Multilevel demand subscription pricing for electric power

Hung-po Chao, Shmuel S. Oren, Stephen A. Smith and Robert B. Wilson

This paper investigates a general demand subscription pricing policy for electric power service, ie a menu of service contracts for assigning different interruption probabilities and prices to different load levels. This interruptibility feature is combined with a non-linear energy price schedule to encourage load flattening. We derive a customer value model and a weighted sum objective function for the supplier and consumers for selecting optimal price structures of this type. The general form of the optimal price policy is derived and discussed, and the implementation methodology is outlined. An illustrative example is solved in detail.

Keywords: Electric power; Optimal pricing; Demand model

Treating electric power service as a family of differentiated products is an approach that has been proposed and tested in a variety of contexts. The economic bases for such differentiation are the variation in the marginal cost of generation and the heterogeneity of customer preferences. Product differentiation with correspondingly adjusted prices can benefit both the consumer and the producer by matching service options more closely to customer preferences and by allowing more efficient use of generation capacity. Recently, the increased cost of new generation capacity and the availability of metering options using microprocessor technology have created new opportunities for which electric power product differentiation is both feasible and desirable. In this introduction, we briefly discuss some previously developed product differentiation methodologies. We then propose and develop a new approach for pricing interruptible service that is designed to exploit some new product differentiation options.

Time-of-use rates

Time-of-use (TOU) rates, which constitute the most common form of product differentiation in the electric power industry, have been a primary focus in the economics literature on electric power pricing. TOU rates are set higher for electric power consumed during peak demand to reflect the higher marginal costs of peak load generation, which thus induces customers to shift their consumption to off-peak times. Research by Steiner [14], Boiteux [2] and Williamson [17] was followed by a series of studies and experiments in the 1970s. The book by Mitchell, Manning, and Acton [9] contains a comprehensive survey of this work, as well as a review of European experiences in implementing TOU pricing and interruptible service rates. The combined problem of optimal TOU pricing and generation capacity planning under demand-and-supply uncertainty has been addressed in a recent paper by Chao [3]. Considerable work has also been done recently on 'spot pricing' mechanisms, in which

Hung-po Chao is with the Energy Analysis and Environment Division, Electric Power Research Institute, 3412 Hill View Avenue, PO Box 10412, Palo Alto, CA 94303. Shmuel Oren is with the Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720. Stephen Smith is with the Leavey School of Business, Santa Clara University, Kenna 201, Santa Clara, CA 95053. Robert Wilson is with the Graduate School of Business Administration, Stanford University, Stanford, CA 94043, USA.

This research was supported financially by the Electric Power Research Institute. Professor Oren's research was partially supported by University of California Energy Research Grant UER-CS-45-18. Professor Wilson's research was partially supported by National Science Foundation Grant PRA-83-142-14. The views expressed in this paper are the authors' and do not necessarily represent the views of the sponsors.

Final manuscript received 12 August 1985.

TOU rates are determined on-line so as to balance supply and demand (see Schweppe, Tabor and Kirtley [13]).

Incentive-based load management

Differential rates based on interruptibility of service, load factor, peak load, voltage, etc, have also been quite common, primarily for large commercial users. However, there have been few economic analyses of such pricing policies. Direct load management options, which allow an electric utility to curtail selectively electric power service in case of a shortage, can reduce the required reserve capacity and hence decrease capital cost. It is, therefore, beneficial to differentiate electric power service based on load management contingencies included in the service contract and to offer incentives in the form of lower rates to customers willing to accept such contingencies.

One of the simplest forms of incentive-based load management is the installation of power limiters (fuses) at customers' sites, accompanied by a capacity charge determined by the fuse size. A similar effect can be achieved through self-rationing by simply imposing a capacity charge based on the customer's peak load over a billing period. Panzar and Sibley [12] analysed such a self-rationing approach when peak capacity and energy use are priced linearly. A more general case involving jointly non-linear pricing of capacity and use has been analysed in a recent paper by Oren, Smith and Wilson [11]. Self-rationing can be easily implemented, but it may curtail consumption even when generation capacity is under-utilized.

A more powerful form of load management is the interruptible service contract, which (in case of a shortage) allows the utility to curtail a customer's power supply (with some advance warning) to a prenegotiated level. Such service agreements are quite common in the USA and abroad among large industrial customers and institutions. Some economic aspects of interruptible service contracts have been analysed by Marchand [8] and by Tschirhart and Jen [15]. Also, relevant to this topic is the work by Harris and Raviv [6] in analysing optimal pricing of supply priorities.

Special incentive programmes for residential users

The general applicability of product differentiation based on delivery conditions in the low-demand portion of the electric power market has been limited until recently by the technology and cost of metering and load control devices. This barrier is disappearing with advancements in microelectronics, which make detailed metering and control of residential loads technologically and economically feasible. (For a survey of recent advances in electronic metering technology, see Gorzelnik [5].) Many utilities now offer residential customers interruptible service for specific devices such as water heaters and pool pumps. Experimental programmes are also under way offering residential customers TOU rates and incentives for accepting various forms of load control contingencies. Demand Subscription Service, a programme offered by Southern California Edison (SCE), allows customers to subscribe for a firm level of power and receive a credit proportional to the difference between this level and their estimated normal peak load. The utility is then allowed, upon a short warning, to limit the customer's power to the subscribed level for up to six hours as many as 15 times per year. This is done remotely by activating a fuse at the customer's site through a radio signal.

Another experimental, incentive-based load management programme currently implemented at SCE involves 'cycling' of residential air conditioners. A device attached to the air conditioner allows the utility to interrupt it, remotely, for a fixed fraction of the time. In exchange, the customer receives a credit depending upon the fraction of time he is willing to be interrupted and proportional to the capacity of his air conditioner.

Rate structures analysed

The optimal product differentiation programme for electric power service requires pricing mechanisms that are easily implemented, that exploit the heterogeneity of customers' preferences, and that are consistent with production cost and technological constraints. For example, the distribution of customers' choices of service reliability must be compatible with the overall reliability of the generation system. Implementation of such policies at the low end of the electric power market, ie residential and small commercial customers, must also rely on self-selection opportunities, since more direct price discrimination would impose unrealistic information requirements.

In this paper we define and analyse a rate structure that differentiates electric power according to both load pattern and service reliability. This rate structure is a generalization of Demand Subscription Service in that it offers multiple levels of interruptible service. The customer can choose increasing probability of interruption for successive increments of his load. The charge for each load increment depends upon its selected interruptibility and the length of time it is utilized. This pricing, like TOU rates, induces customers to flatten their load pattern over time. In addition, it spurs customers to select lower reliability levels for the peak portion of their load. This price structure is similar in some respects to the quality-differentiated tariff discussed by Oren, Smith and Wilson [10], with quality level interpreted as service reliability. Here, however, there is no fixed charge, and the tariff may vary non-linearly with usage.

Demand-layered vs TOU pricing

Demand fluctuations over a time period (eg day, week, month, year) are conveniently represented by a 'load duration curve' $L(t)$, which is the load having duration t or less for each value of t. Normally, t is expressed as a fraction of the total time period; however, for clarity of our subsequent discussion we assume it is measured in hours. This representation is commonly used to describe the generation cost as a function of duration and the optimal mix of generation technologies.

Cost structure

Alternative generation technologies are typically characterized in terms of their capacity cost F_i/kW and energy cost V_i/kWh. Thus, the energy cost for technology i as a function of duration t is given by the function

$$c_i(t) = F_i + V_i t \tag{1}$$

When a diverse set of technologies is available, the minimum cost function is given by the lower envelope of the graphs of the two-part cost functions, as shown in Figure 1. The breakpoints delineate the efficient duration range of each technology. The optimal capacity of each technology can then be determined by projecting these ranges onto the load duration curve (Figure 1).

Load slice approach

In calculating long-run generation cost, one typically assumes an optimal technology mix. A convenient way to represent this cost is to view the load duration curve as a 'stack' of horizontal load slices operating for different durations, as illustrated in Figure 1. The generation cost for a load slice of duration t is thus

$$c(t) = \min_i c_i(t) \tag{2}$$

Hence, the total generation cost for the load duration curve $L(t)$ is

$$C[L(t)] = \int_0^{L(0)} c(t(L))dL \tag{3}$$

where $t(L)$ is the inverse function of $L(t)$, defined as

$$t(L) = \max\{t \mid L(t) \geq L\} \tag{4}$$

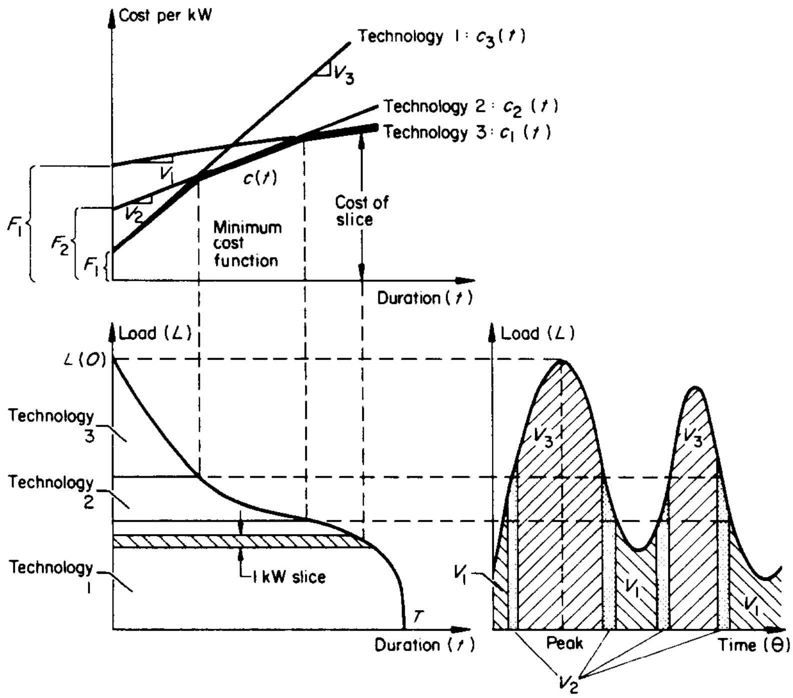

Figure 1. Load slice versus TOU costing with optimal technology mix.

