CAPACITY PRICING

by

Shmuel Oren, Stephen Smith, and Robert Wilson

Technical Report No. 386
October, 1982

Prepared Under
National Science Foundation Grants IST-8108350 and SES-8108226,
and
Contract ONR-N00014-79-C-0685, United States Office of Naval Research

THE ECONOMICS SERIES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
Fourth Floor, Encina Hall
Stanford University
Stanford, California
94305
CAPACITY PRICING* 
by 
Shmuel Oren, Stephen Smith, and Robert Wilson

1. Introduction

A central issue in price theory is how to set prices for services that require the seller to incur a setup cost, such as the cost of capital equipment necessary for production of the services. The technical aspect of the problem arises from the discontinuity of the seller's cost function. In this paper we adopt a novel approach using nonlinear pricing, and address the special case of a monopoly. Restrictive assumptions are imposed on the cost and demand functions, but these enable a complete characterization of the optimal pricing policy. A class of examples analyzed in Section 4 illustrates that the optimal pricing policy assigns a portion of the setup cost to the customer.

The problem we address is formally one of nonlinear pricing for a multiproduct monopoly, which has been studied by Mirman and Sibley [1980]. However, their analysis does not recognize the crucial role of the discontinuity in the cost function and its ramifications for the optimal pricing policy.

In order to appreciate the formulation adopted in Section 2 the reader may find it useful to interpret the setup in terms of the following examples.

*This work was supported by the National Science Foundation, grants IST-8108350 and SES-8108226, and by the Office of Naval Research, contract ONR-N00014-79-C-0685, at the Institute for Mathematical Studies in the Social Sciences, Stanford University.
Leasing Machines: Among the office equipment that is frequently leased from a vendor with a local monopoly for its differentiated product are copiers, small computers, terminals, and WATS lines. The customer typically acquires several units—some to meet a base load and additional units to meet successively higher loads up to the peak load. The customer's load can be described by a load-duration curve \( L(d) \). That is, if \( D(q) \) is the fraction of the time that the usage rate (i.e., the load) exceeds \( q \) then the load exceeds \( L(d) = D^{-1}(d) \) for a fraction \( d \) of the time, and \( L(0) \) is the peak load. The customer's benefits depend upon the load-duration curve that is served, as do the lessor's costs. In addition, the lessor incurs capital costs that depend upon the peak load because that determines the number of units installed. In some cases it might also be that the lessor incurs costs depending on the maximum duration (e.g., the length of the connection period in the case of a terminal).

Quality-Differentiated Services: In some markets 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 many messages and for each one selects a speed of delivery. Both the customer's benefits and the seller's costs depend upon the numbers 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.

These two examples are illustrative of a wide class of circumstances in which the seller delivers a product that is "two-dimensional" in the sense that it can be described as a set of pairs \((x,y)\). In the first case above, for instance, this set comprises those pairs satisfying \(0 < x \) and \(0 < y < L(x)\), circumscribed by the load-duration curve \(y = L(x)\). The customer's benefits and the seller's costs depend upon this set. In addition, the seller incurs costs that depend upon the projections of this set onto the two axes and onto the origin. Continuing the first example, the projection onto the x-axis yields the maximum duration, and the projection onto the y-axis yields the peak load. The projection onto the origin is interpreted as an indicator that a fixed cost is incurred. The cost function is discontinuous because the costs determined by the projections are only incurred if the aforementioned set has a nonempty interior; that is, no equipment is needed if either the maximum duration or the peak load is zero.$^1$

A nonlinear pricing policy in this context specifies a payment schedule \(P\) that assigns to each usage pattern \(Q\) a dollar amount \(P(Q)\) due the seller. In the general case it would be difficult to write down such a schedule, but under the assumptions we make about the customer's benefits and the seller's costs it turns out that the value of the optimal payment schedule for any particular usage pattern can be calculated from a simple formula.
Our criterion for optimality is that the payment schedule maximizes the seller's profit summed over all customers. Some further generality can be obtained by choosing instead to maximize a weighted average of profit and total surplus so as to invoke efficiency criteria, along the lines of Mirman and Sibley's analysis, but we forego that task in this paper.

