NRRI 90-5

RELIABILITY DIFFERENTIATED PRICING OF ELECTRICITY SERVICE

Narayan S. Rau, Ph.D. PE
Research Scientist

Youssef Hegazy
Graduate Research Associate

THE NATIONAL REGULATORY RESEARCH INSTITUTE
1080 Carmack Road
Columbus, Ohio 43210
(614) 292-9404

March 1990

This report was prepared by The National Regulatory Research Institute (NRRI) with funding provided by participating member commissions of the National Association of Regulatory Utility Commissioners (NARUC).

EXECUTIVE SUMMARY

It has been suggested that electricity customers be given the choice of opting for different levels of service reliability. Customers would then subscribe to the level of reliability that best suits them and their processes. Such a choice would unbundle the service. Implicit in the suggestion is the assumption that in the long run, unbundled service improves the lots of both customers and producers. In other words, reliability differentiated service improves efficiency.

The literature is rife with articles analyzing quality differentiated products. The quality of a tangible product is clearly understandable. However, "quality" used to imply the reliability of electricity supply is not an appropriate definition of either quality or reliability.

When the supply is available, it is put to use to achieve satisfaction or to produce some good. When the supply is not usable, either due to voltage or wave form variations, it is not of acceptable quality and, therefore, cannot be used. But, the curtailment of supply means that there is no electricity to be used. The continuity of supply or its curtailment measures the reliability (or otherwise) of supply. Therefore, from the point of view of our analysis, the definition of quality as a synonym for reliability is not acceptable.

Curtailments arise from several causes such as supply (or generation) shortage and outages of transmission and distribution (T & D) related equipment. In power systems, the latter causes lead to a majority of outages. The operation of the power system during emergencies is a complex task. The operator follows the natural human reaction to minimize curtailment (megawatt-hours lost, for example) during an emergency precipitated by supply and T and D shortages.

The purpose of this study is two-fold. First, we examine the practicality of implementing reliability differentiated service. Second, we compare economic welfare measurements as they pertain to reliability-based prices and other methods for setting the prices for electricity services. In essence, the reliability-based prices present customers with the prospect of electing to pay a particular price for a specific service reliability. However, these prices are not optimal in the sense they cause the maximization of consumer plus producer surpluses in a regulated environment. Instead, these prices are constrained by hour-specific revenue requirements for two customer classes. Hence, they are not efficient prices although they possess the potential to increase economic welfare when compared to other economically suboptimal regulated prices.

One concern is that the operation of the system as practiced at present is not conducive to reliability differentiated service. If the shortage is simply due to lack of generation at some location in the network, it is a simple task to trip customers who might not have subscribed to the highest reliability of supply. But, it may be impossible to continue supplying all other customers who might have subscribed to the highest reliability due to unfavorable voltage and reactive power conditions in the transmission network.

In the event of transmission or major distribution outages, the operator might have very little control over the continuity of supply to customers affected by it until the initiating event is cleared. Then, there might be a curtailment to some customers in spite of their subscription to a higher level of reliability while other customers, at another location in the network, might enjoy continuity of supply in spite of their subscription to a lower level of reliability. In a similar vein, during the restoration sequence after an outage, the operator has to follow certain switching sequences to retain integrity and to honor the technical constraints of the network, irrespective of the subscription of customers. Therefore, a mere examination of the interruption record will not tell regulators if the system was indeed operated to honor the customers' subscriptions. The regulator would have to examine the details of operating records for such an assurance. This would be a formidable task.

Another concern is that the evaluation of reliability indices itself is a complex task which is full of network-specific assumptions. The indices documented in the literature are generally intended to compare technological alternatives. They are not meant to give appropriate signals to the consumer as to the probability, time, or magnitude of outages. It would be difficult for consumers to interpret the indices and to decide on an appropriate level of subscription to reliability. Hence, the consumer cannot be expected to make rational decisions based on a set of reliability indices. Therefore, it would be preferable to establish a ranking of priority of service as a surrogate to the indices. During an emergency, the operator would endeavor to supply customers opting for a higher priority level as long as possible on a best-effort basis.

The establishment of a priority system does not allay the concerns outlined above. For example, a customer A in a certain location might opt for a very low priority of service and install local generation to cope with curtailments. His location might be on a major transmission line close to a generating station. It may be necessary to build additional transmission via his location to other customers for their future needs. Then, the reliability of customer A is automatically enhanced as a result of the additional transmission. Customer A will hold that the line would be built to serve others' needs and that the enhancement of reliability to him is incidental. The local generation installed by him cannot be dismantled. Therefore, he would not want to share the cost of the new line. Circumstances like this could make cost allocation an extremely difficult task. Customers desiring higher reliability could end up paying a disproportionately large premium for reliability.

In the literature, several analyses of optimum rationing have addressed the issue of implementing a priority scheme. Some make the assumption, in one form or another, that the probability distribution of the magnitude, duration, and time of outages are known a priori. Such knowledge about distributions is theoretical utopia. Despite this assumption, the goal of the analyses to minimize outage costs by optimizing the rationing process is well founded. Only under circumstances when the shortage can be apportioned is it evident that a proper rationing scheme--perhaps a rotational rationing--minimizes the cost of interruption to society. Rationing and classification of priority levels of service are ideally suited to serve noncore customers. However, it is important to recognize that the rationing schemes have nothing to do with reliability pricing. Under reliability pricing schemes, customers desiring lower reliability would pay lower prices. The price set gives a signal to induce a particular consumption pattern when the supply is available. On the other hand, the rationing scheme is applicable during an outage or a shortage.

In another body of literature, certain customers are referred to as interruptible. Such customers are cut off first from the system in the event of a shortage or other necessity. Since all customers are interrupted at one time or another, we use the terminology of core and noncore customers. Core customers are those whom the industry has an obligation to serve. Their needs and future projections are considered in resource expansion plans. Noncore customers are those who take the supply as and when it is available, after the needs of the core have been met. In other words, the utility has a reduced obligation to serve such customers. The supply of the noncore is similar to the spot market for commodities. Moreover, if there is more than one noncore customer, it is important to decide which is interrupted first. This decision can be based on their subscription to a particular priority of service classification.

The above matters regarding rationing and priority of service address operational aspects. From a planning perspective, one has to examine if reliability differentiated service offers any advantages in the longer term. Such longer-term considerations address the possible improvement in economic welfare or peak load reductions (leading to lesser reserve equipment) attainable by such pricing schemes. These issues lead us to the second major area of analysis.

We examined the welfare and reserve reduction possibilities due to different pricing schemes. The base case for comparison was that of the traditional average-cost pricing of electricity. Certain assumptions regarding price-demand relationships and consumption patterns were made. The analysis simulated the effects of different pricing schemes on two customer classes. Benefits due to these pricing schemes are compared to the base case which signifies the status quo.

The maximization of welfare, of course, is attained under marginal-cost pricing. The marginal cost may include longer term costs of providing the supply such as resource expansion and transmission costs. It has been proposed by some that electricity be priced at the hourly incremental cost of production. Such a pricing scheme is called spot pricing in some journal articles.

First, a simulation of spot pricing was undertaken. For the assumed fixed cost, this simulation indicates that the revenue to the producer would exceed total costs. Therefore, there was an overrecovery of fixed costs.

A method of revenue reconciliation following Ramsey would be to require the deviations from marginal cost to be equal to the inverse ratio of the core consumers' demand elasticities. We are able to follow this method because there are two core customer classes in our analysis. What is not clear, however, is how the revenue reconciliation should be apportioned between these two classes. While we know the total amount of fixed costs that must be recovered over the assumed time period, we do not have any a priori guidance thereafter. In order to overcome this problem, we assumed that the revenue reconciliation would be apportioned in the same proportion as the overrecovery. To clarify this further, let π be the total fixed costs to be recovered. Let the fixed cost recovered under spot pricing be π with $\Pi > \pi$. In a particular hour for a particular class, if the fixed cost recovered under spot pricing was c percent of π, c percent of π was recovered in the same hour from the same class in the Ramsey formulation. This pricing approach is referred to in the text of this report as Ramsey-type prices.

Next, the effects of one form of reliability-based prices were simulated. As with the Ramsey-type prices, the fixed cost was apportioned to each class in proportion to its hourly demand. But the demand allocation for one class was discounted by a factor x for customer class 1 or y for customer class 2 to account for the lower reliability desired by that class. The system would be operated to honor such subscriptions of reliability. When $x < 1$, customer class 1 subscribed to the lower reliability and implicitly customer class 2 subscribed to higher reliability. Conversely, customer class 2 subscribed to the lower reliability when $y < 1$. In those circumstances where both customer class wanted lower reliability, the mantle of lowest reliability was passed to that customer class with the larger increase in expected curtailments. Also, $x = 1$ and $y = 1$ means that neither customer class has made an explicit reliability subscription.

The above price rule means that any class-specific revenue reductions occasioned by the selection of a lower level of reliability has the result that the other class of customers will be obligated to provide more revenues toward the recovery of fixed costs. It is in this sense that the total amount of fixed costs recovered in each hour of the day corresponded exactly to that of the Ramsey-type prices.

Table 1 shows the results obtained as a result of these simulations. As expected, the maximum benefits in terms of economic welfare and peak load reduction occur under spot prices. Also, Ramsey-type prices fare better than the base case. The reliability-based prices show mixed results. In some instances, peak load reductions are greater than that expected to occur under Ramsey-type prices. Mitigating this effect somewhat are the results that in each of these cases the change in economic welfare is less than what occurs under Ramsey-type prices. All the same, for the cases examined reliability-based prices were associated with higher levels of economic welfare when compared to the base case. Note, however, that uncertainty exists with respect to the optimum choice of reliability by consumers. For example, when $x = 0.90$ welfare loss occurs when compared to the case with $x = 1$. Yet when $x = 0.85$ a welfare improvement is noted.

The peak load reduction and economic welfare simulations for reliability-based prices contain some specific numerical relationships. They are the parameters of the price-demand relationship for the two customer classes.

TABLE 1

CHANGES IN PEAK LOAD AND WELFARE
COMPARED TO THE BASE CASE

Pricing Method	Change in Peak Load, MW	Change in Welfare $
Base Case	--	-0-
Peak Load = 2,850 MW		
Total Energy = 51,809 MWh		
Spot Pricing	-454	+16,712
Ramsey-type Pricing	-266	+10,155
Reliability-Based		
x = 1.00	-250	+ 8,260
x = 0.90	-311	+ 7,476
x = 0.85	-368	+ 8,489
y = 0.80	-165	+ 6,241
y = 0.60	- 81	+ 2,007

Partly as a result of these relationships, table 1 shows that x = 0.85 causes a larger increase in economic welfare and peak load reduction than does x = 1. For reliability subscriptions other than x = 0.85, there are welfare losses in relation to x = 1. The detailed computer printout revealed net welfare loss in some hours and gain in others depending on the relationship between the price elasticities of the two customer classes. In general, subscriptions for lower reliability by a class with lower elasticity produces smaller increases in welfare than when the elasticity is higher.

Certainly, an explicit subscription for a certain level of reliability would be dependent on the cost of interruption. It is not related to the price elasticity of demand. Customers with lower expected outage costs and/or access to local standby generation are apt to opt for a lower level of reliability. If these customers also happen to have the lower price elasticities, their election will reduce peak demand and increase economic welfare. But this fortuitous relationship between price elasticities and the election of lower reliability may not happen in practice. It may turn out, instead, that the more elastic customers will opt for lower reliability. (This result occurs in our analysis when $y < 1$.) When this happens, the selection of a lower level of reliability provides welfare losses in relation to the case where no explicit reliability subscription is made by either customer class. Our results also indicate that economic welfare could improve over the base case even when the more elastic customer class chooses the lower level of reliability.

In view of the preceding results of our numerical analyses, the gains in economic welfare due to reliability-based prices are uncertain. It also is the case that reliability-based prices produce uncertain reductions in peak demand. Moreover, in none of the cases examined do reliability-based prices demonstrate better overall performance than Ramsey-type prices. These results suggest that Ramsey-type prices might be preferred to reliability-based prices for purposes of obtaining the best overall improvements in economic welfare and peak load reduction when compared to a base case of existing average-cost prices for electricity services. As a result, our analysis does not suggest any strong engineering or economic reasons to select reliability-based prices over other potential alternatives to average-cost prices.

True, the results of the numerical analyses could vary with the choice of different assumptions, particularly in regard to price-demand relationships. But, ignoring the actual values obtained, the trend of the different pricing approaches investigated herein appears to be that economic welfare can be improved and peak load can be reduced by substituting spot prices, Ramsey-type prices, and reliability-based prices for average-cost prices. In terms of reliability-based prices, they never appear to cause an improvement in economic welfare when compared to Ramsey prices. However, they may or may not represent an improvement over Ramsey-type prices when the predominant decision variable is the expected reduction in peak loads. Therefore, it is our conclusion that reliability-based prices hold few, if any, economic or engineering advantages over Ramsey-type prices.

The preceding conclusion is not altered when reliability-based prices are compared to average-cost prices. Although in all instances the reliability-based prices imply economic welfare and peak load improvements over average-cost prices, our analysis has revealed the potential for significant operational and administrative problems when an attempt is made to implement reliability-based prices. In particular, the operation of the power system to honor reliability subscriptions and regulatory assurances thereof pose important practical problems. Since the costs of rectifying these problems is not known at present, we cannot recommend reliability-based prices over average-cost prices.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF TABLES

LIST OF FIGURES

FOREWORD

There is current interest in offering to electric customers the option of selecting different levels of service reliability. This report examines the engineering feasibility and economic benefits of providing differentiated reliability service. Its conclusions are cautionary as to the workability of such arrangements.

Douglas N. Jones
Director, NRRI
Columbus, Ohio
March 1990

ACKNOWLEDGEMENTS

We would like to express our appreciation to Marilyn Reiss for the painstaking preparation of this manuscript and to Mr. David Wagman for his editorial comments and improvements. We thank Ms. Wendy Windle who prepared the illustrations. Finally, impromptu discussion with Drs. Duann, Graniere, and Rose of NRRI on certain aspects of this work were invaluable.

CHAPTER 1

INTRODUCTION

There is an emerging debate about offering customers the choice of purchasing electric power of differing qualities.[26,27*] Such a move, it is suggested, would enhance choice and efficiency in the marketplace.

In the literature, "quality" has been used to imply a host of attributes of the electricity supply. It is easy to define the quality of a tangible good as, for instance, an automobile. Its color, finish, efficiency, and mechanical reliability are some of the attributes that represent the quality of this product.

Defining the quality of electricity supply is, however, more involved. Electricity is used as the input to perform particular functions such as lighting, heating, entertainment, and to provide motive power. To a particular customer, therefore, the availability of electricity defines the satisfaction attainable from its use. Consequently, it can be argued that the uncurtailed availability constitutes the quality of supply.

For instance, a supply which is unavailable from, say, 4:00 P.M. to 8:00 P.M. every night is usually termed to be of poorer "quality" than another supply that is available throughout the day. Here, one assigns the term quality of supply rather than the more appropriate "continuity" to the supply because of one's experience with mechanical components such as the automobile discussed above. In the case of the automobile, its nonavailability due to breakdown represents a car of poorer quality. Hence, one extends one's perception of nonavailability of mechanical components to electricity so that a curtailment of supply is incorrectly implied to mean poorer quality.

To make this point more concretely, recall the assumption that the voltage, frequency, and wave form of the supplied electricity are acceptable

---

* Superscripts in bold type refer to bibliographic listings.

to end users. This assumption may not hold in some cases. Harmonics in the wave form and voltage surges may render the supply unusable. These occurrences undoubtedly cause some consternation among some customers even though there may be no curtailment of supply.

Of course, some end-use equipment is more tolerant of harmonics and voltage surges than other equipment. For example, the major sources of harmonics in the network are from certain types of switchable electronic loads. Electronic loads such as choppers and invertors inject harmonics into the network, which may result in lost data or even damage to computer loads and other computer-oriented processes connected to the electrical network. Therefore, debate regarding wave form purity is commonplace. The concern is not that the harmonics cannot be reduced or filtered out. Well accepted technical solutions exist such as installing power conditioners to improve the wave form or protect against voltage surges. The discussion mainly centers around the hardware or design aspects of such devices and who should install and pay for such devices, which may prove quite costly in some cases. It is increasingly felt that those responsible for the injection of harmonics and surges should pay for the installation of devices to mitigate such injections, and those that desire pure wave forms for special loads should install the required power conditioners and backup services.

It should now be clear that the word "quality" when applied to electricity does not mean the continuity of the supply of electricity. Instead, it refers to the usability of a given amount of available electricity. The problem of lamps flickering has been familiar to everyone from the start of the electricity industry. This does not mean, however, that the continuity of supply is unimportant. The provision of standby power in hospitals and airports to insure against interruptions is well known. Customers such as these place a high value on the continuity of supply.

A wide spectrum of requirements exists among even customers who value the continuity of supply more than wave form purity. Certain process industries--semiconductor manufacturers, for example--may want advance warning of impending curtailments so they can reschedule production to minimize losses. Residential customers are less concerned with prior notices of these outages because they can accept occasional, but short, interruptions in service. Others who depend on refrigeration or heating in their business activities (poultry producers, for example) are more concerned with the duration of the outage rather than the number of outages. These concerns are summarized in this report under the rubric of the reliability of the electricity supply.

It is not the intention of this report to address either the hardware aspects or the economic concerns associated with the quality of the electricity supply. Rather, the concern over the reliability of supply is the focus of this report. In particular, the two main categories of concerns are the possibilities of offering the customer the choice of service at different reliability, and the effects that reliability-based prices will have on the electric utility and its efficiency.

Chapter 2 examines the work done by others in this and allied areas, and discusses the main results of the literature pertinent to the main points of this report. Chapter 3 outlines the rudiments of networks and power systems, and discusses some illustrative examples of their operation. Some practical operating problems arising from differentiated reliability service are outlined. In chapter 4, the concepts of reliability and the methods used to quantify them are examined. Chapter 5 discusses the results of the numerical analysis of economic welfare developed in the appendix. Conclusions regarding the pros and cons of offering differentiated reliability service also are presented.

CHAPTER 2

PEAK-LOAD PRICING AND RATIONING AS PROXIES TO RELIABILITY PRICING

Several articles in the literature address peak-load or time-of-use (TOU) pricing of electricity and rationing in the event of a shortage. Other reports discuss the reliability-based pricing of the supply of electricity. In the following discussion, some basic aspects of electricity supply and its curtailment are outlined. Subsequently, a brief summary of some of the important contributions by others is presented to set the stage for the points made in later chapters.

Accounting for Curtailment Costs in Planning

The consequence of electricity curtailment is the starting point for electricity supply planning.[1,2] It is argued that the societal damage due to the interruption should be equal to the incremental cost of providing improved reliability by installing adequate equipment to avoid such interruptions. Telson[1] assumed that the maximum damage would be the loss of production in New York City.

At its face value, this concept is sound. We shall see later that from a planning and decision-making point of view, the terms "reliability," "cost of providing reliability," and "damage due to interruption" have to be more clearly defined and focused.

One of the pioneering efforts to assess the costs of interruptions was made by Ontario Hydro[3] which conducted a survey of its customers' perception of interruption costs, which is, in a sense, a measure of the value of reliability. This survey has been updated since 1975 and similar surveys have been conducted by others.[4,5,6]

As can be expected, the cost of an outage is different for different consumers. Table 2-1 shows the result of the survey conducted in Britain.[6] Detailed costs of interruption for North American utilities also can be

TABLE 2-1

ESTIMATED OUTAGE COSTS/ELECTRICITY PREMIUM
(£/kWh in 1977 price level)

	Outage Costs		
Consumer class	Short notice (min.)	Medium notice (1-2h)	Long notice (many hours)
Heavy engineering	0.57	0.44	0.31
Light engineering	0.55	0.43	0.31
Manufacture	1.28	1.00	0.72
Agriculture and food	0.83	0.83	0.75

Source: A. K. David, "The Variation of Electricity Prices in Response to Supply-Demand Conditions and Devices for Consumer Interaction," *Electric Power and Energy Systems* 8, no. 2 (April 1986): 101-14.

found in the literature.[4,5] The purpose of table 2-1 is to underscore the fact that the outage cost depends on the notice available to consumers and that it varies among customer classes. Furthermore, among a particular class, outage costs might vary from customer to customer. For instance, if one considers the agriculture and food industry, those who can reschedule certain activities and processes after the notice of an outage will have a lower outage cost than those with less flexibility to do so. Similarly, in manufacturing and engineering activities, the outage costs are process-specific. Kumar David has categorized certain industrial processes and has examined how the dynamics of the production process affects the costs of interruptions.[7]

Table 2-1 is not the complete story of interruption costs. The cost of interruption (and therefore the value of service) to the customer depends on the time of day, the nature of the interruption, and the characteristics of the customer. For instance, an interruption during the middle of the night usually results in less inconvenience to the residential consumer than a service curtailment occurring during peak-use hours. In addition, a residential customer may more readily tolerate one interruption of a certain duration rather than, say, five interruptions of one-fifth the duration. Either of these tendencies may not hold, however, if the single interruption were of sufficient duration to spoil food in refrigerators. In the case of industrial loads each interruption, in addition to lost production, might entail associated stopping and start-up costs.

Load and Resource Characteristics

A typical load shape of a North American utility for a working day consists of two peak periods of use. The first is around the noon hour and the second is at about 6:00 P.M. In seasons when space heating is required, the evening peak is greater in magnitude than the noon-hour peak. In seasons and areas where air conditioning loads come into play, it is possible that the load at noon and during the afternoon hours is greater than the evening peak. Yet across all seasons, the load demanded is not expected to go below a minimum value. This minimum load is the reason for the running of base-load units at 100 percent load factor. The load above the expected minimum is referred to as the time variant or the stochastic component of demand.

The resources of the utility used to supply these two components of demand consist of generators and purchased power. Under normal circumstances, the sum of the generation capacity and the purchased power in any season exceeds the peak load by an amount called the capacity margin. This margin is defined as the ratio of the difference between the resources and the peak load to that of the resources.

