INTERRUPTIBLE ELECTRIC POWER SERVICE CONTRACTS

by

Chin-Woo Tan and Pravin Varaiya

Memorandum No. UCB/ERL/IGCT M90/108

12 November 1990

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

INTERRUPTIBLE ELECTRIC POWER SERVICE CONTRACTS

by

Chin-Woo Tan and Pravin Varaiya

Memorandum No. UCB/ERL/IGCT M90/108

12 November 1990

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Interruptible Electric Power Service Contracts[1]

Chin-Woo Tan and Pravin Varaiya
Department of Electrical Engineering & Computer Sciences
University of California, Berkeley CA 94720
November 12, 1990

Abstract

A two period model is considered. In period 1 the electric power company offers for sale a set of contracts $(\rho_1, p_1), (\rho_2, p_2)$.... Each consumer must select one contract k and d units of energy for which she pays $p_k d$. The company must deliver d units of energy in period 2 with probability ρ_k; the service may be interrupted with the complementary probability $1 - \rho_k$. The problem is to design the optimal set of contracts to maximize social welfare when demand and supply may be random and when customers suffer a welfare loss due to service interruption. The best design is shown to be a solution to an optimal control problem. The results contrast sharply with previous work on the problem of pricing electric power in the face of random supply or demand. Approximate estimates of welfare loss resulting from selling electric energy at a fixed price are obtained.

[1] Work supported by NSF grants ECS-8715132 and IRI-8902813

1 Introduction

The current method of allocating electric energy among different consumers and end uses on the basis of fixed prices is inefficient, and there are long-standing proposals to increase efficiency by bringing closer together the marginal value of demand for energy and the marginal cost of energy generation. Recent moves towards deregulation and technological innovations favoring small-scale generation plant have renewed interest in these proposals. Although there is no reliable quantitative estimate of the inefficiency, there is a presumption that it is high. This presumption is based on three distinguishing technological features of the electric energy system.

First, unlike natural gas or oil or coal, it is very expensive to store electric energy. So, unlike in those industries, it is not possible to operate the generation plant at a minimum cost point and meet demand fluctuations out of 'inventories' of stored energy. Second, it is not possible for the supplier to curtail consumption by limiting supply (leading to rationing via queues or black markets as in other commodities) since the dynamic stability of the electric energy system can be maintained only if there is always enough generation 'reserve' capacity to meet unexpected increases in demand. Third, generation technology is such that the marginal cost of supply during daily peaks in demand is several times larger than the daily average cost. These three features suggest that with a fixed price scheme the gap between marginal supply and demand, and hence the resulting inefficiency, is quite large.

One scheme to close this gap is that of *spot pricing*. The idea is to revise the price frequently, say each hour, to the value where instantaneous marginal demand and cost are equated. This scheme was generally recommended by Vickrey [1] who called it "responsive pricing." The specifics for the case of electric energy were worked out by Caramanis *et al* [2] and Bohn *et al* [3].

There is regular and frequent spot trading of 'bulk' or 'wholesale' energy among interconnected power generation companies, so that company A will meet its customers' current demand with power imported from company B if the latter is cheaper than A's own current marginal generation cost. (Of course, these transactions are subject to the overriding constraint of maintaining system stability.) However, aside from some very small-scale experiments no attempt has been made to implement spot pricing at the retail level.

A spot pricing scheme is impractical today. It presupposes the means to communicate pricing information in 'real time', as well as the means whereby customers can automatically and correctly adjust, in response to spot price changes, their level of power consumption and its distribution among alternative end uses such as air-conditioning, lighting, etc. The necessary communications infrastructure is not yet in place. And few except the largest customers have installed the 'smart meters' and computers to manage their energy use. Until such infrastructure is built, a more practical scheme might employ *future* prices: the power company announces prices a day (or week) in advance and customers will then have the lead time to adjust their demand. The announced future price would depend on forecasts of some of the determinants of supply (e.g. scheduled generator 'down' times) and demand (e.g. weather).

Future prices can much more easily be implemented than spot prices, see Ahlstrand [4]. For example, tomorrow's prices could today be communicated by telephone, published in newspapers, and broadcast over radio and TV. However, future prices do not resolve an important problem. Recall that a price is announced in period 1 (today) for energy to be delivered and consumed in period 2 (tomorrow). In the interim there can be significant fluctuations in supply or demand that were unanticipated in the forecast. Consequently, a future pricing regime will require maintaining 'on-line' a significant reserve capacity. This will be less than needed under the current fixed price regime, but it will be considerably larger than under spot pricing.

An early proposal that addressed this was made by Brown and Johnson [5]. That proposal combines future prices with rationing. In the scheme the price for period 2 and the level of energy generation is set to equate marginal cost of supply (assuming no uncertainty) to the marginal value of *expected* demand, based on the period 1 forecast. If it turns out, as will often be the case, that the actual period 2 demand exceeds the period 1 forecast, then a sufficient number of customers will have their electricity cut off. Turvey [6] criticized this scheme for failing to recognize the cost of rationing borne by frustrated customers. He argued that higher prices leading to reduced demand and hence reduced rationing are preferable.[1]

Turvey's emphasis on the cost of rationing is well-taken. However, rais-

[1] In this respect Turvey echoes Vickrey's argument for higher prices on grounds of welfare benefits gained from resulting reduction in 'congestion' in telephone systems, transportation, etc.

ing prices introduces its own distortion. There must be a balance between raising prices to reduce rationing-caused losses and lowering prices to increase welfare gains from increased consumption. Crew and Kleindorfer [7] attempt this balance by deducting from the Brown-Johnson social welfare measure the loss due to rationing, given by the term

$$\text{Rationing loss } = E\ L(D(p) - S(p)) \tag{1}$$

Here D and S denote the random aggregate demand and supply for electric energy, L is a loss function, p is the unit price for energy, and E stands for mathematical expectation. The 'welfare optimizing' price maximizes the usual 'consumer plus producer surplus' minus this rationing loss term. This price could be higher than the Brown-Johnson price, in accordance with Turvey's observation, but the need to ration customers will still remain although fewer customers will be rationed.[2] This specification of the loss is theoretically unsatisfactory since there is no reason why rationing losses suffered by individual customers should add up to a function of aggregate excess demand as in (1).[3]

The Brown-Johnson and Crew-Kleindorfer analyses suffers from another, apparently unnoticed defect. Both implicitly assume that when there is excess demand, customers suffering electricity cuts will be those whose marginal utility is less than the marginal utility of those who will continue to receive service. It is quite clear that a power company cannot obtain such information.[4]

In summary, future pricing schemes must take into account rationing loss, and they must ration on the basis of information that can be available. The scheme of *interruptible service contracts* proposed here incorporates both aspects. The operation of the market can be described as a two step process with the help of Figure 1. Throughout our analysis only the supply is considered to be random, and customer preferences are deterministic.