The population of potential customers plays a central role in the analysis. As is always the case with nonlinear pricing, the payment schedule is designed to take advantage of the differences among customers' preferences. In order to illustrate our main points it suffices to allow only a limited heterogeneity among customers, but our method seems to have straightforward extensions to more complicated models of the demand structure. The present model assumes that customers' characteristics are summarized in a one-dimensional index satisfying some monotonicity requirements. As will be seen, this is a stringent assumption since it ultimately implies that the customers respond to the optimal payment schedule with usage sets that are well-ordered by inclusion.

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 neatly illustrate how a portion of the capacity costs are borne by customers. Concluding remarks are offered in Section 5.
2. **Formulation**

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

**Products:** A customer's purchase is represented as a compact and comprehensive subset of the nonnegative orthant of two-dimensional Euclidean space. Typically one interprets one dimension as a measure of the quantity or rate of consumption and the other dimension as a measure of quality or an index of the conditions of delivery (such as the time) but these interpretations are not necessary. Comprehensive- ness is a property reflecting the usual fact that the purchase of the q-th increment of quantity or quality entails purchase of all lesser increments. The "shape" of a purchased set is not otherwise restricted, but as will be seen the rectangular purchase sets play a special role in our analysis. The rectangular set with the corner (i.e., least-upper-bound) \( x \in \mathbb{R}^2 \) will be denoted by \([x]\). For an arbitrary purchase set \( Q \) we use \([Q]\) to denote the smallest rectangular set that contains it. A purchase set with a flat top and side is conveniently represented as the intersection \( Q = [Q] \cap \hat{Q} \), especially if \( \hat{Q} \) has a functional representation as, for example, \( \hat{Q} = \{ x \in \mathbb{R}^2 \mid q(x) > 0 \} \) for some decreasing function \( q \). An example of such a purchase set is depicted in Figure 1. We use \( Q^0 \) to denote the interior of \( Q \) relative to \( \mathbb{R}^2 \).

**Costs:** The seller's costs are generally given as a function \( C \) that assigns a dollar cost \( C(Q) \) to each purchase set \( Q \), with the
convention that \( C(Q) = 0 \) if there is no strictly positive point in \( Q \) (namely \( Q^0 \) is empty). This convention reflects the presumption that capacity costs are incurred only if there is positive consumption along each dimension. Here, however, we make the strong assumption that costs are separable in the following sense:

\[
C(Q) = c_0 + c_1([Q]_1) + c_2([Q]_2) + \int_{Q^0} dC(\xi) ,
\]

where

\[
C(\xi) = C(\lfloor \xi \rfloor) .
\]

Assuming further that the cost function is appropriately differentiable, the Stieltjes integral in this formula can be represented as.
where $d\xi = d\xi_1 d\xi_2$ for Riemann integration in $\mathbb{R}^2$, and

$$\int_Q c(\xi) d\xi,$$

(2.3)

where $d\xi = d\xi_1 d\xi_2$ for Riemann integration in $\mathbb{R}^2$, and

$$c(\xi) = \frac{\partial^2 c(\xi)}{\partial \xi_1 \partial \xi_2}.$$

(2.4)

In spite of the seeming complexity of this representation it has a simple interpretation. The seller incurs a fixed cost $C_0$ and capacity costs $C_1$ and $C_2$ along the two dimensions, depending on the maximum requirements. In addition, a cost $c(\xi)$ is incurred for each little square centered on a point $\xi \in Q^0$. For example, if the two dimensions were to represent the rate and duration of usage of electricity, measured in kilowatts and hours, then the capacity costs reflect the requirements for equipment and manhours to meet the maximum demands, and the remainder might represent the cost of energy to supply the number of kilowatt-hours demanded (measured by the area of $Q$) and could reflect an increasing marginal cost of meeting the load at any one time, or extending the duration of a given load level (such as the base load).

We impose various convexity assumptions on the cost function in order to ensure an interior solution to the optimal pricing problem:

**Assumption 1:** The functions $C_1$, $C_2$, and $C$ are each assumed to be increasing, convex, and twice differentiable.