The resources available to supply electricity are not fixed throughout the year. They may vary from season to season as well as from hour to hour in any particular day because purchased power is available only during certain time slots, maintenance shutdowns, and the forced outage of generating units. Any time the reserve margin is negative, the utility will have to curtail the supply of electricity to its customers if it is to balance supply and demand. Any curtailment due to supply or generation deficiency may be shared or rationed to customers depending on the utility's operating practices and the network conditions. Of course, curtailment also could result from transmission inadequacy even if the generation were adequate. However, these curtailments are more difficult to ration or share among customers. Transmission and generation-related outages could cause a change in the power flow pattern and overloads in certain circuits resulting in unacceptable voltage profiles or unsafe operating conditions. Under such circumstances, rationing may not improve matters.

Even though the sources and causes of supply curtailments vary, it has been suggested that the price of electricity should be set in relation to its reliability of supply. In other words, customers who are willing to put up with more curtailments or interruptions should have the option of choosing such a service and paying a lower price than those who opt for a higher level of reliability. It is argued that the availability of these choices improves efficiency. The remainder of this chapter outlines some important analyses concerning reliability documented in the literature. Subsequent to that, the goals of our analysis and the difference between our focus and those found in the literature are clarified.

Definition of Terms

Several terms have been used to imply reliability in the literature. Among them are: quality, interruptible service, customer damage function, interruption costs, value of service, and peak-load pricing. From our point of view, it is essential to understand clearly what these terms mean and under what circumstances they imply reliability as we see it. Without such precision, our arguments might be misunderstood.

It has been suggested that the higher the load the poorer the quality. The logic behind this claim is that the reliability of supply is lower at higher loads because the reserve generation capacity (resources minus the load) is lower.[1] Therefore, the probability of a negative reserve is higher. It is argued, then, that the quality of power is diminished due to diminished reliability.

This definition of quality is not accepted for three reasons. First, quality is not synonymous with the availability of supply. Electricity is a factor of production used to produce some good or satisfaction. All electrons, irrespective of when they are made available, are just as good at obtaining an end result. In this context, the usability of electricity is the relevant item. The effect of availability, or rather the possibility of electricity being unavailable, causes contrarily cost or damage to the consumer.

A comparison of the characteristics of electricity supply with those of the water supply substantiates this point further. The quality of the water supply is determined by its nontoxicity, clarity, contaminants, and so on; that is, its usability. Therefore, the curtailment of water supply may be called poor service, but it certainly is not poor quality. It is true that the user may choose to store water to mitigate the effects of a curtailment. However, this approach is not economically possible in the case of electricity.

The second reason for not accepting the quality/time relationship of electricity is that its acceptance might lead into some traps of logic. For instance, the literature is rife with articles analyzing products of different quality. For instance, if a widget can be made of two different qualities, the one of poorer quality can be sold at a lower price. Economic literature, among other things, analyzes the optimal mix of production of widgets of differing qualities, the reaction of consumers, and the impact on social welfare. Accepting the wave form purity (quality)/time relationship for electricity would mean that the conclusions regarding pricing and market reactions reached in some of the economic literature could be applicable to electricity as well. In view of the complex nature of the production of electricity, it is extremely difficult to develop a metric for the quality of electricity and therefore, it is our view that the quality/time

[1] The discussion of methods to calculate indices of reliability is deferred to the next chapter.

relationship found in the economic literature is not applicable to the pricing of electricity. That this is true will be evident in subsequent chapters.

The third reason is that the quality/time relationship of electricity is dominated by demand and generation considerations, thereby relegating transmission and distribution concerns to secondary status. In reality, however, curtailments of electricity supply due to transmission and distribution failures constitute the causes for the majority of outages. While it is true that the transmission lines could be loaded to their capacity at the time of peak load, and as a result carry a higher risk of an outage, line failures happen more often due to lightning, ice, wind, contamination, and so on. These more likely sources of line failures can occur at any time of the day.[2] Given that transmission-related failures often make up 70 percent of curtailments and that such curtailments could occur at any time during the day, the acceptance of the quality/time relationship would oversimplify the problem of supply reliability.

Returning to the clarification of terminology, sometimes the term "interruptible service" is used to refer to the degree of reliability of supply at the customer level. Specifically, loads that are to be interrupted due to a shortage are referred to as "interruptible" power or energy in those studies that attempt to obtain an optimal rationing scheme. Yet, all loads are interrupted at one time or another depending on the cause or the severity of an emergency. Therefore, the term "interruptible service" does not clearly distinguish or delineate those classes of customers desiring different service reliabilities. Because of this shortcoming, customers electing interruptible service are deemed to be different from core customers. Core customers are those utility accounts that do not agree to be cut off to avert a shortage situation. However, these customers may elect to take service at different reliability levels. Therefore, every core customer will suffer interruption at one time or another, and some more frequently than others. Noncore customers are those served on an as-and-when-available basis. Utilities often call the loads of

[2] Transmission line failures during higher load periods may produce curtailments, while line failures at low load periods may not produce curtailments due to spare capacity or parallel paths for the flow of energy.

such customers interruptible because their supply is curtailed first in the event of a supply shortage or network-related problems. Presumably, noncore customers suffer more curtailments than core customers. In this report, service to noncore customers is termed interruptible service.

The distinguishing feature between core and noncore customers is that the utility is under no obligation to serve noncore interruptible customers. Of course, this does not mean that only noncore customers suffer service interruptions. Core and noncore customers alike are interrupted depending on the severity of the emergency. The sequencing of these interruptions is what is accomplished by the core/noncore classifications. Driving this sequence is the fact that a particular degree of reliability of supply for noncore customers is not planned or taken into account in a utility's resource expansion plans. The overall default degree of reliability for noncore customers is that which emerges implicitly after serving the core customers. Certain considerations in pricing such interruptible service are outlined in chapter 4.

The term "interruption cost" refers to the cost to the consumer due to the curtailment of electricity. Of course, the direct cost of an interruption depends on the type of consumer, its process, the time of day or year, the duration of interruption, and the number of interruptions. The indirect costs include but are not limited to loss of wages, losses due to looting and rioting, and so on. Damage functions, on the other hand, relate the interruption cost to kilowatts, kilowatt-hours, or other variables.[8,9] For instance, Billinton et al.[8,9] indicate a relationship between interruption cost, duration of curtailment, and kilowatts curtailed. The functional relationship between the above quantities is termed the "damage function."

Some Pricing Schemes

Peak-load pricing is a scheme in which the price of electricity is higher for on-peak hours than off-peak hours. However, the definition of on-peak and off-peak hours is rather arbitrary and depends on the shape of the demand. For instance, hours from 4:00 P.M. to 8:00 P.M. could be considered on-peak hours and the balance off-peak hours. The on-peak and off-peak hours could be different during different seasons and likely are different for different classes of customers.

The motivation for peak-load pricing arises from the fact that the incremental cost of producing energy (ignoring capital costs for the moment) rises with increases in kilowatt demand and in most cases is a concave function.[3] Additionally, this pricing approach is consistent with the cornerstone of economic theory that energy should be priced at the marginal cost of producing it to attain the maximum efficiency possible. Other approximations to marginal-cost pricing include time-of-day (TOD) or time-of-use (TOU) pricing. Under TOD pricing, the tariff is different at different time blocks of the day. At one extreme, there can be just two time blocks which result in peak-load pricing. At the other extreme, one could envision twenty-four blocks or an hourly pricing scheme tracking the marginal cost. Seasonal rates seek the same objective because they attempt to account for the changes in production cost that vary with the seasons.

Approximations to marginal-cost prices raise the issue as to whether prices should reflect short-run or long-run costs. The latter includes the cost of additional equipment and facilities to supply electricity at a future date. Therefore, it is fraught with uncertainties regarding the future costs of equipment, forecast demand, resource expansion, and so on. Furthermore, there is no general agreement regarding the division of the extra revenue, if any, that may result from prices based on long-run marginal cost.

One of the suggestions of Schweppe et al.[10] to set marginal-cost pricing in place has been termed the "spot pricing" of electricity. Simply stated, the price of electricity at every hour is set equal to the operating cost of production at that hour. Such a pricing scheme tends to maximize economic welfare. Its benefits are outlined in other publications.[11] Under this scheme, since the hourly price does not include capital costs, there likely will be a need for a revenue reconciliation adjustment. In other words, the unadjusted revenue to the utility under spot pricing could be more or less than that allowed under the traditional cost-recovery-plus-return method of regulation. Too many dollars can be received by the

[3] The mathematical definition of concavity is given in the appendix.

utilities when the rate base and associated capital costs are low. Too few dollars may accrue to the utilities when the rate base and associated capital costs are high.

The approach that attempts to solve the above dilemma owes much to Ramsey,[12] Boiteux,[16] and Baumol.[13] The suggested tool is to have deviations in prices from marginal cost to different classes of customers in the inverse ratios of their elasticities. It can be shown mathematically that such a scheme maximizes economic welfare subject to the constraint that the revenue recovery is a certain allowed sum. This scheme is not without critics who argue that it is unjust to the most inelastic customers as they would have to bear the biggest price increases or decreases. This pricing approach has been extended to include reliability constraints in the appendix.

The traditional method of pricing electricity is sometimes termed "average-cost pricing." Its impact is that the total cost of production is apportioned to consumers based on their consumption. Therefore, the energy cost component of the tariff is the average operating cost of producing electricity. Capital and other fixed costs usually are apportioned in the ratio of consumers' peak demands.[4] Since the cost of producing electricity is at its highest at the time of peak load, any demand charge, made explicit, can be deemed to be a penalty for use at that time. Note that demand charges may not be averaged over the entire day or month for industrial and commercial customers, but they are set at a certain value per megawatt used at the time of peak. For residential customers, however, the demand charge is usually averaged over the consumption period, thereby making it implicit.

Goal of Pricing Schemes

A pricing scheme should satisfy many attributes. From a regulator's point of view, it should maximize welfare and provide adequate and fair revenue to the supplier. In addition, an economist might argue that it

[4] There are several methods, such as coincident peak, noncoincident peak, and average and excess, for this apportioning. These nuances do not affect the main thrust of the discussion.

should give proper price signals and discourage waste and inefficiency. In certain sections of society and in some countries, the supply of electricity is viewed as an instrument of social and national policy. Its price, therefore, should be examined in a global sense considering industrial economy and unemployment in addition to efficiency and equity.

This report analyzes the maximization of economic welfare and the operational efficiency aspects of several approaches to pricing electricity. It does not address issues of social policy vis a vis these pricing schemes. In the following sections, the highlights of some work in this or allied areas are summarized and briefly critiqued. Subsequently, the character of our analysis as it relates to these works and the sense in which our goal is disparate from them will be made clear.

Summary of Relevant Work by Others on Pricing

Munasinghe has examined the issues of optimal planning by addressing the aspects of welfare gains, effects of shortages and interruptions, and reliability.[2] The incremental cost of changing the reliability of supply has been equated to the incremental change in total benefits. Total benefits include, among other things, changes in supply costs, changes in shortage costs, and the effect of changes in the reliability of demand. The article does not discuss how these costs should be quantified.

Marchand assumes that the loads (power demand) have a certain probability distribution and are noncyclic and nonperiodic.[14] He examines a rationing rule to interrupt loads if they exceed capacity resources. The analysis does not address the identification of optimal resources for the future.

Smith analyzes efficient menu structures for interruptible service.[15] In his analysis, it is assumed that the cost of interruption is expressible in the form of $c + vT$ where c is fixed cost (per kilowatt) per interruption, v is cost per unit time of interruption, T is the duration of interruption and is of a suitable value. A density function for the interruption cost $d(v, c)$ for parameters v and c is assumed. It is assumed that $f(L, T)$, a probability density function of shortfall of magnitude L and duration T, is known.

The analysis examines the magnitude of loads and the duration of interruption for optimal benefits (a priority scheme). The analysis also examines rotations of load interruptions among the customers.

As will be seen in later chapters, it is extremely difficult (perhaps impossible) to obtain the above probability density functions using currently available methods. Furthermore, these density functions are different at different locations in the power system network.

As noted previously, TOD pricing, in a sense, is a proxy to marginal-cost pricing, and it discourages consumption at the time of peak. Several investigations of this are reported in journals. The pioneering effort was due to Boiteux[16] (1949) (which had an impact on the policies of Electricite' de France) followed by many studies in Anglo-American journals. In these analyses, it is generally assumed that the demands in two or more periods are independent. In other words, there is a certain demand-to-price relationship for each period that is independent of any other time period. That is, these relationships are not interrelated.

Such an assumption from a practical standpoint may or may not hold depending on the type of load or customer. Consider, for example, air conditioning or space heating applications. In response to a higher price signal, the user turns the thermostats higher or lower. During periods of lower price, the user reverts to the comfortable use patterns by resetting thermostats to their original position. For this case, therefore, the assumption is tenable that the demand-price relationships in the two periods are independent. However, the following is an example in which such an assumption would be invalid.

Water heater, washing, and drying loads are reduced at the time of higher prices and reappear at the time of lower price due to postponed or shifted activity. Certain industrial processes also possess similar characteristics. It is evident from the above that the assumption of independence implies conservation or a reduction in net energy use from TOD pricing. When total energy usage is constant, as in the case of water heaters, such an assumption would not reflect reality.

Ischirhard and Jen examine a monopoly offering interruptible service.[17] They assume that only the residential class load (or that of only some classes) are stochastic. The other classes with whom the supplier has a contract are considered deterministic. Further, it is assumed that the probability distribution of the random component of the load is known as is the probability distribution of demand as a function of reliability of service. A priority scheme of interruption based on the service reliability is assumed. The goal of the monopolist then is to maximize profit. They then derive the mathematical conditions for profit maximization constrained by the interruption scheme.

Mussa and Rosen examine the situation when a consumer maximizes utility (subject to a budget constraint) when products of different qualities are offered.[18] It is assumed that there are constant costs of producing a good of a given variety and increasing marginal costs for the higher quality items.

For our purposes, the assumptions of the above listed studies are rather simplistic. For example, the loads of industrial and other consumers with whom the supplier has a contract have been assumed to be deterministic. Given that these loads vary with time, they are stochastic in the mathematical sense. Further, the assumption of a known probability distribution for service reliability is an extreme theoretical simplification. It will be shown in subsequent chapters that it is extremely difficult to obtain such a distribution because the methods for doing so are fraught with assumptions. Similarly, the assumptions that f(L, T) is known raises some practical problems.[15] It will be seen in a subsequent chapter that such distributions are impossible to obtain because no known method exists at present to do this. Reliability itself has no meaning unless it is qualified by the number of outages expected and the expected duration of these outages because both of these factors influence demand. As will be seen later, it is difficult (perhaps impossible) to calculate such reliability indices. In summary, therefore, methods that assume a demand function related to price and reliability would be theoretical and simplistic, while yielding results that cannot be implemented practically.

Goal of Analysis

It can be observed from the above that the reviewed analyses are for a static situation in that the characteristics of generation, transmission, and loads are fixed at a given time. Moreover, the above analyses have been set in the context of maximizing profit or optimizing rationing schemes.

The goal of this report is disparate from that of the above analyses in that it examines a dynamic or evolving electricity load and generation capacity from a planning point of view. That is, we want to examine the effects of offering different prices for different service reliabilities on the possible reduction of reserve equipment (or peak load) and on economic welfare. Two questions are addressed: Can the supplier define reliability in a manner understandable to the consumers so that they can make their choices?, and, what are the difficulties in doing so and in operating the power system according to such a definition of reliability and the resulting choices of consumers?

There are two distinct sets of considerations in examining the linking of price to the reliability levels. It is important to distinguish between these considerations and to contrast them. The first set of considerations relate to *planning*. The objective is to extract reductions of peak load (hence the reserve equipment) along with changes in production costs, changes in predicted interruption costs, and changes in economic welfare.[5] The second set of considerations is an optimal allocation of any shortage; that is, rationing. In this instance, the objective is an optimal rotational shutdown sequence and its timing, and the selection of customers for such rationing. The second set of concerns addresses an *operational* aspect in that given that a shortage has arisen, the most economical and efficient rationing scheme has to be chosen. The first set of concerns, on the other hand, plan for an optimum expected shortage, or for its converse optimum reliability. In this report, these distinct considerations are termed "planning and operating considerations." As noted in the introductory chapter, our main concern is that of planning. Therefore, we

[5] The goal of peak-load reduction resulting in an attendant reduction in reserve equipment could cause serious operating problems if carried to the extreme. The major problem would be that of equipment maintenance. Maintenance on generation and transmission equipment is generally carried out during periods of lower load when they can be taken out of service. If the load shape were constant, this "window of opportunity" would be lost. Therefore, the reduction in peak with a corresponding "valley filling" at some point might require installation of redundant facilities to facilitate maintenance.

examine the pros and cons of pricing electricity depending on its reliability level and the longer-term effects of such pricing on the utility industry.

Proponents of reliability-differentiated service argue that the consumer is best served when the total cost of electricity is minimized and the marginal cost of service reliability is equal to the marginal value of the interruption costs. Their premises are that it is inefficient to serve all customers at the same reliability level, and that reliability levels should be unbundled giving customers additional choices. These arguments, sound in theory though they may be, pose many practical problems. The biggest appears to be that of obtaining a universally understandable definition of reliability and an accurate estimate of unreliability. Therefore, this report also sets out to examine the problems associated with defining reliability, establishing the level of reliability, and obtaining defensible estimates of the impacts of alternative levels of reliability on customers.

CHAPTER 3

THE STRUCTURE OF THE POWER SYSTEM AND ITS OPERATION

In this chapter, the topology of the network as it relates to the sources of generation and its operation in relation to reliability are outlined. The causes of outages and the mitigation of their effects are discussed. Such an explanation aids the readers unfamiliar with power systems by helping them to understand the various concepts of reliability discussed in the next chapter.

Example of a Power System

System Structure

Figure 3-1 illustrates a power system. There are three sources of generation at locations A, B, and E and interconnections to a neighboring system at A and F. Generation at these locations is stepped up to 230 KV for transmission to load centers. Two major load centers at locations C and D are connected to the generation sources by lines at voltages indicated in the figure. There is a sizable load near the source of generation at location A that is fed at 69 KV. At location C a major load is supplied at 230 KV and an industrial load at 5 KV. Other loads in the area served at 230/110 V and 400 V are residential and commercial loads respectively. Location D represents a large metropolitan area where major industrial loads are served at 110 KV, 66 KV, and 13 KV. Commercial and residential loads also are shown in the figure.

A real power system is much more complex than this illustration. However, the structure of this illustrative system is useful for discussing certain characteristics of a power system. All the major transmission lines, feeders, and loads are protected by appropriate circuit breakers or switches, not depicted in the interest of simplicity. Also, switchable

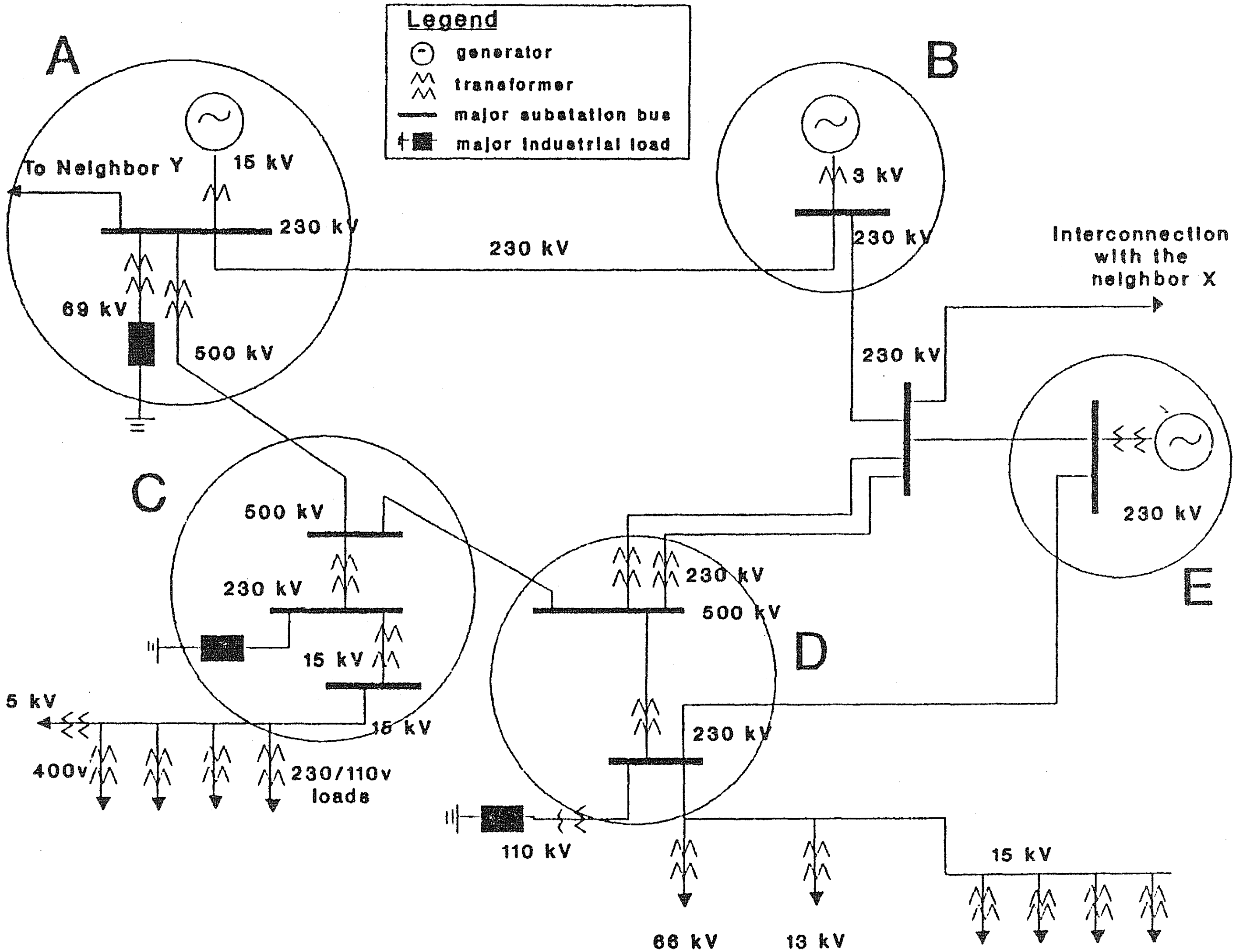


Fig. 3-1. Illustrative power system.

shunt reactors and/or capacitors located at various locations to maintain voltages at acceptable levels are not shown.