Step 1. At the beginning of period 1 the company announces a set of

[2]Moreover, it is argued in §4.1 that if interruptible service contracts are permitted, then it is efficient to *lower* prices.

[3]The situation is not similar to the case in which consumer surplus is identified with the area under the aggregate demand curve; the aggregate loss is indeed the sum of individual losses, but there does not seem to be any practical rationing scheme in which the aggregate loss is a function of excess demand.

[4]Visscher [8] criticized the Brown-Johnson rationing scheme for its lack of theoretical justification; see also their reply [9].

contracts

$$(\rho_k, p_k), \quad k = 1, 2, ...$$

Each customer t chooses one contract $k(t)$ and a quantity $d(t)$ for which she pays $p_{k(t)}d(t)$. (We suppose that energy is measured in kWh and price is in dollars/kWh, so her bill is in dollars.) The understanding is that in period 2 the company will deliver $d(t)$ kWh of energy to our customer with probability $\rho_{k(t)}$. With the complementary probability $1 - \rho_{k(t)}$ she will receive no electricity. Thus ρ_k is the guaranteed reliability or availability of service if the kth contract is purchased.[5] Note that it is immaterial whether the customer selects the quantity $d(t)$ in period 1 or 2; it is important that $k(t)$ is selected in period 1.[6]

Step 2. At the beginning of period 2 the company finds out the actual value of energy supply, $S(\omega)$. Here ω denotes the sample point or *contingency*. The company now decides which customers to ration. This decision is represented by the 0-1 valued function $R_\omega(t)$. If $R_\omega(t) = 0$ customer t will not receive service, if it is 1 she will receive $d(t)$ kWh of energy. The company's decision must meet two constraints:

$$\sum_t R_\omega(t)d(t) \leq S(\omega), \quad \text{for all } \omega \tag{2}$$

$$\text{Prob } \{\omega \mid R_\omega(t) = 1\} = \rho_{k(t)}, \quad \text{for all } t \tag{3}$$

Equation (2) is a physical constraint which says that the total energy delivered cannot exceed the available supply for each contingency. Equation (3) is the social obligation to fulfill every customer's contract.

In §2 we formulate a welfare function and determine the allocation of energy to each customer which maximizes this function. We show that the structure of the optimal contracts is remarkably simple. In §3 we formulate the notion of market equilibrium for our market and relate the equilibrium to the welfare-maximizing allocation. In §4 we draw out some implications of the preceding analysis. Results of extended models that permit randomness in supply and inhomogeneous customers are announced in §5.

Interruptible service contracts and related topics are studied in Chao *et al* [12], Chao and Wilson [13], and Oren *et al* [14]. This study in part is inspired by that work from which it differs in two respects. Our rationing

[5] Although we use 'reliability' and 'availability' interchangeably, these terms are distinguished in power engineering [10, 11].

[6] In §5.1, where we introduce random demand, the customer selects her demand in period 2 after her random preference is revealed.

mechanism is differently specified.[7] Their specification of the customer's rationing loss is not explicit. The different specifications require different technical argument as well.

2 Maximum social welfare

The structure of optimal contracts is obtained indirectly by formulating a welfare maximization problem. In §3 it is shown that the optimum can be sustained as a market equilibrium.

2.1 Problem formulation

We first model the supply side. The total energy available (in period 2) to the company is a random variable taking values $s_i > 0$ with probability $\pi_i > 0$, $i = 1, ..., n$. The set of values $\{(s_i, \pi_i)\}$ is known in period 1, but which contingency s_i occurs is revealed only at the beginning of period 2. We also assume that the variable cost of energy supply (chiefly fuel cost) is zero. Non-zero variable cost is discussed in §5.

We now model consumer welfare. The demand of any individual consumer is assumed to be infinitesimal compared with the total demand of all consumers. This permits us to model the set of customers as a continuum indexed by $t \in [0, 1]$.[8] Suppose consumer t is allocated energy $d(t)$ with reliability $\rho(t)$. The net benefit to customer t, not including her electricity bill, is given by

$$w(t) = \rho(t)\ U(d(t)) - [1 - \rho(t)]\ L(d(t)) \tag{4}$$

The interpretation is that if t actually consumes energy $d(t)$ her utility is $U(d(t))$, and since this occurs with probability $\rho(t)$, the first term in (4) is the expected utility. But if service is interrupted she suffers a disutility of $L(d(t))$, and since that happens with probability $[1 - \rho(t)]$, the second term

[7] Their mechanism is insufficiently precise. Their analysis is conducted assuming each consumer demands exactly one unit of energy, with the understanding that a consumer needing two units (say) is treated as two consumers. But this appears to permit 'cheating'. For instance a customer may purchase one unit of energy with reliability ρ_1 and another unit with reliability ρ_2. The customer can then use the cheaper unit when it is available and the more expensive unit when the cheaper one is not available. This objection is withdrawn if measures are taken to prevent such cheating.

[8] With this convention the total number of customers is 1, so the supplies s_i are measured in average kWh per customer.

measures the expected rationing loss. The disutility will generally depend on $d(t)$ since the customer planned on using that amount.[9] It is assumed that

$$U(0) = L(0) = 0; \; U'(d) > 0, \; U''(d) < 0; \; L'(d) \geq 0, \; L''(d) \geq 0$$

These are standard assumptions: U is strictly concave, L is convex, and both are increasing.

The total social welfare is obtained by summing everyone's welfare,

$$W = \int_0^1 w(t)dt = \int_0^1 \{\rho(t) \, U(d(t)) - [1 - \rho(t)] \, L(d(t))\} dt$$

We are assuming that all consumers have the same utility function. (§5 discusses heterogeneous customers.)

