# Capacity Pricing

Shmuel Oren, Stephen Smith, Robert Wilson

*Econometrica*, Volume 53, Issue 3 (May, 1985), 545-566.

---

*Econometrica*, Vol. 53, No. 3 (May, 1985)

# CAPACITY PRICING[1]

BY SHMUEL OREN, STEPHEN SMITH, AND ROBERT WILSON

We study the problem of optimal pricing for a bundle of services characterized by two attributes (e.g., quantity and quality) and subject to capacity limitations or peakloading. An application is to services that take the form of a load-duration curve. Using separability assumptions on the demand and cost functions, we derive the optimal pricing policy for a monopolist seller. An example is solved completely.

## 1. INTRODUCTION

A CENTRAL ISSUE in pricing theory is how to set prices for time-varying services that require the seller to incur a setup cost, such as the cost of capital equipment necessary for production of the service. Often a consumer of such services can choose, besides the total quantity, such quality attributes as the timing of delivery. The setup and variable costs borne by the seller typically depend on such attributes. The consumer's demand for service is described by his demand rate as a function of time. This rate is bounded above by the capacity of the equipment chosen by the consumer. The seller's costs include both setup and variable costs for meeting the demand of the particular customer. They usually depend on the customer's entire load pattern. The time profile of delivery determines the duration and extent that peak capacity is idle.

Services provided by leased office equipment are examples of products that fit the above description. A photocopy machine, for instance, produces a stream of copies over time at a varying rate that is limited by the machine's capacity (i.e., its maximum rate of operation). The consumer chooses a time pattern of usage, including the peak demand rate that determines the capacity required. In turn, the capacity requirement and the demand pattern determine the most economical system configuration of types and numbers of machines installed by the seller. Besides fixed costs and the amortized cost of capacity, operating costs may include supplies, labor, and maintenance.

These features appear in the markets for a wide variety of products, such as computer systems, WATS lines, electronic mail systems, and various types of industrial equipment. One component of the capacity charges is typically based on the equipment's maximum throughput rate. A second component, described as maintenance and operating charges, is typically based on the duration the equipment is operated. In specifying his choice of equipment and its operating schedule a buyer considers his preferences for time profiles of usage in relation to the price schedule. In selecting his optimal price schedule, the seller considers the costs of capacity and usage, and the distribution of preferences among

[1] Research support was provided by the National Science Foundation, Grants IST 81-08-350 and SES 81-08-226 & 83-08-723, and by the Office of Naval Research, Contract ONR-N00014-79-C-0685. Additional support in the preparation of a revision was provided by the Electric Power Research Institute. The opinions expressed are those of the authors and are not necessarily endorsed by any of the above. This paper is a substantial revision of T.R. 386, IMSSS, Stanford University, August, 1982.

potential customers. The optimal price schedule takes account simultaneously of the effects of capacity and usage charges on buyers' purchases.

The objective of this paper is to develop a methodology for analyzing pricing decisions in markets with the technological features described above. We derive a two-dimensional nonlinear pricing rule for both usage and capacity that discriminates among a heterogeneous population of customers. For each customer type, we obtain the complete two-dimensional demand pattern that the seller prefers to induce. (This demand pattern, or load curve, is the generalized analog of the optimal discrete consumption levels derived in previous work.) The optimal price schedule is then constructed to achieve these demand patterns as the consumers' responses.

Our analysis focuses on a monopoly and we derive the pricing policy that maximizes the seller's profit. In Appendix A we show how to derive the policy that maximizes total surplus subject to a constraint on the seller's profit, as in a regulatory context. Appendix B describes briefly the simple form that the analysis takes in the absence of capacity costs.

We adopt the assumptions common in the literature on nonlinear pricing. Chief among these are that there are no income effects and there is no resale market. Of course it is necessary that the seller can monitor consumers' demand patterns. Further, we assume that the seller knows the distribution of customers' types and the structure of their preferences. Customers' characteristics are summarized in a one-dimensional index satisfying some monotonicity and self-selection properties. (This is a stringent assumption since it ultimately implies that the customers respond to the optimal payment schedule with usage patterns that are well-ordered by inclusion.) Although the seller cannot directly discriminate among the different types, the seller discriminates indirectly by designing the pricing schedule to exploit the differences among consumers' preferences. In addition we impose some separability assumptions on the cost and demand functions that are unique to the two-dimensional characterization of product bundles employed in our formulation. These enable us to obtain a complete characterization of the optimal pricing policy, including a price schedule for incremental units of capacity and a price schedule for the services provided.

Our analysis generalizes the results in the literature on nonlinear pricing (e.g., Goldman, Leland, and Sibley [2], Roberts [6], Spence [8]). In some respects the problem we address is similar to the nonlinear pricing of multiple products studied by Mirman and Sibley [3]. However, their analysis does not consider the role of capacity costs and the corresponding capacity charges for the products' attributes. Panzar and Sibley [5] study a model in which supply costs and prices for both capacity and usage are linear, and the prices are chosen to maximize total surplus. Our results represent an extension to the case that capacity costs and prices can be nonlinear; for example, multiple production technologies are allowed and prices can vary according to the conditions of delivery.

In Section 2 we set up the formulation of the problem and specify the assumptions invoked in the derivation of the optimal payment schedule in Section 3. In Section 4 we solve completely a class of examples that illustrate neatly how a

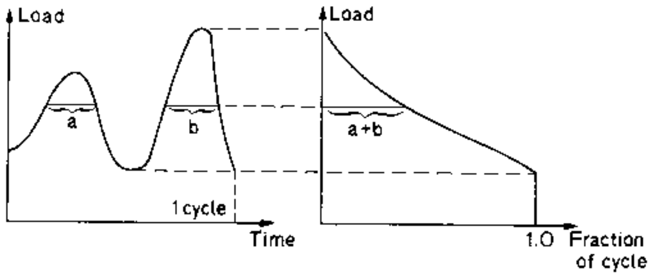


FIGURE 1—Representation of a time-varying load by a load-duration curve.

portion of the capacity costs are borne by customers. Additional applications of the results are discussed in Section 5. These include pricing of quality differentiated products, and the pricing of electric power in markets with special features that enable our formulation to be applicable.

## 2. FORMULATION

In this section we motivate and describe the various parts of the formulation and we specify the assumptions imposed in the subsequent analysis.

THE PRODUCTS: A useful characterization of time-varying demand is the load duration curve, illustrated in Figure 1. The load duration curve describes the fraction of time that a customer's demand exceeds any given rate of usage. The area under the curve is the total usage per unit time. A load duration curve provides a compact representation of a complex load pattern by suppressing details about the timing of usage that are usually unnecessary for production planning, cost calculation, and pricing.

Formally, a customer's load-duration curve (or rather the set it circumscribes) is represented as a compact and comprehensive Borel subset of the nonnegative orthant of two-dimensional Euclidean space.[2] The two axes can be interpreted as the consumption rate and the duration of time, but alternative interpretations are possible. In our analysis the two axes are treated symmetrically and their magnitudes are referred to as the capacities. The property of comprehensiveness reflects the fact that the purchase of the $q$th increment of capacity on either dimension entails purchase of all lesser increments. In the context of load duration curves these capacities are the customer's peak load and the total fraction of time his system operates. As we shall see later, capacity charges induce buyers to truncate their purchase sets in order to reduce their charges for capacity. Thus a purchase set has the shape depicted in Figure 2.

The shape of a purchase set is not otherwise restricted, but as will be seen, rectangular purchase sets play a special role in our analysis. The rectangular set with the corner (i.e., least-upper-bound) $x \in \mathfrak{R}^2_+$ is denoted by $[x]$. For an arbitrary

[2] A set $Q$ is comprehensive if $x \leqslant \hat{x}$ and $\hat{x} \in Q$ imply $x \in Q$.