The seller's objective in choosing a payment schedule will be to maximize the difference between his payment and his costs, summed
over all customers. This assumes, of course, that the seller's costs are additively separable across customers so that there are constant returns to scale with the number of customers.\textsuperscript{3}

**Payment Schedules:** A basic motive for our adoption of nonlinear pricing is to adapt the payment schedule to the structural features of the cost function. Consequently, we allow that a payment schedule \( P \) is conformable to the cost function specified in (2.1):

\[
P(Q) = P_0 + P_1([Q]_1) + P_2([Q]_2) + \int_0^Q p(\xi) d\xi ,
\]

where one obtains as an implication that

\[
p(\xi) = \frac{\partial^2 P(\xi)}{\partial \xi_1 \partial \xi_2} ,
\]

for \( P(\xi) = P([\xi]) \). The other assumptions adopted in the formulation will be sufficient to ensure that the functions \( P_1, P_2, \) and \( P \) are nondecreasing and differentiable. A mild restriction on the maximal degree of concavity of the payment schedule will be imposed later.

**Customers:** The heterogeneity in the customer population has a central role in nonlinear pricing because the payment schedule is designed to induce self-selection among the customers. The different preferences among the customers are reflected in their purchases and realizing this a seller can achieve discrimination 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
The heterogeneity that we allow is restricted to an ordering of the customers' types along a single dimension. Taking the population of customers to be a continuum, one can in general allow the types to be indexed by a real-valued variable \( \tau \) and 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 \( t = F(\tau) \) that \( t \) is uniformly distributed on the unit interval.

As we did with the seller's costs, a key assumption imposed on each customer's utility function is that it is separable:

\[
U(Q,t) = \int_{Q'} u(\xi,t) d\xi ,
\]

where of course the marginal benefit \( u \) is determined by the utility function for rectangular purchases:

\[
u(\xi,t) = \frac{\partial^2 U(\xi,t)}{\partial \xi_1 \partial \xi_2} ,
\]

using \( U(\xi,t) = U([\xi],t) \). Separability imposes a limit on the customer's gains from substitution. In typical examples it means that the customer can choose to purchase more or less on either dimension, but direct substitution between "little squares" differing in both dimensions is excluded.

We impose strong regularity assumptions on the customers' utility functions:
Assumption 2: We assume that $U(x,t)$ is strictly increasing and concave in $x$ and thrice differentiable in $(x,t)$. Furthermore, we assume that the two partial derivatives $\partial U/\partial x_i$ are each concave in the other dimension ($x_j$ for $j \neq i$); that is, the marginal benefit from increasing one attribute flattens out as the other attribute is increased. Finally, we assume that $U(\cdot,t)$ 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.

As is evident from the specification, we presume that customers exhibit negligible income effects. Consequently, each customer $t$ selects that purchase set $Q(t)$ that achieves the maximum difference between his utility and the payment schedule:

\[ (2.9) \quad \max_{Q} U(Q,t) - P(Q) , \]

or he can choose not to make any purchase if this maximum is negative. A convenient way to express (2.9) uses the difference $[Q] - Q$ between the rectangular set $[Q]$ enclosing $Q$, and $Q$, since it is $[Q]$ that determines the parts of the payment dependent on capacity charges. Thus one can think of the customer purchasing $[Q]$ and then relinquishing the portion $[Q] - Q$:

\[ (2.10) \quad \max_{x,Q} U(x,t) - P(x) - \int_{[x]-\hat{Q}} (u(\xi,t) - p(\xi)) d\xi , \]

where the resulting purchase set is then $Q(t) = [x] \cap \hat{Q}$. 
The gist of the 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 hard to characterize in general. As in most studies of nonlinear pricing, however, it will turn out here that the optimal payment schedule has quite regular mathematical properties. As we have seen in (2.6) and (2.10) the payment schedule is wholly determined by its behavior on rectangular purchase sets, namely the function $P(x) = P([x])$, and similarly the customer's utility function is determined by the utility function $U(x,t) = U([x],t)$ for rectangular purchase sets. 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.5/ 

For rectangular purchase sets the customer's surplus is $S(x,t) = U(x,t) - P(x)$. For the following results we assume that Assumption 2 applies also to $S$, and to avoid trivial cases we assume that the first and cross-partial derivatives of $S$ at the origin are positive. According to (2.10) the customer $t$ selects $x$ and $\hat{Q}$ to maximize

$$S(x,t) = \int_{\hat{Q}}^{} S(\xi,t) d\xi,$$

where $s$ 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 three properties

$$(3.2) \quad \hat{Q} = \{\xi \in \mathbb{R}^2 \mid s(\xi,t) > 0\} \ ,$$

$$(3.3) \quad S^1(x_1,y_2,t) = S^2(y_1,x_2,t) = 0 \ ,$$

$$(3.4) \quad s(x_1,y_2,t) = s(y_1,x_2,t) = 0 \ ,$$

where (3.4) determines the point $y \in \hat{Q}$ and $S^1$ and $s^2$ are the partial derivatives of $S$ with respect to its first and second arguments.