We now consider the allocation problem. In period 1 each t is allocated a pair $(\rho(t), d(t))$. At the beginning of period 2, the contingency is revealed. Suppose it is s_i. It must then be decided which if any customers are to be rationed. This is given by a *rationing function* $R_i : [0,1] \to \{0,1\}$ defined as

$$R_i(t) = \begin{cases} 0 & \text{if } t \text{ is rationed in contingency } i \\ 1 & \text{otherwise} \end{cases}$$

The rationing function must satisfy the physical constraint

$$\int_0^1 R_i(t)d(t)dt \leq s_i \text{ for all } i$$

which simply says that supply meets rationed demand. The rationing functions must also meet the obligation to fulfill the contracts

$$\sum_{i=1}^{n} \pi_i \, R_i(t) = \rho(t) \text{ for all } t$$

The welfare maximization problem is to find functions $d, R_1, ..., R_n$ subject to these two constraints so as to maximize total welfare W. This can be reformulated as an optimal control problem. Introduce the 'state' vector x and the 'control' vector z,

$$x(t) = (x_1(t), ..., x_n(t)), \quad z(t) = (d(t), r(t))$$

[9] Imagine that t is running a business. Her failure to receive the planned energy $d(t)$ will cause a loss of $L(d(t))$ from the waste associated with unused labor and materials.

where $r(t) = (r_1(t), ..., r_n(t))$, $r_i(t) = \pi_i R_i(t)$. Then the problem can be reformulated as

$$\max\ W = \int_0^1 w(t)dt = \int_0^1 \{\rho(t)\ U(d(t)) - [1-\rho(t)]\ L(d(t))\}dt \quad (5)$$

subject to

$$\dot{x}(t) = \frac{1}{\pi_i} d(t) r_i(t),\ t \in [0,1],\ i = 1, ..., n \quad (6)$$

$$x_i(0) = 0,\ x_i(1) \le s_i,\ i = 1, ..., n \quad (7)$$

$$d(t) \ge 0,\ r_i(t) \in \{0, \pi_i\},\ \rho(t) = \sum_1^n r_i(t) \quad (8)$$

The Maximum Principle [15] gives necessary conditions for a solution of (5)-(8). However we are interested in sufficiency. Define the *Hamiltonian*

$$H(d, r, \mu) = \sum_1^n r_i U(d) - [1 - \sum_1^n r_i] L(d) - [\sum_1^n r_i \mu_i] d \quad (9)$$

H is defined for all $d \ge 0$, $r_i \in \{0, \pi_i\}$, $\mu_i \ge 0$.

Theorem 1 (Sufficiency) *Suppose there exist $\mu^* \ge 0$ and H^* such that for all d and r*

$$H(d, r, \mu^*) \le H^* \quad (10)$$

Then the maximum social welfare

$$W^* = \max\ W \le H^* + \sum_1^n \pi_i \mu_i^* s_i \quad (11)$$

Moreover, if there is a feasible control $z^* = (d^*, r^*)$ such that

$$H(d^*(t), r^*(t), \mu^*) \equiv H^*,\ \ \mu_i^*[s_i - x_i^*(1)] = 0, \quad \textit{for all } i \quad (12)$$

then this control is optimal.

Proof

Let z be any feasible control and x the corresponding state 'trajectory'. From (5),(6),(9) we get

$$\begin{aligned} W &= \int_0^1 H(d(t), r(t), \mu^*)dt + \int_0^1 \sum \mu_i^* r_i(t) d(t) dt \\ &\le H^* + \sum \pi_i \mu_i^* x_i(1) \le H^* + \sum \pi_i \mu_i^* s_i \end{aligned} \quad (13)$$

The two inequalities follow from (10) and (7). The second part of the assertion follows since (12) gives equality in (13). □

Condition (12) will later be interpreted as a market equilibrium. Let $H(t) = H(d^*(t), r^*(t), \mu^*)$. From (4),(9)

$$H(t) = w(t) - [\sum_i r_i^*(t)\mu_i^*]d^*(t)$$

Interpret $\sum r_i^*(t)\mu_i^*$ as the price of one kWH of energy with reliability $\sum r_i^*(t)$. Then $[\sum_i r_i^*(t)\mu_i^*]d^*(t)$ is the energy bill of customer t and $H(t)$ is t's expected utility net of expenditures on energy. $H(t)$ is t's *consumer surplus.* The first part of (12) asserts that every consumer ends up with the same surplus H^*, and (10) says that at the prevailing prices no customer can purchase energy to obtain a surplus larger than H^*. The second part of (12), $\mu_i^*[s_i - x_i^*(1)] = 0$, is the complementary slackness condition. As we will see later, it implies that at the prevailing prices the power company cannot increase its profits by offering a different set of contracts. Thus the two parts of (12) give conditions for consumer equilibrium and producer equilibrium. Lemma 1 is proved in the appendix.

Lemma 1 *$W^* > 0$ in (11). $H^* > 0$ in (10). If $z^* = (d^*, r^*)$ is optimal, then $d^*(t) > 0$ for all t.*

2.2 Optimal allocation

We give an algorithm to find H^*, μ^*, z^* satisfying (10) and (12). The algorithm also finds the optimal structure of contracts. A contract may offer any reliability level $\sum_{i \in J} \pi_i$ where $J \subset \{1, ..., n\}$ is any set of contingencies. Thus a company may offer up to 2^n different contracts. We will show that an optimal structure includes at most n contracts.

For $d \geq 0$, $\rho \geq 0$, $p \geq 0$, define the function

$$h(d, \rho, p) = \rho\, U(d) - [1 - \rho]\, L(d) - pd \tag{14}$$

This is the surplus derived by a consumer who purchases d units of a (ρ, p) contract (i.e. with reliability ρ and a price of p dollars/kWh).

Fix a level of consumer surplus $H > 0$. Define the *bid price* $p(\rho) = p(\rho; H)$ as the maximum that a consumer is willing to pay per unit of energy with reliability ρ if she is to attain surplus H. That is

$$p(\rho) = \begin{cases} \max\{p \geq 0 \mid \text{there exists } d \geq 0 \text{ with } h(d, \rho, p) \geq H\} \\ \text{undefined if there is no } d \geq 0 \text{ with } h(d, \rho, p) \geq H \end{cases} \tag{15}$$

From (14) $\rho \rightarrow h(d, \rho, p)$ is increasing in ρ for d, p fixed. Hence $p(\rho)$ is defined on a set of the form $1 \geq \rho \geq \rho_{min}(H)$. Also let

$$d(\rho) = d(\rho; H) = \arg\max_{d \geq 0} h(d, \rho, p(\rho)) \tag{16}$$

Thus $d(\rho)$ is the amount of energy the consumer will purchase at the bid price. Lemma 2 is proved in the appendix.

Lemma 2 *(1) $p(\rho)$ is strictly increasing, differentiable, and convex over $\rho \geq \rho_{min}$. (2) $p(\rho; H)$ is strictly decreasing and $d(\rho; H)$ is strictly increasing in H.*

Part (1) says that, with a fixed surplus, consumers will pay more for more reliable service.[10] Part (2) says that, at a fixed reliability level, consumer surplus will increase only if price of energy goes down and their consumption goes up.

For future reference note that

$$p(\rho_{min}(H); H) = 0 \text{ for } H > 0 \tag{17}$$

To see this, let $\hat{\rho} = \rho_{min}(H)$, $\hat{d} = d(\hat{\rho}; H)$, and suppose $\hat{p} = p(\hat{\rho}; H) > 0$. Then

$$\hat{\rho}\, U(\hat{d}) - [1 - \hat{\rho}]\, L(\hat{d}) = \hat{p}\hat{d} + H > H$$

Hence there exists $\rho < \hat{\rho}$, $d = \hat{d} > 0$ and $p = 0$ such that $\rho\, U(d) - [1 - \rho]\, L(d) \geq H$. But this contradicts the definition of ρ_{min}.