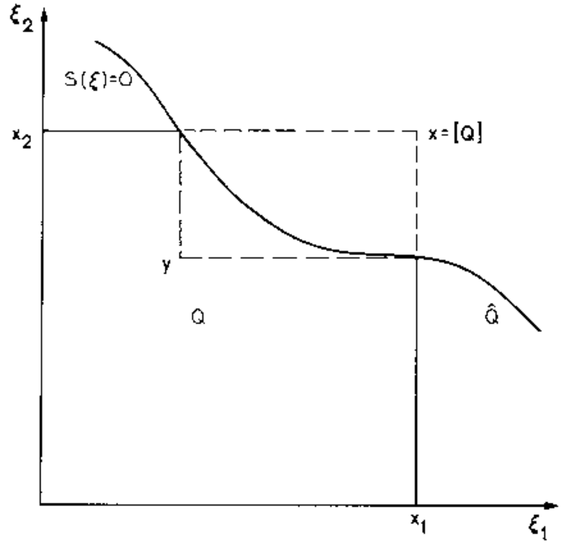


FIGURE 2—Representation of a purchase set.

purchase set $Q$ we use $[Q]$ to denote the smallest rectangular purchase set that contains it. Any decreasing function $s$ on $\mathfrak{R}^2_+$ can be used to define a purchase set, as $\hat{Q} = \{\xi \in \mathfrak{R}^2_+ \mid s(\xi) \geq 0\}$. A purchase set with a flat top and side is conveniently represented as the intersection $Q = [Q] \cap \hat{Q}$, as in Figure 2.

THE SELLER'S COST FUNCTION: The seller's costs are specified by a function $\mathscr{C}$ that assigns a dollar cost $\mathscr{C}(Q)$ to each purchase set $Q$. The cost function $\mathscr{C}$ is a mapping from the collection of purchase sets in $\mathfrak{R}^2_+$ into the nonnegative real numbers. We shall assume that $\mathscr{C}$ is the restriction (to the collection of purchase sets) of an additive measure on the Borel sets. There exists, therefore, a corresponding cumulative distribution function $C: \mathfrak{R}^2_+ \to \mathfrak{R}_+$ defined by $C(x) = \mathscr{C}([x])$, with the property that for an arbitrary purchase set $Q$:

$$(1) \qquad \mathscr{C}(Q) = \int_{x \in Q} dC(x),$$

using Riemann-Stieltjes integration [7].

The distribution function $C(x)$ is absolutely continuous except for discontinuities from below along the two axes; these discontinuities correspond to the capacity costs. The measure in (1) can therefore be expressed in terms of a fixed cost $C_0$, continuous capacity cost functions $C_1(x_1)$ and $C_2(x_2)$ along the two dimensions of capacity (for $x_1$ and $x_2 > 0$), and a density function $c(\xi)$ for usage costs incurred at points $\xi \gg 0$. The density function $c$ is obtained from the relation

$$c(\xi) = \partial^2 C(\xi) / \partial \xi_1 \partial \xi_2,$$

for points $\xi \gg 0$. The total cost of a purchase set $Q$ is therefore given by the formula:

$$(2) \quad \mathscr{C}(Q) = \begin{cases} C_0 + C_1([Q]_1) + C_2([Q]_2) + \int_Q c(\xi)\, d\xi & \text{if } Q \cap \mathfrak{R}^2_{++} \text{ is not empty;} \\ 0 & \text{otherwise;} \end{cases}$$

where $[Q]_i$ is the projection of $Q$ on the $i$th axis, and $d\xi = d\xi_1\, d\xi_2$ for Riemann integration in $\mathfrak{R}^2$.

The purchase set defined by the load duration curve is usefully viewed as a bundle of "little squares," each representing a unit of service produced by a particular unit of capacity at a particular time. The variable cost of such a unit depends on its coordinates under the load duration curve that determine the technology by which it was produced and its marginal production cost with that technology.

The assumption that the cost function $\mathscr{C}$ is a Borel measure is a significant restriction with important practical ramifications. It is satisfied only if usage costs are additively separable along one of the two dimensions (and therefore along both). A more general formulation would allow that usage costs depend on the entire shape of the load-duration curve, but we have not been able to complete the analysis of such a formulation. The formulation adopted here imposes the assumption that each of the components of total cost, namely the fixed cost $C_0$ and each of the capacity cost functions $C_1$ and $C_2$ as well as the usage cost density $c$, is independent of the shape of the specific purchase set $Q$ selected by the customer. This is actually a separability assumption stating that all the cost components of a purchase set are additively separable.

We now consider the meaning of this assumption in a practical context. In a multiproduct environment, the cost of providing a particular purchase set depends on the system configuration used for that purpose. The cost function $\mathscr{C}(Q)$ describes the efficient cost of supplying $Q$ with an optimal mix of production technologies. To illustrate this, consider the solution to a simple technology mix optimization problem. (Our analysis resembles Steiner's [9] and Boiteux' [1] for the planning of electric power generation.)

Suppose, for example, that the three infinitely divisible technologies illustrated in Figure 3 are available to meet the service requirements described by a particular load duration curve. The fixed and variable costs are such that each technology is efficient in some range of the duration of utilization, as shown in Figure 3a. This is a plausible assumption since in the absence of effects due to discrete sizes of units (i.e., with infinite divisibility) an efficient system cannot include technologies that are dominated with respect to both fixed and variable costs. Projecting the efficiency range of each technology onto the load duration curve enables one to determine the optimal technology mix and the dispatch policy for each technology (i.e., the usage rates at which it is operated). As shown in Figure 3b, technologies with low capacity costs are used to satisfy the peak load while those with low marginal costs are used for the base load, and an intermediate technology meets the shoulder load. The total cost of satisfying the load duration curve with the optimal technology mix can be calculated layer by layer using the cost functions corresponding to the dispatched technologies in each range. However, since the

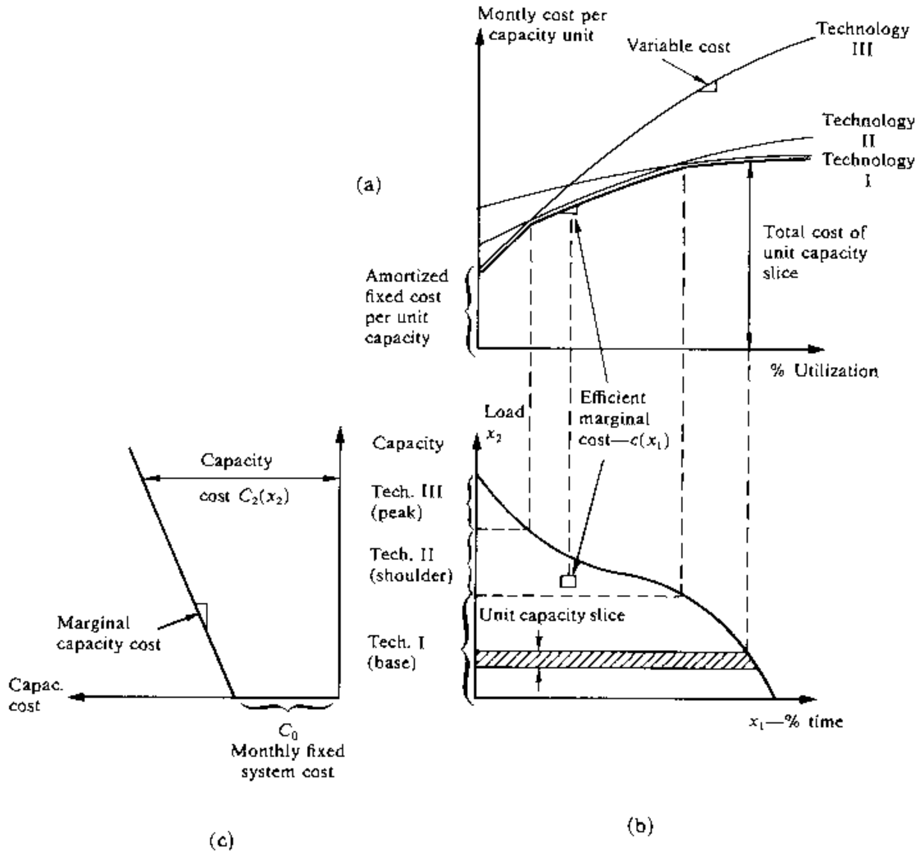


FIGURE 3—Illustration of the optimal technology mix determination and efficient cost calculation.