Time slice approach

An alternative representation of the generation cost is obtained by viewing the load duration curve as an array of vertical time slices representing the load in each time interval. The time slice cost formula follows from integrating Equation (3) by parts, which yields

$$C[L(t)] = c(0^+)L(0) + \int_{0^-}^{t(0)} c'(t)L(t)dt \qquad (5)$$

Here $c'(t)$ is the marginal energy cost of the most efficient technology per kWh for the time slice $[t, t+dt]$. An additional capacity cost $c(0^+)$ per kW, which equals the capacity cost of the peaking technology, is assigned to the load $L(0)$ occurring during the system's peak.

Implications for pricing

The two representations of the generation cost described above suggest two alternative pricing approaches. The time slice approach corresponds to the common TOU pricing combined with a demand charge. An ideal cost-based pricing policy of this type will impose a TOU energy rate that equals the marginal energy cost at that time, plus a demand charge equal to the marginal capacity cost on every kW used during the system's peak. Such a scheme exactly recovers generation cost and is best. However, predetermining the chronology of the different rating periods and the time of the peak (which is necessary for implementation) requires a perfect forecast of the equilibrium load pattern. Such a forecast would have to anticipate load shifts in response to the rate structure, which would be exceedingly difficult when demands in different time slices are interdependent. Main [7] discusses such difficulties.

The load slice approach suggests a pricing policy, often referred to as demand-layered (DL) pricing, that imposes on every load slice a tariff tthat is non-linear in duration. In particular, this non-linear tariff could be the minimum cost function $c(t)$ mentioned earlier. In such a pricing scheme, each customer's payment is based on his individual load duration curve. The volume discount on duration induces customers to flatten their load duration curves. Demand-layered pricing and its comparison to periodic pricing of electric power is the subject of a PhD thesis by Main [7]. A more recent article on this topic is by Ballonoff [1]. Oren, Wilson and Smith [11] present a theoretical analysis of a related pricing scheme in which price varies non-linearly in both capacity and usage.

The main disadvantage of DL pricing is that it does not account for complementarities in the customer load patterns. Hence, it may be socially inefficient in the sense that it does not equate marginal benefit of consumption across users at every point in time. This is because the marginal price faced by different customers at any time depends on their entire load pattern which thus allows Pareto improvements through redistribution of power. This shortcoming should be weighted, however, against the fact that DL pricing requires only a forecast of the system's equilibrium load duration curve (rather than the chronological load pattern), which is not susceptible to peak shifting problems. We will demonstrate that if customers' load patterns are synchronous, then DL pricing will be as efficient as TOU pricing. Assumptions for customer preference structures that yield such synchronization are presented later.

In this paper we adopt the demand-layered approach partly because it provides a natural framework for differentiation of electric power service according to reliability. In this framework, each load increment (or slice) can be assigned a level of reliability and the non-linear tariff for a load slice can be varied according to its service reliability. Thus, a customer's total charge will be determined by his load duration curve and his choice of service reliability for each load increment.

The demand model and customer self-selection

We consider a price plan that offers the option of selecting a different duration time and reliability level for each load slice. The general form of the price plan is

$p(r, t)$ = total charge for a load slice of duration t at specified reliability level r

Each load slice is priced independently, in the sense that its price depends only upon the duration t and the reliability r selected for it. This type of pricing can be non-linear in duration and reliability, but it is linear in capacity, since each load slice of duration t and reliability r is priced the same. The substitution of longer duration consumption for peaked, short duration consumption can be encouraged, for example, by a price plan in which

$$p_t(r, t) > 0 \quad \text{and} \quad p_{tt}(r, t) < 0$$

where the subscripts denote partial derivatives. That is, the marginal charge for duration time is decreasing.

The reliability r for a load slice is defined as follows:

r = the long-run average fraction of a load slice that will be serviced when reliability r is selected

Consumer preference model

Aggregate consumer response to price is defined by the function $L(p, t)$, which describes a family of system load duration curves, parametric on a uniform price p/kWh. Alternatively, the inverse function $t(L, p)$, defined as in Equation (4) for any value of p, may be interpreted as a family of demand functions for energy, parametric on the load level L. The peak load pricing literature often characterizes demand in terms of a function $L^*(\theta, p)$, which gives the load demanded at time θ as a function of price p/kWh. The duration demand function $t(L, p)$ can be related to $L^*(\theta, p)$ as follows:

$$t(L,p)=M\{\theta|L^*(\theta,p)\geq L\} \qquad (6)$$

where $M\{\Omega\}$ denotes the Lebesgue measure of the set Ω. We assume that consumption decisions for different load slices are independent, but that the duration and reliability for a particular load slice are determined by a single decision entity (one user or a coalition of users coordinating their consumption). This assumption is analogous to that of independent time slices in the TOU pricing literature.

Given this assumption, the inverse of $L(p, t)$ with respect to p may be interpreted as a marginal value measure for the load slice at load level L. Thus, if we let

$v(L, t)$ = the maximum willingness to pay for the load slice at level L when its duration is t

then $\partial v(L, t)/\partial t$, denoted $v_t(L, t)$, is the inverse of $L(p, t)$ with respect to p.

Assuming no satiation, we have that v_t is positive. Because the values of the various load slices are independent and $L_t(p, t) < 0$, it must also be the case that lower load slices (ie those corresponding to smaller values of L) are more valuable than higher load slices. That is, for any fixed price p, longer durations are assigned to lower L values for the load duration curve. Thus, we have $v_L < 0$ for all L and t. We assume that value functions corresponding to different values of L do not cross, or equivalently, that the demand curves do not cross. This implies that $v_{Lt} < 0$.

Consumer surplus formula

The consumer surplus depends explicitly upon how random fluctuations in demand are modelled. If the load cycle represents the 24-hour day, for example, the load duration curve $L(t)$ would actually be an average of many daily load duration curves, observed under a variety of conditions. We analyse the effects of random fluctuations about the average load curve by assuming that there is a scaling function $h(w)$ for which

$h(w)L(t)$ = the actual system load duration curve under the random conditions w, where $\int h(w)dw = 1$

It is natural to expect that the likelihood of supply interruption would be dependent upon the random conditions w as well. If w is determined by temperature during the summer months, for example, both $h(w)$ and the likelihood of supply interruption would increase and decrease together. We assume that w is a one-dimensional variable, uniformly distributed between zero and one, that has been defined such that both $h(w)$ and the likelihood of a supply interruption increase monotonically with w. This rescaling can be achieved without loss of generality for one-dimensional monotonic variables. (If temperature T has a cumulative probability distribution $M(x)$, for example, the random variable $w = M(T)$ is uniformly distributed between zero and one.) We then define the indicator random variable $R(r, w)$ as follows:

$$R(r,w)=\begin{cases}1 \text{ if } w\leq r\\ 0 \text{ otherwise}\end{cases} \qquad (7)$$

The thickness dL of the load slice in Figure 1 is viewed as an average value. Under the random conditions w, each load slice is scaled as follows:

$h(w)R(r, w)dL$ = the thickness of a load slice under the random conditions w

Thus, L serves as an index of the same load slice in all circumstances but is equal to the height of the load slice only for the average system load duration curve.

This model assumes, in effect, that the service for a particular load slice is either off or on (depending on the value of R), and that the thickness of the load slice is scaled to account for the random demand fluctuations. Because of the definition of w, the set of events $w \leq r$ corresponds exactly to receiving service with probability r. The product of $h(w)$ and $R(r, w)$ introduces a correlation that can be used to reflect the increased likelihood of supply interruptions occurring during random conditions that produce unusually high demand.

Optimal consumption selection

If we assume that income effects can be neglected, the consumer's optimal choices of r and t for each load slice are determined independently so as to maximize expected consumer surplus. The expected consumer surplus is given by the integral over w, which simplifies to

$$S(r,t,L)=[v(L,t)-p(r,t)]H(r) \quad (8)$$

where

$$H(r)=\int_0^r h(w)dw, \text{ with } H(0)=0 \text{ and } H(1)=1$$

The price function $p(r,t)$ is assumed to be twice continuously differentiable except for a discontinuous jump upward at $t=0^+$ to capture capacity charges. It is also assumed that $v(L,0^+)=0$ and $p(0,t)=p(r,0)=0$. The optimal pair $\{r(L),t(L)\}$ is determined by solving the maximization problem

$$\max_{\substack{0\leqslant r\leqslant 1\\ 0\leqslant t\leqslant T}} \{S(r,t,L)=H(r)[v(L,t)-p(r,t)]\} \quad (9)$$

In addition to the constraints on r and t in Equation (9), the reliability options available to the consumer may be constrained further as a result of technological considerations. These are introduced explicitly later when we consider the supplier's optimal pricing problem. The supplier's optimal price function automatically induces consumer choices that satisfy the constraints imposed by technological considerations. This is a practical necessity, because the deliverable reliability options are determined by the aggregate system load curve and cannot be specified in terms of an individual customers' load duration curve.

A solution method for the consumer is now proposed so that the constraints $t\geqslant 0$ and $r\geqslant 0$ are not binding in the above optimization. Basically, the solution begins with the load slice $L=0$, which is assumed to satisfy $r(0)$, $t(0)>0$, and shows, by continuously increasing L, that the non-negativity constraints on r and t do not become active as long as the consumer surplus remains positive.