**Proof:** The point $y$ is illustrated in Figure 1. We show first that $s(x,t) < 0$. 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 (3.1) is minimized by choosing the domain of integration to be the region where the integrand is nonnegative, thus verifying (3.2). Equations (3.3) and (3.4) then state the first-order necessary conditions for the choice of $x$ subject to this choice of $Q$. See Figure 2.

A customer who makes a purchase chooses the purchase set specified in Lemma 1. Those customers electing to purchase will be the ones with types satisfying $U(Q(t), t) \geq P(Q(t))$, meaning that their consumer surpluses are nonnegative. These types are characterized by the following results.

**Lemma 2:** The set of purchasers is an interval $T = [0, T]$ and the purchase sets are ordered by inclusion:

$$t > \tilde{t} \Rightarrow Q(t) \subset Q(\tilde{t}) \text{ and } [Q(t)] << [Q(\tilde{t})].$$

Further:

$$\frac{dS(t)}{dt} = \int_{Q(t)} \frac{\partial u(\xi, t)}{\partial t} d\xi < 0.$$  \hspace{1cm} (3.5)

**Proof:** The ordering of the purchase sets by inclusion is a simple consequence of the monotonicity assumptions imposed. Similarly,
\[ dS(t) = \partial Q S \cdot dQ + \partial S \cdot \partial t S = \partial_t S , \]

using Lemma 1, and then (3.5) follows from the monotonicity of \( u \).

Thus the set of purchasers is an initial segment of the unit interval.

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

\[ t(\xi) = \sup_\tau \{ \tau \mid \xi \in Q(\tau) \} . \]

One can then define \( t_1(x_1) \) so that \( t_1(x_1,y) \) 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

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

The marginal purchaser is identified by the condition \( S(T) = 0 \). 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.

**Optimal Payment Function:** The seller offers the same payment schedule to every customer and his profit from each one who purchases is the difference between the payment received and the cost incurred. Alternatively, his profit is the difference between the total surplus \( W(t) = U(Q(t),t) - C(Q(t)) \) and the consumer surplus \( S(t) \). Thus, the total profit can be expressed as
\[ \Pi(P) = \int_0^T \{ W(t) - S(t) \} dt, \]

depending on the payment function \( P \) for rectangular purchase sets (which in turn determines the full payment schedule \( P \)). Integrating by parts using (3.5) and calculating terms puts this objective into a more tractable form for analysis:

\[
\begin{align*}
\Pi(P) &= TW(T) - \int_0^T \{ W'(t) - \int Q(t) \frac{\partial u(\xi,t)}{\partial t} d\xi \} dt \\
&= TW(T) - \sum_{i=1,2} \int_{x_i(0)}^{x_i(T)} t_i(x_i) \cdot C_i(x_i) dx_i \\
&\quad + \int_{Q(0)}^{Q(T)} t(\xi) \cdot \{ u(\xi,t(\xi)) - c(\xi) \} d\xi \\
&= TW(T) - \sum_{i=1,2} \int_{x_i(0)}^{x_i(T)} t_i(x_i) \cdot C_i(x_i) dx_i \\
&\quad + \sum_{i=1,2} \int_{x_i(0)}^{x_i(T)} t_i(x_i) \cdot \int_0^{y_i(t_i(x_i))} \{ u(x_i,y,t_i(x_i)) \\ - c(x_i,y) \} dy dx_i \\
&\quad + \int_{Q^*} t(\xi) \cdot \{ u(\xi,t(\xi)) - c(\xi) \} d\xi \\
&= TW(T) + \sum_{i=1,2} \int_{x_i(0)}^{x_i(T)} t_i(x_i) \cdot \frac{3}{3x_i} S_i(x_i) dx_i \\
&\quad + \int_{Q^*} t(\xi) \cdot \{ u(\xi,t(\xi)) - c(\xi) \} d\xi.
\end{align*}
\]
where \( t(\xi) \) and \( t_1(x_1) \) are the corresponding inverses of the map \( t \to (\hat{Q}, x) \) from Lemma 1 which we will characterize below, and

\[
S_1(x_1) = U(x_1, y_1(t_1(x_1)), t_1(x_1)) - C(x_1, y_1(t_1(x_1)))
\]