dispatch boundaries are determined by the shape of the load duration curve, it might appear that the usage and capacity costs at a point $(x_1, x_2)$ also depend on the shape. This would violate our assumption that the cost function is a Borel measure. Fortunately that is not the case. This becomes clear if we note that the efficient cost of operating any one unit of capacity, as a function of operating time, is given by the lower envelope of the individual technologies' cost functions, as depicted in Figure 3a. This envelope can be interpreted as a nonlinear cost function on capacity utilization. The effect of technology mix optimization is manifested through the concavity of this cost function, reflecting the fact that technologies with lower operating cost are assigned to capacity units that are more heavily utilized. A general cost formula based on this latter approach is constructed in Appendix C. It is shown there that integrating by parts converts the cost function into a form satisfying our assumptions. We only note here that under this construction the marginal cost of any "little square" under the load duration curve depends only on its first coordinate $x_1$. Furthermore, the capacity cost is linear, and there may be an additional system cost per cycle (e.g., overhead) as illustrated in Figure 3c. This cost function clearly meets all the conditions assumed for the cost functional $\mathscr{C}(Q)$.

We impose the following regularity assumption on the cost function:

ASSUMPTION [A1]: *The functions $C_1$, $C_2$, and $C$ are each increasing, convex, and twice differentiable.*

PAYMENT SCHEDULES: A basic motive for the adoption of nonlinear pricing is to adapt the payment schedule to the structural features of the cost function. Thus we assume that the payment schedule $\mathscr{P}$ is also an additive measure on $\Re^2_+$ with a corresponding cumulative distribution function $P$. Thus, by analogy with the cost function in (2), the payment schedule can be expressed in the form:

$$\mathscr{P}(Q)=P_0+P_1([Q]_1)+P_2([Q]_2)+\int_Q p(\xi)\,d\xi, \tag{3}$$

where

$$p(\xi)=\partial^2 P(\xi)/\partial\xi_1\,\partial\xi_2, \tag{4}$$

and $P(\xi)\equiv\mathscr{P}([\xi])$. The other assumptions adopted in the formulation will be sufficient to ensure that the seller's optimal functions $P_1$, $P_2$, and $P$ are nondecreasing and differentiable.

The components of the payment schedule are interpreted by the customer in the same fashion that the seller interprets the analogous cost components. Thus, $P_0$ is a fixed installation charge; $P_i(x_i)$ is the charge for $x_i$ units of the $i$th capacity, and the total usage charges are accumulated by integrating the usage prices over the purchase set $Q$. The capacity required to provide the usage in $Q$ is $[x]=[Q]$.

CUSTOMERS: The heterogeneity in the customer population has a central role in nonlinear pricing because the optimal payment schedule is designed to induce self-selection among the customers. The different preferences among the customers are reflected in their purchases. Realizing this, a seller can achieve discrimination indirectly by adjusting the payment schedule so that, for example, customers making small purchases pay small fixed fees and relatively high marginal charges, whereas customers making large purchases pay larger fixed fees in return for smaller marginal charges.[3]

The heterogeneity that we allow is restricted to an indexing of the customers' types along a single dimension. Taking the population of customers to be a continuum, one can allow the types to be characterized by a real-valued index $\tau$ distributed according to a distribution function $F$ that is strictly increasing and differentiable. However, it suffices to take as the index of the type $\tau$ its rank-order $t=F(\tau)$ so that $t$ is uniformly distributed on the unit interval.

[3] For expositions of the basic ideas underlying nonlinear pricing see Goldman, Leland, and Sibley [2], Mirman and Sibley [3], Roberts [6], or Spence [8].

As with the seller's costs, we assume that each customer's utility function is an additive measure on $\Re^2_+$ having an integral representation

$$\mathcal{U}(Q; t) = \int_Q u(\xi; t)\, d\xi, \tag{5}$$

where the marginal benefit $u$ is determined by the utility function for rectangular purchase sets:

$$u(\xi; t) = \partial^2 U(\xi; t)/\partial\xi_1\, \partial\xi_2, \tag{6}$$

using $U(\xi; t) \equiv \mathcal{U}([\xi]; t)$ to specify the utility for rectangular purchase sets. Discontinuities along the two axes are excluded in this case since ordinarily customers do not value capacity except as it provides usage; however, the later analysis will make clear that one could easily allow that customers value capacity directly.

The assumption that utility is an additive measure imposes a limit on the customer's gains from substitution. Typically it means that the customer can choose to purchase more or less on either dimension, but direct substitution between points differing in both dimensions is excluded. This reflects the essential feature that utility is additive over one (and therefore both) of the dimensions.

We impose strong regularity assumptions on the customers' utility functions.

ASSUMPTION [A2]: *$U(x; t)$ is strictly increasing and concave in $x$ and thrice differentiable in $(x; t)$. Furthermore, the two partial derivatives $\partial U/\partial x_i (i = 1, 2)$ are each concave in the other dimension ($x_j$ for $j \neq i$); that is, the marginal benefit for increasing one attribute flattens out as the other attribute is increased. Finally, $U(\cdot\,; t)$ and its first and cross partial derivatives with respect to $x$ are all decreasing with the customer's type index $t$; thus, a higher index corresponds to a uniformly lower willingness-to-pay for incremental purchases.*

OPTIMAL SELECTION OF THE PURCHASE SET: Assuming negligible income effects, each customer $t$ selects that purchase set $Q(t)$ that achieves the maximum difference between his utility and the required payment:

$$\max_Q \mathcal{U}(Q; t) - \mathcal{P}(Q), \tag{7}$$

or the customer can choose not to make any purchase if this maximum is negative. A convenient way to express (7) uses the difference $[Q] - Q$ between the rectangular set $[Q]$ enclosing $Q$, and $Q$, since it is $[Q]$ that determines the capacity charges. Thus one can think of the customer as purchasing $[Q]$ and then relinquishing the portion $[Q] - Q$:

$$\max_{x, \hat{Q}} U(x; t) - P(x) - \int_{[x]-\hat{Q}} [u(\xi; t) - p(\xi)]\, d\xi, \tag{8}$$

where the resulting purchase set is then $Q(t) = [x] \cap \hat{Q}$.

This formulation is the simplest structure that captures the main features we want to emphasize. These are the two-dimensional character of the product space, capacity costs depending on the maximal attributes purchased along each dimension, and heterogeneity among the customers. Our aim in the subsequent analysis is to characterize the optimal pricing policy and to show how the payment schedule reflects these structural features.

### 3. DERIVATION OF THE OPTIMAL PAYMENT SCHEDULE

In this section we characterize the seller's optimal payment schedule. The task divides into two parts. First we determine the customers' demand behavior in response to a selected payment schedule, and then we analyze the seller's choice of the profit-maximizing payment schedule.