Lemma 1. The solution of the first-order necessary condition $S_r=0$ or ($S_r>0$ and $r=1$) can never have $r\geqslant 0$, $t=0$, or $t>0$, $r=0$.

Proof. If $t=0$, we have, using $v(L,0)=0$, that

$$S_r(r,0,L)=-h(r)p(r,0^+)-H(r)p_r(r,0^+)\geqslant 0$$

Since the sum of the terms on the left is negative for any $r\geqslant 0$, there is a contradiction and $t\neq 0$. Similarly, if $r=0$, we have, using $H(0)=0$ and $p(0,t)=0$, that

$$S_r(0,t,L)=0=h(0)v(L,t)>0 \text{ for all } t>0$$

which again is a contradiction, since in this case the first-order necessary conditions require that $r=1$. QED.

The constraints $r\leqslant 1$ and $t\leqslant T$ may be active, however. Thus the necessary conditions for a maximum of Equation (9) may be written as

$$S_r(r,t,L)=0 \quad \text{or} \quad \{r=1 \quad \text{and} \quad S_r(1,t,L)>0\} \quad (10)$$

and

$$S_t(r,t,L)=0 \quad \text{or} \quad \{t=T \quad \text{and} \quad S_t(r,T,L)>0\} \quad (11)$$

where

$$S_r(r,t)=h(r)v(L,t)-\{H(r)p(r,t)\}_r \quad (12)$$

and

$$S_t(r,t)=H(r)\{v_t(L,t)-p_t(r,t)\} \quad (13)$$

From Equations (10) and (11), it can be seen that any solution $r(L)$, $t(L)$ of Equations (10) and (11) produces a non-negative consumer surplus, since

$$h(r)\{v(L,t)-p(r,t)\}-H(r)p_r(r,t)=0 \quad (14)$$

or, multiplying through by $H(r)/h(r)$, we have

$$S(r,t,L)=[H(r)^2/h(r)]p_r(r,t)\geqslant 0 \quad (15)$$

Furthermore, the surplus for each load slice L with $r(L)$ and $t(L)$ optimally selected is monotonically decreasing in L. To demonstrate this, we define $S^*(L)=S[r(L),t(L),L]$ to be the surplus evaluated along the trajectories determined by Equations (10) and (11). We assume that $S^*(0)>0$ to avoid degeneracy. By the chain rule, we have

$$dS^*(L)/dL=r'S_r+t'S_t+H(r(L))v_L(L,t(L))<0$$

since the first two terms are zero by Equations (10) and (11) and v_L is assumed to be negative for all values of t and L. Thus, there is at most one value L_0 such that $S^*(L_0)=0$. The solution for the consumer would thus proceed monotonically in L until L_0 is reached, at which point it is optimal to consume no further load slices. Thus, for $L\leqslant L_0$, the internal conditions (10) and (11) determine the maximum of Equation (9). For $L>L_0$, 'no consumption' is specified by selecting $t=0$. The parameter L_0 is determined from $S^*(L_0)=0$.

The modified load duration curve

The average load duration curve for the system implied by the selection $t(L)$ is its inverse $L(t)$. However, the actual average system load curve will be truncated at the top as a result of the implementation of the interruptible service options. The resulting truncated curve can be determined analytically by integrating over all load slices under random conditions w and with reliability selection $r(L)$. This gives

$$\bar{L}(t)=\int_0^{L(t)}\int_0^1 h(w)R(w,r(L))dwdL$$

$$=\int_0^{L(t)} H(r(L))dL \qquad (16)$$

Regularity conditions

It is necessary for practical reasons that $t(L)$ and $r(L)$ be monotonically decreasing functions of L in the optimal solution. Since $t(L)$ is the inverse of the consumer's load duration curve, $t'(L)\leqslant 0$ is required by definition. Furthermore, since partial interruption of electric power service reduces the load from the top down, reliability selections must also decrease monotonically with load level L. Fortunately, both of these conditions can be guaranteed 'for free' with the price structure $p(r,t)$ that we develop for the supplier. It is shown in Lemma 3 that, without loss of generality, the supplier's price function can be specified as an additively separable function $p(r,t)=g(r)+f(t)$. Lemma 2 shows that for this form of the price function, the monotonicity of $t(L)$ and $r(L)$ are equivalent to the second-order necessary conditions for the solution of Equation (9).

Lemma 2. If the function $p(r,t)$ is additively separable, ie $p(r,t)=g(r)+f(t)$, then the conditions $r'(L)<0$ and $t'(L)<0$ guarantee that second-order necessary conditions for an interior solution to Equation (9) are satisfied. Furthermore, if $r'(L)$ and $t'(L)$ are also bounded from below, then the second-order sufficiency conditions for Equation (9) are satisfied.

Proof. By Equations (10)–(13), an interior solution $\{r(L),t(L)\}$ must satisfy

$$S_r(r,t)=h(r)v(L,t)-\{H(r)p(r,t)\}_r=0 \qquad (17)$$

and

$$S_t(r,t)=H(r)\{v_t(L,t)-p_t(r,t)\}=0 \qquad (18)$$

At a point satisfying Equations (17) and (18), the Hessian $\nabla^2 S$ is given by

$$\nabla^2 S=\begin{vmatrix} h'v-\{Hp\}_{rr} & -Hp_{rt} \\ -Hp_{rt} & H(v_{tt}-p_{tt}) \end{vmatrix} \qquad (19)$$

Taking the total derivative of Equations (17) and (18) with respect to L yields the vector equation

$$\nabla^2 S=\begin{vmatrix} r' \\ t' \end{vmatrix}+\begin{vmatrix} hv_L \\ Hv_{tL} \end{vmatrix}=0 \qquad (20)$$

When $p(r,t)$ is additively separable, then $p_{rt}=0$ and thus $\nabla^2 S$ becomes diagonal. Hence, Equation (20) can be written as

$$\{Hp\}_{rr}r'=hv_L \qquad (21)$$

$$(p_{tt}-v_{tt})Ht'=Hv_{Lt} \qquad (22)$$

Since $v_{Lt}<0$ and $v_L<0$, the terms on the right in Equations (21) and (22) are strictly negative (for strictly negative r' and t'). It thus follows that if r' and t' are negative, then the diagonal elements of the Hessian $\nabla^2 S$ must be non-positive. Since the Hessian is diagonal, this implies its negative semidefiniteness. When r' and t' are bounded below (ie no jumps in $r(L)$ and $t(L)$), these diagonal terms are strictly negative and hence the Hessian is negative definite, which satisfies the sufficiency conditions for a maximum. QED.

To guarantee both the required monotonicity and the second-order conditions for Equation (9), we require in the case of a separable price function that $v_{tt}-p_{tt}<0$ and $\{Hp\}_{rr}<0$. The first condition is equivalent to the common assumption in the non-

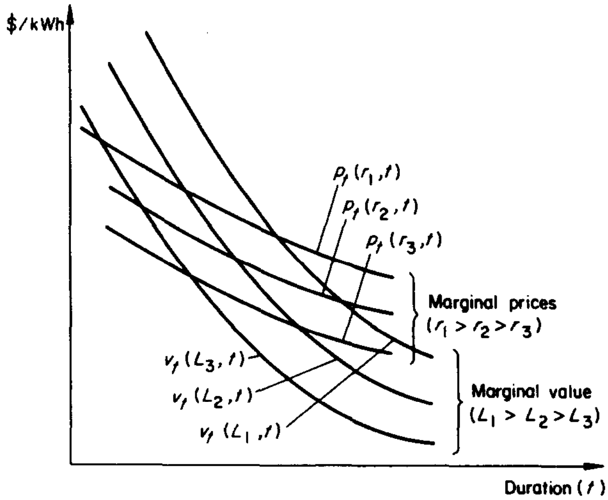

Figure 2. Marginal price and marginal value functions for duration.

linear pricing literature (eg Goldman, Leland and Sibley [4]) that the marginal price function should cross each marginal value function exactly once from below (see Figure 2). The second condition states that the marginal price *paid* by the consumer should be a decreasing function of r. This can be argued intuitively, if reliability is valued only for its ability to increase the probability of consumption of each given load slice.

The primary assumptions required for the consumer's preference structure can be summarized as follows:

- Random scaling of demand and random fluctuations in reliability, applied uniformly to all load slices.
- Independently optimized decisions on load slice consumption.
- Non-crossing marginal demand curves, each intersected at most once from below by the marginal price function.
- Decreasing marginal payments for reliability.

The supplier's problem

We now consider the problem of choosing optimal prices to be charged by an electric utility. For practical reasons, pricing cannot discriminate directly on the basis of the identity of a particular load slice. However, through pricing that introduces customer self-selection of both duration time and reliability for each load slice, the price plan can be designed to encourage both flattening of the system load duration curve and the selection of lower reliability service options for peak demand.

Pricing objective function and constraints

The supplier's objective is defined as the maximization of a weighted sum of total surplus and net revenue. This includes as special cases net revenue and total surplus maximization as well as the maximization of total surplus, subject to a net revenue constraint. An important element in this objective function is the production cost. It is assumed that the generation system has a fixed capacity configuration, and hence capacity costs are treated as sunk costs, which appear in the net revenue constraint. The optimized operating cost, on the other hand, is assumed to be attributable to individual load slices through a function $c(r, t, L)$, the average cost per kW for a load slice of type L with duration t and service reliability r. An explicit formula for $c(r, t, L)$ illustrating this assumption is derived in the Appendix.