The formula (3.7) has a natural interpretation that is worth emphasizing. The last term adds up the contribution to the total profit from each "little square" centered on a point \( \xi \) — each such square is included in the purchase set of a fraction \( t(\xi) \) of the potential customers if \( \xi \) is on the boundary of \( Q(t(\xi)) \). Similarly, the terms within the summation account for the capacity costs, i.e., the marginal cost of an increment of capacity along the \( i \)-th dimension is incurred for a fraction \( t_1(x_1) \) of the potential customers. 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 \( Q(t(\xi)) \), and again there is a unique customer index \( t_1(x_1) \) such that that customer selects the capacity \( x_1 \) along the \( i \)-th dimension. The construction ensures that \( t_1([Q(t(\xi))],_1) = t(\xi) \) if \( \xi \) is on the boundary of \( [\hat{Q}(t(\xi))] \). The "differential" associated with an incremental customer, which has the key role in the integration by parts in (3.6) and in (3.7) is shown in Figure 2.

**Theorem:** The optimal marginal price schedule is

\[
p(\xi) = u(\xi, t(\xi))
\]
where $t(\xi)$ solves
\begin{equation}
  t \cdot \frac{\partial u(\xi, t)}{\partial t} + u(\xi, t) = c(\xi) \quad .
\end{equation}

The optimal marginal capacity prices are determined from the conditions
\begin{equation}
  t \cdot \frac{\partial^2 u(x^i(t), t)}{\partial x_i \partial t} + \frac{\partial u(x^i(t), t)}{\partial x_i} = \frac{\partial c(x^i(t))}{\partial x_i} \quad ,
\end{equation}

where $x^i(t) = (x_1(t), y_2(t))$, and symmetrically for $x^2(t)$ as in (3.3) and (3.4), and $t = t(x^i)$. (Below we show how to compute the capacity charges.) The optimal market penetration is determined by the condition
\text{(3.11)} \quad T \cdot \frac{\partial U(Q(T), T)}{\partial t} + U(Q(T), T) = C(Q(T)) + \lambda ,

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

\textbf{Proof:} We use (3.6) and (3.7) 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 (3.7), which yields (3.9), and then (3.8) is implied by (3.2). Similarly, optimizing the integrand in the second term in (3.7) yields (3.10). 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 (3.3) yields the marginal capacity charges:

\text{(3.12)} \quad P_1'(x_1) = \frac{\partial U(x_1, y(x_1), t(x_1))}{\partial x_1} - \int_0^{y(x_1)} p(x_1, y) dy ,

and similarly for $P_2'$. The first-order necessary condition for the optimal market penetration $T$ is (3.11), obtained by differentiating with respect to $T$ in (3.6) after adding a Lagrangian term $\lambda \cdot \{1 - T\}$. One sees from (3.11) that

\text{(3.13)} \quad \lambda = \max \{0, \frac{\partial U(Q(T), T)}{\partial t} + U(Q(T), 1) - C(Q(T))\} ,
and if \( \lambda > 0 \) then \( T = 1 \). The fixed charge \( P_0 \) that achieves \( T \) is then constructed as

\[
(3.14) \quad P_0 = \lambda + u(Q(T), T) - \sum_{i=1,2} \int_0^{x_i(T)} p_i'(\xi_1) d\xi - \int_0^{Q(T)} p(\xi) d\xi.
\]

Note that in each of these constructions the seller first determines the customer's purchase set that the seller prefers, subject to the demand behavior imposed by Lemma 1. The payment schedule can then be 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.

Taken together these results yield an "algorithm" for constructing the optimal payment function \( P \) for rectangular purchase sets. In brief, one uses (3.9) to find which customer type is to have \( \xi \) on the boundary of its purchase set and then one uses (3.8) to set the marginal payments for "little squares" to obtain the desired result; similarly, (3.10) determines which customer type chooses each capacity level and then (3.12) sets the right marginal capacity charge to achieve this result; and lastly, (3.11) identifies the desired market penetration and then (3.14) specifies the fixed charge that achieves it. Specifying the various marginal charges is sufficient since the various 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 quite clearly the assignment of capacity costs to customers.