Algorithm

Step 0 Order the contingencies in decreasing order of severity,

$$o < s_1 < \dots < s_n$$

and define reliabilities $1 = \rho_1 > \dots > \rho_n = \pi_n$ by

$$\rho_m = \sum_{i \geq m} \pi_i$$

Step 1 Pick a trial consumer surplus $H > 0$. Let $1 \leq k \leq n$ be such that

$$\rho_1 > \dots > \rho_k \geq \rho_{min}(H) > \rho_{k+1}$$

[10] The lemma does *not* say whether the consumption of energy goes up or down with reliability; indeed examples with both behaviors can be given.

and consider the k contracts $(p_1, \rho_1), ..., (p_k, \rho_k)$ where

$$p_m = p(\rho_m; H), \quad m = 1, ..., k \tag{18}$$

By Lemma 2(1) we must have $p_1 > ... > p_{k-1} > p_k \geq 0$. Let

$$d_m = d(\rho_m; H), \quad m = 1, ..., k \tag{19}$$

Observe that although these k contracts are all distinct (each offers a different reliability), every consumer is *indifferent* between them since they all yield the same surplus H. See Figure 2.

Step 2 We calculate the number of customers that can be assigned one of these k contracts in amounts given by (19) without violating the supply constraint.

Define numbers $0 = t_0 < t_1 < ... < t_k$ as follows.

$$\begin{aligned} t_m &= t_{m-1} + \frac{s_m - s_{m-1}}{d_m}, \quad m = 1, ..., k-1 \\ t_k &= t_{k-1} + \frac{s_k - s_{k-1}}{d_k} \text{ if } t_{k-1} > 1 \end{aligned}$$

But if $t_{k-1} \leq 1$, then

$$t_k = \begin{cases} t_{k-1} + \frac{s_k - s_{k-1}}{d_k} & \text{if } p_k > 0 \\ \min\{1, t_{k-1} + \frac{s_k - s_{k-1}}{d_k}\} & \text{if } p_k = 0 \end{cases}$$

Above $s_0 = 0$. Thus consumer $t \in [t_{m-1}, t_m)$ is assigned contract (ρ_m, p_m) in the amount d_m, $m = 1, ..., k$. This assignment is feasible, i.e. all customers are assigned, if and only if $t_k = 1$. As the next lemma asserts we can increase t_k by decreasing H.

Lemma 3 *There is a unique consumer surplus H^* for which $t_k(H^*) = 1$.*

Proof

By Lemma 2(2), $d_m = d(\rho_m; H)$ is strictly decreasing in H. Hence the t_m are also strictly decreasing. Also $p_k = p(\rho_k; H)$ is strictly decreasing in H. The result follows from the construction of the t_m. (Note: In the lemma, the number of contracts $k = k(H)$ depends on H and decreases with H.) □

The proof of Theorem 2 is in the appendix.

Theorem 2 *Let H^* be such that $t_k(H^*) = 1$. Define the control $z^*(t)$ for $t \in [t_{m-1}, t_m)$, $m = 1, ..., k$, by*

$$d^*(t) \equiv d_m, \quad r_i^*(t) \equiv \begin{cases} \pi_i, & i \geq m \\ 0, & i < m \end{cases} \tag{20}$$

Let

$$\mu_m^* = \frac{p_m - p_{m+1}}{\pi_m}, \quad m = 1, ..., n \tag{21}$$

where $p_{k+1} = ... = p_n = 0$. Then H^*, μ^*, r^* satisfy (10) and (12) so that z^* is optimal.

We summarize the results of this section. Suppose $s_1 < ... < s_n$ are the supply contingencies. Then

1. The optimal contracts are $(\rho_1, p_1), ..., (\rho_n, p_n)$. The reliabilities are fixed by the supply contingencies, $\rho_m = \sum_{i \geq m} \pi_i$. That is, the mth contract guarantees delivery under contingencies $m, m+1, ..., n$.

2. The prices are determined by a single parameter H, $p_m = p(\rho_m; H)$, which also determines the demand $d_m = d(\rho_m; H)$. For a given H, these prices and demands depend only on customer preferences, U and L.

3. If k is such that $\rho_k \geq \rho_{min}(H) > \rho_{k+1}$, then customers will only demand the first k contracts. The remaining prices $p_{k+1} = ... = p_n = 0$.

4. The company 'produces' $D_m = (t_m - t_{m-1})\, d_m$ units of energy with reliability ρ_m. Moreover $D_m = s_m - s_{m-1}$ if $m \leq k-1$, $D_k \leq s_k - s_{k-1}$ (with equality if $p_k > 0$), and $D_{k+1} = ... = D_n = 0$. The revenue received by the company is

$$\sum_{m=1}^{n} p_m D_m = \sum_{i=1}^{n} \pi_i \mu_i^* s_i \tag{22}$$

   This follows from (21).

3 Market equilibrium

A market equilibrium is a feasible allocation of contracts that simultaneously maximizes consumer surplus and company profits. We have discussed consumer surplus already. We need to model company profit opportunities.

3.1 Company profit

Suppose the market consists of the n contracts $(\rho_1, p_1), ..., (\rho_n, p_n)$. The company can 'package' its contingent supplies into many different 'bundles' of contracts. It will choose that bundle which maximizes its revenue which equals profit since we have assumed zero cost of supply.

We determine the feasible bundles that the company can supply. A bundle is denoted $\Delta = (\Delta_1, ..., \Delta_n)$ and contains, for each m, Δ_m kWh of energy with reliability ρ_m. This reliability can be achieved by providing Δ_m kWh from s_1 with *conditional probability* $R(1, m)$, Δ_m kWh from s_2 with conditional probability $R(2, m)$, and so on. To obtain reliability ρ_m, it must be that

$$\pi_1 R(1, m) + ... + \pi_n R(n, m) \geq \rho_m = \sum_{i \geq m} \pi_i, \quad m = 1, ..., n \tag{23}$$

and $0 \leq R(i, m) \leq 1$. (This permits 'randomized' rationing; if this is excluded then $R(i, m) = 0$ or 1.) The bundle Δ must also satisfy the supply constraint,

$$R(i, 1)\Delta_1 + ... + R(i, n)\Delta_n \leq s_i, \quad i = 1, ..., n \tag{24}$$

The revenue received from Δ is

$$\text{Revenue } = \sum_{m=1}^{n} p_m \Delta_m$$

The company will produce that bundle which maximizes its revenue subject to (23), (24). This is a nonconvex programming problem.