In maximizing his objective function, the supplier must take into account the relationship between his pricing policy and the customer's selection of $t(L)$ and $r(L)$ based on Equation (9). The monotonicity constraints $t'(L)<0$ and $r'(L)<0$ will be enforced, which guarantees that the solutions of the customer's first-order conditions satisfy the second-order conditions for Equation (9). In addition to the self-selection constraints, the supplier may have further restrictions on the reliability options due to technological considerations. For example, a fixed total generation capacity Y may limit the deliverable reliability for a load slice at level L because of the requirement that $h(r)L \leqslant Y$. This would imply a shifting upper bound on deliverable reliability of the form

$$r(L) \leqslant R(L) = \min\{1, h^{-1}(Y/L)\}$$

Note that in this definition, $R(L)$ is monotonically decreasing in L. We include the constraint $r(L) \leqslant R(L)$ explicitly in the problem, with the understanding that it may express other reliability considerations as well.

The supplier's optimization problem may then be expressed as follows:

$$\max_{\substack{p(r,t), r(L)\\ t(L), L_0}} \int_0^{L_0} [H(r)v(L,t) - c(r,t,L) + a\{H(r)p(r,t) - c(r,t,L)\}]dL \tag{23}$$

subject to $r = r(L)$ and $t = t(L)$ satisfying

$$H(r)[v_t(t, L) - p_t(r, t)] - \lambda(L) = 0$$

with

$$\lambda(L) \geqslant 0,\ t \leqslant T, \text{ and } \lambda(L)[t - T] = 0 \tag{24}$$

$$h(r)v(t, L) - \{H(r)p(r, t)\}_r - \mu(L) = 0$$

with

$$\mu(L) \geqslant 0,\ r \leqslant 1, \text{ and } \mu(L)[r - 1] = 0 \tag{25}$$

$$H(r)\{v(t, L_0) - p(r, t)\}\Big|_{\substack{r = r(L_0)\\ t = t(L_0)}} = 0 \tag{26}$$

$$r \leqslant R(L),\ t \leqslant T \tag{27}$$

$$r'(L) < 0,\ t'(L) < 0 \tag{28}$$

In this formulation, a is the weighting parameter for net revenue. If we wish to maximize total surplus subject to a new revenue constraint, then a is the Lagrange multiplier, which needs to be determined endogenously to satisfy the constraint. Equations (24) and (25) are equivalent to the first-order conditions for consumer surplus maximization. We have introduced slack variables $\lambda(L)$ and $\mu(L)$ to account for the boundary conditions $t=T, S_t>0$, and $r=1, S_r>0$. The monotonicity constraints and the upper bounds on $r(L)$ and $t(L)$ have been included as well.

The optimal induced consumer behaviour

We obtain the solution to the above problem by using Equations (24)–(26) to eliminate $p(r,t)$ from the objective function. This reduces the supplier's problem to one of determining the optimal selection $r(L)$, $t(L)$ for each load level L and the cutoff level L_0. A price function $p(r,t)$ that will induce these optimal selections can then be determined through the self-selection constraints Equations (24)–(26). The mechanics are as follows.

If we integrate the term apH in Equation (23) by parts, we obtain

$$aHpL\Big|_0^{L_0} - \int_0^{L_0} aL[r'\{Hp\}_r + t'Hp_t]dL$$

Using Equations (24)–(26) to replace the terms $\{Hp\}_r$, Hp_t, and the term HpL evaluated at L_0, we may rewrite the objective as

$$\max_{\substack{r,L_0\\ t\leq T}} aHvL\Big|_0^{L_0} + \int_0^{L_0} \{vH-(1+a)c-ar'L[hv-\mu] - at'L[Hv_t-\lambda]\}dL \tag{29}$$

where the independent variables have been suppressed. We first note that $t'\lambda = r'\mu = 0$ for all L, since either the slack variables vanish or the corresponding functions $t(L)$ or $r(L)$ are constant. Hence, the terms containing λ and μ in Equation (29) can be eliminated.

We assume for now that the monotonicity constraints on r and t are not binding, but we do explicitly account for the upper bounds on r and t. If any of the monotonicity constraints are violated, it is possible, in principle, to modify the solution using a procedure similar to that employed by Goldman, Leland and Sibley [4]. This procedure may, however, become quite cumbersome if both monotonicity constraints are violated simultaneously. In order to simplify our discussion, we do not deal with the case in which the monotonicity constraints are binding. We introduce the Lagrange multipliers $\eta(L)$ and $\zeta(L)$ for the respective constraints $r(L)\leq R(L)$ and $t\leq T$. This causes us to subtract from the integrand in Equation (29) the term $\eta(L)[r-R(L)]+\zeta(L)[t-T]$. These multipliers must be non-negative and satisfy the standard complementary slackness conditions (ie either the multiplier is zero or its corresponding constraint is binding).

The Euler conditions for the above maximization are then given as follows:

$$\begin{aligned}\partial/\partial r &= hv-(1+a)c_r - aL(r'h'v+t'hv_t)-\eta\\ &= (d/dL)\{-aLhv\}\\ &= -a(v_LhL+hv+r'Lh'v+t'hv_tL)\end{aligned} \tag{30}$$

and

$$\begin{aligned}\partial/\partial t &= Hv_t-(1+a)c_t - aL(r'hv_t+t'Hv_{tt})-\zeta\\ &= (d/dL)\{-aLHv_t\}\\ &= -a(Hv_{tL}L+Hv_t+r'hv_tL+t'Hv_{tt}L)\end{aligned} \tag{31}$$

Also, since L_0 is unconstrained, the total derivative of the objective function with respect to L_0 must vanish at the optimum. Thus, at $L=L_0$, we have

$$a[Hv_L+Hv_tt'+hvr']L+(1+a)(Hv-c)-aLr'hv - at'Hv_t=0$$

which simplifies to

$$H(r)v(t,L_0)+bH(r)v_L(t,L_0)L-c(r,t,L_0)\Big|_{\substack{r=r(L_0)\\ t=t(L_0)}}=0 \tag{32}$$

Equations (30) and (31) can be simplified to the form

$$h(r)v(L,t)+bh(r)v_L(L,t)L-c_r(r,t,L)=\eta(L) \tag{33}$$

where

$$b=a/(1+a)$$

and

$$H(r)v_t(L,t)+bH(r)v_{tL}(L,t)L-c_t(r,t,L)=\zeta(L) \tag{34}$$

Using Equation (33), we can simplify Equation (32) to

$$\{H(r)[c_r(r,t,L_0)+\eta(L_0)]-h(r)c(r,t,L_0)\}\Big|_{\substack{r=r(L_0)\\ t=t(L_0)}}=0 \tag{35}$$

The solution of these equations is carried out as follows. First, we note that no derivatives of the optimal trajectories appear in the equations, so that they can be solved simultaneously, pointwise in r, t, and L. It is probably best to begin by determining L_0, $r(L_0)$, and $t(L_0)$ from Equations (33), (34) and (35). The value of $\eta(L_0)$ is determined from these three equations because either $\eta > 0$ and $r = R(L_0)$ or else $\eta = 0$ and r is arbitrary. The constraint $r(L) = R(L)$ may be active in several places. However, each of these is determined in sequence as the solution is developed by solving Equations (33) and (34). By the monotonicity of $t(L)$, the constraint $t \leqslant T$ becomes active at most once. Thus we continue our solution in the direction of decreasing L until the optimal t value touches this constraint. We then have $t = T$ from that point onward, and we continue to solve for $r(L)$ using $t(L) = T$.

Elasticity interpretation of the necessary conditions

Equations (33)–(35) can be interpreted as variants of the classical constrained monopoly pricing rule

$$1 + b[D(p)/p]/D'(p) = c/p \tag{36}$$

which holds simultaneously at each marginal price level. The second term on the left-hand side is the reciprocal of the demand elasticity, scaled by the factor $b = a/(1+a)$, where a is the weighting of net revenue in the objective function.

To bring Equations (33) and (34) to this form, we can interpret $p^r = \{Hp\}_r$ as the average marginal price of reliability per load slice and $p^t = Hp_t$ as the expected marginal price of duration per load slice with reliability r. Using conditions (24)–(26) for interior r and t, we then have

$$p^r = hv \quad \text{and} \quad p^t = Hv_t$$

Consequently, it follows that

$$dp^r/dL = hv_L \quad \text{and} \quad dp^t/dL = Hv_{tL}$$

Equations (33) and (34) can thus be rewritten as

$$1 + b[(L/p^r)/(dL/dp^r)] = c_r/p^r \tag{37}$$

$$1 + b[(L/p^t)/(dL/dp^t)] = c_t/p^t \tag{38}$$

Since $r(L)$ and $t(L)$ are monotonically decreasing in L, the index L is also the demand for reliability r or lower and the demand for duration t or less. Clearly L is the demand for load slices of level L or less. Hence, the term in brackets in Equations (37) and (38) is the reciprocal of the appropriate demand elasticity. In Equation (37) it is the elasticity of demand for reliability improvement, parametric on a fixed duration t. In Equation (38) it is the elasticity of demand for duration, parametric on fixed reliability level r.

Determining the optimal price function $p(r, t)$

Given the optimal trajectory $\{r(L), t(L)\}$ satisfying Equations (23)–(28) and the boundary level L_0, the solution is completed by determining the corresponding price plan $p(r,t)$. It is useful first to determine $P(L) = p[r(L), t(L)]$, which is the price for a load slice at level L when duration and reliability are optimal. Using the chain rule, we have

$$d(HP)/dL = [\{Hp\}_r r' + Hp_t t']|_{r=r(L),\, t=t(L)} \tag{39}$$

Next we substitute into Equation (39) the consumer self-selection constraints (24) and (25). Note that $\lambda = 0$ unless $t = T$, in which case $t' = 0$. Similarly, $\mu = 0$ unless $r = 1$ and correspondingly $r' = 0$. Thus, we always have $r'hv = r'\{Hp\}_r$ and $t'Hv_t = t'Hp_t$, ie

$$d(HP)/dL = hvr' + Hv_t t' = [d(Hv)/dL - Hv_L]|_{r=r(L),\, t=t(L)} \tag{40}$$

The resulting function HP is thus monotonically decreasing in L.