DEMAND BEHAVIOR: A customer's response to an *arbitrary* payment schedule is difficult to characterize in general. As in most studies of nonlinear pricing, however, it will turn out here that the response to the *optimal* payment schedule has quite regular mathematical properties. Typically the optimal payment function $P$ is concave, but less so than is $U$, and this allows one to assert that the first-order differential condition for an optimum is necessary and sufficient. This property is satisfied in a wide class of models, including the example studied in Section 4, and we shall assume it here although we have not identified the exact sufficient conditions.[4]

For a rectangular purchase set $[x]$ the customer's surplus is $S(x;t)=U(x;t)-P(x)$. For the following results we assume that the Assumption [A2] applies also to $S$ when the optimal payment schedule is used by the seller, and to avoid trivial cases we assume that the first and cross-partial derivatives of $S$ at the origin are positive. According to (8) the customer $t$ selects $x$ and $\hat{Q}$ to maximize

$$S(x;t)-\int_{[x]-\hat{Q}} s(\xi;t)\,d\xi, \tag{9}$$

where $s(\xi;t)=u(\xi;t)-p(\xi)$ is the cross-partial of $S$, and the chosen purchase set is $Q(t)=[x]\cap\hat{Q}$.

LEMMA 1: *The optimal purchase set is characterized by the following properties*:

$$\hat{Q}=\{\xi\in\mathfrak{R}^2_+\,|\,s(\xi;t)\geqslant 0\}, \tag{10}$$

$$S^1(x_1,y_2;t)=S^2(y_1,x_2;t)=0, \tag{11}$$

$$s(x_1,y_2;t)=s(y_1,x_2;t)=0, \tag{12}$$

[4] In one-dimensional problems the sufficient condition is a version of [A2]; see Goldman, Leland, and Sibley [2].

*where*

$$(13) \qquad S^1(x_1, y_2; t) = \int_0^{y_2} s(x_1, y; t)\, dy - P_1'(x_1),$$

*and similarly for $S^2$, are the partial derivatives of $S$ with respect to its first and second arguments. (The point $y$ is illustrated in Figure 2.)*

PROOF: We show first that $s(x; t) \leq 0$ outside $\hat{Q}$. Supposing the contrary, the concavity of $S^1$ and $S^2$ imply that $s(\xi; t) > 0$ for all $\xi \in [x]$ and therefore it is optimal to have $Q(t) = [x]$. The optimality of $x$ then requires that $S^1 = S^2 = 0$ at $x$. But the concavity of, say, $S^1$ implies that $s(x; t) \leq S^1(x; t)/x_2 = 0$, contradicting the supposition. It follows that the integral in (9) is minimized by choosing the domain of integration to be the region where the integrand is nonnegative, thus verifying (10). Equations (11) and (12) then state the first-order necessary conditions for the choice of $x$ subject to this choice of $\hat{Q}$. *Q.E.D.*

This characterization of the optimal purchase set is intuitively appealing. It implies that in the absence of capacity charges the buyer would simply select the set $\hat{Q}$ (shown in Figure 2) containing those points $x$ for which $u(x; t) \geq p(x)$—namely those points providing benefits exceeding the payment required. The capacity charges $P_1$ and $P_2$ induce the buyer to truncate this purchase set by eliminating those segments of usage for which the net benefits do not cover the incremental capacity charges.[5] Because we have assumed that the utility distribution function $U(x; t)$ is more concave in each argument of $x$ than the optimal payment distribution function $P(x)$, the truncation occurs where the marginal consumer surplus is zero along each dimension. The last relationship (12) determines the coordinates $y_1$ and $y_2$ below which the truncation due to the capacity charges is effective in curtailing demand.

SUBSCRIBERS: A customer who makes a purchase chooses the purchase set specified in Lemma 1. The customers who elect to purchase, called the subscribers, will be the ones with indices $t$ satisfying $\mathcal{U}(Q(t); t) \geq \mathcal{P}(Q(t))$, so that their consumer surpluses are nonnegative. The subscribers are characterized by the following result, which relies on the assumption that the indices are ordered along a single dimension.

LEMMA 2: *The set of subscribers is an interval $\mathcal{T} = [0, T]$ and their purchase sets are ordered by inclusion*:

$$(14) \qquad t > \hat{t} \Rightarrow Q(t) \subset Q(\hat{t}) \quad \text{and} \quad [Q(t)] \ll [Q(\hat{t})].$$

[5] Such truncations are similar to the *ex ante* self-rationing of capacity studied by Panzar and Sibley [5]. The variable usages of the various capacity increments described here by the load duration curve is captured in their model by random fluctuations in demand for capacity.

*Further*:

$$(15) \qquad d\mathcal{S}(t)/dt = \int_{Q(t)} \partial u(\xi; t)/\partial t \, d\xi < 0.$$

PROOF: The ordering of the purchase sets by inclusion is a simple consequence of the monotonicity assumptions imposed. Similarly,

$$(16) \qquad d\mathcal{S}(t) = \partial_Q \mathcal{S} \cdot dQ + \partial_t \mathcal{S} = \partial_t \mathcal{S},$$

using Lemma 1, and then (15) follows from the monotonicity of $u$. Thus the set of purchasers is an initial segment of the unit interval. *Q.E.D.*

In view of Lemma 2 it is useful to characterize the inverse of the map $t \mapsto (\hat{Q}, x)$ in Lemma 1 in terms of the function

$$(17) \qquad t(\xi) = \sup_\tau \{\tau \mid \xi \in Q(\tau)\}.$$

One can then define $t_i(x_i)$ so that $t_1(x_1) = t(x_1, y)$ for any $y \leq y_2$ and analogously for $t_2$. Also associated with each customer type $t$ is a value $(y_1(t), y_2(t))$ of the point $y$. The subset $Q^*$ of $Q(0) - Q(T)$ that is not constrained by capacity choices plays an important role later: it is defined as

$$(18) \qquad Q^* = \{\xi \in Q(0) - Q(T) \mid \xi \ll [Q(t(\xi))]\}.$$

The marginal subscriber is identified by the condition $\mathcal{S}(T) = 0$. Since the type index $t$ is uniformly distributed on the unit interval, $T$ is also the market penetration, namely the fraction of potential customers who are induced to purchase.

THE SELLER'S OBJECTIVE FUNCTION: The seller offers the same payment schedule to every customer and his profit from each subscriber is the difference between the payment received and the cost incurred. Alternatively, his profit is the difference between the total surplus $\mathcal{W}(t) = \mathcal{U}(Q(t); t) - \mathcal{C}(Q(t))$ and the consumer surplus $\mathcal{S}(t)$. Thus, the total profit can be expressed as[6]

$$(19) \qquad \Pi(P) = \int_0^T [\mathcal{W}(t) - \mathcal{S}(t)]\, dt,$$

depending on the payment function $P$ for rectangular purchase sets (which in turn determines the full payment schedule $\mathcal{P}$). Integrating by parts using (15)

[6] The integrand would be $(1+\mu)\cdot\mathcal{W} - \mu\cdot\mathcal{S}$ in a welfare problem subject to the constraint $\Pi \geq \Pi^0$ for which $\mu$ is a Lagrange multiplier; here we take $\mu = \infty$ but the methods employed do not depend on this restriction. Appendix A indicates how the methods in the text can be applied to this more general case. See Mirman and Sibley [3] for related results. Roberts [6] studies the welfare aspects of nonlinear pricing in detail.