Operation of the System

The operation of a power system involves several aspects such as load dispatch, maintenance of facilities, procuring fuel, billing, and so on. The following is a description of one such activity, that of load dispatch, as it impinges on reliability.

Load Balance

Since the generated electricity should match the load or the demand at every instant,[1] the energy generated within the power system plus electricity imports must be equal to consumption within the area plus electricity exports plus system losses, at all times. The operator dispatches energy from A, B, and C in the most economical manner using first the unit with the least expensive incremental cost of production and then those with higher incremental costs. In other words, all sources of generation are loaded to levels so that the incremental costs of production from all the machines are the same. This is termed "economic dispatch." In addition, the operator ensures that the lines are not loaded beyond acceptable limits and that the voltage at each bus bar is at a specified value. Following are two examples of how an operator performs this function. In these examples, the times of occurrence of certain events and the amount of generation from each source are important.

Assume the electricity imported from neighbor X is 100 MW and the electricity exported to neighbor Y is zero. What would happen if the transmission line to neighbor X was lost? By automatic tripping, the line from neighbor Y would suddenly be loaded to 100 MW of electricity imports to make up for the loss of supply from X. The operator has to reduce this uncontracted for electricity import to zero within a specified time by

[1] This ignores storage capabilities since today's technology does not allow large storage economically.

increasing generation from locations A, B, or E. Under these crisis circumstances, it may not always be possible to increase generation from the most efficient source due to many reasons, including loading limits on the machines, unacceptable voltage profiles arising from increased generation from a particular source, nonavailability of fuel or water (including frozen coal) at one or more locations, and/or certain transmission lines or line sections and/or generators being on scheduled outage (maintenance). Any decision to increase generation from not necessarily the cheapest source to cope with an emergency situation is termed "going off economic dispatch."

Now assume that the generation at location A is lost, say due to an outage. The transmission line from location E to location D could become overloaded due to increased importation of electricity from neighbors Y and X to feed loads at locations C and D. Even if the lines between locations E and D are not loaded to their thermal limit, a further loss of any line between these locations could overload the remaining line leading to automatic tripping and system failure (blackout).[2] The operator therefore has to take immediate action to increase generation from either location B or location E to make up for the shortfall due to the failure at location A. If it is not possible to increase generation within the system, it may be necessary to reduce load by any of the methods discussed below.

Load Reduction

There is more than one reason for an inability to match demand and generation. Shortage of fuel/water, more-than-anticipated demand due to adverse weather, and outages of machines, transformers or transmission lines are reasons for such a mismatch. Under these circumstances, one can resort to several actions depending on the magnitude of mismatch. Small, predictable mismatches are generally handled by purchases from neighbors and perhaps by appeals for voluntary reductions in consumption. A more severe mismatch may require the reduction of voltage followed by or accompanied by

[2] This is called second contingency due to the first being the loss of generation at A.

a frequency reduction.[3] A reduction of network-wide voltage, of 5 percent say, would result in end-use equipment working below their rated capacities.[4] The most severe cases induced either by large mismatches or sudden failures of equipment that do not allow sufficient time to put the above actions in place produce curtailments through automatic underfrequency load shedding or as the central operator switches off load clusters. From a reliability standpoint, these curtailments are generally viewed as interruptions whereas the earlier actions of voltage and frequency reductions and appeals to customers to lower usage may not be perceived per se as an interruption. On the basis of these examples, it is evident that a system failure may cause load curtailment if it occurs at certain times, whereas the same event occurring at another time might not result in a curtailment.

The above matters are of importance in identifying a proper metric to measure and quantify reliability and in designing methods to evaluate reliability.

The Contradiction in the Subscription of Different Reliabilities of Service

Before we turn our attention to the methods of evaluating reliability, it is beneficial to examine the reliability of supply in a qualitative sense.

Consider the 69 KV load supplied at location A. If this load is less than or equal to the generation at this location, it is correct, in theory, to treat the generation at location A as if it were dedicated to supplying this load. In other words, the reliability of serving this load could be unaffected by failures of the transmission and distribution facilities in other parts of the system as long as a suitable operating strategy was employed. Whether this is the operating practice at the time of a capacity shortage is another question.

[3] Reduction in frequency is of the order of 0.1 or 0.2 HZ. Most networks cannot be operated continuously without damage to equipment below about 59.6 HZ.

[4] Lamps burn slightly dimmer and induction motors work at a lower power output, etc.

Now assume there is a transmission outage or a transmission and generation outage at some location in the power system other than location A. Some loads at, say, location D may have to be curtailed or dropped off automatically because unacceptable voltage conditions have arisen as a result of either of these system failures. In other words, system-wide rationing of the capacity according to a preestablished priority schedule based on contracted reliabilities which includes the 69 KV load at location A may not be suitable for restoring the network's voltage conditions. Therefore, certain loads at locations C and D, irrespective of their subscription to a certain level of reliability, will have to be curtailed before the load at location A, and they could be curtailed automatically due to relaying action. Suppose now that the customer with the 69 KV load at location A has opted for the lowest level of reliability, and the customers at locations C and D have opted for a higher degree of reliability. Because the loads at locations C and E have been curtailed by assumption and by construction of the problem, the load at location A can still be supplied. This example illustrates that the reliability of loads is dependent on their location; that is, their topological position in the network.

Is it possible to avoid the above set of circumstances, and to meet the expectations of those customers that have opted for a higher degree of reliability? One way is to build additional lines to those locations. An alternative is to install standby generation at locations C and D. Another option is to encourage customers at these locations to install emergency power generation equipment,[5] as hospitals and other customers with essential loads do. What is important in this context is that the utility honors its contracted for levels of reliability to different customers. But obviously, the costs of doing so may not be inconsequential.

There are additional concerns in relating price to the reliability of service. For example, let an industrial customer at location C opt for a lower reliability of supply. Then this customer may undertake the expense of installing local standby emergency generation to cope with the expected larger number of interruptions. Assume that the utility's operating

[5] An aware customer (particularly industrial customers) would examine the economics of upgrading the reliability by any of these means vis a vis the option of subscribing to a higher reliability of supply.

practice also reflects the wish of the customer at location C to obtain lower service reliability. Now consider what happens as the demand at location D increases over time. This increase could be met either by the installation of additional generation at location A or purchases from neighbor Y. In terms of transmission, a line from location A to location D via location C would be built.

Clearly in the above example, the reliability of supply to the customer located at C would also increase. Because it would be foolish to think of an operator deliberately curtailing service to location C to maintain the old level of reliability, the customer at location C might argue that the cost associated with the additional line should not be allocated to him since the new line was not to move additional power to him. He might hold that the enhanced reliability is incidental to his needs and desires. He would argue therefore that his reliability has increased due to the nature of the network's evolution. Consequently, this improvement in reliability is not relevant from his perspective because he has installed local emergency generation.

Such arguments and contentions induced by reliability-based prices might make the customer service representative's lot a contentious one. At present, additional network costs are allocated whenever possible to customers in proportion to their demand or consumption depending on whether these costs are demand or energy related. Under reliability pricing, a customer (a large one in particular) who does not increase demand or consumption and has no demand for additional reliability may not want to share the costs of new construction. As an allied matter, if a customer satisfies his additional demand by locally installed generation, he feels that he owes the utility nothing more on the basis that he needs neither more energy nor reliability. In this instance also, the customer may believe that the burden of additional costs should not fall on him.

On the other side of the coin is the question of the avoided cost to the utility. If customers with local generation sell power and energy to the utility, it could enhance reliability to other customers. What is the reliability of such a supply worth to the utility? How should one account for reliability and worth under conditions of competitive bidding? These questions do not have a ready answer and would have to be pondered if reliability-differentiated price is put in place.

Another area of concern is that of practical implementation. Since a utility generally serves more than one customer, each could subscribe to different levels of reliability. Even if the subscriptions were restricted to two levels of reliability, the combination of all possible events of outages and contingencies lead to a prohibitively large number of different operating decisions, much more than under the existing scheme of one level of reliability. One might consider computer-aided intelligence to solve this problem, but the technology in its present state would require a program providing instructions for several hundred if not thousands of combinations for making reliability-related decisions.

There also is the problem of the regulator ensuring that the contracted reliability has been delivered. Our example of the 69 KV load at location A has illustrated that a comparison of the curtailment records alone may not reveal if the contracted reliability was delivered. Recall that in the example, the customer with the 69 KV load at A subscribed for a lower reliability and yet is likely to have a lower number of interruptions. Therefore, if all curtailments have to be examined, it would mean a mammoth task of examining the details of all operating records and practices.

Commercial and residential customers when compared to industrial customers are more distant (electrically) from the source of generation. Many components could fall between these customers and the sources of generation including fuses, transformers, distribution and transmission lines, and so on. Therefore, the interruptions to these customers are expected to be greater in number than to those severed at a higher voltage. Again, this shows that a mere comparison of interruption records of all customers without due attention to their topological position will not indicate to the regulator if contracted loads of reliability were indeed delivered. In a similar vein, it is clear that it would cost more to enhance the reliability of supply to such consumers due to the larger numbers of components involved and the redundancies that have to be built into distribution lines, transformers, and the like. Viewed another way, reliability quantified as an index (see the next chapter) will be lower for residential and commercial customers than for those supplied at a higher voltage. Correspondingly, the cost of energy to residential and commercial customers will be higher than to those at the higher voltage because of the allocation of component costs. These relationships might be perceived as inconsistent or as anomalies by customers who may not understand the "ins and outs" of system operation.

Another concern is that of complicating the day-to-day operations of the power system. The installation of computer aided or computer controlled logic for tripping loads would involve a large number of machine instructions that may go astray and surely would initially increase the risk of operating the power system successfully. In addition, the computer assisted approach requires communications channels to send signals to and receive confirmation from the remote loads to be tripped. The receipt of these messages would require, furthermore, the installation of electronic equipment that also would require periodic maintenance and tests to ensure proper operation and to avoid malfunction. Additionally, the computer logic must be updated and changed periodically to account for additional subscriptions for reliability of service or to account for altered subscription levels.

On top of this is the concern that the operation of the power system would not follow the natural instinctive reactions of human beings. It is normal for the system operator to strive to keep the power system up and to provide the best possible service. But, when emergencies arise, alarms sound and lights blink in the control room. The operator is under stress to make split-second decisions that minimize megawatt-hours interrupted or to localize a cascading outage. Additionally, they are trained to restore the supply of electricity to customers at the earliest opportunity. Meeting these objectives might require a certain sequence of load shedding or restoration based on technical considerations that are unrelated to the reliability subscriptions of customers. Therefore, forcing the operator to interrupt or restore some loads before others because of the character of their reliability subscriptions might produce errors of judgment or wrong decisions even if aided by a computer.

What appears as a contrast to the above description of the potential effects of reliability-based prices is that the present day operation of the power system does indeed include a priority of service of some sort. For instance, during system emergencies, the operator strives to continue serving some "important" loads. Major primary metal industries with no or little back-up generation, petrochemical industries, airports, and so on might be considered "important" loads. In this practice, the operator has a qualitative ranking of outage costs for some important customers. It is important to note here, however, that the above practices are possible because the network has been planned to provide adequate transmission capacity and voltage support when these contingencies arise. In other words, the planner has included a qualitative assessment of outage losses in the design of the power system which has evolved over the years. And in many instances, the utility is likely to have encouraged certain major customers to locate their activity near certain nodes in the network topology, thereby assisting the utility in its contingency planning.

Such fine tuning becomes more problematic if customers, irrespective of their position in the network, make reliability choices. As a result, it may be difficult and costly to honor the chosen priority classification because any modifications to the power system are likely to be system-specific. Hence, it is prudent to compare the gains attainable from such modifications with the expense involved.

In this chapter, certain complications that arise due to network topology and operation have been delineated. This is not to say that the idea of reliability-differentiated price should be abandoned. On the contrary, the above discussion is meant to alert readers that the benefits due to reliability pricing should be properly quantified and compared to its shortcomings. Only after such a comparison is one able to decide on the pros and cons of reliability pricing. Toward that end, certain basic concepts of quantifying reliability will have to be understood first, followed by an analysis of economic welfare. We now turn to these.

CHAPTER 4

QUANTIFICATION OF POWER SYSTEM RELIABILITY AND ITS RELATION TO PRICE

In this chapter, certain concepts of power system reliability evaluation are presented as background material that will enhance our understanding of reliability indices. Subsequently, the methods of relating a reliability index to price are examined and the difficulties in establishing such a relationship are considered.

Concepts of Reliability Evaluation

It is important to recognize that evaluating the reliability of a bulk power system is an extremely complex task. Thus, it is not surprising that several techniques are available for measuring the reliability of subsections or components in the system. It is also not surprising that there is no generally acceptable method for evaluating or predicting the reliability of supply at the end-users' bus bar. Research in this area is vast and continuing. Therefore, it is impossible even to attempt to summarize all the methods. The purpose here is to give a general outline of certain philosophies and methods. Such an outline will aid us in understanding the difficulty associated with defining a projected reliability level of service to end-users.

Hierarchical Levels and Meshed Networks

Figure 4-1 shows the organization of a power system in three hierarchical levels[20] (HL) based on functional zones. The bulk system is defined as HLII which consists of integrated generation and transmission facilities. Therefore, HLII is of interest to customers supplied from the major transmission system. Those customers supplied from distribution

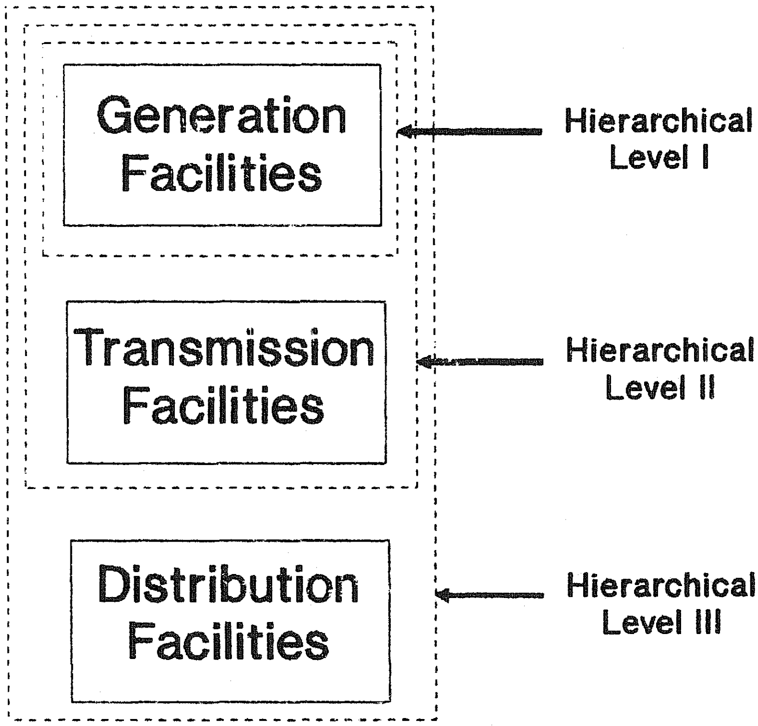


Fig. 4-1. Hierarchical levels.

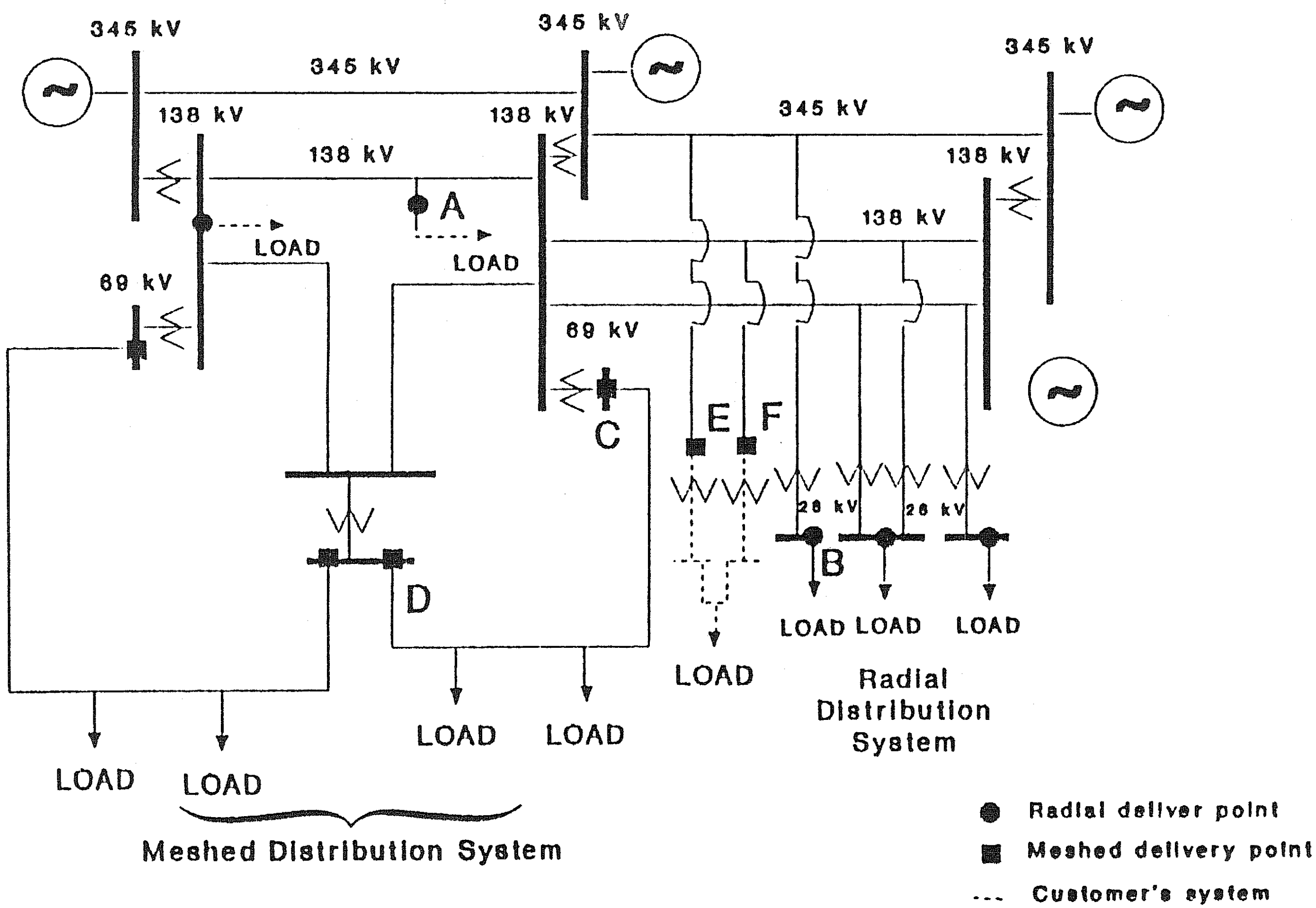
networks are concerned with HLIII which contains the generation and transmission facilities plus distribution lines and associated equipment.

Chapter 3 has shown that the pattern of outages does not mesh perfectly with the three hierarchical levels shown in figure 4-1. That is, system failures are not simply generation failures, or generation plus distribution failures, or generation plus transmission plus distribution failures. They can occur singly in any one of the components of HLII and HLIII. Yet, even if one were to calculate separate reliability indices for each functional zone such as generation, transmission, and distribution, it is not a direct addition or even a simple combination of such indices that defines system outages.

Furthermore, chapter 3 makes it clear that the system operator takes several actions to avoid and/or confine a system failure. When these actions are successful in averting an outage, they prevent changes in the actual measurement of the power system's reliability. Consequently, it is only those curtailments of supply by automatic relaying action that affect the perceptions of the system's unreliability. Therefore, these outages have been of primary interest to modelers seeking to predict the indices of reliability. Hence, it must be noted that any model of reliability does not take into account all the manual actions taken by the operator in hopes of averting system failures. This modeling characteristic is a shortcoming because these actions depend on the relationship between network and demand conditions.

Figure 4-2 taken from reference 21 shows the delivery points in a network. For example, point A radially supplies a customer-owned system directly connected to the bulk system. It is, therefore, a radial delivery point. Point B is a low-voltage bus radially supplying a distribution system, and also a radial delivery point. Point C is a meshed delivery point. It supplies a meshed distribution system which is also supplied from another meshed delivery point, D. Points E and F are meshed delivery points supplying a meshed system owned by the customer. Note that for the meshed system the outage of certain components does not result in curtailment--as for example the loss of transformers at C and D when the two loads could still be supplied from the other functioning 69 KV transformer.

In terms of evaluating reliability indices for the radial and meshed delivery points for this power system, one has to consider the network



Source: "Bulk System Reliability--Measurement & Indices." Measurement Indices Working Group Report, *IEEE Transaction on Power Systems* 4, no. 3 (August 1980).

Fig. 4-2. Power delivery points.

topology as including all combinations of contingencies that could result in a curtailment. This illustration also shows that neither the index nor the probability distribution of shortfalls, assuming this measure can be calculated by defensible methods, would be the same for all the delivery points.[1] Furthermore, the duration of interruption could also be different for different delivery points.

Reliability Evaluation Methods

In the following sections, the reliability evaluation techniques for the three functional zones are examined. The reader may refer to several books and articles for a deeper understanding of the subject.[22,23] Subsequently, our goal is to examine whether methods exist which enable the combination of these zones to obtain delivery point indices that can be used to relate the price of supply to the reliability of supply.

Reliability Evaluation of Generation

Evaluating the reliability of generation is perhaps the most developed and perfected technique. The starting point is the outage statistic of machines similar to those one wishes to model. An important parameter of the outage of a generating unit is its forced outage rate. The forced outage rate of a unit is defined as the ratio of the number of hours of outage divided by the sum of the number of hours of outage plus the hours in service. For base-loaded units, service hours would include all the hours in a period since such units are constantly in use. For cycling units, however, the definition of service hours is rather involved.

The theory of repairable components due to Markov starts from the mean up time (MU) and mean down time (MD) of machines or components. These are obtained by dividing the total hours of available time and the down time by the total number of available states and the total number of down states of the component (or generating unit in our case). The state of the unit is

[1] Recall that Smith[15] assumed a single probability distribution of shortfall for the whole system. He also assumed that such a distribution also included the duration of interruption.