Integrating Equation (40) along the optimal trajectory $\{r(L), t(L)\}$, using the boundary condition (26), and then dividing through by H gives

$$P(L) = v[L, t(L)] + \{1/H(r(L))\} \int_L^{L_0} [H(r(l)) v_L(l, t(l))] dl \tag{41}$$

Note that this form always holds, even when $r(L) = 1$ or $t(L) = T$.

While Equation (41) completely specifies $P(L)$, this pricing function cannot be directly implemented, since L is not observable for individual customers. Instead, the function $p(r,t)$, satisfying $p(r(L), t(L)) = P(L)$, must be determined. The function $p(r,t)$ and its first partial derivatives are specified only along the trajectory $\{r(L), t(L)\}$ in the r–t plane. This leaves considerable freedom in the choice of $p(r,t)$. We show in Lemma 3 that it is always possible to meet the required conditions with an additively separable function of the form $p(r,t) = f(t) + g(r)$.

Lemma 3. Let $P(r,t)$ be an arbitrary function of r and t, and let $r(L)$ and $t(L)$ be monotonically decreasing functions of L defining a trajectory in the r–t plane, parametric on L. Then there exists an additively separable function $p(r,t) = f(t) + g(r)$ such that, for

$r=r(L)$, $t=t(L)$, and all values of L,

$$P(r,t)=p(r,t) \tag{42}$$

$$P_r(r,t)=p_r(r,t)=g'(r) \tag{43}$$

and

$$P_t(r,t)=p_t(r,t)=f'(t) \tag{44}$$

Proof. Allowing for discontinuities in $P(r,t)$ along the axes, let

$$f(t)=P(0^+,0^+)+\int_{0^+}^{L(t)} P_t(r(l),t(l))t'(l)dl \tag{45}$$

and

$$g(r)=\int_{0^-}^{L(r)} P_r(r(l),t(l))r'(l)dl \tag{46}$$

where

$$L(r)=\min\{L|r(L)=r\} \text{ and } L(t)=\min\{L|t(L)=t\} \tag{47}$$

Differentiating Equations (45) and (46) with respect to t and r along the trajectory $\{r(L),t(L)\}$ verifies Equations (43) and (44). Then, adding Equations (45) and (46) and invoking the chain rule yields

$$f(t(L))+g(r(L))=P(0^+,0^+)+\int_{0^+}^{L} dP(r(l),t(l))$$

$$=P(r(L),t(L)) \tag{48}$$

QED.

The components of the additively separable price function $p(r,t)$ can be uniquely determined as follows. For $r>0$ and $t<T$, Equation (24) implies

$$f'(t)=P_t(r,t)=v_t(t,L) \tag{49}$$

Furthermore, at the boundary $L=L_0$, we have by Equation (26) (assuming $r(L_0)>0$) that

$$f(t(L_0))+g(r(L_0))=v(L_0,t(L_0)) \tag{50}$$

We may arbitrarily set $g[r(L_0)]=0$, ie zero reliability charge for the lowest level of reliability offered. Then, along the trajectory $\{r(L),t(L)\}$, we have by Equations (49) and (50)

$$f(t)=v(L_0,t(L_0))-\int_{t}^{t(L_0)} v_t(L(\tau),\tau)d\tau \tag{51}$$

where $L(t)$ is the inverse function of $t(L)$ defined by Equation (47).

The function $g(r)$ can then be determined from $P(L)$ and $f(t)$ as follows

$$g(r)=P(L(r))-f(t\{L(r)\}) \tag{52}$$

where $L(r)$ is the inverse function of $r(L)$ defined in Equation (47).

Implementation

Individual customer preferences

The load slice model assumes in effect that each type of load slice is priced and consumed independently. Each customer's load duration curve is then composed of a collection of predetermined load slice types. This does not assume that all customer load duration curves have the same shape, because the distribution of load slice types can vary by customer. The constraint that is imposed is that all load slices that have equal durations should have equal reliability levels across all customers. In this section, we develop the precise assumption that guarantees this for the customer demand functions. This is the 'weak separability' of the demand function family indexed by customer type, which was used in a different context by Panzar and Sibley [12].

For individual customer type u, let us define the following:

$L^*(\theta,u)$ = the load curve (measured in real time) for customer type u
$L^u(t)$ = the load duration curve for customer type u
$r^u(l)$ = the reliability selections of customer type u
$t^u(l)$ = the duration time selections of customer type u
L_0^u = the maximum load of customer type u
$G(u)$ = the probability distribution of customer types

The load level l used above refers to the height of the customer's load curve, as opposed to L, the height of the system load curve. Clearly, however, the following relationships must hold:

$$t^u(l)=M\{\theta|L^*(\theta,u)\geqslant l\}$$
$$L^u(t)=(t^u)^{-1}(t) \tag{53}$$

If the preferences for load slices are uniform and independent across customers (a condition that corresponds to weak separability), then

L corresponds to $l=(t^u)^{-1}[t(L)]$

and

$$r^u(l)=r\{L(t^u(l))\} \tag{54}$$

where $r(L)$ and $L(t)$ are the system functions derived earlier. The relationship in Equation (54) implies that load slices having equal durations will have equal reliabilities.

Pricing and interruption of service

Interruption of service in case of shortage can be easily implemented. The supplier would simply select a given reliability level r, and all customer load slices l such that $r^u(l)\leqslant r$ would be interrupted. The level r would be selected, so that the total supply and demand are brought into balance. A customer's monthly bill can also be computed in a straightforward manner. The fact that the price function is additively separable means that the bill is the sum of an energy charge, which is increasing in reliability. These would be computed as follows:

$$\text{total charge}=\int_0^{L_0^u} H(r^u(l))g(r^u(l))dl \quad \text{(demand charge)}$$
$$+\int_0^{L_0^u} H(r^u(l))f(t^u(l))dl \quad \text{(energy charge)} \tag{55}$$

In Equation (55) it is assumed that all customers u share the same scaling function $H(r)$. This restriction could be relaxed without affecting the applicability of the formula.

In applying the above approach, only a few discrete reliability levels would typically be offered, so that the function $g(r)$ is approximated by a vector of demand charges per kW corresponding to the reliability options. Demand Subscription Service at SCE, mentioned in the introduction, is an example of two reliability levels (interruptible/non-interruptible). Also, $f(t)$ would be approximated by a piecewise linear function.

When the individual loads are not synchronized, the billing formula (55) will generate excess revenue because of the cost savings from the complementarities in customer loads. Furthermore, the marginal energy price at a given point in time will differ across customers, according to their respective load patterns, so that marginal benefits are not equated across users. This is clearly not socially efficient, since a costless reallocation of power could improve social welfare. There is, however, a class of preference structures under which individual loads are synchronized. Proposition 1 defines a general class of demand functions $L^*(\theta,p,u)$ that lead to synchronized load curves.

The billing and load control approach described above does not require explicit knowledge of the system load slice index L that corresponds to the load slices making up the customer's load duration curve. The system's average load duration curve $L(t)$ is related to the individual load curves as follows. Letting $t(L)$ be the inverse of $L(t)$, we have

$$t(L)=M\left\{\theta\left|\int_u L^*(\theta,u)dG(u)\geqslant L\right.\right\} \tag{56}$$

If the individual loads are synchronized so that the same mapping from real time to duration holds for all customers, then the individual load duration curves may be added directly, ie

$$L(t)=\int_u L^u(t)dG(u) \tag{57}$$

Weakly separable demand functions

The class of demand functions that yield load curves and reliability selections satisfying Equations (54) and (57) can be characterized as follows.

Proposition 1. Let $L^*(\theta,p,u)$ be a weakly separable function of the form

$$L^*(\theta,p,u)=\Lambda(W(\theta,p),u) \tag{58}$$

and assume that $\Lambda(W,u)$ is monotonically increasing in W. Then for any fixed p, the mapping from real time θ onto duration t implied by Equation (53) is independent of the index u.

Proof. Let $t(\theta,p,u)$ denote the duration value for customer u, corresponding to real time θ under price p. The mapping $t(\theta,p,u)$ is given by

$$t(\theta,p,u)=M\{\tau|L^*(\tau,p,u)\geqslant L^*(\theta,p,u)\} \quad (59)$$

If Equation (59) holds, however, then clearly for any u, we have

$$t(\theta,p,u)=M\{\tau|W(\tau,p)\geqslant W(\theta,p)\} \quad (60)$$

which implies that $t(\theta,p,u)$ is independent of u. QED.

We now show that the assumption of weak separability is also sufficient to guarantee that preferences for load slices will be uniform across all customer types, as discussed earlier.

Proposition 2. Let $v(L,t,u)$ denote the willingness of consumer u to pay for a unit load slice of duration t at system load level L. Then the weak separability condition (58) implies that $v(L,t,u)$ is independent of u.

Proof. Let $t(\theta,p)$ be as in Equation (60) and let $L(t,p)$ represent the system's average load duration curve as a function of price per kWh. To be consistent with the consumer's demand function $L^*(\theta,p,u)$, the value function $v(L,t,u)$ must satisfy the condition

$$v_t(L[t(\theta,p),p],t(\theta,p),u)=p \quad (61)$$

for all u. Differentiating both sides of Equation (61) with respect to u proves that $v_{tu}(L,t,u)=0$. This implies that the function $v(L,t,u)$ must be of the form

$$v(L,t,u)=v^1(L,t)+v^2(L,u) \quad (62)$$

Clearly, however, $v(L,0,u)=0$ for all values of L and u. Hence $v^2(L,u)=0$ for all values of L and u, which proves that $v(L,t,u)$ must be independent of u. QED.