No further restrictions are placed on the cost function, but we assume that the customers' utility functions take the very special form

\[ U(x,t) = [k - t^a] \cdot U(x) \]

where \( a > 0 \) and \( k > 0 \).

**Proposition:** The optimal price schedule is determined by the price function

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

for rectangular purchases, where \( a = a/[1 + a] \) and

\[ \lambda = \max \{0, KU(\xi,1) - C(\xi)\} \]

for \( K = k - [1 + a] \). If \( k < 1 + a \) then \( \lambda = 0 \). This payment schedule induces the purchase sets

\[ Q(t) = \{\xi \in \mathbb{R}_+^2 \mid \xi \preceq x, \frac{c(\xi)}{U(\xi)} \leq k - [1 + a]t^a\} \]

where for each type \( t \) the selected capacities are determined by the conditions:

\[ \frac{u(x_1,y_2)}{c(x_1,y_2)} = \frac{U^1(x_1,y_2)}{C^1(x_1,y_2)} = \frac{u(y_1,x_2)}{c(y_1,x_2)} = \frac{U^2(y_1,x_2)}{C^2(y_1,x_2)} = k - [1 + a] \cdot t^a \]
where $U^t = \frac{\partial U}{\partial x_1}$ and similarly for $C^t$.

**Proof:** Express the optimal price function as

$$P(x) = P_0 + P_1(x_1) + P_2(x_2) + P_3(x)$$

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

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

To do this we employ the optimality condition (3.9) to obtain

$$[k - (1 + a)t^a] \cdot u(C) = c(E)$$

where $u$ is the cross partial of $U$. Solving this for $t$ yields $t(\xi)$. Then (3.8) yields

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

yielding the desired result. Next we show that

$$P_1(x_1) = [1 - a] \cdot C_1(x_1)$$

and similarly for $P_2$. To do this we invoke the optimality condition (3.10) and obtain

$$[k - (1 + a)t^a] \cdot \frac{\partial U}{\partial x_1} = \frac{\partial C}{\partial x_1}$$

or equivalently
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, (3.12) yields, at \((x,t) = (x_1,y_2(x_1),t_1(x_1))\):

\[
P'_1(x_1) = \left[k - t_1(x_1)^a\right] \cdot \frac{\partial U}{\partial x_1} = \int_{0^+}^{y_2(x_1)} \left(ak + [l - a]c\right) dy
\]

\[
= ak \cdot \frac{\partial U}{\partial x_1} + [l - a] \cdot \frac{\partial C}{\partial x_1} - \left(ak \cdot \frac{\partial U}{\partial x_1} + [l - a] \cdot \frac{\partial C}{\partial x_1}\right)_{y_2(x_1)}
\]

\[
= ak \cdot \frac{\partial U}{\partial x_1} (x_1,0^+) + [l - a] \cdot \frac{\partial C}{\partial x_1} (x_1,0^+)
\]

\[
= [l - a] \cdot C'_1(x_1)
\]

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 (2.7).) The form of the purchase set follows directly from the derivations so far. The last step is to establish that

\[
P_0 = [l - a] \cdot [C_0 + \lambda]
\]

For this we apply the optimality condition (3.11) which takes the form

\[
[k - (1 + a)t^a] \cdot U(Q(T)) = C(Q(T)) + \lambda
\]
or equivalently,
\[ [k - T^0] \cdot U(Q(T)) = ak \cdot U(Q(T)) + [l - a] \cdot \{ C(Q(T)) + \lambda \} \]

with
\[ \lambda = \max \{ 0, (k - [l + a]) \cdot U(Q(1)) - C(Q(1)) \} \]

where \( U(Q) \) is defined analogously to (2.7). Using this condition in (3.14) yields
\[ P_0 = ak \cdot U(Q(T)) + [l - a] \cdot \{ C(Q(T)) + \lambda \} = [l - a] \cdot [\lambda + C_0] \]
as required. Clearly, if \( k < l + a \) then \( \lambda = 0 \) and \( T \) will be an interior solution, namely \( T < l \).