and calculating terms puts this objective into a more tractable form for analysis:

$$\Pi(P) = T\mathcal{W}(T) - \int_0^T t \cdot \left[ \mathcal{W}'(t) - \int_{Q(t)} \partial u(\xi; t)/\partial t \, d\xi \right] dt \tag{20}$$

$$= T\mathcal{W}(T) - \sum_{i=1,2} \int_{x_i(T)}^{x_i(0)} t_i(x_i) \cdot C_i'(x_i)\, dx_i$$

$$+ \int_{Q(0)-Q(T)} t(\xi) \cdot [u(\xi; t(\xi)) - c(\xi)]\, d\xi,$$

$$= T\mathcal{W}(T) - \sum_{i=1,2} \int_{x_i(T)}^{x_i(0)} t_i(x_i) \cdot C_i'(x_i)\, dx_i$$

$$+ \sum_{i=1,2} \int_{x_i(T)}^{x_i(0)} t_i(x_i) \cdot \int_0^{y_i(t_i(x_i))} [u(x_i, y; t_i(x_i)) - c(x_i, y)]\, dy\, dx_i$$

$$+ \int_{Q^*} t(\xi) \cdot [u(\xi; t(\xi)) - c(\xi)]\, d\xi$$

$$= T\mathcal{W}(T) + \sum_{i=1,2} \int_{x_i(T)}^{x_i(0)} t_i(x_i) \cdot \frac{\partial}{\partial x_i} W_i(x_i)\, dx_i$$

$$+ \int_{Q^*} t(\xi) \cdot [u(\xi; t(\xi)) - c(\xi)]\, d\xi, \tag{21}$$

where $t(\xi)$ and $t_i(x_i)$ are the corresponding inverses of the map $t \mapsto (\hat{Q}, x)$ from Lemma 1 which we will characterize below, and

$$W_i(x_i) = U(x_i, y_i(t_i(x_i)); t_i(x_i)) - C(x_i, y_i(t_i(x_i))) \qquad (i = 1, 2). \tag{22}$$

Figure 4 provides a useful interpretation of equation (21). The volume depicted in Figure 4 can be regarded as a "stack" of horizontal slices in the shape of the purchase sets of the various customer types. The differential of the original profit integral is the profit contributed by the purchase sets of customers in the interval $[t, t+dt]$. Consequently the total profit is obtained by adding up the profit contributions of all the slices from $t = 0$ to the marginal buyer $t = T$. The key step in the derivation of (21) amounts to redefining the profit differential so that the total profit associated with the entire volume is obtained by accumulating "vertical shells" rather than horizontal "slices". These vertical shells can be broken up as shown in the figure into flat vertical "slabs" representing differential profits from capacity increments and "columns" representing differential profits from incremental usage within the capacity limits. These can be accumulated independently by integration over the appropriate domain in the $(\xi_1, \xi_2)$ plane. The usefulness of the above transformation stems from the fact that the pricing policy does not discriminate by customer type. Thus, the profit contribution of any infinitesimal square of usage at $\xi$ depends only on its position in the $(\xi_1, \xi_2)$ plane and is proportional to $t(\xi)$, the number of customers who purchase it, which is the

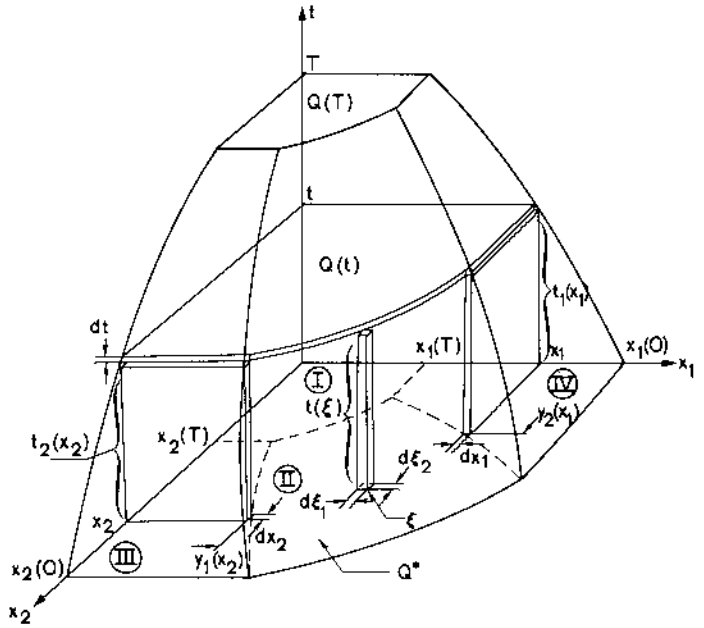


FIGURE 4—Evaluation of the profit function.

height of the slab or column in that position. This height specifies the market share for that capacity or usage increment.

The terms in equation (21) have a natural economic interpretation. The four terms in (21) correspond to the four domains delineated with a dashed line in the $(\xi_1, \xi_2)$ plane of Figure 4. The first term in (21) gives the profit contributed by the smallest purchase set $Q(T)$ which is included in the purchase set of all subscribers making up a share $T$ of the potential market. Since the marginal subscriber (whose index is $T$) breaks even on his total purchase (i.e., his consumer surplus is zero), the seller's profit from that purchase equals the total surplus $\mathscr{W}(T)$ resulting from consumption of purchase set $Q(T)$ by buyer $T$. However, since the supply cost and payment for this purchase set are the same for all buyers, the total profit contribution (normalized to a unit market size) from domain I amounts to $T\mathscr{W}(T)$.

The last term in (21) corresponds to domain II. Each infinitesimal square in that region represents a consumption increment whose cost and price depend only on its coordinates. If we define $t(\xi)$ such that $\xi$ is on the boundary of $Q(t(\xi))$ then by Lemma 2, the point $\xi$ is in the purchase sets of all subscribers of types $t < t(\xi)$. Consequently the market share for a consumption unit in the square $[\xi, \xi + d\xi]$ is $t(\xi)$ which is also the index of the marginal buyer of that square. The monotonicity assumptions we impose are sufficient to ensure that there is a unique customer index $t(\xi)$ for which $\xi \in Q(0) - Q(T)$ is on the boundary of $Q(t(\xi))$.

The optimal purchase behavior characterized in Lemma 1 implies that the marginal buyer of a square at $\xi$ breaks even on this incremental purchase, i.e., his marginal consumer surplus is $s(\xi)=0$. Consequently the total surplus $[u(\xi;\, t(\xi))-c(\xi)]\,d\xi$ resulting from that incremental consumption by customer $t(\xi)$ equals the profit extracted from him (for that purchase) by the seller. However, since this profit must be the same for all $t(\xi)$ customers buying that square it follows that the seller's total profit for that square is given by $t(\xi)[u(\xi;\, t(\xi))-c(\xi)]\,d\xi$. The last term in (21) is obtained by integrating the latter expression over all values of $\xi$ in the set $Q^*$ which gives the total profit associated with domain II.

Similar reasoning to the above can be used to derive the second and third terms in (21) giving the profit contributions of domains III and IV. Here again $t_i(\xi)$ is the market share for capacity level $x_i$ or higher and $\partial W_i(x_i)/\partial x_i$ times $dx_i$ is the incremental profit from capacity increment $[x_i, x_i+dx_i]$. Again there is a unique customer index $t_i(x_i)$ such that that customer selects the capacity $x_i$ along the $i$th dimension. The construction ensures that $t_i([Q(t(\xi))]_i)=t(\xi)$ if $\xi$ is on the boundary of $Q(t(\xi))$. Finally, the profit from an incremental capacity unit equals the total surplus corresponding to the marginal buyer of that increment.

THE OPTIMAL PAYMENT SCHEDULE: We now turn to the characterization of the optimal price schedule which results from maximizing the expression in (21) for the seller's profit.

