating a Monopolist with Unknown Cost", Econometrica 50, pp. 911-930.
- Baumol, W.J., and D.P. Bradford, 1970, "Optimal Departures from Marginal Cost Pricing", American Economic Review 60, pp. 265-283.
- Coase, R.H., 1945, "Price and Output Policy of State Enterprise: A Comment", Economic Journal 55, pp. 112-113.
- , 1946, "The Marginal Cost Controversy", Economica 13, pp. 169-182.
- Finsinger, J., 1979, Wohlfahrtsoptimale Preisstrukturen von Unternehmen unter Staatlicher Regulierung, Doctoral Dissertation, University of Bonn.
- , and I. Vogelsang, 1982, "Performance Indices for Public Enterprises", in L.P. Jones (ed.), Public Enterprise in Less-developed Countries, Cambridge University Press, pp. 281-296.
- Hotelling, H., 1938, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", Econometrica 6, pp. 242-269.
- Loeb, M., and W.A. Magat, 1979, "A Decentralized Method for Utility Regulation", Journal of Law and Economics 22, pp. 399-404.
- Ng, Y., and M. Weisser, "Optimal Pricing with a Budget Constraint - The Case of the Two-part Tariff", Review of Economic Studies 41, 1974, pp. 337-345.
- Sappington, D., 1980, "Strategic Firm Behavior Under a Dynamic Regulatory Adjustment Process", Bell Journal of Economics 11, pp. 360-372.
- Sappington, D., and D. Sibley, 1985, "Regulatory Incentive Schemes Using Historic Cost Data", Working Paper at Bell Communications Research, August.
- Schmalensee, R., "Monopolistic Two-part Pricing Arrangements", Bell Journal of Economics 12, 1981, pp. 445-466.
- Vogelsang, I., "Two-part Tariffs as Regulatory Constraints", Boston University mimeo, February 1987.
- Vogelsang, I., and J. Finsinger, 1979, "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms", Bell Journal of Economics 10, pp. 157-171.