For purposes of discussion we will assume that \( k = l \) so that it is assured that \( \lambda = 0 \). The simple 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 a fraction \( 1 - a \) of the seller's actual costs, 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. To our knowledge the literature includes three approaches. First there is the view that fixed and capacity-dependent charges are costs of doing business and that if a pricing policy based on demand and marginal-cost considerations yields enough net revenue to cover these costs then the firm will remain profitable and stay in business. A second view assumes U-shaped average cost curves (ignoring evidence to the contrary in many cases) and envisages entry sufficient to equate marginal and average cost. Associated with this view is an ambiguous distinction between short-run costs and long-run costs (presumably fixed and capacity costs are included in the latter), but we have not found any analysis that implies the same distinction will be manifested in pricing policies. The third view is that industries with declining average costs are natural monopolies that are appropriately analyzed as regulated or state-managed firms. The large literature on Ramsey prices and related approaches is aimed primarily at assigning charges to customers (in the case of Ramsey prices, on the basis of demand elasticities) that will be sufficient in the aggregate to cover fixed and capacity costs. In some regulated industries, such as the electric power industry, explicit charges (called "demand" charges, as opposed to "energy" charges for variable costs) are made for capacity costs based on the theory of peakload pricing, and in recent years pricing policies have
been adopted that discriminate among customers based on their usage patterns (for example, in California prices for industrial customers depend upon the time-of-use and the maximum kilowatt rate in a period).\(^8\)

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 somewhat to the cost schedule, as in (2.5), so that the formulation incorporates the seller's opportunities to invoke self-selection among heterogeneous customers who do not otherwise directly value the capacity services provided, as reflected in the omission of capacity benefits in (2.7), which we have done partly to emphasize this point.\(^2\)

From a technical viewpoint our work addresses the case of a firm with differentiated products sold in bundles, and a discontinuous cost function. 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 their customers.
Footnotes

1/ The discontinuity is in terms of the Hausdorff topology on compact sets in $\mathbb{R}$.  

2/ A set $Q$ is comprehensive if $x \preceq \hat{x}$ and $\hat{x} \in Q$ imply $x \in Q$.

3/ This is one way in which the problem we address does not encompass the traditional peakload pricing problem in its full complexity.

4/ For expositions of the basic ideas underlying nonlinear pricing, see Goldman, Leland, and Sibley [1977], Mirman and Sibley [1980], Roberts [1979], or Spence [1977]. However the form that the discrimination takes in the example studied in Section 4 is somewhat different.

5/ In one-dimensional problems the sufficient condition is a version of [A2]; see Goldman, Leland, and Sibley [1977].

6/ The integrand would be $(1 + \mu)W - \mu \cdot S$ in a welfare problem subject to the constraint $\Pi > \Pi'$ for which $\mu$ is a Lagrange multiplier; here we take $\mu = \infty$ but the methods employed do not depend on this restriction. See Mirman and Sibley [1980]. Roberts [1979] studies the welfare aspects of nonlinear pricing in great detail.

7/ Although (3.10) may appear mysterious at first, one can see from Figure 2 that it arises from considering increments in capacity treated as incremental slices on the sides of the purchase set of a customer. As we have expressed it in (3.10) the operation is actually reversed: one considers which customer type is best assigned to each capacity.

8/ We hope to address such pricing problems in subsequent work. An adequate formulation must include the features that the seller's costs depend on the aggregate load-duration curve, different customers may have their peakloads occurring at different times in a cycle, and customers have greater opportunities for substitution than we have allowed here.

9/ Although here we have confined attention to the monopoly case, one can envision that this approach might ultimately provide some alternative justification for the cost-based pricing policies prevalent among regulated public utilities.
References


REPORTS IN THIS SERIES


Reports in this Series


377. "The Role of Reputation in a Repeated Agency Problem Involving Information Transmission" by W.P. Rogerson

378. "Unemployment Equilibrium with Stochastic Rationing of Supplies" by Ho-mou Wu.


380. "On the NTU value" by R.J. Aumann.


382. "Informational Equilibrium" by Robert Kast.

383. "Cooperative Oligopoly Equilibrium" by Mordecai Kurz.


386. "Capacity Pricing" by S. Oren, S. Smith and R. Wilson