THEOREM: *The optimal marginal price schedule is*

$$p(\xi)=u(\xi;\, t(\xi)), \tag{23}$$

*where $t(\xi)$ solves*

$$t\cdot \partial u(\xi;\, t)/\partial t+u(\xi;\, t)=c(\xi). \tag{24}$$

*The optimal marginal capacity prices are determined from the conditions*

$$t\cdot \partial^2 U(x^i(t);\, t)/\partial x_i \partial t+\partial U(x^i(t);\, t)/\partial x_i=\partial C(x^i(t))/\partial x_i \qquad (i=1,2) \tag{25}$$

*where $x^1(t)=(x_1(t), y_2(t))$, and symmetrically for $x^2(t)$ as in* (11) *and* (12), *and $t=t(x^i)$.* (*Below we show how to compute the capacity charges.*) *The optimal market penetration is determined by the condition*

$$T\cdot \partial \mathcal{U}(Q(T);\, T)/\partial t+\mathcal{U}(Q(T);\, T)=\mathcal{C}(Q(T))+\lambda, \tag{26}$$

*where $\lambda$ is a Lagrange multiplier on the constraint $T\leqslant 1$.* (*Below we show how to compute the fixed charge.*)

PROOF: We use (20) and (21) to justify these results. The optimal marginal price schedule is the one that achieves the pointwise maximization of the integrand in the last term of (21), which yields (24), and then (23) is implied by (10). Similarly, optimizing the integrand in the second term in (21) yields (25). The various monotonicity properties enable one to combine all these conditions to solve for $t_i(x_i)$, namely the customer type selecting capacity $x_i$ on the $i$th dimension, and $y_j(x_i)$, namely the usage on the other dimension at which the upper boundary of the purchase set intersects the capacity limitation selected. Using these and invoking (13) yields the marginal capacity charges:

$$(27) \qquad P_1'(x_1) = \int_0^{y_2(x_1)} [u(x_1, y; t_1(x_1)) - p(x_1, y)]\, dy,$$

and similarly for $P_2'$. The first-order necessary condition for the optimal market penetration $T$ is (26), obtained by differentiating with respect to $T$ in (20) after adding a Lagrangian term $\lambda \cdot [1-T]$. One sees from (26) that

$$(28) \qquad \lambda = \max[0, \partial \mathcal{U}(Q(1); 1)/\partial t + \mathcal{U}(Q(1); 1) - \mathcal{C}(Q(1))],$$

and if $\lambda > 0$ then $T = 1$. The fixed charge $P_0$ that achieves $T$ is then constructed as

$$(29) \qquad P_0 = \lambda + \mathcal{U}(Q(T); T) - \sum_{i=1,2} \int_0^{x_i(T)} P_i'(\xi_i)\, d\xi_i - \int_{Q(T)} p(\xi)\, d\xi.$$

Note that in each of these constructions the seller first determines the customer's purchase set that the seller prefers to induce, subject to the demand behavior imposed by Lemma 1. The payment schedule is then determined to achieve these results. In each case the fact that the ordering of customers' types is reflected in the demand behavior enables the parts of the payment function to be obtained by integration operators once the desired induced demand behavior has been established. *Q.E.D.*

ELASTICITY INTERPRETATION: As in the case of one-dimensional nonlinear tariffs (cf. Goldman, Leland, and Sibley [2]), one can interpret equation (24) as a classic monopoly pricing condition for incremental units of consumption. Here the incremental unit is a square at $\xi$ and as explained before the total demand (market share) for that square is $t(\xi)$. Thus in view of (23), we can express the price elasticity $\varepsilon(\xi)$ for units of consumption at coordinates $\xi$:

$$(30) \qquad \varepsilon(\xi) = \frac{u(\xi; t(\xi))}{t(\xi)\, \partial u(\xi; t(\xi))/\partial t}.$$

Condition (24) then becomes the classical monopoly pricing condition:

$$(31) \qquad 1 + \frac{1}{\varepsilon(\xi)} = \frac{c(\xi)}{p(\xi)},$$

where each square $[\xi, \xi + d\xi]$ is regarded as a separate market. Conditions (25) and (26) can be represented in the same fashion as monopoly pricing conditions with the proper definitions of price elasticities for capacity and membership (in the subscribers' "club"). Note however, that the demand functions needed for these conditions are defined recursively. The demand function for capacity is defined in terms of the optimal prices for usage. Similarly the demand function for the service as a whole (membership) is defined in terms of the optimal capacity and usage prices.

Taken together these results yield an algorithm for constructing the optimal payment function $P$ for rectangular purchase sets. In brief, one uses (24) to find which customer type $t(\xi)$ is to have $\xi$ on the boundary of its purchase set and then one uses (23) to set the marginal payments for squares (i.e., the payment density function $p$) to obtain the desired result; similarly, (25) determines which customer type chooses each capacity level and then (27) sets the right marginal capacity charge to achieve this result; and lastly, (26) identifies the desired market penetration and then (29) specifies the fixed charge that achieves it. Specifying the various marginal charges is sufficient since the constants of integration associated with integrating these marginal charges can be lumped into the subsequent calculation. We illustrate the method in the next section.

## 4. AN EXAMPLE

In this section we apply the results from Section 3 to solve completely a class of examples that illustrate the assignment of capacity costs to customers. No new restrictions are placed on the cost function, but we assume that the customers' utility functions take the special form

$$U(x; t) = [k - t^{\alpha}] \cdot U(x), \tag{32}$$

where $\alpha > 0$ and $k > 0$.

PROPOSITION 4: *The optimal price schedule is determined by the price function*

$$P(x) = ak \cdot U(x) + [1 - a] \cdot [C(x) + \lambda] \tag{33}$$

*for rectangular purchases, where* $a = \alpha/[1+\alpha]$ *and*

$$\lambda = \max[0, K\mathcal{U}(Q(1); 1) - \mathcal{C}(Q(1))]$$

*for* $K = k - [1+\alpha]$. *If* $k \leq 1 + \alpha$ *then* $\lambda = 0$. *This payment schedule induces the purchase sets*

$$Q(t) = \{\xi \in \mathfrak{R}^2_+ \mid \xi \leq x(t);\ c(\xi)/u(\xi) \leq k - [1+\alpha]t^{\alpha}\}, \tag{34}$$

*where for each type $t$ the selected capacities $x(t)$ are determined by the conditions*:

$$\frac{u(x_1, y_2)}{c(x_1, y_2)} = \frac{U^1(x_1, y_2)}{C^1(x_1, y_2)} = \frac{u(y_1, x_2)}{c(y_1, x_2)} = \frac{U^2(y_1, x_2)}{C^2(y_1, x_2)} = k - [1+\alpha] \cdot t^{\alpha}, \tag{35}$$

*where* $U^i = \partial U/\partial x_i$ *and similarly for* $C^i$.

PROOF: Express the optimal price function as

$$P(x) = P_0 + P_1(x_1) + P_2(x_2) + P_3(x), \tag{36}$$

and write the cost function similarly. We now construct each term separately. The first step is to show that

(37) $$P_3(x) = ak \cdot U(x) + [1-a] \cdot C_3(x).$$

To do this we employ the optimality condition (24) to obtain

(38) $$[k-(1+\alpha)t^\alpha] \cdot u(\xi) = c(\xi),$$

where $u$ is the cross partial of $U$. Solving this for $t$ yields $t(\xi)$. Then (23) yields

(39) $$p(\xi) = ak \cdot u(\xi) + [1-a] \cdot c(\xi),$$

yielding the desired result. Next we show that

(40) $$P_1(x_1) = [1-a] \cdot C_1[x_1], \quad \text{and} \quad P_2(x_2) = [1-a] \cdot C_2(x_2).$$

To do this for $P_1$ we invoke the optimality condition (25) and obtain

(41) $$[k-(1+\alpha)t^\alpha] \cdot \partial U/\partial x_1 = \partial C/\partial x_1,$$

or equivalently

(42) $$[k-t^\alpha] \cdot \partial U/\partial x_1 = ak \cdot \partial U/\partial x_1 + [1-a] \cdot \partial C/\partial x_1,$$

evaluated at $(x_1, y_2)$ for the associated $t$. Solving this equation for $y_2$ defines $y_2(x_1; t)$; or together with the previously derived expression for $t(x)$ it determines $t_1(x_1)$ and $y_2(x_1)$. Consequently, (27) yields, at $(x; t) = (x_1, y_2(x_1); t_1(x_1))$:

(43) $$\begin{aligned} P_1'(x_1) &= [k - t_1(x_1)^\alpha] \cdot \partial U/\partial x_1 - \int_{0^+}^{y_2(x_1)} [aku + [1-a]c]\, dy \\ &= ak \cdot \partial U/\partial x_1 + [1-a] \cdot \partial C/\partial x_1 \\ &\quad - [ak \cdot \partial U/\partial x_1 + [1-a] \cdot \partial C/\partial x_1]_{0^+}^{y_2(x_1)} \\ &= ak \cdot \frac{\partial U}{\partial x_1}(x_1, 0^+) + [1-a] \cdot \frac{\partial C}{\partial x_1}(x_1, 0^+) \\ &= [1-a] \cdot C_1'(x_1), \end{aligned}$$

which implies the desired result. (In the above the first equality uses the previous result about the form of $P_3$, the second uses the optimality condition, and the last uses (6).) The form of the purchase set follows directly from the derivations so far. The last step is to establish that

(44) $$P_0 = [1-a] \cdot [C_0 + \lambda].$$

For this we apply the optimality condition (26) which takes the form

(45) $$[k-(1+\alpha)T^\alpha] \cdot \mathcal{U}(Q(T)) = \mathcal{C}(Q(T)) + \lambda,$$

or equivalently,

(46) $$[k-T^\alpha] \cdot \mathcal{U}(Q(T)) = ak \cdot \mathcal{U}(Q(T)) + [1-a] \cdot [\mathcal{C}(Q(T)) + \lambda],$$

with

(47) $$\lambda = \max\{0, (k-[1+\alpha])\cdot \mathcal{U}(Q(1)) - \mathcal{C}(Q(1))\},$$

where $\mathcal{U}(Q)$ is defined analogously to (5). Using this condition in (29) yields

(48) $$\begin{aligned} P_0 &= ak\cdot \mathcal{U}(Q(T)) + [1-a]\cdot[\mathcal{C}(Q(T))+\lambda] \\ &\quad -[P_1(x_1(T)) + P_2(x_2(T)) + P_3(Q(T))] \\ &= [1-a]\cdot\left[\lambda + \mathcal{C}(Q(T)) - C_1(x_1(T)) - C_2(x_2(T)) - \int_{Q(T)} c(\xi)\, d\xi\right] \\ &= [1-a]\cdot[\lambda + C_0], \end{aligned}$$

as required. Clearly, if $k < 1+\alpha$ then $\lambda = 0$ and $T$ will be an interior solution, namely $T < 1$. *Q.E.D.*

For purposes of discussion we will assume that $k = 1$ so that it is assured that $\lambda = 0$. The conclusion from this example is that the payment schedule is a weighted average of the benefit to the heaviest user ($t = 0$) and the seller's cost. All customer types pay capacity charges that are the same fraction $1-a$ of the seller's actual cost, and they pay usage charges in between the costs and the maximal benefit among the customers.

## 5. REMARKS

The direct assignment of capacity costs to customers is a common practice but economic theory includes few attempts to address the subject as an inherent part of an optimal pricing policy. In this paper we have attempted to show that a direct analysis of the pricing problem yields immediately the key feature that capacity charges are an important component of the payment schedule. We have addressed the problem using fully nonlinear prices, but presumably comparable results would follow from more restrictive types of payment schedules found more commonly in practice, such as two-part tariffs and block tariffs. The key feature is that the payment schedule conforms to the cost schedule, as in (3), so that the formulation incorporates the seller's opportunities to invoke self-selection among heterogeneous customers. From a technical viewpoint our work addresses the case of a firm with differentiated products sold in bundles, and a discontinuous cost function reflecting capacity costs. Casual observation suggests that such firms are sufficiently prevalent to justify a thorough treatment of their pricing strategies, recognizing their opportunities to adopt elaborately complex payment schedules that reflect their cost structures and that exploit optimally the heterogeneity among customers.

This paper focuses on leasing of systems which provide a metered service characterized by individual customers' load duration curves. In this case the capacities ordinarily correspond to the usage rate and the service duration. Our results can be reinterpreted for an analogous pricing problem arising in markets where a customer purchases quantities of various services differentiated by a measure of quality. One such market is the market for electronic mail. A customer sends messages and for each one selects a speed of delivery. The two-dimensional purchase set will define, in this case, the distribution of messages with respect to required service quality (delivery speed). Both the customer's benefit and the seller's costs depend upon the number of messages sent at the various speeds. In addition, the seller incurs costs that depend upon the maximum speed selected for any message and the maximum transmission rate (number of messages per period) at each speed, since these characteristics of the customer's usage pattern determine the type and capacity of the transmission equipment that the seller installs.

Because of the decentralization of production capacity and the exclusion of a resale market in our formulation, the analysis does not consider the relative timing of customers' demands. In particular, the additivity of costs across customers implicitly assumes that individual loads are synchronized, and therefore that capacity must be reserved for each customer to meet his peak load. With some limited exceptions, this restriction rules out applications where production is centralized. The salient example, of course, is the utility company that sells electric power. In such markets, asynchronous peak loads can be beneficially exploited to reduce requirements for generation capacity. To deal with situations of this type one needs to add a second customer index characterizing a relative shift in load pattern. Such a modification complicates the analysis significantly and is beyond the scope of this paper. A special case, however, of electric utility pricing to which our results could be applied arises when dealing with a restricted market where demand patterns are highly correlated (e.g., power for air conditioners). In that case the system's load duration curve is highly synchronized with the individual load duration curves. Hence one could consider a pricing approach which would be based on leasing capacity shares of the central generation facility to customers. The analysis presented in this paper provides a mechanism for pricing according to the individuals' load duration curves. (One could, of course, modify the objective function as indicated in Appendix A to reflect regulatory constraints on revenue.) With the rapid diffusion of electronic metering this method is now feasible. It does not require any *a priori* information about demand, it is computationally simple, and most important from a practical viewpoint it fully allocates generation costs. A legitimate criticism of such a scheme might be that it is less efficient than an ideal time-of-use pricing scheme which charges instantaneous marginal costs. This is because the exclusion of a resale market prevents inter-customer trading to equate their marginal benefits.[7] However, exact time-of-use pricing schemes are impractical and are usually approximated by second best schemes based on a few discrete time-of-use periods; moreover, in

[7] The inherent inefficiency of *ex ante* self-rationing schemes, as opposed to *ex post* rationing through spot pricing, is discussed by Panzar and Sibley [5]. They also obtain conditions on the customers' preferences under which *ex ante* self-rationing via marginal cost pricing is as efficient as spot pricing at marginal cost. These conditions on the customers' preferences induce all customers to demand their peak loads under the same environmental circumstances (i.e., simultaneously).

practice time-of-use pricing also departs from marginal costs (by imposing "demand charges" to meet capacity costs) in order to meet revenue requirements. Finally, to determine the price in each period one needs elaborate market data reflecting the intertemporal cross elasticities. In light of these considerations, utility pricing of synchronous load via direct capacity and usage charges may be desirable.

*Stanford University,*
*University of Santa Clara,*
*and*
*Stanford University*

*Manuscript received August, 1982; final revision received July, 1984.*

## APPENDIX A

### Welfare Optimization

In this Appendix we indicate the amendments to the formulation in the text that are required to obtain payment schedules that maximize total surplus. In the absence of a revenue constraint for the seller it is immediately obvious that the optimal payment schedule imposes an ordinary linear pricing scheme in which each capacity and usage at each time are priced according to marginal cost. We therefore address the case in which the payment schedule must be designed to obtain a minimum net revenue for the seller. This case is effectively summarized by adopting a Lagrangian formulation in which total surplus is augmented by the seller's net revenue weighted by a Lagrange multiplier, say $\lambda > 0$. That is, the payment schedule $\mathcal{P}$ is to be selected to maximize

$$\int_0^T \mathcal{W}(t)\,dt + \lambda \cdot \int_0^T [\mathcal{W}(t) - \mathcal{S}(t)]\,dt. \tag{49}$$

By varying the value of $\lambda$ one can adjust the seller's net revenue to meet the constraint imposed. The analysis in the text corresponds to the special case that $\lambda = \infty$, in which case the seller's net revenue is maximized.