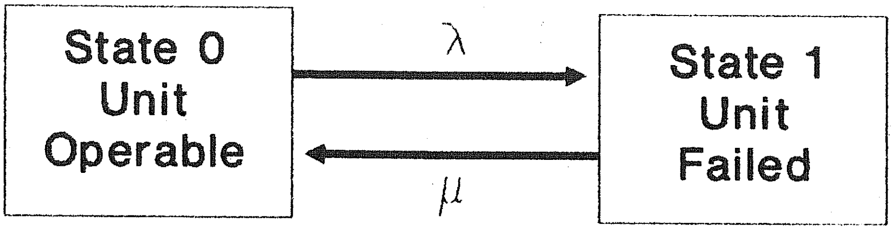


Fig. 4-3. Two-state model of unit failure.

modelled as in figure 4-3 where the failure rate λ = 1/MU and μ, and the repair rate = 1/MD. Then, assuming that the probability of failure during any time interval t is independent of the prior operating time[2] (or the reliability is constant for equal operating periods), it can be shown [22,23] that

$$P_U(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} , \tag{4-1}$$

and

$$P_D(t) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} , \tag{4-2}$$

[2] This assumption implies that the failure density of the component under examination is an exponential distribution; i.e., the probability of a component showing or being in the up state at a time t is $e^{-\lambda t}$. Whether this assumption is valid or not for the generating units is debatable.

where $P_U(t)$ and $P_D(t)$ are the probabilities associated with finding the unit in operational and down states, respectively, at time t given that at time zero it started in an operable state.

At $t = \infty$, one finds from (4-1) and (4-2) that

$$P_U(\infty) = \frac{\mu}{\lambda + \mu} = \frac{m}{r + m} , \tag{4-3}$$

and

$$P_D(\infty) = \frac{\lambda}{\lambda + \mu} = \frac{r}{r + m} , \tag{4-4}$$

where m and r are the mean time to failure and mean repair time. Note that P_U and P_D indicate the conventional unit availability rate and forced outage rate.

The above method of modelling unit failures in figure 4-3 is termed a Markov model or a frequency and duration method (FAD). The FAD method can be used to calculate probabilities associated with multiple units such as two or more generators on simultaneous outage in a system. However, the

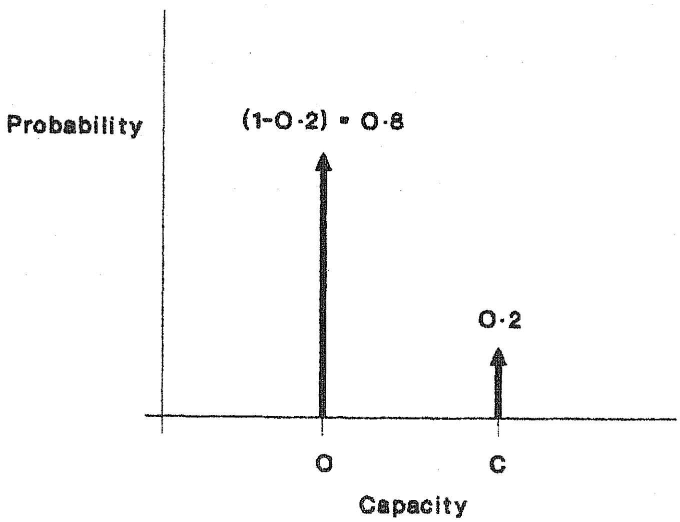


Fig. 4-4. Binary probability density of capacity on outage.

following methods using the definition of forced outage of the unit as in equation (4-4) are in common usage.

The probability density distribution of capacity available from a unit is represented in a binary fashion as in figure 4-4. Refinements to model partial outages are common in the literature.[22] Then, a convolution of such unit distributions gives the joint outcomes and yields a probability density distribution of available capacity which can be used to obtain a capacity outage distribution or table.[3]

Another technique of establishing capacity outage probabilities is from a Monte Carlo draw. In this procedure, the outage or otherwise of each machine is decided by the outcome of a draw from a random number generator. From a series of multiple draws, one for each generating unit, the probabilities associated with losing specified amounts of capacities can be determined.

The Load Model

In addition to the capacity outage model described above, one has to define a load level or levels to calculate the probability of load curtailment. In any hour during the study period, at any specified load level, reserve generation is the operating capacity minus the load.[4] The probability of load curtailment refers to states when reserve is less than zero. This can be obtained by adding the appropriate probabilities in the capacity outage distribution that correspond to events making the reserve less than zero. This probability is expressed by an index, for example 0.3×10^{-3}, and is called loss of load probability or LOLP.

In a method originally due to Baleriaux,[23] the loads are assumed to be random and the LOLP at each load level can be calculated.[5] The hourly

[3] The capacity outage table has entries of capacity ranging from zero to total installed capacity with the corresponding probabilities of losing that amount of capacity.
[4] Operating capacity is the installed capacity minus the capacity on planned outage such as maintenance, refueling, etc.
[5] One could argue whether or not loads are really random. The loads are almost cyclic from day to day and can be predicted with good accuracy for the short-term future such as a day or two in advance.

LOLPs are then weighted by the probability associated with the occurrence of that particular load and are summed to obtain a probability weighted LOLP.

In another approach practiced by some North American utilities, the worst case scenario--the annual peak load--is used to obtain the LOLP. This, in some instances, is modified to include all the daily peak loads of the peak month to calculate the system LOLP as the average of the daily LOLP for the peak month. Other variations to these methods exist in utility practice, such as multiple load level models and the like,[22] which the interested reader can find in the source documents.

Cautionary Note

It is important to make the following observations from the above two attributes of the LOLP.

Two analyses or two utilities may obtain the same LOLP index, say, $0.3\ 10^{-3}$. That does not mean, however, that the risks of curtailment would be the same in the two utilities. The same index could represent different risks due to different assumptions in modelling load and generation in the analyses. On the other hand, analyses making different assumptions could give different LOLP indices for the same system.

Note in the above analyses that all the generation is considered to be aggregated at one bus bar and that the sum of all the loads is connected or presented to the generation sources at that bus bar. This means that the effect of the transmission and distribution network is completely ignored. It was shown earlier that the transmission system has a profound effect on the risk of system failure and that the risk of curtailment depends on the location of loads. All the same, it must be said in defense of these methods that they are designed for longer-term resource expansion planning and that they are appropriate for that purpose.

Loss of Load Expectation

Expectations can be calculated from any reliability index that is derived from probability distributions. For instance, if one considers a time span of 365 days (some ignore weekends in their analysis), an index of $0.11\ 10^{-3}$ translates to an expectation of 0.1 days lost per year, or the familiar one day lost in ten years. This loss of load expectation is termed LOLE and is generally expressed in days or hours per year. There are indices other than loss of load expectation. The reader is referred to reference 20 for a definition and list of these indices such as interruption frequency, duration, annual unsupplied energy, interruption severity, and so on.

Transmission Reliability Evaluation Methods

The inclusion of transmission outages into the generation outages to obtain HLII indices is a more involved task. From a statistical examination of outages of existing lines, transformers, switches, and so on of different voltage classes, one can establish a two-state (up and down) or a multistate model for such components of the transmission system. Such a model can be analyzed by the Markovian or other methods. But the combination of the outage representation of transmission with that of generation is complex and involves several assumptions. The following are four major complexities.

The first is that the outage of selected transmission or generation facilities either simultaneously or singularly may or may not result in curtailment at any bus bar. Depending on the time-of-day/season-of-occurrence of the outage, the resulting "load flow"[6] in the network produces certain voltage conditions. If the resulting voltage profile is unacceptable or if the reactive power requirements to maintain voltage profiles are excessive, loads at some delivery points or buses may have to be curtailed.

The second complication is that of assuming the time of the occurrence of outages. They might be assumed to be random events or as events that occur more frequently in adverse weather. Several weather models for line outages are proposed in the literature, but it is open to debate as to whether weather is a random phenomenon or cyclic in nature.

The third is that of modelling common mode failures. Common mode means a common cause ensuing in the outage of one or more components. An example

[6] A load-flow study is one in which one calculates the voltages at bus bars and the flow of power, vars and losses in lines.

is lightning hitting both lines in a common right-of-way causing an outage of both lines. This is in contrast to the outage of one line due to some cause giving rise to another outage (say, due to ensuing overload) called the resulting outage. The statistics or data for such events are hard to collect and modelling such events and integrating them into the generation module is cumbersome, rife with assumptions and approximation, and subject to contention.

The fourth concern is that of the correlation of loads at buses. Should the loads at all buses be modelled as correlated or as independent events? If they are correlated, what is the correlation coefficient between them and are they likely to change in the years to come?

These are some of the major complexities that face analysts. There are a host of other variations and complications in the analytical methods. However, it is easy to see that the index obtained could vary from method to method depending upon the assumptions therein. There are several competing and disparate methods in the literature[21] and a variety of indices are obtained.

Reliability of Distribution Network

The model selected to represent the distribution outage has to be integrated to the HLII models to obtain HLIII indices. The distribution outage model is not all that dissimilar from those of transmission. They could include, perhaps, more lines, cables, and switches found in distribution networks. As in the case of HLII, several approaches for calculating delivery point indices have been proposed, but they are intensive in machine computation.

Shortcomings of Reliability Indices

The above discussion has shown that the calculation of reliability indices involves several modelling assumptions. Any particular method, in spite of its inherent assumptions, is useful for comparing relative reliabilities of alternative power supply schemes that are being considered by a planner. For example, the cost associated with two engineering alternatives to improve reliability can be estimated and compared. Consider the alternatives of using underground lines instead of overhead lines for improving distribution reliability. Using the same model, method, and assumptions makes the analyses comparable across alternatives. But, this modelling approach is not useful when the objective is to propose an index of supply reliability and relate it to price. It is not clear how the indices translate into the number of outage occurrences, their duration, and the time that they can be expected to arise. On top of these ambiguities, there is the complication that the indices obtained from different models could have different hidden assumptions. Hence, the trade off between the supply reliability (expressed as an index) and price would not be clear to the modeler or to the consumer. Therefore, it is expected that the consumer offered such a reliability-index-based service would find it difficult to make intelligent choices.

An Example

The following illustrates the enigmas that might arise by relating the price of electricity to an index representing the reliability of supply. This problem can, of course, be overcome by using a priority classification of service rather than a reliability index.

It was mentioned earlier that conservation often results in a net reduction of energy consumption and that certain types of loads may demand the same total energy. Consider, however, a customer whose total energy requirement is constant. The customer can change his use pattern somewhat depending on the price signals; that is, if the price of electricity is the highest at the time of peak, there would be a reduction in peak-time use followed by an increased use at other times.[7]

For illustrative purposes, consider the use pattern during a twenty-four hour period. Let the total energy consumed during this period be E_o,

[7] As mentioned earlier, peak-load pricing is a proxy to pricing based on reliability of supply. This assumption is different from that of independence between demand patterns of off-peak and on-peak periods. Implicit in this assumption is a form of dependence or coupling between the demand patterns during the on-peak and off-peak periods.

invariant under any use pattern. Let the demand L(t) as a function of time be represented by the function

$$L(t) = L_{po} \exp - \frac{(t - t_p)^2}{2\sigma_o^2} , \tag{4-5}$$

where t_p is the time of peak load or demand, L_{po} is the magnitude of peak demand and σ_o a parameter of the function (4-5). A demand pattern that does not resemble (4-5) can be approximated by (4-5) using a "best fit" technique.

If the original load factor of the consumer is μ_o, clearly

$$L_{po} = E_o \,/\, 24\mu_o \ . \tag{4-6}$$

The purpose is to examine the effect of different reductions of peak load which is of interest because of the long-term benefits due to reserve reduction. L_{po} is decreased to a new peak load of, say, L_{p1} (see figure 4-5). The amount of reduction is dependent on the peak load price and the demand elasticity of the user at the time of peak load. If the same time of peak demand t_p is maintained, the energy invariance constraint requires that

$$\int_o^{24} L(t)\, dt = E_o , \tag{4-7}$$

where L(t) is the demand as a function of time.

Since discreet time intervals are considered, one has

$$\Sigma\, L(t)\, dt = E_o \ . \tag{4-8}$$

Since $L_{p1} < L_{po}$, the above would be possible by increasing σ_o to an appropriate value σ_1 to result in the modified load L_1 in figure 4-5. The increase of σ_o to σ_1 models the assumption that the energy not consumed during the peak hour is consumed at hours straddling the peak. The load factor of the modified load would, of course, be greater than μ_o.

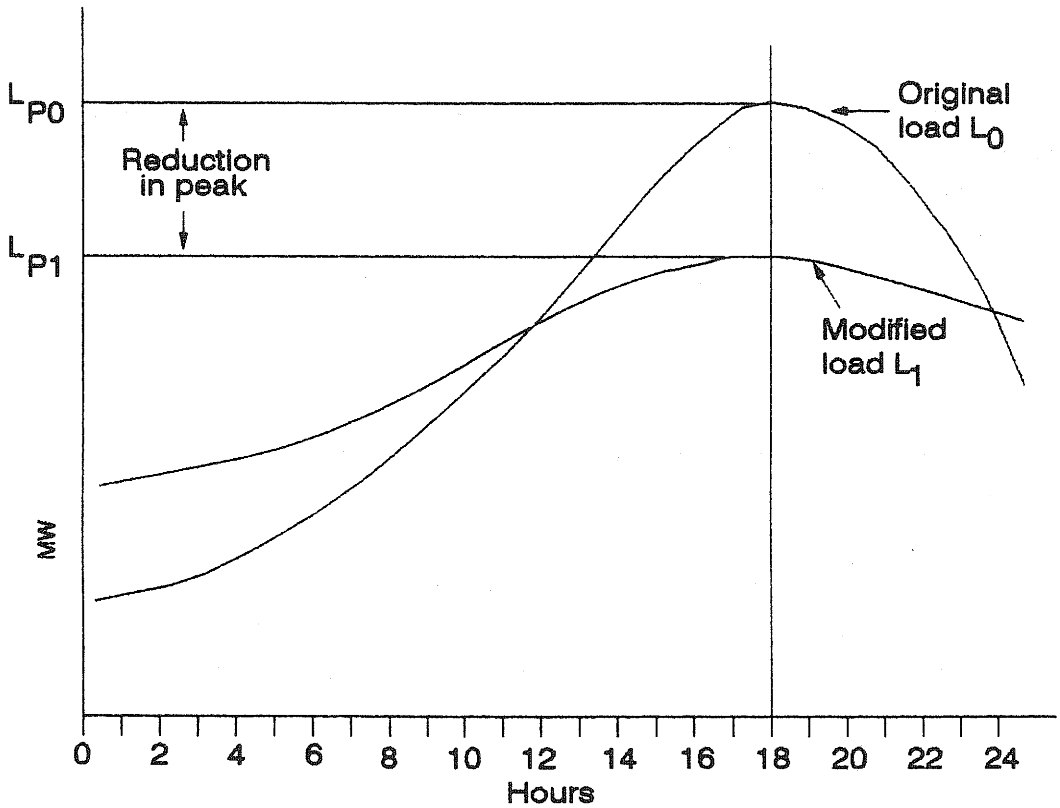


Fig. 4-5. Functional approximation to model energy invariance.

The computational procedure was to assume an initial peak load L_{po} and load factor μ_o (and therefore E_o). As a result of the consumer response to peak-load pricing signals, several reductions in peak load were assumed. Thus, the value of L_{po} in the numerator of (4-5) is reduced to the assumed value of L_{p1}. The corresponding value of σ_1 ($\sigma_1 > \sigma_o$) to satisfy (4-8) was calculated iteratively on a digital computer.

Figure 4-6 portrays the results obtained. Note from the figure that the lower the load factor initially, the greater the potential for peak reduction by proper price signals, an expected result. However, for each load factor, there is a limiting value of peak-load reduction. For instance, with μ_o = 0.6 and a starting peak of 200 MW, price signals at the time of peak cannot reduce the peak below about 116 MW if the energy invariance condition is to be satisfied. This is as a result of σ_1 reaching a limiting value of ∞ and yet not satisfying energy invariance.

Effect on Reliability Index

It is clear that energy invariance requires that a reduction in demand at periods of high prices must be accompanied by a corresponding increase in demand at other times. This can be seen in figure 4-5 as well. If the reliability index is calculated on an hourly basis for a given installed capacity, a reduction in peak demand is accompanied by a reduction of the LOLP index at the peak hours.

Table 4-1 shows the results of an illustrative study consisting of two classes of loads, one with a peak at 3:00 P.M. and the other at 6:00 P.M. The assumed load factor, peak load, the resulting load of classes according to figure 4-5, and the total demand profile are shown. The total generation resource was assumed to be 3,105 MW (comprised of fossil oil, 941 MW; fossil coal, 1,274 MW; nuclear, 800 MW; and four combustion turbines of 20 MW each totalling 80 MW) as in reference 19. The reader can obtain the unit sizes and their forced outage rates from this source.

The loss of load probability index was calculated at each hour for this demand and generation model as indicated in table 4-1. Note that the highest risk is at the time of system peak which is at 4:00 P.M. The effect of the transmission network was ignored in the calculation of these indices.

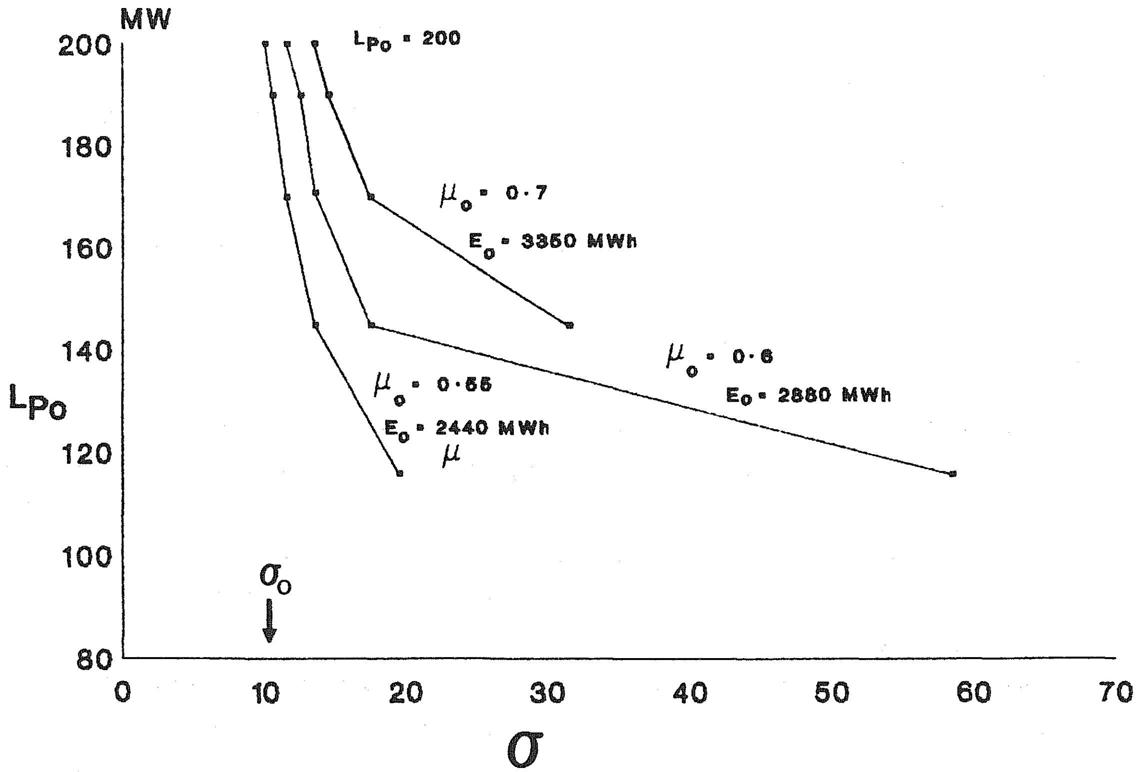


Fig. 4-6. Relation between peak load and σ for energy invariance.

TABLE 4-1

HOURLY DEMANDS AND LOSS OF LOAD PROBABILITY

Hour	Demand, MW Class 1 $L_{po} = 1,800$ $\sigma_o = 9.15$ $t_p = 15$	Class 2 $L_{po} = 970$ $\sigma_o = 7.25$ $t_p = 18$	Total	LOLP 10^{-3}
1	445	26	472	--
2	539	40	580	--
3	644	59	704	--
4	759	85	845	--
5	882	107	1,001	--
6	1,010	162	1,173	--
7	1,140	216	1,356	--
8	1,269	280	1,550	--
9	1,392	355	1,747	--
10	1,506	438	1,944	1.330
11	1,605	528	2,184	8.300
12	1,688	620	2,308	21.860
13	1,749	711	2,460	46.690
14	1,787	795	2,582	101.850
15	1,800	867	2,667	166.550
16	1,787	923	2,710	205.440
17	1,749	958	2,707	202.760
18	1,688	970	2,658	158.460
19	1,605	958	2,564	90.460
20	1,506	923	2,429	38.870
21	1,392	867	2,260	17.490
22	1,269	795	2,064	4.710
23	1,140	711	1,852	0.390
24	1,010	620	1,631	--
				1,065.060

Operating capacity 3,105 MW
Average LOLP over 24 hours = 1,065.06/24 = 0.04437

Next, it was assumed that due to appropriate price signals, the peak demands of classes 1 and 2 fell to 1,720 MW and 890 MW from the previous 1,800 MW and 970 MW, respectively. Assuming energy invariance in each class of consumption, the resulting total system load, as well as the hourly and average LOLP, are shown in table 4-2.

Note that the system peak has declined to 2,566 MW from 2,710 MW, a reduction of 144 MW. Also note that the operating capacity has been reduced to 2,768 MW. This is a reduction of 140 MW from the earlier case and is obtained by eliminating four 20 MW combustion turbines and five 12 MW coal units. This reduction in operating capacity models the reaction of the utility to the reduced peak load. After all, the purpose in reducing the peak load is to reduce the reserve equipment. Therefore, on a longer term basis, the utility would reduce the installed capacity in reaction to the reduced peak. Observe that the risk index at the time of peak demand is about the same in tables 4-1 and 4-2; that is, 205.44 and 193.14 respectively.

Problems of Relating Price to Reliability of Supply

Observe in table 4-2 that compared with table 4-1 the risk index is higher every hour except from 4:00 to 6:00 P.M. This is due to the total load in the hours other than 4:00 to 6:00 P.M. being higher than in the case studied in table 4-1. The average risk index in the case of peak-load reduction has increased rather than decreased as one might expect at first.