Lemma 4 *For every feasible bundle* Δ,

$$\sum_{m=1} p_m \Delta_m \leq \sum_{i=1}^{n} \pi_i \mu_i^* s_i \tag{25}$$

Moreover, equality is achieved by the bundle $(D_1, ..., D_n)$ *in (22).*

Proof

Rewrite (23) as

$$\sum_{i<m} \pi_i R(i, m) \geq \sum_{i \geq m} \pi_i [1 - R(i, m)]$$

Since $0 \le R(i,m) \le 1$, each term above is nonnegative, and since $\mu_1^* \ge ... \ge \mu_n^* \ge 0$, we can conclude that

$$\sum_{i<m} \mu_i^* \pi_i R(i,m) \ge \sum_{i \ge m} \mu_i^* \pi_i [1 - R(i,m)]$$

or

$$\sum_{i=1}^{n} \mu_i^* \pi_i R(i,m) \ge \sum_{i=m}^{n} \mu_i^* \pi_i = p_m$$

where the equality follows from (21). Multiply (24) by $\mu_i^* \pi_i$ and add,

$$\sum_{i=1}^{n} \mu_i^* \pi_i s_i \ge \sum_{m=1}^{n} \Delta_m \sum_{i=1}^{n} \mu_i^* \pi_i R(i,m) \ge \sum_{m=1}^{n} \Delta_m p_m$$

to prove (25). The second assertion follows from (22). □

3.2 Equilibrium

We now define market equilibrium.

Definition Contracts $\{(\rho_m, p_m)\}$, consumer demand $t \to (m(t), d(t))$, and a producer bundle $\Delta = \{\Delta_m\}$, form a *market equilibrium* if:

1. Δ meets the aggregate consumer demand, i.e. for each m

$$\Delta_m \ge \int_{\{t:m(t)=m\}} d(t)dt$$

2. $(m(t), d(t))$ maximizes consumer surplus, i.e. for each t

$$(m(t), d(t)) = \arg\max_{m,d} \ \rho_m \ U(d) - [1 - \rho_m] \ L(d) - p_m d$$

3. Δ maximizes company revenue $\sum p_m \Delta_m$.

Theorem 3 *The contracts, consumer allocation, and producer bundle given in Theorem 2 form a market equilibrium.*

4 Implications

We draw out some theoretical and practical implications of the preceding analysis.

4.1 High prices to prevent rationing

Vickrey and Turvey argue for higher prices to reduce demand and rationing loss. Suppose that demand fluctuations can be reflected as supply contingencies as we have done here.[11] Then their argument calls for offering a single contract (ρ_v, p_v), where $\rho_v = \rho_1 = 1$. The company can produce only s_1 of this 'firm' energy. So each consumer's demand must be such that

$$U_d(s_1) = p_v$$

By contrast, with interruptible service contracts, the number of customers with the firm energy contract (ρ_1, p_1) is t_1 and their demand is $d_1 = s_1/t_1 > s_1$. Moreover the price is given by

$$U_d(\frac{s_1}{t_1}) = p_1$$

Comparing these two relations shows that $p_1 < p_v$. Thus interruptible service contracts leads to a *lower* price than the Vickrey-Turvey price.

Notice also that under their recommendation, in contingencies $2, ..., n$, the company will have unused excess capacity of $s_2 - s_1, ..., s_n - s_1$. The expected unused capacity is[12]

$$\text{Excess } = \sum_{i=2}^{n} \pi_i (s_i - s_1)$$

For electric power, the Excess can be large since it seems that π_1 is much smaller than 1, and s_1 is much smaller than the average capacity, $\sum \pi_i s_i$.

Of course an interruptible service contract presupposes the means to provide such service. For electric power this seems easy: appropriate signals to turn off and on a customer's power supply could be sent over a telephone or by modulating a power line carrier wave.

4.2 Meaning of interruptible service

A contract (ρ, p) stipulates that power may not be delivered with *probability* $1 - \rho$. Service contracts with such specification are not to be found in the real economy. This seems to be so for two reasons. First, a customer can

[11]This is shown in §5.1.

[12]This formula suggests that a practical measure of inefficiency is $p \times$Excess, where p is some price per kWh.

not tell how frequently in fact her service will be cut off since probability is an *ensemble* concept. Second, such contracts may encourage the company to turn off service with probability greater than $1-\rho$ since it will be very difficult to challenge.

Much more frequent in the real economy are *contingency* contracts stipulating that service will not be delivered under such and such objectively verifiable conditions. Fortunately, the contracts proposed here are of this type: service interruption depends on the realization of supply $s(\omega) = s_i$ which can be specified in terms of such and such failure in generation.[13]

4.3 Alternative implementations

In considering a (ρ, p) contract, a customer presumably evaluates her expected surplus

$$\rho\, U(d) - [1-\rho]\, L(d) - pd$$

Typically $1-\rho$ will be quite small, $L(d)$ will be quite large, and it will be difficult to calculate the expected loss $[1-\rho]\, L(d)$ with confidence. In any case, after the unlikely event of service interruption the customer will feel very differently than before it. One way of compensating against this is for the customer to purchase an insurance contract whose face value is $L(d)$ and whose premium is $[1-\rho]\, L(d)$. Such a contract is actuarially sound and the customer is less likely to feel differently before and after an interruption.[14]

A similar effect can be reached by a market in which an interruptible contract is replaced by a pair of contracts: a firm energy contract sold to the customer and a 'call option' sold by the customer to the company under which the latter can purchase back the energy in period 2 (tantamount to an interruption) at a fixed 'strike' price.

4.4 Role of reserve capacity

In the current fixed price regime, a power company maintains 'dispatchable' capacity on-line in order to meet unexpected demand increases.[15] This capacity is expensive. With interruptible service contracts, the power company can instead turn off service to some customers. In effect, a fraction of the consumer load becomes dispatchable, the amount of reserve capacity needed

[13] For an account of such an experimental program by Pacific Gas & Electric Company, see Ahlstrand [4].

[14] For deeper discussion of insurance schemes see Oren [16].

[15] 'Dispatchable' means that power can be increased quickly.

is correspondingly decreased, and the lower cost is passed on to consumers as lower prices. Some power companies have introduced analogous schemes: for example, lower prices are charged to consumers who permit the company to turn off their air-conditioners.

4.5 Contract design

This involves selecting the number n of contracts, and the reliability and price of each contract. From its generation plant data the power company can obtain the next period's availability curve

$$S(\rho) = \max \{s \mid \text{Prob } \{\text{supply } > s\} \geq \rho\}, \quad 0 \leq \rho \leq 1$$

$S(\rho)$ decreases with ρ. The number n and the reliability levels $1 = \rho_1 > \ldots > \rho_n$ are obtained by selecting a decreasing 'staircase' curve under the availability curve. See Figure 3 for an example with $n = 3$. A measure of inefficiency is given by the area between the two curves. As n is increased, this area can be reduced. But this must be balanced against the cost of implementing a large number of contracts.