Value function estimation

The value function $v(L,t)$ can be estimated from the system load duration curve $L(t)$ that results in response to a uniform energy price. The load duration curve resulting from charging a fixed energy price p per kWh (call it $L(p,t)$) satisfies the following relationship:

$$v_t(L(p,t),t)=p \qquad \text{for all } t \quad (63)$$

Thus, for any fixed p, the load duration curve $L(p,t)$ is a level set of the function $v_t(L,t)$. Based on demand elasticity estimates, the sensitivity of the load curve to the price p, and thus other level curves of $v_t(L,t)$, could be estimated. Segmenting the market into customer classes with separate value function estimates could be expected to improve the accuracy of the estimation process. As a first cut, however, the value function can be inferred from the system load curve with no additional market research information.

TOU rate with a block declining-demand charge

It is interesting to note that the pricing policy based on individuals' load duration curves, can alternatively be implemented as a TOU rate with demand charges. The relationship between the two implementations is similar to that between demand-layered and TOU pricing.

Let us assume for simplicity that $f(t)$ is piecewise linear with breakpoints at $0^+,t_1,t_2,\ldots,t_n$ and corresponding slopes $f'_0,f'_1,\ldots,f'_n$. Given a forecast of the system load pattern, it is possible to map the duration intervals to their corresponding chronological time intervals, as illustrated in Figure 3. This defines the rating periods for a TOU rate structure, with the energy rate corresponding to duration interval $[t_i,t_{i+1}]$ equal to f'_i. The reliability price and the fixed charge $f(0^+)$ will be imposed as a demand charge on kW used during the system's peak. A customer's service contract will specify his load level breakpoints L_i^u, at which the respective reliability level r_i and corresponding demand charge rate g_i take effect. This determines the customer's successive interruption levels under increasing shortage conditions, with a corresponding block declining-demand charge applied to the customer's load during the system's peak. Figure 3 illustrates the two alternative rate structures discussed above. They are equivalent if each customer's load is synchronized with the total system load.

Solved example

To illustrate the solution techniques discussed for the optimal price function, a specific example is solved here. The required functional forms are given as follows:

$$h(w)=2w \qquad 0\leqslant w\leqslant 1 \quad (64)$$

$$c(r,t,L)=\mathrm{K}+t\mathrm{V}r^2 \quad (65)$$

$$v(L,t)=\underline{a}L^{-\beta}t^{\alpha} \qquad \beta>0,\ 0<\alpha<1 \quad (66)$$

where K, V, $\underline{a}$, α, and β are constants. By integrating $h(w)$, we obtain

$$H(r)=r^2 \qquad 0\leqslant r\leqslant 1 \quad (67)$$

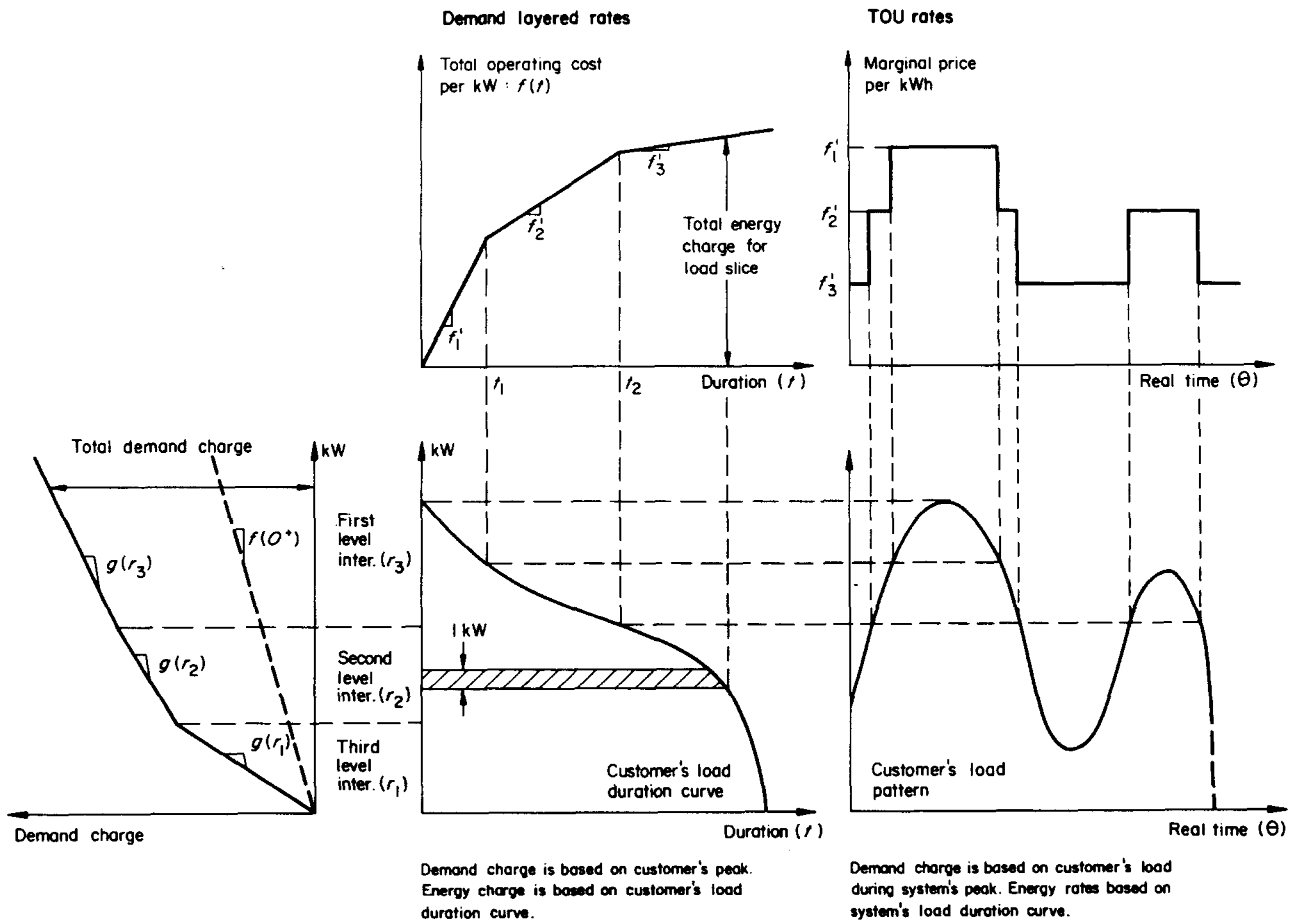

Figure 3. Alternative implementations of multilevel demand subscription pricing.

The cost function $c(r,t,L)$ is consistent with Equation (96) for a single technology, but with a capacity cost K per unit of average load. Assuming that total capacity is Y, we also obtain the reliability and duration constraints

$$r(L) \leqslant \min\{1, Y/2L\} = R(L) \quad \text{and} \quad t(L) \leqslant T = 1 \tag{68}$$

From Equations (65) and (66), we obtain

$$c_t = \mathrm{V} r^2, \quad c_r = 2\mathrm{V} tr \tag{69}$$

$$v_L = -a\beta L^{-\beta-1} t^{\alpha} = -\beta v/L \tag{70}$$

$$v_t = a\alpha L^{-\beta} t^{\alpha-1} = \alpha v/t \tag{71}$$

$$v_{Lt} = -a\alpha\beta L^{-\beta-1} t^{\alpha-1} = -\alpha\beta v/tL \tag{72}$$

From Equation (34), we have

$$\alpha(v/t)[1-b\beta] \begin{array}{ll} = \mathrm{V} & \text{for } t<1 \\ \geqslant \mathrm{V} & \text{for } t=1 \end{array} \tag{73}$$

This yields

$$t(L) = \min\{1, AL^{-\beta/(1-\alpha)}\} \tag{74}$$

where

$$A = [a\alpha(1-b\beta)/\mathrm{V}]^{1/(1-\alpha)}$$

Clearly, by monotonicity, $t(L)=1$ for all $L \leqslant L^*$, where

$$L^* = A^{(1-\alpha)/\beta} \tag{75}$$

From Equations (33) and (73) we have

$$\eta = 2r(v - b\beta v) - 2\mathrm{V}tr \geqslant 2r\mathrm{V}t(1-\alpha)/\alpha > 0 \tag{76}$$

Thus, the upper bound on reliability is always binding. Therefore, we have

$$r(L) = \min\{1, Y/2L\} \tag{77}$$

We assume here that $L^* \leqslant Y/2$, ie that the base load with duration $t(L)=1$ always has $r(L)=1$. We also assume that $t(L_0)<1$ to rule out a completely rectangular load. To determine L_0, we now use Equation (35) which yields

$$\{r^2[1-b\beta]v-tVr^2-K\}|_{r=Y/2L_0,t=t(L_0)}=0 \tag{78}$$

Using Equation (73), we can reduce this to

$$(Y/2L_0)^2Vt(L_0)(1-\alpha)/\alpha=K \tag{79}$$

and by applying Equation (74), we obtain

$$L_0=[AY^2V(1-\alpha)/(4\alpha K)]^{(1-\alpha)/[\beta+2(1-\alpha)]} \tag{80}$$

To obtain the total price function $P(L)$, we first substitute Equation (74) and (75) into Equation (66) to obtain

$$v(L,t(L))=\begin{cases}\underline{a}A^{\alpha}L^{-\beta/(1-\alpha)} & \text{for } L\geqslant L^*\\ \underline{a}L^{-\beta} & \text{for } L\leqslant L^*\end{cases} \tag{81}$$

By Equations (41), (67), (70), (77) and (81) it follows that

$$P(L)=\underline{a}A^{\alpha}L^{-\beta/(1-\alpha)}+(2L/Y)^2\int_L^{L_0}(Y/2l)^2(-\beta/l)\,\underline{a}A^{\alpha}l^{-\beta/(1-\alpha)}dl$$

$$=\begin{cases}\{[\alpha\beta+2(1-\alpha)]/[\beta+2(1-\alpha)]\}\underline{a}A^{\alpha}L^{-\beta/(1-\alpha)}+\underline{K}(2L/Y)^2 & \text{for } L_0\geqslant L\geqslant Y/2 \quad (82)\\ \{\alpha L^{-\beta/(1-\alpha)}+[2(1-\alpha)^2/[\beta+2(1-\alpha)](Y/2)^{-\beta/(1-\alpha)}\}\underline{a}A^{\alpha}+\underline{K} & \text{for } Y/2\geqslant L\geqslant L^* \quad (83)\\ P(L^*) & \text{for } 0\leqslant L\leqslant L^*\end{cases}$$

where $\underline{K}=aK\alpha\beta/\{V[\beta+2(1-\alpha)]A^{1-\alpha}\}$

For $0\leqslant L\leqslant L^*$, $P(L)=P(L^*)$ is constant. As was noted earlier, the price function $P(L)$ must also be constant when $t(L)$ and $r(L)$ are constant. In this case, the customer's choices are determined by the constraints, rather than by consumer surplus maximization.