A comparison of tables 4-1 and 4-2 raises several important questions. When attempts are made to link the price of electricity and reliability, should one be concerned with reliability at the hour of system peak load or the average system reliability? Is the index of concern that of class average risk or that of the whole system at the time of system peak demand? How should one include the duration of outages in such indices?

For instance, an outage occurring at 9:00 A.M. and lasting four hours could be more severe than an outage at 5:00 P.M. and lasting one hour. How does the operating practice and network topology affect the calculation of such indices? How are transmission-related outages included in the above index?

TABLE 4-2

DEMANDS AND LOSS OF LOAD PROBABILITY AFTER PEAK REDUCTION

	Demand, MW			
Hour	Class 1 $L_{p1} = 1,720$ $\sigma_1 = 10.85$ $t_p = 15$	Class 2 $L_{p1} = 890$ $\sigma_1 = 8.45$ $t_p = 18$	Total	LOLP 10^{-3}
1	533	56	590	--
2	626	77	704	--
3	727	104	832	--
4	835	137	972	--
5	946	178	1,124	--
6	1,060	226	1,286	--
7	1,173	281	1,455	--
8	1,283	343	1,627	0.090
9	1,387	411	1,799	1.100
10	1,481	484	1,965	6.170
11	1,563	558	2,121	16.840
12	1,629	631	2,261	31.990
13	1,769	701	2,381	64.160
14	1,709	764	2,474	116.630
15	1,720	816	2,536	166.550
16	1,709	856	2,566	193.140
17	1,679	881	2,560	188.010
18	1,629	890	2,519	152.090
19	1,563	881	2,444	97.170
20	1,481	856	2,338	48.880
21	1,387	816	2,204	24.490
22	1,283	764	2,047	11.190
23	1,173	701	1,875	2.640
24	1,060	631	1,692	0.250
				1,121.460

Operating capacity 2,865 MW
Average LOLP over 24 hours = 1,121.46/24 = 0.04672

Such concerns raise the related issue of how the price should be related to reliability. Should the tariff for demand include in it a component for reliability? If so, should the demand rates, in the conventional demand-cost-allocation ratemaking methods, be reduced for less reliable service at the expense of a proportionate increase for a more reliable service? If there is a premium for higher reliability service that is added only to the demand charge at the time of class peak, the reliability at off-peak periods would be ignored. Alternative pricing schemes are possible; for example, one in which the demand charges are not related to peak class demand but are spread over all hours of consumption. In that case, should the hourly demand charges be discounted by a certain percentage for a lower priority of service class at the expense of a corresponding increase to the higher priority class?

The above questions are difficult to answer within the scope of this report. In fact, the existing techniques of reliability evaluation are not adequate to answer all the above questions, particularly those that address the time of expected outages and their expected duration. It is important to note, therefore, that the development of more sophisticated reliability evaluation models would be a prerequisite to answering the above questions. In addition, one has to consider operating and planning aspects of utility systems, and the economic implications of selecting any particular pricing scheme.

Relation Between Reliability Index and Shortage

At first thought, one would assume that the greater the probability of shortage, the higher the risk index. This is true if one deals only with capacity shortages. Then, the lower the reserve, the higher the risk of loss of load. Figure 4-7, taken from reference 24, shows the distribution of capacity outage in a 23,000 MW system. Note in the figure, if the reserve is 2,500 MW, the LOLP is about $0.5\ 10^{-3}$. Note also the nonlinearity in the risk index. In other words, if the reserve is reduced by, say 10 percent to 2,250 MW, the risk does not increase by 10 percent. In this instance, the risk increases by more than 10 percent to $0.7\ 10^{-2}$. Therefore, linking the price to a reliability index might be unwise. The

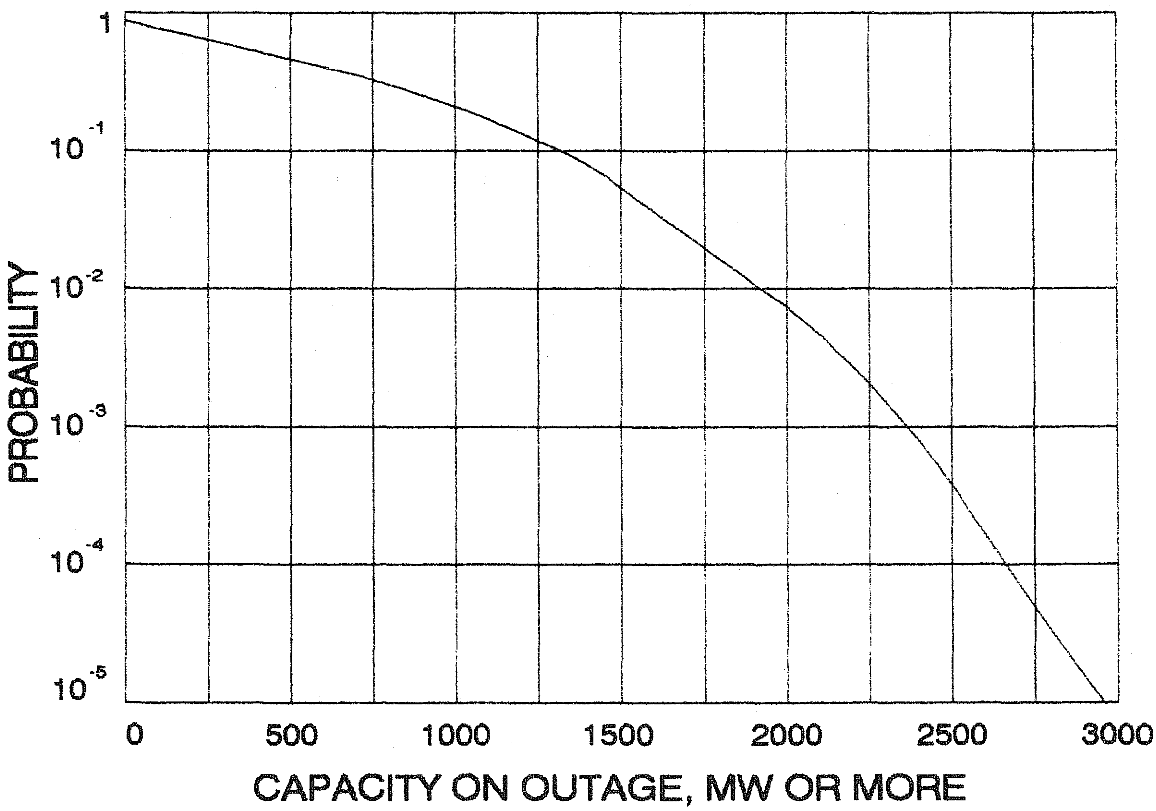


Fig. 4-7. Nature of the reliability index.

high degree of nonlinearity characterizing this reliability index could easily confuse consumers causing them to make decisions with unintended effects.

An alternative without the above nonlinearity problem is to use reserve equipment as a proxy for the reliability index. Instead of offering a certain index of reliability to a customer, one could hold certain reserve capacities to supply a particular customer. The disadvantage of such a choice is that it would ignore transmission failures which constitute the majority of outages. Furthermore, it is unclear as to how transmission allocations could be melded into this alternative because of their complex nature as shown earlier.

In view of the above complications regarding reliability indices, in our view the best method for incorporating reliability considerations into a pricing scheme is to establish a ranking for the priority of service (such as priority 1, priority 2, etc.). The system operator would make best efforts to continue supplying priority 1, followed by 2, and so on. In the event of generation shortages, for example, every attempt would be made not to cut off priority 2 customers until the shortage exceeded a certain amount, and priority 1 customers until the shortage exceeded an even greater amount. While the regulator might be in the position to ensure that proper reliability of service was delivered during a capacity shortage, similar activities under transmission-related failures present a formidable task to the regulator. Recall that it is difficult to quantify indices for transmission failures, and, therefore, it would be extremely difficult for the user and the regulator to measure and ensure that the system was indeed operated to honor the priority levels. As a result, one has to trust that the utility and its operators made every effort to maintain the required priority of service.

The question as to what planning benefits emerge as a result of priority service remains to be answered. There should, of course, be a price differentiation between such service categories. Whether such pricing schemes result in either reductions of reserve or increases in welfare can be examined following the analysis outlined in the appendix.

Pricing Service to Noncore Customers

A discussion of pricing the service to noncore customers is now in order.[8] It is shown later that noncore customers are good candidates for reliability pricing. Figure 4-8 represents a typical daily demand of core customers and the total operable capacity for a particular day. The following arguments based on a day's load are illustrative. In practice, of course, one would consider demand over a longer period of time such as a season or a year.

Fixed or Capacity-Related Costs

It is clear from figure 4-8 that the index of reliability every hour is changing due to changing loads. The highest risk (LOLP) is at the time of peak load. At other times the value of LOLP is lower (indicating lower risk) resulting from greater reserve capacity than at the time of peak load. Therefore, one can earmark sufficient capacity as reserve to keep the risk level the same at every hour. This would free some capacity in the off-peak hours to noncore customers. The amount of reserve required at each hour depends on the size of units and their forced outage rates. This can be calculated using appropriate methods employing probability theory.

Figure 4-8 shows the required reserve for constant risk. Also shown is the capacity remaining from the total installed capacity which is available for commitment to noncore customers. Note that this balance capacity varies from hour to hour. In this simple illustration, capacity on maintenance has been ignored.

The above illustration should not be misunderstood to mean that there is no capacity available at the time of system peak to serve noncore customers. Nor should one infer that it is only the capacity represented by the hatched area in figure 4-8 that is available to serve the noncore customers. In fact, at any hour, the difference between the installed

[8] As discussed earlier, noncore customers are commonly referred to as interruptible customers by the utility.

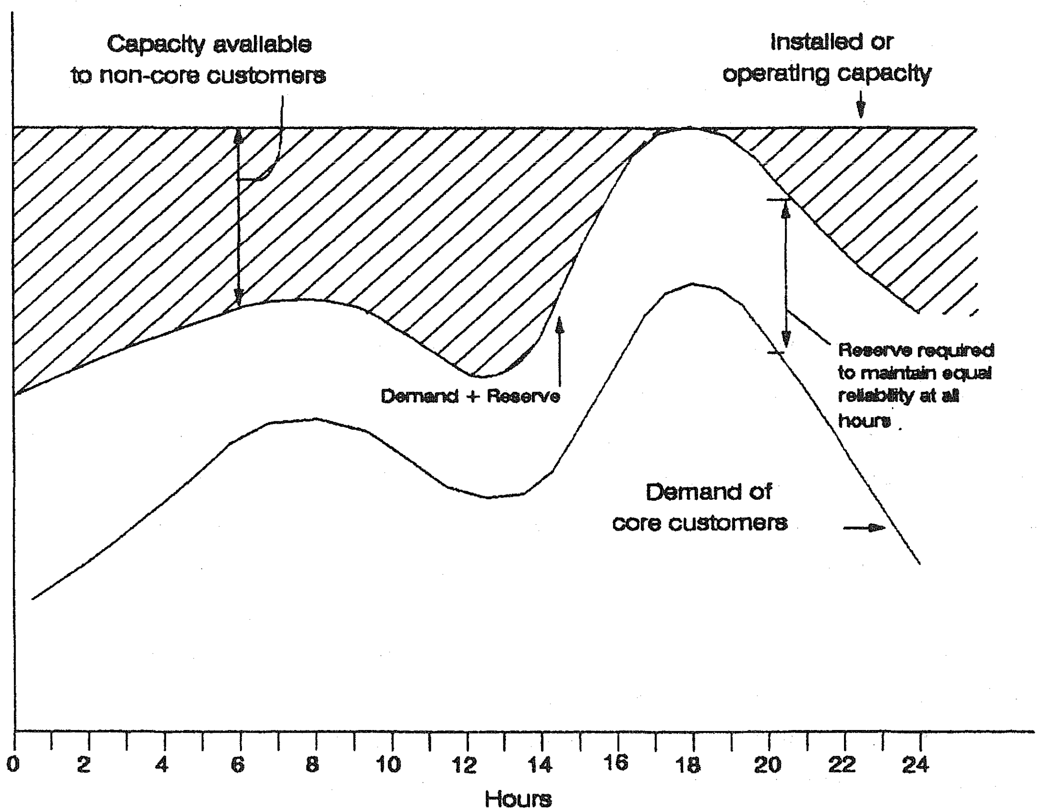


Fig. 4-8. Service to noncore customers.

capacity and the core customer demand is available to serve the noncore customers.

The purpose of identifying the capacity in the hatched zone as available to serve the noncore customers is to highlight the following pricing considerations. The amount of reserve capacity required to maintain a particular reliability level to the core customers is not kept idle. It is used to supply the noncore customers under normal operating conditions, but it is "clawed at" when forced outages of units serving the core customers occur. Therefore, the availability of reserve capacity for noncore customers is subject to probability laws determined by the outage of other units. Of course, the capacity in the hatched areas is also subject to random outages. Hence, it is possible using well-understood methods to establish a probability distribution of the capacity available to the noncore customers. This distribution will have two components: one due to the capacity in the hatched zone and another from the reserve generation for the core customers. From this distribution, an expected available capacity (average available capacity) can be calculated and offered to the noncore customers. Of course, it is to be hoped that not all the expected capacity will have takers.

In terms of pricing the supply to noncore customers, the capacity costs associated with the hatched area would be fully allocated to them. If there is only one noncore customer, he would have to bear the above capacity costs. However, rationing would not be an issue. If there is more than one noncore user, the question of a priority or a rationing scheme in the event a shortage arises. In this case, the utility could resort to a bidding scheme from the noncore users. Those who value this would pay higher prices and would be served at higher reliability. Of course, the maximum reliability would be less than that of the core customers. A priority ranking scheme of pricing the supply, therefore, is ideally suited to noncore customers.

The capacity costs of the reserve to supply the core are allocated to both core and noncore customers. The proportion of allocation would depend on the probabilities or the expected value of the reserve capacity to be used by the respective customers; noncore customers having access to this capacity only when it is not needed by core customers. Note that the more noncore customers there are, the lesser the cost borne by each, core and noncore customers alike as long as a noncore customer was not previously a core customer. The allocation of transmission-related costs is somewhat more involved, but it could embrace principles similar to the above. It is assumed in the above argument that the system is operated to correspond to such cost allocation policies.

Energy-Related Costs

In terms of energy-related costs, the most economical units would be used to dispatch energy to the core customers. Therefore, energy to the noncore customers would be supplied from less efficient units, thereby raising their per-unit energy costs.

Whether the total cost of energy for noncore customers is less than that charged to core customers is an interesting question. The answer lies in the capacity-related costs associated with the units of lesser efficiency in the hatched zone. It is expected that these peaking and cycling units have a smaller capital investment than that of the base-load units whose costs are allocated among core customers. Therefore, it seems that the fixed or capacity-related costs could be lower for the noncore customers. Hence, it might be that the average cost per kilowatt-hour to a noncore customer would be less than to core customers. If, however, the higher cost of energy production from the less efficient units outweighs the lower capital cost of such peaking units, it may well be that the average cost per kilowatt-hour would be higher for the interruptible customer than for the core customers. In such a situation, it is to be expected that there will be no subscriptions to the noncore service as one could easily obtain a cheaper price for a better service by being a core customer. Alternatively, noncore customers might choose to obtain supply from other suppliers, perhaps with mandated wheeling via the utility's transmission, or the customer might choose to use alternate sources of energy.

If there are no subscribers to the noncore service, the utility and its regulators must examine ways to attract such customers. A lack of customers would mean that the costs associated with all the capacity would be recovered from the core customers. In order to attract noncore customers, it may well be that the costs allocated to this customer class might have to be reduced to a value wherein the average total cost of energy would be less than that charged to core customers. If and only if additional noncore customers come from a body of firms that is not currently part of the existing customer base does it follow that the revenue requirement from the core customers would be less than in the case of having no noncore customers.

In any event, the above shows an application of reliability-based pricing to noncore customers can be used to enhance the gains to the utility and its customers.

CHAPTER 5

PRICE RELATED TO RELIABILITY

It is suggested that giving customers the option of choosing their reliability of supply would unbundle the service and, therefore, improve efficiency. The goal now is to bring together the analyses of the preceding chapters and the subsequent appendix to examine if such unbundling offers any advantages. This effort is conducted in four major areas. The first area is on methods of evaluation of reliability indices. The second area is the consumers' understanding of reliability indices and how this understanding will translate into adjusted consumption patterns. The third area is the suppliers' ability to operate the system in a fashion to deliver the contracted reliability level. The final area of investigation is longer-term considerations such as economic welfare gains and reductions in the cost of resource expansion arising from reduced reserve requirements.

The following is a summary of chapters 2 to 4 and the appendix. Subsequent to this is a discussion regarding the pros and cons of reliability pricing from which conclusions are drawn.

Reliability Indices

Rudimentary concepts of evaluating reliability are described in chapter 4. It is shown there that the probability of outage (LOLP) or expectations of loss (LOLE) are well-developed indices. However, it also is shown that many assumptions and approximations are necessary to approximate the effect of transmission failures on these delivery point indices of reliability. Further complicating matters is the fact that there is no universal or commonly accepted method to obtain delivery point indices. In addition, due to different modelling assumptions in the methods, identical indices may not denote the same risk exposure.

Another major shortcoming of reliability indices is that they cannot predict the time that an outage will occur or the duration of outage at particular delivery points. However, methods have been proposed which calculate the frequency and mean duration of outages at delivery points, albeit after substantial computational effort. But, these indices express the probability of outage or shortage, and they are more meaningful for describing shortages due to generation. Conversely, they are not particularly useful for predicting the effects of transmission- and weather-related outages. Therefore, most reliability indices do not tell consumers how to alter their consumption patterns to maximize their benefits.

A final shortcoming is that it is not clear whether reliability should be measured at the time of peak or calculated as an average index based on demand over all the hours. If the former path is taken, some indices may not capture the duration of outage or whether the outage spills into nonpeak hours. If the other path is taken, again the duration of outage as it straddles different use periods is unclear. The enigma of increasing average risk by minimizing the risk during peak consumption has been shown by an example in chapter 4. Also shown in that chapter is that different elasticities of demand for different classes of consumers have an impact on the reliability of service for these customer classes.

The above comments are not meant to portray reliability indices as either faulty or unusable. They simply are meant to indicate that reliability indices do not signal the expectations of interruption and thereby cause changes in the consumption pattern of the end-user. What function these indices do perform is the comparison of the relative reliabilities of engineering or technological alternatives.

Reaction of Consumers to Indices

That the reaction of consumers to reliability index-based prices can be nothing but imperfect is evident from the above discussions of these indices. In particular, it will be difficult for consumers to comprehend and untangle the nonlinearity that is embedded in these indices. In view of the lack of clarity and meaning of reliability indices to the consumers, it is best that reliability be translated into a priority of service ranking. Thus, in the event of curtailment either due to generation shortage and/or transmission shortage, the system operator would endeavor to continue supplying a higher priority class longer than a lower priority class.

System Operation

The difficulties of operating the system to deliver the contracted reliability are outlined in chapter 3. The problem is that an operator cannot foresee when a curtailment situation will arise. In fact, only in some instances can the degree of the initial and additional curtailments be predicted after the initiating event. What this means is that in a majority of cases the magnitude and duration of curtailments is not known to the operator even after the initiating event. Therefore, even if the curtailment situation has arisen in response to a supply shortage, the operation of the system to maintain service according to subscribed priority levels is a complex task. This task is only made more difficult by the complications of transmission outages, double contingencies, line loading, voltage profiles, and reactive power loading of lines and equipment. Using computer intelligence to aid in the operation under such emergencies is a solution to this problem, but it would add another layer of complexity to system maintenance and monitoring. In view of these facts and the complex nature of the outages, it is expected that regulators will face a formidable task as they attempt to ensure that a particular contracted level of reliability was indeed delivered. It is shown in chapter 3 that a mere examination of the interruption record is inadequate for such regulatory assurance.

Longer-Term Considerations

A distinction between operating and planning considerations is made in earlier chapters. Operating considerations, generally, address optimal rationing given that a curtailment is imminent. Planning considerations embrace a longer-term view in terms of optimal resource expansion where costs could be minimized due to welfare gains and reduction in reserve requirements.

Assumptions

The details of the price signal analysis are presented in the appendix. The matters addressed and the underlying assumptions are described below.

The gains one expects from any alternative pricing scheme for electricity are improvements in economic welfare and/or a reserve reduction. The latter translates into reduced costs that would benefit all the consumers in the longer term. Furthermore, any substantial changes in the load factor resulting from better usage would allow the substitution of more efficient base load units for peaking units in the longer-term resource expansion plans.

In this report, an economic welfare analysis is conducted based on the assumption of hourly price-demand relationships for two classes of consumers. Each class of consumers is assumed to have an exponential price-demand function of the form αe^{-Kq}. The parameters α and K were derived from the assumed base case consumption pattern of the two classes (see tables A-3 and A-4). In the base case studies, it was assumed that the hourly price of electricity was equal to the average cost. The appendix shows the details of this assumption. The demand-price relationships obtained thus were fixed for all the ensuing analysis. From these price-demand functions, it was possible to determine hourly price elasticities, as in table A-4.

The first pricing scheme studied is that of spot pricing where the hourly price is set at the incremental cost of the production of energy. In our example, spot pricing produced revenues greater than the cost of production plus fixed costs giving rise to a situation of overrecovery. To simulate the case of a regulated monopoly, it was necessary to make revenue equal to cost (ignoring return for the present). The revenue to the producer was adjusted in the following fashion using the spot-pricing scheme as the basis for reconciliation.

The traditional formulation of the revenue reconciliation problem is due to Ramsey. It dictates that the variation of price from the marginal cost be equal to the inverse ratios of the price elasticities for different services or different classes of customers. In our simulations, there is no a priori guidance regarding how much of the total cost should be recovered in each hour although the total fixed costs to be recovered were known. If the intention is to use Ramsey pricing for every hour, presumably one would choose different sets of hourly fixed cost recoveries and examine each set for the obtainable gains in economic welfare. Such efforts to optimize welfare gains can be guided by the price elasticities or the shadow prices (the Legrangian multiplier). However, the goal of our analysis was neither to optimize welfare nor to determine the optimum rates to charge consumers. Therefore, the hourly allocations of fixed cost can be made according to a rather arbitrary method as follows.