To find the correct price for each contract is much more difficult. One could try estimate the preference functions U and L. Alternatively, the company selects a trial set of prices and adjusts them until an equilibrium is established. Such adjustment schemes need study.

A more accurate estimate of the loss in welfare resulting from a smaller set of contracts can be obtained by doing more work. Let $S(\rho)$ be the availability curve defined above; fix n and probabilities $\pi_1, \ldots, \pi_n$. Say that a supply list $s = (s_1, \ldots, s_n)$ is feasible if for each m

$$s_m \leq S(\pi_m + \ldots + \pi_n)$$

Let $W^*(s)$ and $\mu_1^*(s) \geq \ldots \geq \mu_n^*(s) \geq 0$ be the optimum values corresponding to the supply contingencies given by s. It is not difficult to see that W^* is concave in s and increasing in each coordinate; moreover, the vector $(\pi_1\mu_1^*(s), \ldots, \pi_n\mu_n^*(s))$ is the super-differential of W^* at s, see Tan [17]. Hence if s' is another feasible list we have

$$\sum_i \pi_i\mu_i^*(s)(s_i - s_i') \leq W^*(s) - W^*(s') \leq \sum_i \pi_i\mu_i^*(s')(s_i - s_i')$$

Now fix m and consider s' with $s_j' = s_j$ for $j \neq m$ and $s_m' = s_{m-1}' = s_{m-1}$. s' is feasible; moreover, the contract with reliability ρ_m is not offered under

s' since its supply $s'_m - s'_{m-1} = 0$. Thus there is one less contract in s' and the inequality above simplifies to

$$\pi_m \mu^*_m(s)(s_m - s_{m-1}) \leq W^*(s) - W^*(s') \leq \pi_m \mu^*(s')(s_m - s_{m-1})$$

which provides a better estimate of the welfare loss due to dropping contract m.

5 Extensions

We consider several extensions of the simple model analyzed above.

5.1 Random demand

Suppose the supply is fixed at s_0 but the preferences during period 2 are random with n possible sample values: (U_ω, L_ω), $\omega = 1, ..., n$.[16] The market operates as follows. In period 1, the company announces contracts $\{(\rho_k, p_k)\}$ as before. Customer t selects one contract $k = k(t)$. At the beginning of period 2 her random preference ω is revealed to customer t, and she demands d kWh so as to

$$\max_d \ \rho_k \ U_\omega(d) - [1 - \rho_k] \ L_\omega(d) - p_k d$$

The company then selects a rationing function $R_\omega(t)$ as before.

The general analysis with arbitrary preferences is complicated, see Tan [17]. However, one special case is easy. Suppose the preferences are given by the functions:

$$U_i(d) \equiv U(d - \xi_i), \ \ L_i(d) \equiv L(d - \xi_i), \quad i = 1, ..., n$$

where the random values

$$\xi_1 > \xi_2 > ... > \xi_n \geq 0$$

occur with probabilities $\pi_1, ..., \pi_n$, and the functions L and U are as before.[17] It can be shown that again the optimal set of contracts is of the form $(\rho_1, p_1), ..., (\rho_n, p_n)$ where $\rho_m = \sum_{i \geq m} \pi_i$. Suppose, using the notation of

[16] Imagine that the shift in preferences depends on the weather in period 2.

[17] Under this specification the demand functions corresponding to each pair (U_i, L_i) are 'horizontal' shifts of each other.

Theorem 2, that $t_m - t_{m-1}$ customers are assigned contract (ρ_m, p_m). It is easy to see that $t_1\xi_1 \geq t_2\xi_2 ....$

It now follows that the structure of the optimal contracts is identical to the case with *deterministic* preferences (U, L) and random supply with values

$$s_1 = s_0 - t_1\xi_1 < ... < s_n = s_0 - t_n\xi_n$$

occurring with probabilities $\pi_1, ..., \pi_n$.

The cases of random demand and independent random supply can be combined. Suppose the random supply takes values $s_1, ..., s_m$ with probabilities $\eta_1, ..., \eta_m$. Then the problem is equivalent to deterministic preferences (U, L), and random supply with values $s_i - t_{ij}\xi_j$, $i = 1, ..., m$, $j = 1, ..., n$, occurring with probabilities $\eta_i\pi_j$.

5.2 Heterogenous consumers

So far all customers had identical preferences (U, L). Suppose instead that there are T^j customers with deterministic preferences (U^j, L^j), $j = 1, ..., l$ and $T^1 + ... + T^l = 1$. Suppose supply is random with values $s_1 < ... < s_n$ with probabilities $\pi_1, ..., \pi_n$ as before. It again turns out that the optimal structure contains n contracts $(\rho_1, p_1), ..., (\rho_n, p_n)$ where $\rho_m = \sum_{i \geq m} \pi_i$ depends only on supply conditions as before. The prices $p_1 > ... > p_n$ are determined by consumer bid prices. However, consumers in different groups have different bid price curves, and the algorithm for determining the correct prices is much more complex, see Tan [17].

5.3 Variable supply cost

We have assumed that the variable cost of supply is zero. (Variable cost consists largely of fuel cost.) Suppose now that there is a constant variable cost of c dollars per kWh. This introduces only a minor change in the formulation. The welfare function (5) is replaced by

$$\int_0^1 \{[\rho(t)\ U(d(t)) - [1 - \rho(t)]\ L(d(t))\}dt - \int_0^1 c\rho(t)d(t)dt$$

where the last term is the variable cost. The rest of the argument of §2 goes through with obvious changes. The optimal contracts now take the form $(\rho_1, \hat{p}_1), ..., (\rho_n, \hat{p}_n)$, where $\hat{p}_m = p_m + \rho_m c$. Again p_m is given by (21) and $\rho_m c$ is the expected variable cost. It is customary to call p_m the 'scarcity' price.

Suppose more generally that the variable cost during contingency i is c_i.[18] The situation is now more complicated. The optimal structure still contains n contracts. Moreover, the reliabilities of these contracts is again given by an ordering of the random events except that it may no longer coincide with the ordering of the magnitudes s_i. Indeed the ordering now also depends on the demand, see Tan [17].

6 Conclusions

The airline and telecommunications industries were dramatically reconfigured by the increased competition following deregulation. Few predicted, for example, that the minimal cost airline operations would be based on the now familiar 'hubbing' strategy. It would be foolhardy to suggest how competitive forces might reshape the electric power industry. Judging from these two cases, however, it seems safe to expect that one outcome will be the increased differentiation of electric power service. This paper contributes to the study of one dimension of differentiation, namely the reliability of service delivery. It seems obvious that there is a large efficiency gain to be made by recognizing the diversity of customers and end uses and 'packaging' electric power to exploit this diversity.