From Equations (71) and (73) we have (for $t<1$)

$$v_t[L(t),t]=V/[1-b\beta] \tag{84}$$

Also from Equation (73) it follows that

$$v[L_0,t(L_0)]=[V/\alpha(1-b\beta)]t(L_0) \tag{85}$$

Substituting Equations (84) and (85) into Equation (51) yields

$$f(t)=(V/\alpha)\{(1-\alpha)t(L_0)+\alpha t\}/[1-\underline{b}\beta] \tag{86}$$

From Equation (79), however, we have

$$t(L_0)(1/\alpha)/\alpha=(K/V)(2L_0/Y)^2 \tag{87}$$

Substituting Equation (87) into Equation (86) yields (for $t<1$)

$$f(t)=[V/(1-b\beta)][t+(K/V)(2L_0/Y)^2] \tag{88}$$

Finally, by Equations (52) and (82), it follows that (for $r<1$)

$$g(r)=\underline{a}A^{\alpha}\{2(1-\alpha)^2/[\beta+2(1-\alpha)]\}(2r/Y)^{\beta/(1-\alpha)}+\underline{K}\{(1/r^2)-[1+2(1-\alpha)/\beta](2L_0/Y)^2\} \tag{89}$$

As was remarked earlier, the separable form $p(r,t)=f(t)+g(r)$ holds even when the price is constant in r or t. This can, in fact, be verified directly in the above expressions. It can be seen that $P(L^*)=f(1)+g(1)$ and that $P(L)=f(t(L))+g(1)$ for $L^*\leqslant L\leqslant Y/2$. The supplier can then enforce the constraints $r\leqslant 1$ and $t\leqslant 1$ for the consumer directly.

Specific illustration

The example discussed above can be further illustrated by giving its exact form for specific choices of the parameter values. Let us use the following:

$$\underline{a}=1,\ \alpha=1/2,\ \beta=1,\ K=1/4,\ V=1 \text{ and } Y=1$$

This gives from Equations (74) and (80)

$$A=[(1-b)/2]^2,\ A^{\alpha}=A^{1-\alpha}=(1-b)/2 \text{ and } L_0=[(1-b)/2]^{1/2}$$

Substitution into Equations (74) and (77) gives

$$r(L)=\begin{cases}1 & \text{for } 0\leqslant L\leqslant 1/2\\ 1/2L & \text{for } 1/2\leqslant L\leqslant[(1-b)/2]^{1/2}\end{cases}$$

$$t(L)=\begin{cases}1 & \text{for } 0\leqslant L\leqslant(1-b)/2\\ [(1-b)/2L]^2 & \text{for } (1-b)/2\leqslant L\leqslant[(1-b)/2]^{1/2}\end{cases}$$

From Equation (81) we obtain

$$v(L,t(L))=\begin{cases}1/L & \text{for } 0\leqslant L\leqslant(1-b)/2\\ (1-b)/2L^2 & \text{for } (1-b)/2\leqslant L\leqslant[(1-b)/2]^{1/2}\end{cases}$$

From Equation (82) it follows that $\underline{K} = 1/[8(1-b)]$. The price function $P(L)$ must be defined in the three regions in Equation (82). This gives

$$P(L) = \begin{cases} 3(1-b)/8L^2 + L^2/[2(1-b)] & \text{for } 1/2 \leqslant L \leqslant [(1-b)/2]^{1/2} \\ [1 + 1/2L^2](1-b)/2 + 1/[8(1-b)] & \text{for } (1-b)/2 \leqslant L \leqslant 1/2 \\ (1-b)/2 + 9/[8(1-b)] & \text{for } 0 \leqslant L \leqslant (1-b)/2 \end{cases}$$

From Equations (88) and (89) we can also obtain the components of the price function:

$$f(t) = 1/2 + t/(1-b) \quad \text{for } 0 < t \leqslant 1$$

$$g(r) = -1/2 + (1-b)r^2/2 + 1/[8(1-b)r^2] \quad \text{for } 0 \leqslant r \leqslant 1$$

These last two expressions show how the optimal price components $f(t)$ and $g(r)$ are related to the marginal cost V and to the weighting factor b ($b=0$ corresponds to welfare maximization). It can be verified directly for this example that $f(t(L)) + g(r(L)) = P(L)$ for all L, as required.

It is interesting to graph the cost, price and value functions for this example. This is most directly done as a function of the load slice index L. In Figure 4, for $b=0.1$, the three functions $H(r(L))v(L,t(L))$, $H(r(L))P(L)$, and $c(r(L),t(L),L)$ are plotted as a function of L. The value and price functions are scaled by $H(r(L))$ to make them correspond to the realized values after interruption effects are included. Note that the consumer surplus decreases monotonically to zero, as required by Equation (16). As L approaches zero, the consumer surplus becomes infinite.

The consumer's optimal selections are plotted in Figure 5. Note that $r(L) = 1$ for $L \leqslant 1/2$ and $t(L) = 1$ for $L \leqslant \bar{L} = 0.45$. This means that $S_r > 0$ for $L < 1/2$ and $S_t > 0$ for $L < 0.45$, ie the upper bounds on the consumer choices are binding. For $L < 0.45$, the price and cost functions in Figure 4 are constant, while the value function approaches infinity as L approaches zero.

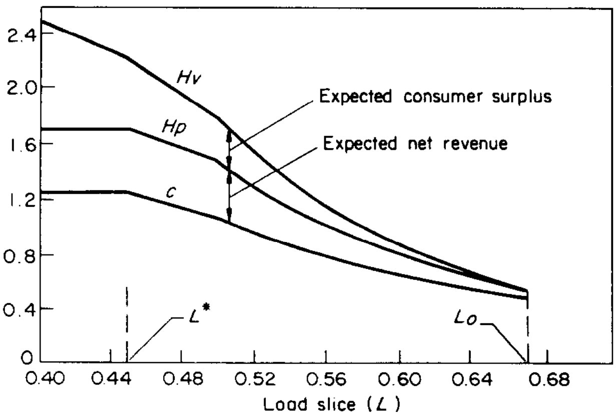

Figure 4. Cost, price and value functions.

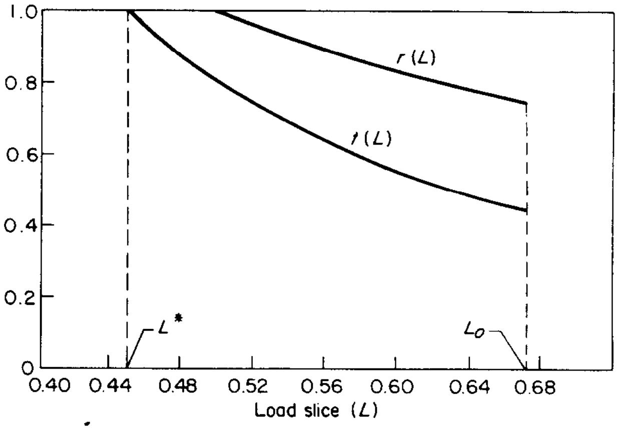

Figure 5. Optimal selections.

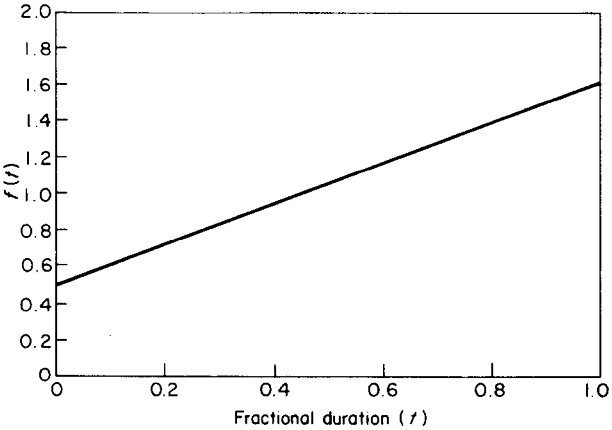

Figure 6. Duration charge per kW.

We can also compute the expected load duration curve that will result from these selections. The nominal load duration curve is $L(t)$, the inverse of $t(L)$, which is given by

$$L(t) = \begin{cases} [(1-b)/2]^{1/2} & \text{for } 0 \leqslant t \leqslant (1-b)/2 \\ (1-b)/2t^{1/2} & \text{for } (1-b)/2 \leqslant t \leqslant 1 \end{cases}$$

From Equation (16), for $H(r(L)) = \min\{1, (1/2L)^2\}$, we have

$$\bar{L}(t) = \int_0^{L(t)} H(r(L))dL$$

$$= \begin{cases} L(t) & \text{for } 0 \leqslant L(t) \leqslant 1/2 \\ 1 - 1/[4L(t)] & \text{for } 1/2 \leqslant L(t) \leqslant [(1-b)/2]^{1/2} \end{cases}$$

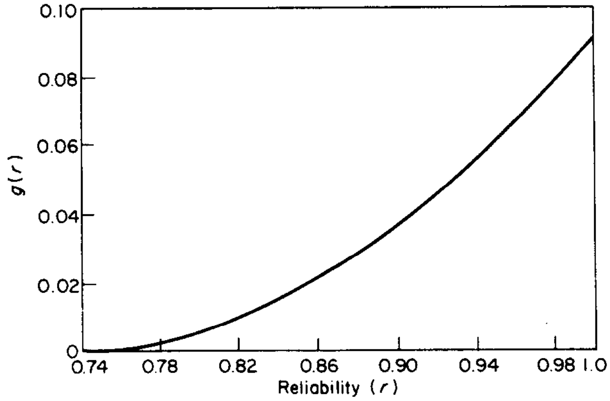

Figure 7. Reliability charge per kW.