The basis of the cost allocation method for the Ramsey formulation, as said earlier, is the spot pricing scheme. The producer surplus (revenue - cost of production) at every hour represents the contribution to fixed charges (since profit or returns is ignored). While the sum of the producer surplus is, of course, greater than the assumed fixed cost for the spot price case, the hourly recovery as a percent of the total recovery of fixed costs are useful statistics. They can be employed to allocate total fixed costs to each hour of the day. Because such cost allocations are not meant to maximize economic welfare, the resulting pricing scheme may be termed Ramsey-type pricing.

The final pricing scheme introduces reliability into the Ramsey-type pricing formulation. The hourly total cost recoveries are combined for the two classes in the Ramsey formulation. In the reliability formulation, the hourly cost recovery is determined separately for each class, and they are specified as constraints. In other words, at every hour of the day, the cost of production and the fixed or common costs are allocated to each class in proportion to their consumption. This represents an allocation based on the demand of the two classes.[1]

To simulate a priority scheme of operation by the subscription of class 1 to lower reliability, a discount x ($x < 1$) in the demand charges is assumed. Naturally, since the total revenue to the regulated producer is fixed, the balance ($1 - x$) remaining after the simulation of consumption has to be recovered from class 2. Similarly, subscriptions of class 2 is represented by y. For example, if $x = 0.8$, in principle, only 0.8 of the

[1] This simulates the ratemaking procedure in which fixed costs are allocated to consumers in proportion to the KW demand.

allocated hourly fixed costs is recovered from class 1. The balance is recovered from class 2. It is clear that when x or y equals 1.0, there is no explicit reliability differentiation between the two classes of service.

As in the case of Ramsey-type pricing, no attempt was made to optimize the reliability-based pricing. Presumably, the balance of the discount (x for class 1, and y for class 2) offered to one class need not be recovered from the other class in the same hour. One could envision a myriad of alternative methods. For instance, the fixed cost allocations to the class desiring lower reliability could be discounted only during the peak hours. The loss of revenue due to discounting could be recovered from either class 1 or class 2 during off-peak hours. Such exercises in optimization are left for further study.

Results

The results obtained are indicated in tables 5-1 and 5-2. They permit the drawing of certain conclusions that are of course limited by the assumptions outlined above.

Table 5-1 indicates the economic welfare gains arising from different pricing schemes with average-cost pricing as the base case. As expected, spot pricing yields the maximum economic welfare. Ramsey-type pricing is the second best solution. For the cases studied in this report, these reliability-based prices outperform average cost prices. However, as the analysis in the appendix shows the subscription to a lower reliability service by one class (for example, class 1, x = 0.85) results in a lower price to class 1 with a corresponding higher price to class 2. Hence, the amount of economic welfare lost depends on whether the customer with higher price elasticity of demand and consumption opts for lower reliability or vice versa. Note also that table 5-1 suggests that the welfare changes between different reliability-based prices follow an uncertain pattern. Recalling that x = 1 and y = 1 means no explicit choice of reliability by either customer class, but also specifies hourly revenue requirements for each customer class, table 5-1 shows a slight gain in economic welfare when

TABLE 5-1

GAINS IN WELFARE COMPARED TO BASE CASE

Pricing Method	Change in Welfare $	Welfare $/MWh
Base Case Welfare $3,431,009	-0-	66.22
Spot Pricing	+16,712	70.36
Ramsey-type Pricing	+10,155	65.49
Reliability-Based†		
x = 1.00	+ 8,260	65.25
x = 0.90	+ 7,476	65.77
x = 0.85	+ 8,489	66.23
y = 0.80	+ 6,241	63.12
y = 0.60	+ 2,007	61.13

*See test in the appendix.

†See appendix for explanation of parameters x and y.

TABLE 5-2

REDUCTION IN PEAK LOAD AND ENERGY CONSUMPTION
COMPARED TO THE BASE CASE

Pricing Method	Change in Peak Load, MW	Energy Consumption MWh
Base Case,	--	--
Peak Load = 2,850 MW		
Total Energy = 51,809 MWh		
Spot Pricing	-454	-2,919
Ramsey-type Pricing	-266	+ 709
Reliability-Based†		
x = 1.00	-250	+ 898
x = 0.90	-311	+ 466
x = 0.85	-368	+ 120
y = 0.80	-165	+2,646
y = 0.60	- 81	+4,343

*See test in the appendix.

†See appendix for explanation of parameters x and y.

customer class 1 with higher consumption and lower elasticity[2] opts for lower reliability (x = 0.85). However, a similar comparison when x = 0.9 shows a welfare loss.

When the customer class 2 with its higher elasticity and lower consumption opts for reduced reliability (y = 0.8) there is a welfare loss as compared to the absence of reliability selection by either class which is denoted in our model by x =1 and y = 1.

Table 5-2 shows the peak reduction and the reduction in energy consumption expected by different pricing schemes. As a first approximation, it can be assumed that the amount of reserve reduction to maintain the same level of reliability is the amount of peak-demand reduction. As in the case of the economic welfare analysis, spot pricing and Ramsey-type pricing can offer the most benefits. Note, however, that consumption relative to the base case would increase under certain circumstances with Ramsey-type and reliability-based pricing.[3] But, consumption within the set of reliability-based prices could increase or decrease as compared to x = 1 and y = 1. This erratic pattern emerges because the reduction in price to the class desiring lower reliability induces consumption while the opposite is true for the other class because of the increased price. Therefore, the increase in consumption by the lower-reliability class could more than offset the reduced consumption by the higher-reliability class whenever the class with higher elasticity opts for lower reliability.

The peak load reduction pattern is the opposite of the energy consumption pattern caused by reliability prices. These reductions are the lowest when the class with higher elasticity subscribes to lower reliability. These reductions are similar in magnitude to the Ramsey-type pricing effects because the total hourly recovery of fixed costs under reliability pricing is identical to the total hourly recovery under the Ramsey-type scenario.

---

[2] In our simulation (see table A-6), the price elasticities of demand of the two classes were equal during the peak hours. During the nonpeak hours, class 1 was less elastic than class 2.

[3] This increase is due to the revenue under spot pricing being higher than the production cost plus fixed costs. If the opposite were true, the consumption would decrease.

Conclusions

This and preceding chapters have enabled us to reach several conclusions with respect to the delivery of differentiated reliability service to consumers. They are:

- Optimal rationing schemes should not be confused with reliability pricing. Reliability-based prices are used to change consumption and demand patterns. Rationing under an imminent or actual shortage to all customers minimizes outage costs.

  Priority of service classifications are ideally suited to minimize interruption costs. Yet, an extension of this procedure to formalize this current practice would require network modifications and computer-aided logic to determine the sequence of interruptions. This could entail large system-specific expenditures which can only be justified by other benefits such as economic welfare gains and reserve reductions, both leading to a more efficient resource expansion plan.

  While outage cost analysis of different consumers should aid in the planning of an overall level of reliability for the power system and outage cost surveys should aid in establishing a priority scheme of interruption during the operation of the system, they bear no relation to the economic welfare and reserve reduction benefits that may or may not be obtained by linking the price of electricity to the reliability of supply.

- The methods of calculating reliability indices at the delivery point are full of assumptions and do not give the consumer information regarding the time of expected outages and their duration. Hence, the consumer will be faced with difficult and abstract choices. It is possible that as a result of this complexity the actual choices of different grades of service may not reflect the value of reliability to these customers. But assuming for the moment that the consumers' choices are efficient, effective regulation would require a detailed examination of operating records to obtain assurance that the system was operated to honor the choices of reliabilities would be required. Clearly, this regulatory effort is not warranted if the consumers' choices are not efficient.

- The operation of an integrated generation and transmission network is complex. Outages and the contingencies of multiple outages are hard to anticipate. The operator must make immediate and irreversible decisions regarding the integrity of the system. Therefore, discriminating operation based on reliability subscription of classes may not be possible during system emergencies. Even if the operator is aided by computer intelligence, the disadvantage is that the operation of the system would become complex.

- In the event of a curtailment, a rationing or rotational sharing of shortages can be optimized to minimize the cost of interruptions. In a simplistic scheme, customers who value supply less (or whose worth of reliability is less) are rationed first. Rationing addresses minimization loss due to interruption. But optimum rationing, as a "damage control" function, has nothing to do with relating the price of electricity to the reliability of supply. Reliability-based pricing should implement the longer-term goals of the maximization of economic welfare and obtaining cost reductions due to a peak-load reduction and/or an improved load factor.

- In terms of longer-term gains in economic welfare and peak-demand reduction, the analysis indicates some advantages due to reliability-based pricing compared to the base case. The advantages depend on the elasticities and consumptions of the consumers opting for lower reliability in relation to the elasticities of other consumers.

- As expected, improvements over the present average-cost pricing are highest under spot pricing. The results of this study also indicate that significant economic and operational efficiency gains can be expected from Ramsey-type pricing. The benefits due to reliability pricing are less favorable and uncertain in terms of the level of reliability offered to consumers. It is to be expected that industrial customers are the most likely to subscribe to lower reliability. However, their demands are generally more price elastic due to fuel switching and local standby generation capabilities. Therefore, a welfare loss would result in comparison to no explicit reliability selection by either customer class. But recall that the economic welfare would improve in relation to average cost prices.

- Another complication with respect to reliability-based prices is anomalies in cost allocation and ratemaking. The peak demand reduction is due to the class not desiring lower reliability of service or desiring a priority service. Therefore, in the longer term, the savings due to more efficient resource expansion plans resulting from peak reduction and improved load factor has to be passed on to consumers not requiring a lower reliability of service. But, it will be very difficult to determine what would have been the peak demand without differentiated service. Peak demand reduction could arise from several other causes as well. Therefore, the evaluation of the savings to be passed on to the appropriate class of consumer would be difficult or almost impossible. In any event, such savings negate the premium charged for higher reliability of service. Therefore, in the longer term, the prices for different service reliabilities are expected to converge.

The studies shown in the appendix have assumed certain price-demand relationships. The benefits are dependent on these relationships and are sensitive to the elasticities as indicated by the x and y results earlier. Therefore, any attempt to install service reliability differentiation must be preceded by defensible methods of determining price-demand relationships and analysis. Such relationships are dynamic in the sense that they could change in response to price signals and in time.

In view of the likelihood that the customers with relatively high price elasticities are the most likely to chose the lower reliability services thereby mitigating the welfare gains against a backdrop of cost allocation anomalies, practical operational complications, and unknown implementation costs, it is not difficult to assert that reliability-based prices, as defined in this study, do not unequivocally improve economic welfare or reduce the total cost of producing electricity. On these grounds, we conclude at this time that there are no economic compulsions for reliability-based prices.

APPENDIX

WELFARE ANALYSIS

The purpose of the analysis is to examine the welfare gains obtainable by the price signals inherent in different methods of pricing electricity. The analysis assumes that the reliability of service is reflected in the hourly prices or in the demand charges. This is in contrast to the establishment of a priority scheme for the interruption of customers during an emergency.

The pricing based on average cost, marginal cost, deviation from marginal cost for revenue reconciliation (Ramsey pricing), two-part tariff (peak and off-peak use), and reliability-based pricing are examined. Our goal is to compare the welfare gains and reserve equipment reductions obtainable by different pricing schemes in order to examine if reliability pricing offers any advantages.

The Social Welfare Implications of Various Pricing Schemes

The traditional measure of welfare employed in evaluating public utility policies has been the following:[25]

$$W = TR + S_c - TC , \qquad \text{(A-1)}$$

where W = net social benefit, TR = total revenue, S_c = consumers' surplus, and TC = total costs.

Since our attention is on one product, namely that of electricity considered to be the same quality at all reliability levels of supply,[1] the net benefits accruing at a given output level Q can be expressed as

[1] See the text of the report for the argument in support of this.

$$W = \int_{o}^{Q} u'(q)dq - C(Q) \ , \qquad \text{(A-2)}$$

where $u'(q)$ is the (inverse) demand function relating the quantity demanded to price (see figure A-1) and $C(Q)$, the total cost $= \int_{o}^{Q} c(q)dq$. The integral in (A-2) is the 'gross surplus.' It encompasses both total revenue $TR(q) = P.Q$ as well as consumers' surplus S_c which is the shaded area in figure A-1 given by

$$S_c(q) = \int_{o}^{Q}[u'(q) - P]dq \ , \qquad \text{(A-3)}$$

where P is the price. TR - TC includes any profit (or loss) by the producer called the producer surplus S_p which is given by

$$S_p(q) = \int_{o}^{Q}[P - C(q)]dq \ . \qquad \text{(A-4)}$$

It is easy to see that maximizing social welfare requires that $dW/dq = o$. Representing $u'(q)$ as $P(q)$, which is the common symbol in economic literature for the inverse demand function, $dW/dq = o$ implies from (A-2)

$$u'(q) = P(q) = \frac{dC(q)}{dq} \ ; \qquad \text{(A-5)}$$

that is, maximizing W in (A-2) leads to price = marginal cost which is the intersection of $c(q)$ and $u'(q)$ in figure A-1.

The price elasticity of demand, η, defined as the percent change in consumption resulting from a percent change in the price is

$$\eta = \frac{\partial q}{\partial P} \frac{P}{Q} \ . \qquad \text{(A-6)}$$

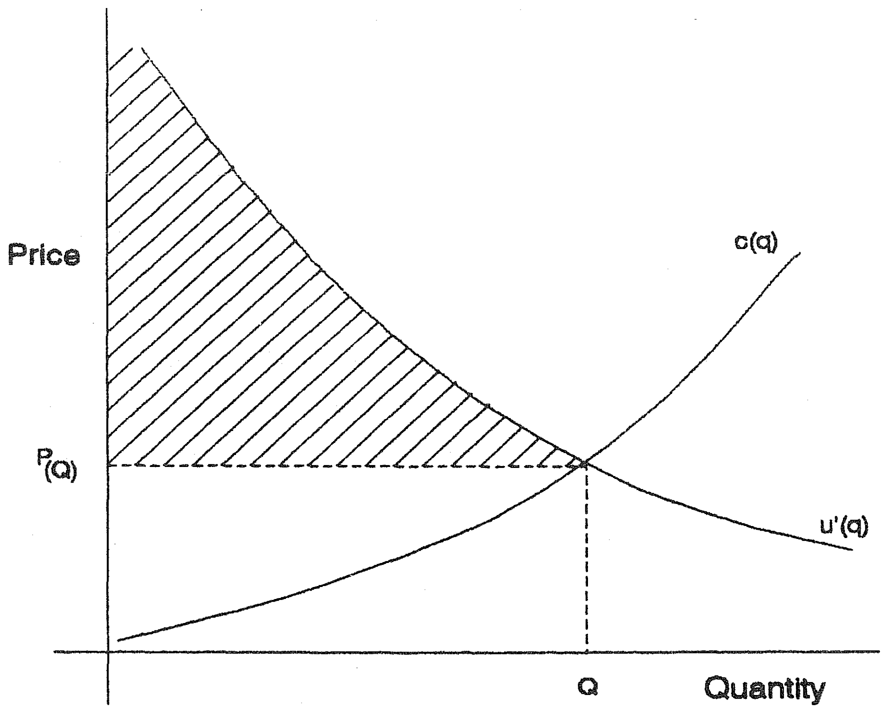


Fig. A-1. Demand-supply equilibrium.

Assumptions Regarding Demand, Generation, and Price

It is evident from the above that assumptions regarding u'(q), the demand function, and the cost function C(q) have to be made to conduct a welfare analysis. The following is a description of the example system used in this study.

Demand Pattern

Two classes of consumers were assumed: industrial and a combination of residential and commercial demands. Although the analysis could be conducted for multiple classes, it was restricted to two classes to make it simpler and to illustrate the main points of this study.

The consumption patterns of the classes were obtained from a typical midwestern utility. The demands of residential and commercial classes were combined. This combined demand and that of the industrial class are labeled class 1 and 2, respectively. The demands were appropriately scaled to match the generation model discussed below. The hourly demands for a typical day of classes 1 and 2 and the total system load are shown in table A-1.[2] Note from the table that the peak demands of the two classes occur at different times; class 1 at 6:00 P.M. and class 2 at 2:00 P.M.

Generation

Table A-2 shows the generation mix and the incremental cost of generation from each generating unit. These values are taken from reference 19. Note that the total system peak load (table A-1) is made identical to the system peak suggested in reference 19. Therefore, the peak load reliability index of our model should be identical to that of the Reliability Test System of reference 19.

[2] The study over a day is illustrative. Realistic study should consider demands over a year, or during a season, at least.

TABLE A-1

DEMANDS (MW) UNDER TRADITIONAL AVERAGE-COST PRICING

Hour	Class 1	Class 2	System
1	871	764	1,635
2	781	755	1,536
3	713	758	1,472
4	677	745	1,423
5	670	739	1,410
6	699	762	1,402
7	826	819	1,645
8	1,045	885	1,931
9	1,293	908	2,201
10	1,491	915	2,407
11	1,632	923	2,556
12	1,732	913	2,645
13	1,688	922	2,610
14	1,753	926	2,680
15	1,769	904	2,673
16	1,838	864	2,703
17	2,009	840	2,850
18	1,896	819	2,715
19	1,781	817	2,598
20	1,580	802	2,385
21	1,552	800	2,353
22	1,323	815	2,139
23	1,182	809	1,991
24	995	784	1,779

Total Energy Demand = 51,809 MWh

TABLE A-2

DETAILS OF GENERATION MODEL

Unit size MW	No. of units	Forced outage rate	Fuel	Fuel cost, \$ per MWh
400	2	0.12	Nuclear	5.59
350	1	0.08	Coal	11.40
197	3	0.05	#6 Oil	19.87
155	4	0.04	Coal	11.16
100	3	0.04	#6 Oil	22.08
76	4	0.02	Coal	14.88
20	4	0.10	#2 Oil	37.50
12	5	0.02	#6 Oil	28.56

Total capacity = 3,105 MW

From the cost of generation of each generating unit, the incremental cost of generation at various system demands can be obtained with the usual assumption that the more efficient units are fully dispatched prior to the dispatch of lesser efficient units. A least square error analysis to approximate the incremental cost versus total demand curve in the form ae^{bq_T} (q_T is the total consumption) gave values of a = 4.56284 and b = 0.000617 for this model (the goodness of fit, R = 0.916). In view of this function for c(q), the total cost function C(q) of (A-2) for any total consumption q_T works out to

$$C(q) = \frac{a}{b}\left(e^{bq_T} - 1\right). \tag{A-7}$$

The total cost curve of (A-7) is convex and the marginal cost c(q) given by $c(q) = ae^{bq_T}$ is also convex.[3]

Price of Electricity

The generation and demand models described above were used for a base case study to benchmark certain parameters. The cost of production to meet the demand indicated in table A-1 was obtained by a production costing program.[23] In this illustrative study, downtimes due to maintenance were not considered in calculating the production cost. The cost of supplying the 51,809 MWh of energy demand was $564,718, or an average of $10.90 per mWh.

In terms of capital and fixed costs, we had to resort to assumptions. For the utilities across the nation, the fuel component in the price of electricity varies from about 30 percent to 70 percent. The fuel component is dependent on the mix of generation and fuel prices. The fixed costs include the undepreciated capital component of equipment and is higher for nuclear utilities than for nonnuclear utilities. In this study, it is assumed that the fixed costs were $282,359, which is half of the production cost. Thus, the fuel component in the cost of electricity translates to 66 percent.

The average cost of electricity for the above assumed fixed costs works out to $16.30 per mWh. This is rather low compared to the national average. The recourse is to increase the fixed costs to include return and other costs or to increase the cost of fuel. In the absence of specific data from a particular utility, such adjustments are arbitrary. Besides, the fact that the price of electricity is rather low does not affect the arguments regarding the welfare analysis that follows. Therefore, no further adjustments to fuel prices or the fixed changes were warranted.

[3] Convex functions are functions for which the chord connecting any two points on the function lies above the function. For differentiable functions, strict convexity may be described by saying that f(x) always lies above the tangent line at any point. Also, the function must be convex if its second derivative is positive ($f''(x) \geq 0$), and concave if ($f''(x) \leq 0$).

In order to mimic standard ratemaking procedure, the fixed cost of \$282,359 was allocated to the two classes in the ratio of their demands at the time of system peak (coincident peak method). The allocations to classes 1 and 2 work out to (see table A-1) \$199,038 and \$83,320, respectively.

The starting point of any economic analysis is a suitable demand-price relation. Such a relationship is system-specific and can be approximated after conducting tests on consumers, using appropriate sampling techniques. In such tests, groups of customers of different classes are chosen to obtain demand-price relationships with a required degree of confidence. Due to the nascency of this field, it is difficult to find the results of conclusive tests to establish these functions. Needless to say that a proper design and conduct of such tests is of paramount importance in the future.

It is the opinion of some that the demand for electricity is relatively inelastic. That may be true in the short term, but long-term consumption, in our opinion, is certainly affected by price. The affect is more on certain categories of load than others. The inelastic nature of demand, if it is indeed so, does not affect the following analysis. Elasticities of magnitude much less than unity are used for certain hours of the analysis.

The assumption of a linear relationship between price and quantity with a certain negative slope is common in literature. We assumed a nonlinear relationship for u' of the form

$$u'_i(q, t) = \alpha_i(t)K_i(t)[e^{-K_i(t)q_i(t)}], \tag{A-8}$$

where q and t are the quantity consumed at any hour t by the i^{th} class of customer (i = 1, 2 in our case) and α and K are constant at hour t. Presumably, other forms of nonlinear relationships could be assumed. Then the final result could, of course, vary from that obtained. The motivation for assuming the function as in (A-8) was the following.

Note that an integration of (A-8) with respect to q between zero and q gives

$$u_i(q, t) = \alpha_i(t)[1 - e^{-K_i(t)q_i(t)}] . \tag{A-9}$$

Equation (A-9) includes the consumer and producer surpluses of (A-3) and (A-4), and the total cost C. Also, (A-9) is a saturating function in that it does not increase proportionately as consumption increases. This has the characteristics of diminishing returns. Therefore (A-9) can be viewed as proxy to a utility function. The derivative of (A-9) as in (A-8) represents the willingness to pay or the inverse demand function. The choice of (A-8) was as a result of the need to model saturating benefits or satisfaction with increasing consumption.

Another way of expressing the above is in terms of the price elasticity of demand. It is a trivial exercise to show that $-1/K(t)\ q(t)$ is the price elasticity of demand. This means that as the consumption increases, elasticity decreases making the consumer less sensitive to price signals.