The traditional approach of building new generation capacity to meet increased demand is not as viable and this may lead to perceived 'shortages'. Interruptible services, spot prices, and time of use prices are related approaches to 'demand management' that can relieve these shortages. This is an additional reason for studying such schemes.

Power system operations have the goal of meeting demand efficiently without jeopardizing system security. In this formulation demand is assumed *exogenous*, efficiency essentially means increasing the output power of the generators with least variable cost, and security is the ability of the system to maintain stability following a fault. The dropping of customer load is a measure of last resort taken to maintain security. Under an interruptible service regime, this would be common practice and it will induce a major change in power system operations. In effect, customer load will enter economic dispatch (and security) calculations on the same footing as generation since a change in either can be measured in dollar values. These topics need to be carefully researched so that future changes can be met

[18] This is usually the case since generation plant consists of different technologies with varying fuel cost.

with knowledge and confidence.

Appendix

Proof of Lemma 1

The control $d(t) \equiv \min_i\ s_i > 0$ and $\rho(t) \equiv 1$ or $r_i(t) \equiv \pi_i$, is feasible and gives welfare $W = U(\min\ s_i) > 0$. Hence $W^* > 0$.

If $d^*(t) \equiv 0$, then $W^* = 0$. So there must exist at least one customer τ with $d^*(\tau) > 0$. Also, since $H(0, 0, \mu^*) = 0$, we must have $H^* \geq 0$. The map $d \rightarrow H(d, r, \mu^*)$ is strictly concave. Since $d^*(\tau) > 0$, this implies

$$H^* = H(d^*(\tau), r^*(\tau), \mu^*) > 0$$

Finally, since $H(d, r^*(t), \mu^*) = 0$ if $d = 0$, we must have $d^*(t) > 0$. □

Proof of Lemma 2

For $\rho \geq \rho_{min}$ (15) and (16) imply the two relations

$$H \equiv \rho\ U(d(\rho)) - [1 - \rho]\ L(d(\rho)) - p(\rho)d(\rho)$$

$$\rho\ U_d(d(\rho)) - [1 - \rho]\ L_d(d(\rho)) - p(\rho) \equiv 0$$

Differentiating the first relation with respect to ρ and using the second relation gives

$$p_\rho(\rho) = \frac{U(d(\rho)) + L(d(\rho))}{d(\rho)} > 0$$

which shows that p is differentiable and strictly increasing. (Here and below subscripts denote partial derivates, $U_d = \partial U/\partial d$, etc.)

For each $d > 0$ define the function $\rho \rightarrow P(\rho; d)$, with d as parameter, by

$$P(\rho; d) = \frac{\rho\ U(d) - [1 - \rho]\ L(d) - H}{d}$$

This is an affine function of ρ, and since by (15), $p(\rho) = \sup_d P(\rho; d)$, it follows that $p(\rho)$ is convex. This proves (1).

It is obvious that $p(\rho; H)$ is strictly decreasing in H. Also, from (14) we see that $h(d, \rho, p)$ is strictly concave in d and so $d(\rho; H)$ is the unique solution of

$$\rho\ U_d(d) - [1 - \rho]\ L_d(d) = p(\rho; H)$$

The left hand side is strictly decreasing in d. Since p is strictly increasing in H, it follows that $d(\rho; H)$ is strictly increasing in H, proving (2). □

Proof of Theorem 2

Under z^* customer $t \in [t_{m-1}, t_m)$ obtains reliability

$$\rho^*(t) = \sum_1^n r_i^*(t) = \sum_{i \geq m} \pi_i = \rho_m$$

and energy $d^*(t) = d_m$. Hence by (19) she gets surplus H^*. Moreover, from (20) and the construction of the t_m,

$$\begin{aligned} x_i^*(1) &= s_i, \quad i < k \\ x_k^*(1) &= \begin{cases} = s_k & \text{if } p_k = \mu_k^* > 0 \\ \leq s_k & \text{if } p_k = \mu_k^* = 0 \end{cases} \\ x_m^*(1) &= x_k^*(1) \leq s_m, \quad \mu_m^* = 0, \quad m > k \end{aligned}$$

Hence (12) is satisfied. It remains to show that (10) holds, i.e.

$$H^* = \max_{d,r} H(d, r, \mu^*)$$

By Lemma 2(1) $p(\rho)$ is convex in ρ, hence

$$\mu_{m+1}^* = \frac{p(\rho_{m+1}) - p(\rho_{m+2})}{\rho_{m+1} - \rho_{m+2}} \leq \frac{p(\rho_m) - p(\rho_{m+1})}{\rho_m - \rho_{m+1}} = \mu_m^*$$

and so

$$\mu_1^* \geq \ldots \geq \mu_n^* \tag{26}$$

Now suppose (d^+, r^+) maximizes

$$H(d, r, \mu^*) = \sum_i r_i \, U(d) - [1 - \sum_i r_i] \, L(d) - [\sum_i \mu_i^* r_i] d \tag{27}$$

Since the term involving r_i in (27) is linear in r_i, it follows that

$$U(d^+) + L(d^+) - \mu_i^* d^+ \begin{cases} > 0 \Rightarrow r_i^+ = \pi_i \\ < 0 \Rightarrow r_i^+ = 0 \end{cases} \tag{28}$$

Now let $m = \min\{i \mid r_i^+ = \pi_i\}$ and consider the control value $(\hat{d}, \hat{r})$ given by

$$\hat{d} = d^+; \;\; \hat{r}_i = \pi_i, \; i \geq m \text{ and } \hat{r}_i = 0, \; i < m$$

It is straightforward to verify using (26), (28) that

$$H(d^+, r^+, \mu^*) = H(\hat{d}, \hat{r}, \mu^*)$$

But from (20) we see that

$$(\hat{d}, \hat{r}) = (d^*(t), r^*(t)), \quad t \in [t_{m-1}, t_m)$$

which shows that H^* is the maximum value of the Hamiltonian. □

References

[1] W. Vickrey, "Responsive pricing of public utility service," *Bell Journal of Economics and Management Science*, vol. 2, no. 1, pp. 337–346, 1971.

[2] M. C. Caramanis, R. E. Bohn, and F. C. Schweppe, "Optimal spot pricing: Practices and theory," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 9, pp. 3234–3245, 1982.

[3] R. E. Bohn, M. C. Caramanis, and F. C. Schweppe, "Optimal pricing in electrical networks over space and time," *Rand Journal of Economics*, vol. 15, pp. 360–376, Autumn 1984.