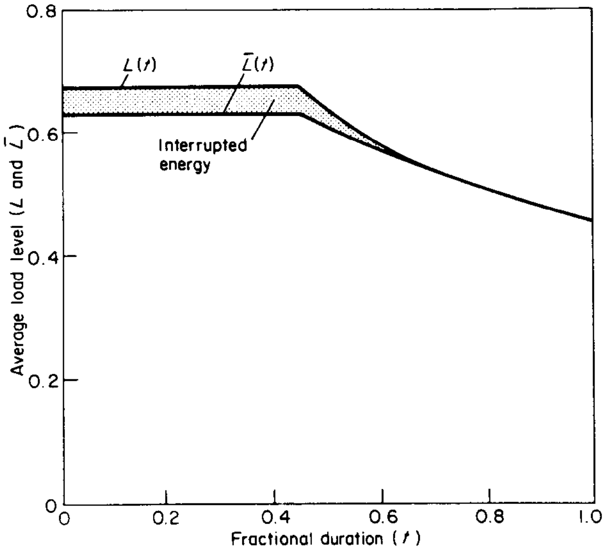

Figure 8. Selected and realized average load duration curves.

This shows how the nominal load duration curve $L(t)$ is partially curtailed by the selected interruption function $r(L)$.

Conclusion

In this paper we have investigated a price structure that permits non-linear pricing in duration and the offering of interruptible service contracts based on load level. We believe that our analysis is unique in that it includes both of these features, which simultaneously encourage flattening of the customers' load duration curves and permit a direct load management programme to be offered as a menu of options for the customer. In particular, the inclusion of the correlation between random demand level and the likelihood of supply interruption is an important realistic feature. The assumptions imposed for the purposes of our analysis do not constrain its applicability, but they do cause a degradation in the social efficiency of the resulting pricing policies when they are violated. The objective function is flexible in that it accommodates any selected weighting of net revenue and social welfare objectives. The solved example illustrates explicitly the dependence of the optimal price policy upon this weighting factor.

A very useful insight derived from our analysis is the additively separable form of the optimal pricing function. This property simplifies the implementation of the resulting price schedules. As remarked in the implementation section, no individual preference data need be gathered to implement such a plan, because the essential trade-offs between load slice value and duration selection can be inferred from the average system load duration curve. A more accurate model of customer preference could, however, be obtained by applying market research techniques to segment the customer base into similar classes and estimate separate value functions for the model in each class. Different price schedules would then be offered to meet the needs of each of the separate classes. A similar variety of price plans is characteristic of electric power service pricing today in that residential and various classes of commercial users are offered different schedules. This paper's development can provide the basis for designing experiments with such price plans and for future research to investigate the attractiveness of specific pricing functions.

References

1 P. A. Ballonoff, 'Demand layered costing and pricing of electricity', *Public Utility Fortnightly*, 3 March 1983, pp 30–37.
2 M. Boiteux, 'Peak-load pricing', in J. R. Nelson, ed, *Marginal Cost Pricing in Practice*, Prentice-Hall, Englewood Cliffs, 1964, pp 59–90.
3 H. P. Chao, 'Peak load pricing and capacity planning with demand and supply uncertainty', *Bell Journal of Economics*, Vol 14, No 1, Spring 1983, pp 179–190.
4 M. B. Goldman, H. E. Leland and D. S. Sibley, 'Optimal nonuniform prices', *Review of Economic Studies*, Vol 51, 1984, pp 305–319.
5 E. F. Gorzelnik, 'Electronic meter carves a niche for itself', *Electrical World*, January 1983, pp 119–126.
6 M. Harris and A. Raviv, 'A theory of monopoly pricing schemes with demand uncertainty', *American Economic Review*, Vol 71, No 3, 1981, pp 347–365.
7 R. S. Main, *Periodic vs Demand Layered Pricing for Utility Loads*, PhD dissertation, Department of Economics, UCLA, 1973.
8 M. G. Marchand, 'Pricing power supplied on an interruptible basis', *European Economic Review*, Vol 5, 1974, pp 263–274.
9 B. M. Mitchell, W. G. Manning and J. P. Acton, *Peak Load Pricing*, Ballinger, 1978.

10 S. S. Oren, S. A. Smith and R. B. Wilson, 'Linear tariffs with quality discrimination', *Bell Journal of Economics*, Vol 13, No 2, Autumn 1982, pp 455–471.
11 S. S. Oren, S. A. Smith and R. B. Wilson, 'Capacity pricing', *Econometrica*, Vol 53, No 3, May 1985, pp 545–566.
12 J. C. Panzar and D. S. Sibley, 'Public utility pricing under risk: the case of self-rationing', *American Economic Review*, Vol 68, No 5, 1978, pp 888–895.
13 F. C. Schweppe, R. D. Tabor and J. Kirtley, 'Homeostatic control for electric power usage', *IEEE Spectrum*, July 1982, pp 44–48.
14 P. O. Steiner, 'Peak loads and efficient pricing', *Quarterly Journal of Economics*, Vol 71, 1957, pp 585–610.
15 J. Tschirhart and F. Jen, 'Behavior of monopoly offering interruptible service', *Bell Journal of Economics*, Vol 10, Spring 1979, pp 244–258.
16 J. Vardi, J. Zahavi and B. Avi-Itzhak, 'Variable load pricing in face of loss-of-load probability', *Bell Journal of Economics*, Vol 6, Autumn 1977, pp 270–288.
17 O. E. Williamson, 'Peak-load pricing and optimal capacity under indivisible constraints', *American Economic Review*, Vol 56, 1966, pp 810–827.

Appendix

The generation cost function

We derive here an expression for $c(t,r,L)$ under the assumption that the generation system consists of n technologies with fixed capacities $X_1, X_2, \ldots, X_n$. Each technology is characterized by an amortized fixed cost F_i per kW capacity and a variable cost V_i per kWh. We assume that the technologies are indexed in the optimal dispatching order, ie, $V_1 < V_2 < \ldots < V_n$. Let Y_i denote the cumulative capacities, ie

$$Y_i = \sum_{j=1}^{i} X_j$$

The effect of uncertainty in the available capacity can be roughly approximated, as in Chao [3], by assuming that the installed generation capacity of each technology is composed of a large number of small generation units, each of which fails independently with a probability a_i characteristic of that technology. Each of the capacities X_i can then be regarded as the expected available capacity of technology i, which is the installed capacity times the probability a_i. Similarly, the fixed cost F_i may be interpreted as the amortized fixed cost per available kW, which is the cost per installed kW divided by a_i.

Since we are dealing with short-term pricing decisions, we may regard the total fixed cost of the system as a sunk cost that will only enter as a constant in a net revenue constraint and that will not affect the supplier's objective function. According to our demand model, a load slice of type L will be served at load level $h(w)L$, where w is a random variable controlling the scaling function $h(w)$. The energy cost per kWh for serving a load slice of type L is therefore a discrete random variable that can assume values $V_1, V_2, \ldots, V_n$ and whose probability distribution depends on L, w, and r.

In particular,

$$V(w,L,r) = \begin{cases} V_i & \text{if } Y_{i-1} \leqslant h(w)L \leqslant Y_i \text{ and } w \leqslant r \\ 0 & \text{otherwise} \end{cases} \tag{90}$$

Since $h(w)$ is monotonic by assumption, we can define

$$w_i(r,L) = \min\{h^{-1}(Y_i/L), r\} \tag{91}$$

Then, the random variable V can be defined by

$$V(w,L,r) = \begin{cases} V_i & \text{for } w_{i-1}(r,L) \leqslant w \leqslant w_i(r,L) \\ 0 & \text{for } w > r \end{cases} \tag{92}$$

The width of a load slice under conditions w is $h(w)$ times its average width. Consequently, the expected cost of serving a load slice of type L, with duration t and chosen reliability r, is given by

$$c(r,t,L) = t\int_0^r V(w,L,r)h(w)dw$$

$$= t\sum_{i=1}^{n} V_i\{H(w_i(r,L)) - H(w_{i-1}(r,L))\} \tag{93}$$

Let

$$I(r,L) = \max\{i | h(r)L > Y_{i-1}\} \tag{94}$$

Then, according to Equation (91), it follows that

$$w_i(r,1)=\begin{cases} r & \text{for } i \geqslant I(r,L) \\ n^{-1}(Y_i/L) & \text{for } i < I(r,L) \end{cases} \tag{95}$$

Consequently, Equation (93) can be rewritten as

$$c(r,t,L)=t\{v_I H(r) - \sum_{i=1}^{I-1} [h^{-1}(Y_i/L)](V_i - V_{i-1})\} \tag{96}$$

The above cost function conforms to the general form assumed in the text. Note that $c(r,t,L)$ is linear in t. We also obtain from Equation (96)

$$c_r = tV_I h(r) \tag{97}$$

where I is given by Equation (94). Finally, we note that the reliability level of any load slice L is constrained by the relations $h(r)L \leqslant Y_n$ and $r(L) \leqslant h^{-1}(Y_n/L)$.