Evaluation of Constants α and K

At the base level of consumption shown in table A-1, a value for K was chosen to reflect that the consumer has attained a certain percent of the maximum utility attainable. For instance, at any hour for a given q (if one wishes to assume that 90 percent of the maximum utility attainable was attained) one would set K such that $e^{-Kq} = 0.1$. In other words, one has implicitly modelled a saturation effect in that the increase in consumption due to price reduction to attain a certain incremental benefit would be larger than in an hour when only 60 percent of the utility was attained.

Another way of viewing this is the following. If at a price P and consumption Q (see figure A-1) the value of K is set such that $e^{-KQ} = 0.1$, it implies that the consumer has attained 90 percent or 0.9 of the maximum surplus that could be attained by further reduction of P to induce consumption. This also means that price elasticity of demand given by 1/KQ is equal to -0.434.

In this study, K(t) at each hour was set at values to correspond to the following logic. Let $e^{-K(t)q(t)} = \tau(t)$ at any hour. The hourly price of electricity to each class is the cost of energy production plus one twentyfourth of the allocated fixed cost.[4] This works out to an average hourly cost of \$17.16 per mWh for class 1 and \$14.96 per mWh for class 2. For the base case demands of table A-1 with the hourly price set as above, it was assumed that class 1 had attained 99 percent of the consumers' surplus during the hours 11:00 P.M. to 6:00 A.M. That is, in the natural course of events, the consumers' use pattern reflects the attained satisfaction. During these hours of sleep, the consumer is fully satisfied by the quantity of electricity used and therefore has attained 99 percent utility. Thus, a suitable value of K(t) can be obtained by setting $\tau(t) = 0.01$ for these hours. This corresponds to a price elasticity of demand equal to -0.217.

During the hours 7:00 A.M. to 2:00 P.M. and 8:00 P.M. to 9:00 P.M., the shoulder hours on either side of peak consumption hours, it was assumed that $\tau(t) = 0.4$ or that 0.6 of the consumer's surplus was attained by class 1. During the peak hours of 3:00 P.M. to 6:00 P.M., $\tau(t)$ was put equal to 0.2.

Table A-3 shows the values of $\tau(t)$ and K(t) assumed for each class of consumer. In addition, it shows the values for $\alpha(t)$ of the inverse demand relation in equation (A-8). Having selected K(t) by the above means, $\alpha(t)$ in (A-8) was calculated to result in a consumption corresponding to that of table A-1 when the price of electricity was set at the average price of \$17.16 per mWh for class 1 and \$14.96 per mWh for class 2.

Two aspects are noteworthy. First, as mentioned earlier, the assumption of suitable values of K(t) and $\alpha(t)$ automatically determines the demand elasticities. Due to the assumption of exponential functions, the demand elasticities are inversely proportional to consumption and are given by $-1/Kq$. Table A-4 shows the demand elasticities η_1 and η_2 of the two classes at the consumption levels of table A-1.

The second aspect is that the demand at every hour has been assumed to be independent of the demand at other hours. As discussed in chapter 2,

[4] This arises from the assumption that the customers' reactions to price changes are due to the monthly electricity bills. The fixed cost is recovered as demand charges based on the maximum demand from some customers as, for instance, from the industrial users. It is assumed in this study that the hourly demand is influenced by the average price of energy. This means that the use pattern is influenced by the monthly electricity bill that includes either implicit or explicit demand charges in it.

TABLE A-3

PARAMETERS OF INVERSE DEMAND FUNCTION

	Class 1			Class 2		
Hour	$\tau_1(t)$	$K_1(t) \times 10^{-3}$	$\alpha_1(t) \times 10^4$	$\tau_2(t)$	$K_2(t) \times 10^{-3}$	$\alpha_2(t) \times 10^4$
1	0.01	5.28	32.51	0.1	3.01	4.99
2	0.01	5.89	29.13	0.1	3.04	4.93
3	0.01	6.45	26.62	0.1	3.03	4.95
4	0.01	6.79	25.28	0.1	3.08	4.87
5	0.01	6.86	25.02	0.1	3.11	4.83
6	0.01	6.58	26.10	0.4	1.20	3.13
7	0.40	1.10	3.80	0.4	1.11	3.36
8	0.40	0.87	4.90	0.4	1.03	3.63
9	0.40	0.70	6.06	0.4	1.00	3.72
10	0.40	0.61	6.94	0.4	1.00	3.75
11	0.40	0.56	7.65	0.4	0.99	3.79
12	0.40	0.52	8.11	0.4	1.00	3.75
13	0.40	0.54	7.91	0.4	0.99	3.78
14	0.40	0.52	8.22	0.4	0.98	3.80
15	0.20	0.90	9.44	0.4	1.01	3.71
16	0.20	0.87	9.81	0.4	1.05	3.55
17	0.20	0.80	10.72	0.4	1.08	3.45
18	0.20	0.84	10.12	0.4	1.11	3.36
19	0.20	0.90	9.51	0.4	1.12	3.35
20	0.40	0.57	7.40	0.4	1.13	3.30
21	0.40	0.59	7.27	0.4	1.14	3.28
22	0.40	0.69	6.20	0.4	1.12	3.34
23	0.01	3.89	44.10	0.1	2.84	5.28
24	0.01	4.62	37.12	0.1	2.93	5.12

TABLE A-4

DEMAND ELASTICITIES AT BASE CASE CONSUMPTION

Hour	η_1	η_2
1	-0.21714724	-0.43429448
2	-0.21714724	-0.43429448
3	-0.21714724	-0.43429448
4	-0.21714724	-0.43429448
5	-0.21714724	-0.43429448
6	-0.21714724	-1.09135667
7	-1.09135667	-1.09135667
8	-1.09135667	-1.09135667
9	-1.09135667	-1.09135667
10	-1.09135667	-1.09135667
11	-1.09135667	-1.09135667
12	-1.09135667	-1.09135667
13	-1.09135667	-1.09135667
14	-1.09135667	-1.09135667
15	-0.62133493	-1.09135667
16	-0.62133493	-1.09135667
17	-0.62133493	-1.09135667
18	-0.62133493	-1.09135667
19	-0.62133493	-1.09135667
20	-1.09135667	-1.09135667
21	-1.09135667	-1.09135667
22	-1.09135667	-1.09135667
23	-0.21714724	-0.43429448
24	-0.21714724	-0.43429448

this assumption may not be valid for all types of loads. In a later section of the appendix, the results obtained by assuming certain cross-coupling between the hourly demands is outlined.

Marginal-Cost Pricing

It was mentioned earlier that the maximization of welfare requires the product to be priced at the marginal cost. The following is the derivation of this result considering two classes of customers.

The welfare maximization problem at any hour t can be formally written as:

$$\text{Maximize } W = \sum_{i=1,2} \int_{o}^{q_i} \alpha_i K_i e^{-K_i q_i} dq_i - \text{Cost} . \tag{A-10}$$

The cost function is approximated by (A-7) as

$$C(q) = \frac{a}{b} (e^{bq_T} - 1) , \tag{A-7}$$

$$\text{where } q_T = \sum_{i=1,2} q_i . \tag{A-11}$$

Using (A-7) for the cost in (A-10) one has the welfare problem as

$$\text{Max } W = \sum_{i=1,2} \int_{o}^{q_i} \alpha_i K_i e^{-K_i q_i} dq_i - C(q_T) . \tag{A-12}$$

The first order conditions (FOC) give

$$\frac{\partial W}{\partial q_1} = \alpha_1 K_1 e^{-K_1 q_1} - \frac{\partial c}{\partial q_1} = 0, \tag{A-13}$$

and

$$\frac{\partial W}{\partial q_2} = \alpha_2 K_2 e^{-K_2 q_2} - \frac{\partial c}{\partial q_2} = 0. \tag{A-14}$$

From (A-7) and (A-11), one has

$$C' = \frac{\partial c}{\partial q_1} = \frac{\partial c}{\partial q_2} = ae^{bq_T} \quad . \tag{A-15}$$

Therefore, from (A-13) and (A-14) one gets

$$\alpha_1 K_1 e^{-K_1 q_1} = C' \ , \qquad \text{and} \tag{A-16}$$

$$\alpha_2 K_2 e^{-K_2 q_2} = C' \tag{A-17}$$

indicating that the price for both consumers is set equal to marginal cost.

In order to determine the quantities, one has from (A-16) and (A-17)

$$\alpha_1 K_1 e^{-K_1 q_1} = \alpha_2 K_2 e^{-K_2 q_2} \quad . \tag{A-18}$$

Taking logarithms on both sides gives

$$K_2 q_2 - K_1 q_1 = \ln \frac{\alpha_2 K_2}{\alpha_1 K_1} \quad . \tag{A-19}$$

Substitution of (A-11) into (A-19) results in

$$q_2^* = Rq_T + S \ , \tag{A-20}$$

where q_2^* is the optimal consumption of class 2 to maximize welfare, and

$$R = K_1/(K_2 + K_1) \ , \tag{A-21}$$

and

$$S = \ln \frac{\alpha_2 K_2}{\alpha_1 K_1} \ / \ (K_2 + K_1) \ . \tag{A-22}$$

Using (A-20) in (A-17), one has

$$\alpha_2 K_2 e^{-K_2(Rq_T+S)} = ae^{bq_T} \ . \tag{A-23}$$

Taking logarithms on both sides yields upon simplification

$$q_T^* = \frac{\ln \ \alpha_2 K_2 \quad - K_2 S}{b + K_2 R} \ , \tag{A-24}$$

where q_T^* is the optimal total consumption. By a procedure similar to the above, one can get

$$q_1^* = R' q_T^* - S \ , \tag{A-25}$$

where

$$R' = K_2/(K_1 + K_2) \ . \tag{A-26}$$

Hence, the quantities of consumption for welfare maximization resulting from marginal-cost pricing can be obtained.

The literature is rife with discussion regarding the marginal cost, if it should include longer term costs to maximize welfare in the long run, and the methods of calculation of such costs. For the purpose of present discussion, the concern is only with regard to the marginal cost of production (sometimes referred to as incremental cost). The pricing of energy equal to the marginal cost of production at each hour has been referred to in the literature as spot pricing.[10]

Results

The consumption of the classes at each hour was calculated using the above procedure in a digital computer program.[5] Table A-5 shows the results obtained. Note that the peak load has been reduced compared to the base case and the welfare is higher than that of the base case (as expected). The total revenue to the producer is also indicated in the table to be $817,085. The total cost of production is shown as $456,940. The cost which is a summation of hourly costs is calculated by integrating the marginal-cost function of equation (A-15) between zero and the hourly consumption. Therefore, it is somewhat less than what is obtained from a production costing simulation[23] as the latter accounts for forced outages of generating units.

Note that the revenue exceeds the cost of production by $360,145. Whether the revenue exceeds the total costs is another question. If the fixed costs are $282,359 as assumed, the revenue exceeds total costs. The producer then has a situation of overrecovery.[6] However, if the fixed

---

[5] The consumptions, welfare, cost, and revenue can be calculated from closed-form expressions without the use of a digital computer. For instance, since q_1, q_2, and q_T are known from (A-25), (A-20), and (A-24), their substitution in (A-10) gives the expression for W at each hour. Also, since revenue is $\sum_j P_j q_{Tj}$ for j = 1, twenty-four hours, where P_j is the hourly price and q_{Tj} is the hourly consumption, for total revenue one gets the expression

$$\text{Revenue} = \sum_{j=1,24} \gamma a e^{b\gamma} ,$$

where

$$\gamma = \frac{\ln \dfrac{\alpha_2 K_2}{a} - K_2 S}{b + K_2 R} .$$

[6] Recall that return and other costs have been ignored.

TABLE A-5

SPOT PRICING: WELFARE MAXIMIZATION BY P = MC

Hour	Demand MW Class 1	Demand MW Class 2	Demand MW Total	Price ($/mWh) = MC	S_p $	W $
1	920.94	805.29	1,727.24	13.24	14,980	354,137
2	833.54	813.06	1,646.61	12.60	13,237	321,379
3	766.78	827.23	1,594.01	12.20	12,173	297,510
4	731.89	821.29	1,553.19	11.89	11,381	284,051
5	725.21	817.05	1,542.27	11.81	11,179	281,251
6	746.23	905.75	1,652.99	12.65	13,370	266,785
7	990.64	862.93	1,853.57	14.31	18,001	30,813
8	1,151.72	846.67	1,998.39	15.65	21,951	34,375
9	1,320.88	795.89	2,116.77	16.84	25,611	37,506
10	1,443.44	753.34	2,196.78	17.69	28,323	39,720
11	1,522.32	727.02	2,249.35	18.28	30,218	41,209
12	1,579.54	700.19	2,279.74	18.62	31,356	42,107
13	1,553.16	711.57	2,267.74	18.48	30,903	41,713
14	1,586.78	703.51	2,290.29	18.74	31,759	42,795
15	1,652.29	668.65	2,320.94	19.10	32,951	53,104
16	1,706.19	630.11	2,336.30	19.28	33,561	69,507
17	1,818.10	578.53	2,396.64	20.01	35,041	73,323
18	1,753.63	591.71	2,345.35	19.39	33,924	70,601
19	1,680.45	616.78	2,297.23	18.82	32,026	57,877
20	1,531.18	663.58	2,194.76	17.67	28,253	39,851
21	1,517.10	665.74	2,182.85	17.54	27,837	39,526
22	1,370.35	725.10	2,096.45	16.63	24,951	37,113
23	1,205.80	791.73	2,000.53	15.67	22,017	165,383
24	1,036.40	803.70	1,840.11	14.20	17,564	393,979
Total			48,980.00		322,057	3,446,291

Revenue = $817,085 Cost of production = $456,940

S_{c1} = 2,734,468 S_{c2} = 351,681

costs are higher than $360,145, spot pricing results in an underrecovery to the producer. Therefore, the price has to be modified to result in appropriate revenue reconciliation. The following analysis addresses the welfare maximization problem with the constraint of revenue reconciliation.

Revenue Reconciliation

There is a major concern in setting the price equal to marginal cost of production. When the price is set equal to marginal cost, the revenue to the producer should not exceed (or fall below) a value reflecting the allowable costs and a fair return. Thus, in addition to the welfare maximization resulting from Price = MC, one has the constraint of revenue reconciliation. A rigorous method of revenue reconciliation when the price could vary at each hour is presented below.[7]

Ramsey Pricing

The object is to

$$\text{Maximize } W = \sum_i \int_o^{q_i} P_i(q)dq_i - C(q) - F \; ; \; i=1,2 \qquad \text{(A-27)}$$

subject to

$$\sum_{i=1}^{2} q_i P_i(q) - C(q) \geq \pi \; , \qquad \text{(A-28)}$$

where P(q) is the inverse demand function, C(q) is the production cost function, and π is the revenue reconciliation component which, in our case, should be equal to the fixed costs F.

[7] This method is originally due to Ramsey. See reference 12.

The Legrangian function using the multiplier λ is

$$L = W + \lambda\left(\sum_{i=1}^{2} q_i P_i(q) - C(q) - \pi\right) \quad . \tag{A-29}$$

The FOC yields

$$\frac{\partial L}{\partial q_1} = P_1(q) - MC_1(q) + \lambda[P_1(q) + q_1 P_1'(q) - MC_1] = 0 \tag{A-30}$$

$$\frac{\partial L}{\partial q_2} = P_2(q) - MC_2(q) + \lambda[P_2(q) + q_2 P_2'(q) - MC_2] = 0 \tag{A-31}$$

$$\frac{\partial L}{\partial \lambda} = P_1 q_1 + P_2 q_2 - C(q) - \pi = 0 \ . \tag{A-32}$$

In the above, MC represents marginal cost which is the derivative of C(q), P_1' represents the derivative of the demand function Wrt q_1, and P_2' the derivative Wrt q_2.

Equations (A-30) to (A-32) represent three nonlinear equations with three unknowns, namely q_1, q_2, and λ. Their solution proceeds as follows. From (A-30), one has

$$-\lambda = (P_1 - MC_1)/(P_1 - MC_1 + q_1 P_1') \tag{A-33}$$

which can be rewritten as

$$-\frac{1}{\lambda} = \frac{P_1 - MC_1}{P_1 - MC_1} + \frac{q_1 P_1'}{P_1 - MC_1} + 1 + \frac{q_1 P_1'}{P_1 - MC_1} \quad , \tag{A-34}$$

or

$$\frac{q_1 P_1'}{P_1 - MC_1} = -1 - \frac{1}{\lambda} = \frac{-\lambda - 1}{\lambda} = -\frac{1 + \lambda}{\lambda} \quad . \tag{A-35}$$

Writing (A-35) in its reciprocal manner obtains

$$\frac{P_1 - MC_1}{q_1 P_1'} = - \frac{\lambda}{1 + \lambda} \qquad \text{(A-36)}$$

or,

$$P_1 - MC_1 = - \frac{\lambda}{1 + \lambda} q_1 P_1' , \qquad \text{(A-37)}$$

which when divided by P_1 yields

$$\frac{P_1 - MC_1}{P_1} = - \frac{\lambda}{1 + \lambda} \frac{q_1}{P_1} \frac{\partial P_1}{\partial q_1} . \qquad \text{(A-38)}$$

Writing (A-38) in terms of demand elasticities η_i, and recognizing that (A-30) is identical to (A-31) but with a change of subscripts, one gets

$$\frac{P_i - MC_i}{P_i} = \frac{-\lambda}{1 + \lambda} \frac{1}{\eta_i} \qquad \text{(A-39)}$$

in which $\lambda/(1 + \lambda)$ is called the Ramsey number ψ. Then,

$$\frac{P_i - MC_i}{P_i} = - \frac{\psi}{\eta_i} , \qquad \text{(A-40)}$$

where when $\psi = 1$ results in the profit maximization solution.[8] Equation

[8] The proof of this is evident from the fact that the marginal revenue $MR = P + q \frac{\partial P}{\partial q}$ where P is the price. When $\psi = 1$, (A-39) yields

$$\frac{P - MC}{P} = -1/\eta = - \frac{q}{P} \frac{\partial P}{\partial q}$$

hence

$$P - MC = -q\partial P/\partial q, \quad \text{or}$$

Marginal cost = Marginal revenue, the condition for profit maximization.

(A-40) is called the inverse elasticity rule since it says that the percent deviation of price from marginal cost should be inversely proportional to elasticity.

We now substitute the assumed inverse demand function in (A-40) as follows. Note that (A-40) can be simplified to give

$$P_i = \frac{\eta_i MC}{\eta_i - \psi} \tag{A-41}$$

substituting the expression for P_i in (A-41), one gets

$$\alpha_i K_i e^{-K_i q_i} = \eta_i MC/(\eta_i - \psi). \tag{A-42}$$

Taking logs results in

$$\ln \alpha_i\ K_i - K_i q_i = \ln \frac{\eta_i MC}{\eta_i + \psi} \tag{A-43}$$

yielding the solution for the optimal quantities as

$$q_i^* = \frac{\ln \alpha_i K_i}{K_i} - \frac{\ln \dfrac{\eta_i MC}{\eta_i + \psi}}{K_i}, \tag{A-43}$$

which when used in the constraint equation (A-28), gives (using equality)

$$\sum_i q_i^* \alpha_i K_i e^{-K_i q_1^*} - ae^{b(q_1^* + q_2^*)} = \pi . \tag{A-44}$$

Computational Procedure

The computational procedure starts with a positive value of λ. It is decreased in small quantums in each iteration until the relation of (A-32) is satisfied. To understand the solution procedure and to improve the efficiency of the computation, it is necessary to examine the behavior of the three critical functions. They are the Ramsey number ψ related to λ according to (A-39), revenue R, and the cost of production.

Nature of ψ and λ

Figure A-2 represents the relation between λ and ψ as given by (A-30). From (A-43), since the argument of logarithm is positive, one has

$$\frac{\eta_i MC}{\eta_i - \psi} \geq 0 \ . \tag{A-45}$$

Since η_i, the elasticity is negative, (A-45) requires that

$$\eta_i + \psi < 0$$

or,

$$\psi < -\eta_i \quad .$$

The iteration procedure, therefore, starts with ψ (and a corresponding λ) equal to the lesser of the two elasticities. Let η_{im} be the minimum of the two elasticities. In each iteration, λ is decreased by a small value. This procedure is tantamount to tracing a locus along the direction of the arrows shown in figure A-2.

In figure A-2, the regions where price is greater than and less than marginal cost are shown. These arise from (A-40) from which it is clear that P = MC when $\psi = 0$. Also, when P > MC, (A-40) requires that ψ and λ to be positive and when P < MC, ψ and λ should be negative. When P >> MC, $\psi >$ 1 and, therefore, $\lambda < -1$. This last condition leads to unrealistically high prices and low consumptions and is ignored. From the above, it is clear that the range of search for λ is from a value that corresponds to $\psi = -\eta$ to $\lambda = -1.0$.