Exploiting the linearity of this construction enables one to derive the conditions for an optimal payment schedule by simple modifications of the proof of Theorem 3. For example, the condition (24) now takes the form:

$$u(\xi; t) - c(\xi) + \lambda[t\,\partial u(\xi; t)/\partial t + u(\xi; t) - c(\xi)] = 0. \tag{50}$$

This amended condition is a simple linear combination of the optimality conditions for the unconstrained total surplus maximization problem and for the seller's profit maximization problem. For example, the first two terms are obtained from the total surplus maximization condition and from (23) that $u(\xi; t) = c(\xi) = p(\xi)$ for the marginal customer $t$ who purchases $\xi$. The same form carries over to the optimal capacity conditions (25) and the optimal market penetration condition (26). For example, (25) now has the form:

$$\begin{aligned} &\partial U(x^i(t); t)/\partial x^i - \partial C(x^i(t))/\partial x_i \\ &\quad + \lambda[t\,\partial^2 U(x^i(t); t)/\partial x_i\,\partial t + \partial U(x^i(t); t)/\partial x^i - \partial C(x^i(t))/\partial x_i] = 0. \end{aligned} \tag{51}$$

## APPENDIX B

### Nonlinear Pricing of Usage Only

In this Appendix we describe briefly the derivation of an optimal nonlinear payment schedule for a monopolist selling usage only, there being no capacity costs or charges. In this situation it is possible to allow customers' indices to be multidimensional vectors with no complication of the analysis. As in the text we assume that the customers' types $\tau$ are vectors in a Euclidean space, and that they are described according to a nonatomic measure with the distribution function $F$.

If the price for usage at the point $\xi$ is $p$ then a subscriber of type $\tau$ buys $\xi$ if and only if $u(\xi;\tau)\geqslant p$. The fraction of the customer population who will buy $\xi$ if the price is $p$ is therefore

$$(52)\qquad n(\xi;p)=\int_{B(\xi,p)} dF(\tau),$$

where the set of buyers is

$$(53)\qquad B(\xi;p)=\{\tau\,|\,u(\xi;\tau)\geqslant p\}.$$

The seller's optimal price $p(\xi)$ for $\xi$ is therefore the value of the price $p$ that maximizes the net revenue for the "local" market at $\xi$:

$$(54)\qquad n(\xi;p)[p-c(\xi)].$$

Under the assumptions made in the text this construction suffices to determine the optimal payment schedule. With weaker assumptions this construction is insufficient if the resulting purchase sets selected by the customers are not comprehensive because the price function $p(\xi)$ is not monotone; in this case one must address the more complicated problem in which the price function is constrained to be nonincreasing, and use the methods developed by Mussa and Rosen [4].

*Example*: Consider a one-dimensional product space, so that $\xi\in\mathfrak{R}^1_+$, and assume that the customers' types $\tau=(a,b)$ are uniformly distributed on the unit square. If the customer of type $(a,b)$ has the marginal benefit $u(\xi;a,b)=a-b\xi$ then

$$n(\xi;p)=\{a,b\in[0,1]\,|\,a-b\xi\geqslant p\}$$
$$=\frac{1}{2\xi}\begin{cases}(1-p)^2 & \text{if } p+\xi\geqslant 1,\\ (1-p)^2-(1-p-\xi)^2 & \text{if } p+\xi\leqslant 1.\end{cases}$$

For instance, if $c(\xi)\equiv 0$ then

$$p(\xi)=\begin{cases}\frac{1}{2}-\frac{1}{4}\xi & \text{if } \xi\leqslant\frac{2}{3},\\ \frac{1}{3} & \text{if } \xi\geqslant\frac{2}{3}.\end{cases}$$

## APPENDIX C

### The Cost of a Load-Duration Curve

In this appendix we briefly describe the construction of a cost function of the form that is common in applications, and in particular in the electric power industry. We use the interpretation of the power context.

The load duration curve specifies the load $L(t)$ as a function of the duration $t$, in the sense that the load exceeds $L(t)$ for a duration $t$. The peak load is $L(0)$ and the base load is $L(1)$, on the interpretation that duration is expressed as a fraction of a periodic cycle, and the dispatch rate is $l(t)=-L'(t)$, indicating the rate at which new power capacity is turned on as the load increases from the base to the peak. We assume that there are no ramping costs associated with dispatch.

The technologies available are indexed by $i$ and $c_i(t)$ is the cost per kilowatt using technology $i$ for a duration $t$. In the linear case, for example, $c_i(t)=F_i+V_it$ where $V_i$ is the energy cost per kilowatthour and $F_i$ is the amortized cost of generating equipment. The efficient cost envelope is $c(t)=\min_i c_i(t)$, and $i(t)$ indicates the efficient technology for the duration $t$.

The basic cost formula integrates the costs of horizontal slices under the load-duration curve:

$$\mathscr{C}(L)=c(1)L(1)+\int_{L(1)}^{L(0)} c(T(L))\,dL,$$

where $T(L)$ is the smallest duration $t$ for which $L(t)\leqslant L$. In practice it is customary to convert this load-slice formula into a time-slice formula by integrating by parts:

$$\mathscr{C}(L)=c(0^+)L(0)+\int_0^1 c'(t)L(t)\,dt.$$

One may, of course, add a term $C_0$ indicating a fixed cost of operations. This cost function has the form specified in (2) using the identifications $C_1 \equiv 0$, $C_2([Q]_2) = c(0^+)L(0)$, and for a point $\xi = (t, L)$ below the load-duration curve $c(\xi) = c'(t)$. In particular, $c(\xi) = V_{i(t)}$ in the linear case.

Just as the costs can be expressed in terms of either load slices or time slices, so also prices can be billed on the basis of either load slices (called demand-layered pricing) or time slices (called time-of-use pricing).

## REFERENCES

[1] BOITEUX, MARCEL: "Peak Load Pricing," *The Journal of Business*, 33(1960), 157–179.

[2] GOLDMAN, M. BARRY, HAYNE E. LELAND, AND DAVID S. SIBLEY: "Optimal Nonuniform Prices," *Review of Economic Studies*, 51(1984), 305–319.

[3] MIRMAN, LEONARD J., AND DAVID S. SIBLEY: "Optimal Nonlinear Prices for Multiproduct Monopolies," *The Bell Journal of Economics*, 11(1980), 659–670.

[4] MUSSA, MICHAEL, AND SHERWIN ROSEN: "Monopoly and Product Quality," *Journal of Economic Theory*, 18(1978), 310–317.

[5] PANZAR, JOHN C., AND DAVID S. SIBLEY: "Public Utility Pricing under Risk: The Case of Self-Rationing," *American Economic Review*, 68(1978), 888–895.

[6] ROBERTS, KEVIN W.: "Welfare Considerations of Nonlinear Pricing," *Economic Journal*, 89 (1979), 66–83.

[7] ROYDEN, HALSEY L.: *Real Analysis*. London and New York: Macmillan Company, 1968.

[8] SPENCE, A. MICHAEL: "Nonlinear Prices and Welfare," *The Journal of Public Economics*, 8(1977), 1–18.

[9] STEINER, PETER O.: "Peak-Loads and Efficient Pricing," *Quarterly Journal of Economics*, 72(1957), 585–610.