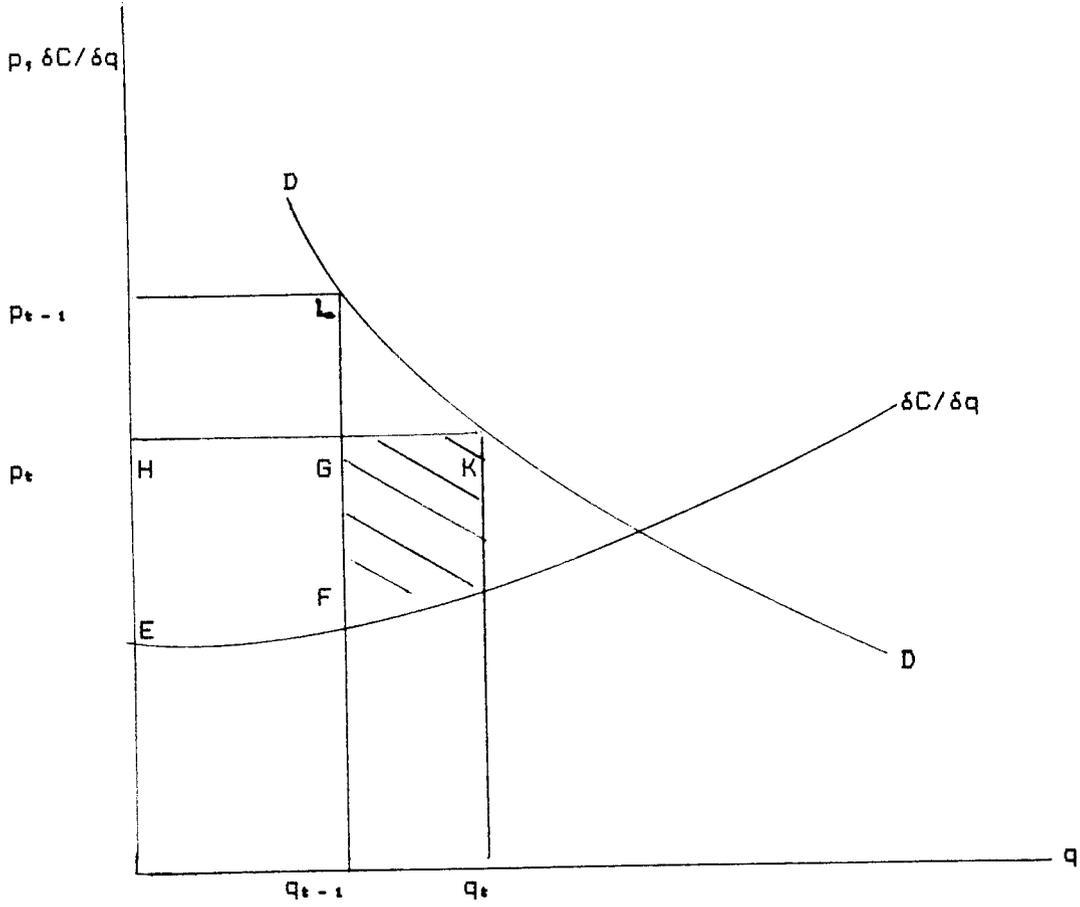


Figure 1: Profits and the Fixed Fee with Constant N

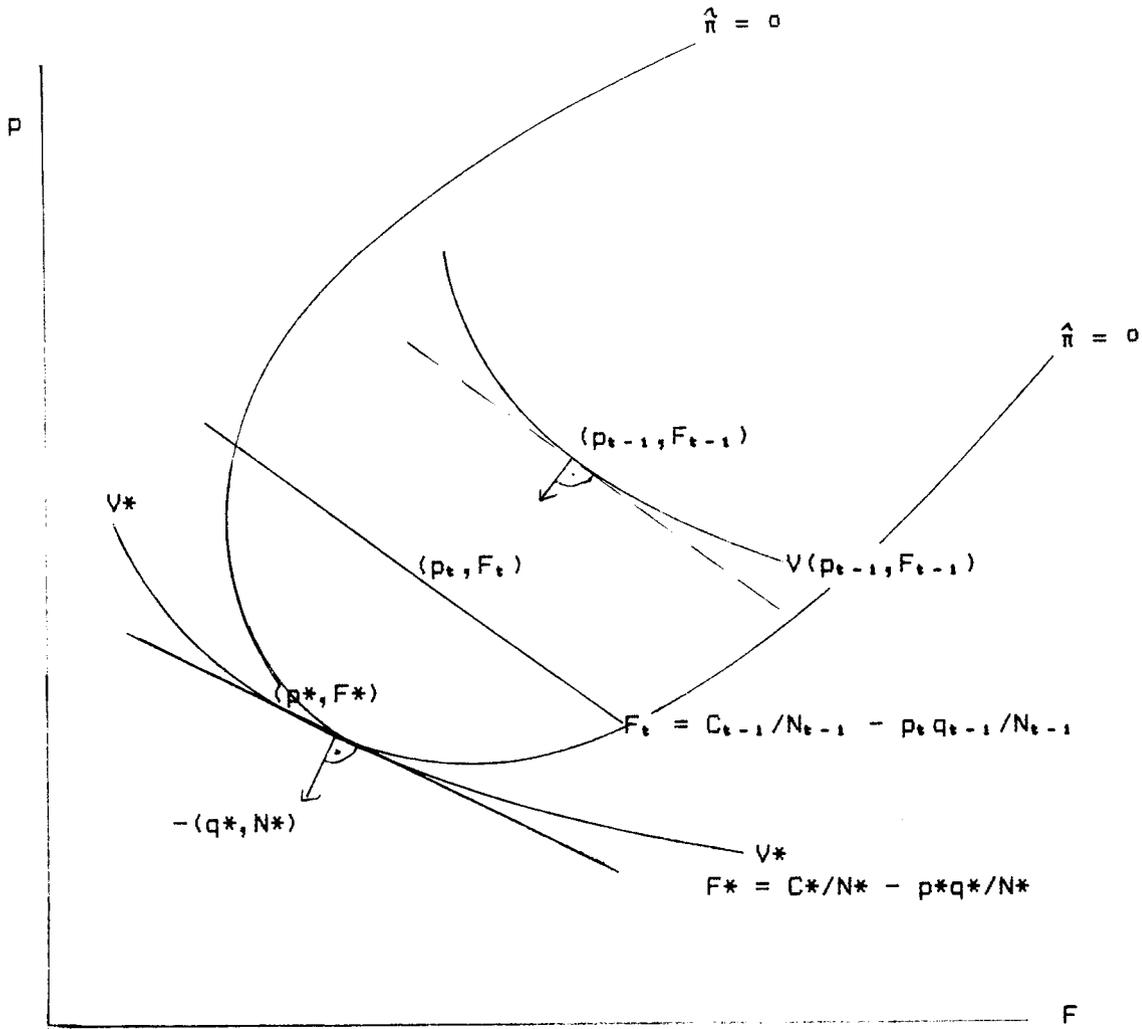


Figure 2: The Constraint, Profits and Consumer Welfare