[4] P. T. Ahlstrand, "Demand-side real-time pricing," in *1986 Annual Report*, Pacific Gas & Electric Co., 1987.

[5] G. Brown and M. B. Johnson, "Public utility pricing and output under risk," *American Economic Review*, vol. 59, pp. 119–128, March 1969.

[6] R. Turvey, "Public utility pricing and output under risk: Comment," *American Economic Review*, vol. 60, pp. 485–486, June 1970.

[7] M. A. Crew and P. R. Kleindorfer, *The Economics of Public Utility Regulation.* MIT Press, 1986.

[8] M. Visscher, "Welfare-maximizing price and output with stochastic demand: Comment," *American Economic Review*, vol. 63, pp. 224–229, 1973.

[9] G. Brown and M. B. Johnson, "Welfare-maximizing price and output with stochastic demand: Reply," *American Economic Review*, vol. 63, pp. 230–231, 1973.

[10] C. K. Pang and A. J. Wood, "Multi-area generation system reliability calculations," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-94, pp. 508–517, March/April 1975.

[11] S. Kumar and R. Billinton, "Adequacy equivalents in composite power system evaluation," *IEEE Transactions on Power Systems*, vol. 3, pp. 1167–1173, August 1968.

[12] H. P. Chao, S. S. Oren, S. A. Smith, and R. B. Wilson, "Multilevel demand subscription pricing for electric power," *Energy Economics*, vol. 4, pp. 199–217, October 1986.

[13] H. P. Chao and R. B. Wilson, "Priority service: Pricing, investment, and market organization," *American Economic Review*, vol. 77, pp. 899–916, December 1987.

[14] S. S. Oren, S. A. Smith, R. B. Wilson, and H. P. Chao, "Priority service: Unbundling the quality attributes of electric power," in *EPRI Interim Report, EA-4851*, Electric Power Research Institute, Palo Alto, CA, 1986.

[15] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes.* Interscience Division of John Wiley & Sons, Inc., 1962.

[16] S. Oren and J. Doucet, "Interruption insurance for generation and distribution of electric power," *Journal of Regulatory Economics*, vol. 2, pp. 5–19, 1990.

[17] C.-W. Tan, *Prices for Interruptible Electric Power Service.* PhD thesis, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, December 1990.

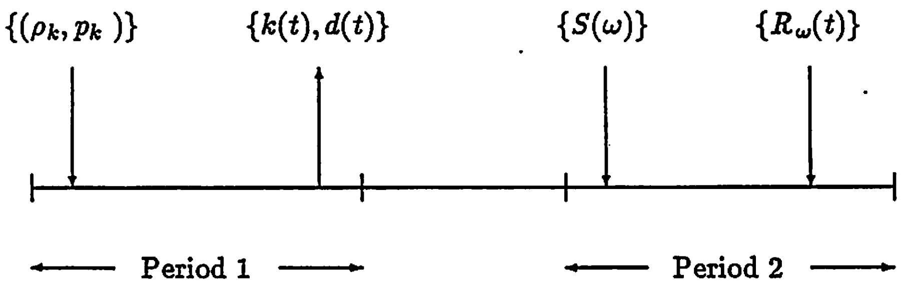

Figure 1: Market operation

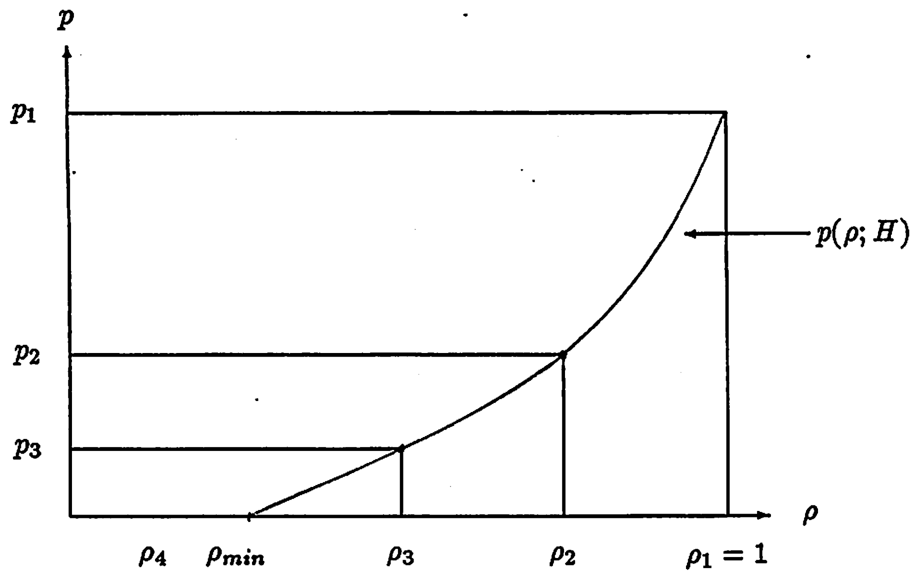

Figure 2: Contracts with $k = 3$, $n = 4$

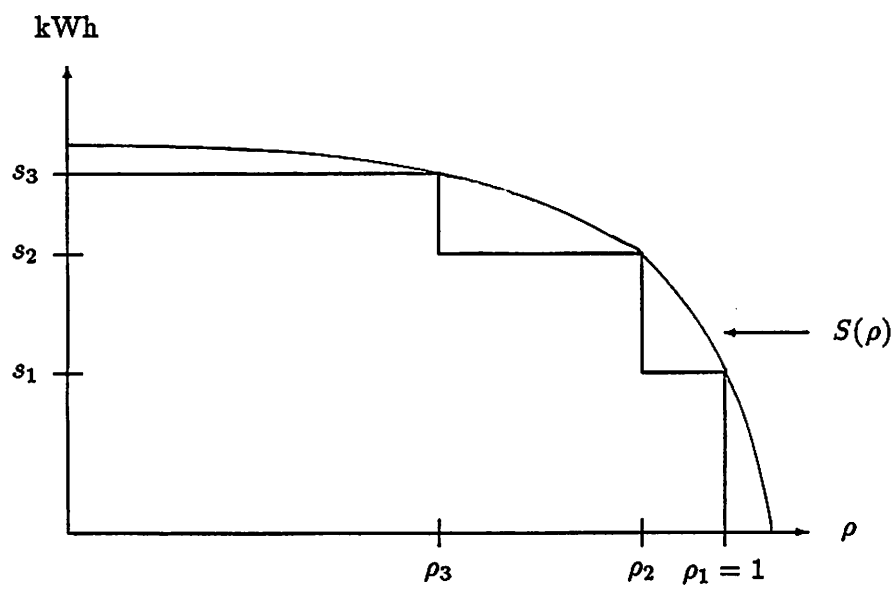

Figure 3: Selecting reliabilities