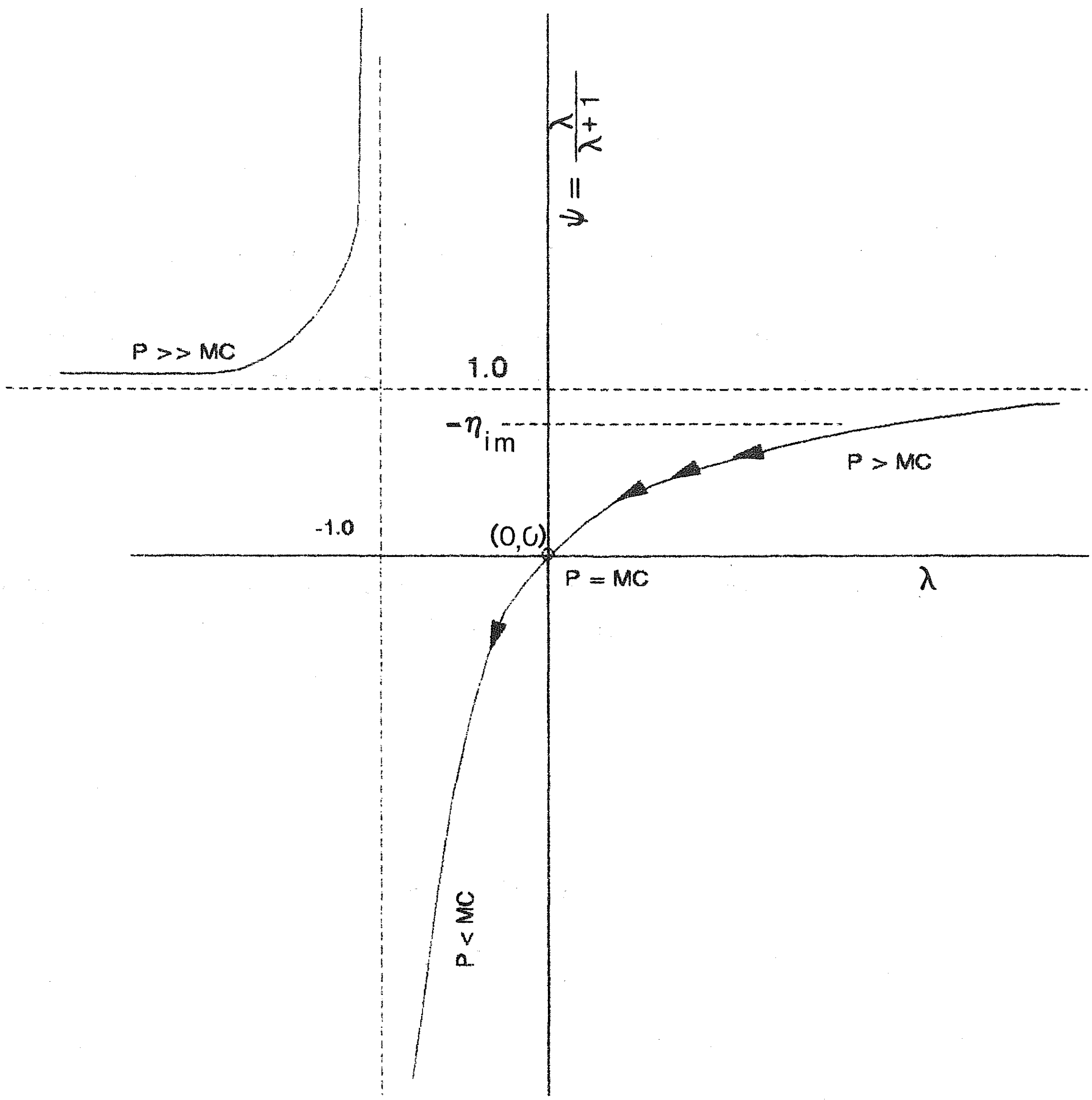


Fig. A-2. Relation between ψ and λ.

Nature of Revenue

It is evident that the revenue has a maximum value at some q_T as shown in figure A-3.[9] Along the trajectory of λ traced by the digital computer program, price is progressively decreasing resulting in increased total consumption. Therefore, at some value of λ_1, the revenue has a peak value after which it starts to decrease. Note that the value of λ at which the revenue is a maximum depends on the elasticities of demand. Since the demand elasticities vary from hour to hour, this value of λ can be to the right or left of the origin in figure A-2.

Cost of Production

The cost of generation increases as λ (and, therefore, q_T) increases. This is shown in figure A-3a with a vertical offset equal to the fixed cost.

---

[9] The total change in revenue dR is given by

$$dR = \sum_i \frac{\partial R}{\partial q_i} \Delta q_i \text{ where } R = \sum_i q_i \alpha_i K_i e^{-K_i q_i} .$$

The condition for maximum revenue is obtained by putting dR = 0 for i=1,2 as

$$\frac{\Delta q_1 \alpha_1 K_1 e^{-K_1 q_1}}{\Delta q_2 \alpha_2 K_2 e^{-K_2 q_2}} = \frac{-(1 - k_2 q_2)}{(1 - K_1 q_1)} .$$

Since $\eta = -1/Kq$, the above condition can be written as

$$\frac{\Delta q_1 \alpha_1 K_1}{\Delta q_2 \alpha_2 K_2} e^{(K_2 q_2 - K_1 q_1)} = - \frac{(1 + 1/\eta_2)}{(1 + 1/\eta_1)} .$$

If there is only one consumer, i.e., i = 1, it is a trivial exercise to prove that the maximum revenue results at $\eta = -1.0$. Recall that $\eta = -1/Kq$.

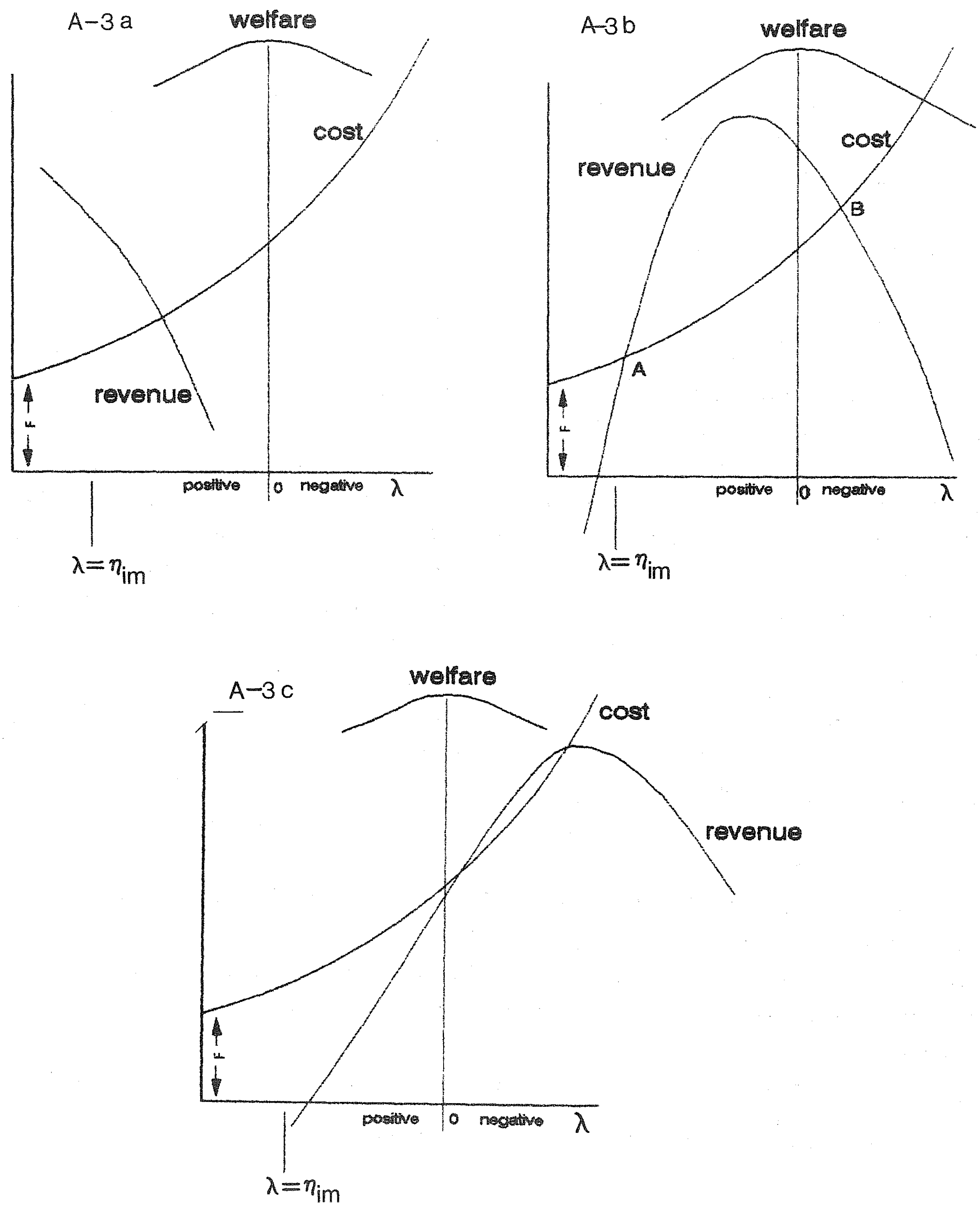


Fig. A-3. Revenue and cost functions.

There are two possible intersections[10] between cost and revenue (solutions) at A and B for the Ramsey pricing problem. The following logic identifies one of the intersections as the correct solution.

Solution to the Problem

Proceeding along the trajectory from a starting value of $\lambda = -\eta_{im}$ where $-\eta_{im}$ is the minimum of the two elasticities of the base case consumption, the revenue curve (figure A-3) starts to increase or decrease. Since the peak of the revenue can be either to the left of the origin, it might be to the right of $-\eta_{im}$ (see footnote 8). Then, at the start of iterations, $\partial W/\partial\lambda$ is positive or the revenue decreases along the trajectory traced by the iteration. If both elasticities are <1.0, the peak of the revenue is to the left of $-\eta_{im}$. In that case, at the start of the iteration, $\partial W/\partial\lambda$ is negative, or the revenue increases along the trajectory traced. The peak value of welfare is, of course, at the origin.

The cost function, as said earlier, increases along the trajectory traced by the program. The first intersection (Cost = Revenue), if revenue is decreasing, is the solution since no further intersections are possible. This situation is shown in figure A-3b. If the revenue is increasing at the first intersection, an additional check is necessary to determine if the intersection is the desired solution. This intersection could be either to the left or right of the origin in figure A-2. If the welfare is increasing in such a first intersection (signifying positive values of λ) the intersection is ignored. The second intersection of cost and revenue gives higher welfare and is the desired solution. This situation is portrayed in figure A-3a.

If the first intersection is at a negative value of λ, the welfare is decreasing ($\frac{\partial W}{\partial \lambda}$ is + ve). Hence, at the second intersection, one would

---

[10] Note that if F is very large, there will be no intersection and, therefore, the solution to the problem does not exist. For some particular value of F, there is only one intersection which gives a unique solution to the problem.

obtain a smaller value of welfare. Therefore, the first intersection gives the correct solution. This is shown graphically in figure A-3c.

It is clear from the above that the solution chosen could be at values of λ to the left or the right of the origin in figure A-2. Therefore, depending on the elasticities of demand, the price could be more or less than the marginal cost. The logical diagram for detecting the correct solution in the computer program is shown in figure A-4.

Results

In obtaining a solution to the problem, it is important to decide how much of F has to be recovered in each hour. There appears to be no a priori guidance to select the hourly values for F. Once could consider different hourly divisions of F and examine the resulting welfare to decide on the best hourly allocation of F. Such hourly allocations can be related to the elasticities. But, to reiterate, it is not our intention to establish optimum rates. The goal is to examine the effect of reliability pricing on welfare and peak load reduction. Therefore, only two possible hourly allocations were considered. One was that of dividing F equally between all the twenty-four hours. The other was to recover the same percent of total fixed costs in each hour as in table A-5. The latter would offer a comparison between Ramsey pricing and spot pricing schemes. Table A-6 shows the results obtained when π was set at an hourly value of $11,764.90 corresponding to a total fixed cost recovery of $282,359 spread equally over twenty-four hours. Also shown are the values of surplus and welfare.

It is intuitive to expect that the price will be set below marginal cost to both the consumers (in the ratios of inverse elasticities) to reduce the revenue to the producer. But this may not be true at all hours. As was explained earlier, the solution depends on the elasticities of demand. During the hours when the elasticities are low, there is indeed an increase in price compared to the case of welfare maximization. This can be seen in table A-6. Hours 1 to 7 and hour 24 have higher prices than under the MC pricing case.

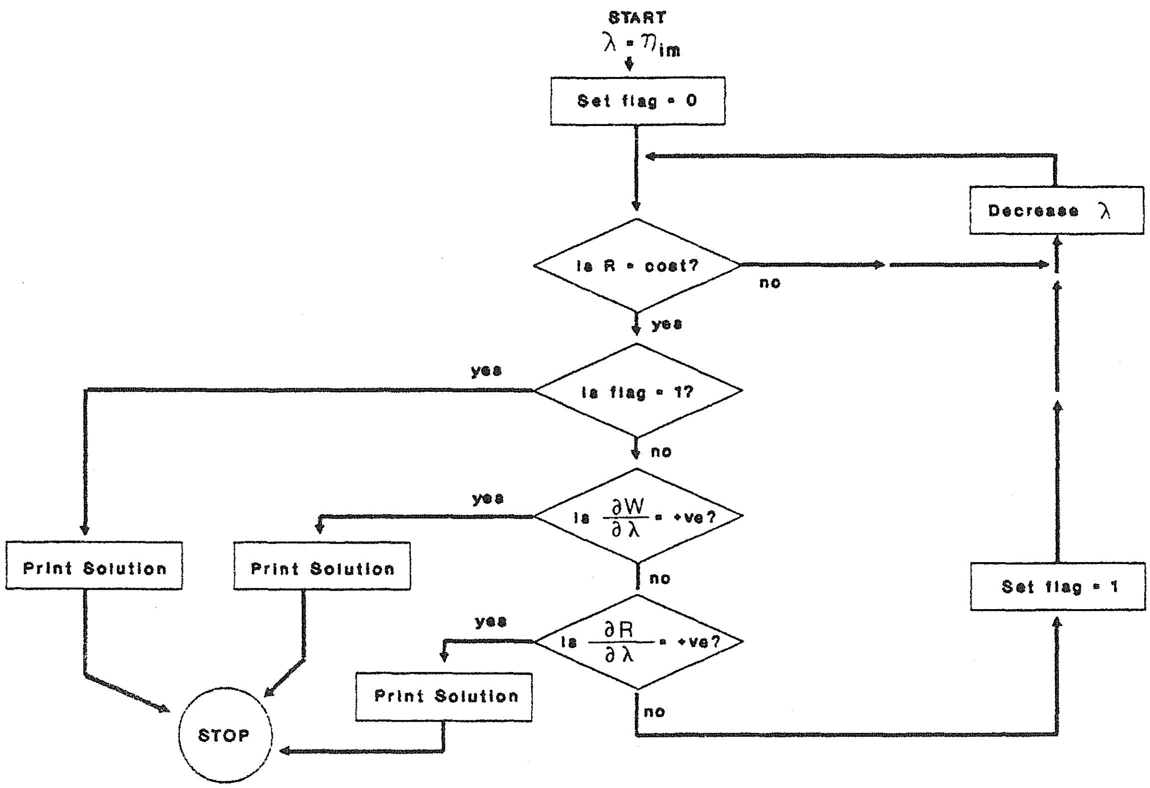


Fig. A-4. Flow chart of computer program.

TABLE A-6

RAMSEY PRICING WITH F = \$282,359, HOURLY π = \$11,764
DEMAND AND PRICES OF ENERGY FOR THE TWO CLASSES

Hour	Demand MW			Price \$/MWh		* MC \$/MWh
	Class 1	Class 2	Total	Class 1	Class 2	
1	888	779	1,657	15.7	14.3	13.2
2	791	774	1,566	16.1	14.1	12.6
3	721	781	1,503	16.3	13.9	12.2
4	677	765	1,443	17.2	14.1	11.9
5	673	764	1,437	16.8	13.9	11.8
6	698	855	1,554	17.2	13.4	12.6
7	948	826	1,775	15.0	14.9	14.4
8	1,205	886	2,092	14.9	15.0	15.6
9	1,467	885	2,352	15.1	15.3	16.8
10	1,656	865	2,522	15.5	15.8	17.7
11	1,783	853	2,636	15.7	16.1	18.3
12	1,869	830	2,700	15.9	16.3	18.6
13	1,830	843	2,673	15.9	16.2	18.5
14	1,884	836	2,720	16.0	16.4	18.7
15	1,886	769	2,655	15.4	17.2	19.1
16	1,951	726	2,677	15.5	17.4	19.2
17	2,107	676	2,783	15.8	17.9	20.0
18	2,010	683	2,693	15.5	17.5	19.4
19	1,902	702	2,605	15.4	17.0	18.8
20	1,751	760	2,512	15.5	15.8	17.6
21	1,728	759	2,488	15.4	15.7	17.5
22	1,507	799	2,306	15.1	15.3	16.6
23	1,226	808	2,035	14.4	15.0	15.6
24	1,022	793	1,815	15.1	14.6	14.2
Total:			53,224			

W = \$3,428,959 S_{c1} = \$2,783,433 S_{c2} = \$363,609 S_p = \$281,463

* This represents the marginal cost of production (and price) for the demands under the welfare maximization case.

Table A-7 shows the results obtained when hourly recovery of F corresponded to the spot pricing scheme shown in table A-5. To elucidate this further, the percent of fixed costs recovered in hour 1 of table A-5 is \$14,980 ÷ 322,057 = 4.65 percent. The same percent of the allowable fixed costs (\$282,359) was set equal to π in (A-28) for that hour. For other hours, similar calculations established the value of π. Note that the welfare has decreased compared to the spot pricing case. The peak load has increased from 2,396 MW in the spot pricing case to 2,584 MW.

A comparison of tables A-6 and A-7 shows that the welfare in table A-7 is higher than in table A-6. This substantiates that an unequal recovery of fixed costs in each hour could enhance the welfare. However, there is no guarantee that the hourly allocation used in A-7 produces the highest welfare possible. Perhaps there could be other allocations that would enhance the welfare even further. It is clear that our intention is not to optimize rates. Therefore, the allocations of F used in table A-7 were also used in the yet to be discussed reliability pricing scheme in order that comparisons between Ramsey-type pricing and reliability pricing could be made.

Reliability-Based Pricing

In the above formulation of Ramsey prices, while the constraint stipulated a certain total revenue recovery, there was no explicit stipulation of recovery from each class. Consider the following formulation with a constraint for class revenue recovery as

$$\text{Maximize } W = \sum_{i=1,2} \int_{o}^{q_i} \alpha_i K_i e^{-K_i q_i} dq_i - C(q) - F \tag{A-45}$$

subject to

$$P_1 q_1 = \frac{q_1}{q_T} C(q) + \frac{q_1}{q_T} F \tag{A-46}$$

TABLE A-7

RAMSEY PRICING WITH π = \$282,359:
HOURLY PERCENTAGE OF F SAME AS IN TABLE A-5

Hour	Demand MW			Price \$/MWh		* MC \$/MWh
	Class 1	Class 2	Total	Class 1	Class 2	
1	941	824	1,766	11.9	12.5	13.2
2	850	830	1,681	11.4	11.9	12.6
3	783	846	1,629	10.9	11.5	12.2
4	747	838	1,586	10.7	11.3	11.9
5	740	834	1,575	10.6	11.2	11.8
6	768	935	1,703	10.9	12.2	12.6
7	1,065	928	1,993	13.2	13.3	14.4
8	1,249	919	2,169	14.3	14.5	15.6
9	1,447	872	2,320	15.4	15.6	16.8
10	1,591	831	2,423	16.2	16.4	17.7
11	1,681	802	2,482	16.7	16.9	18.3
12	1,748	775	2,524	17.0	17.3	18.6
13	1,720	791	2,512	16.9	17.1	18.5
14	1,757	779	2,537	17.1	17.4	18.7
15	1,777	720	2,498	17.0	18.1	19.1
16	1,834	678	2,513	17.2	18.3	19.2
17	1,959	625	2,584	17.9	19.0	20.0
18	1,881	636	2,517	17.4	18.4	19.4
19	1,802	662	2,465	16.9	17.9	18.8
20	1,683	729	2,413	16.2	16.4	17.6
21	1,665	731	3,396	16.0	16.3	17.5
22	1,495	793	2,289	15.2	15.4	16.6
23	1,235	814	2,056	13.9	14.8	15.6
24	1,060	823	1,883	12.7	13.4	14.2
Total:			52,518			

W = 3,439,848

* This represents the marginal cost of production (and price) for the demands under the welfare maximization case.

$$P_2 q_2 = \frac{q_2}{q_T} C(q) + \frac{q_2}{q_T} F , \qquad \text{(A-47)}$$

where C(q) is the cost of energy production, and F is the demand-related fixed cost. Note that in every hour the cost of production as well as the demand charges are spread to each class in proportion to their consumption. This assumption of pricing is but one possibility. Of course, other variations are possible.

With the above assumption regarding pricing, one way of introducing reliability considerations into the constraint equations (A-46) is (A-47) as follows:

$$P_1 q_1 = \frac{q_1}{q_T} C(q) + \frac{q_1}{q_T} F x , \qquad \text{(A-46a)}$$

$$P_2 q_2 = \frac{q_1}{q_T} C(q) + \frac{q_2}{q_T} F + \frac{q_1}{q_T} F (1 - x). \qquad \text{(A-47a)}$$

In the above, x is a factor to account for the reliability of service achieved by an appropriate operating practice. For instance, x = 0.9 represents a 10 percent reduction in the class 1 demand charges to account for a lower level of reliability. Naturally, since it is assumed that the total revenue to the producer is fixed, class 2 demand allocation has to be increased by a sum equivalent to the last term in (A-47a). Similarly, x = 1.2 indicates an enhanced level of reliability to class 1 with a corresponding reduction in the demand allocation for class 2. When x = 1, (A-46a) and (A-47a) degenerate to (A-46) and (A-47) and the two classes are served without any wanton differentiation of reliability.[11] In the above formulation, the fact that class 1 opts for a lower level of reliability is reflected in a lower cost allocation to that class.

[11] Equations (A-46) and (A-47) are not the same as the previous Ramsey pricing case where there is a constraint on total revenue. Here there is a constraint for revenue recovery from each class.

Class 2 can also subscribe to a different reliability of supply. To represent such a subscription by either class, (A-46) and (A-47) can be written as

$$P_1 q_1 = \frac{q_1}{q_T} C(q) + \frac{q_1}{q_T} F x + \frac{q_2}{q_T} F(1 - y) \qquad \text{(A-46b)}$$

$$P_2 q_2 = \frac{q_2}{q_T} C(q) + \frac{q_2}{q_T} F y + \frac{q_1}{q_T} F(1 - x) \qquad \text{(A-47b)}$$

where x and y account for the subscription of reliability by classes 1 and 2 respectively.[12]

In the interest of simplicity of expressions, the following deviations use (A-46a) and (A-47a). The additional terms due to y can be easily worked out by the reader and will not alter the Jacobian matrix derived later.

The FOC for solution of (A-45), (A-46a), and (A-47a) yield (λ_1 and λ_2 are the Legrangian multipliers)

$$f_1 = P_1(1 + \lambda_1 + \lambda_1/\eta_1) - C' + \frac{q_2}{q_T^2} [C + F_2] (\lambda_2 - \lambda_1) - \frac{C'}{q_T} (\lambda_1 q_1 + \lambda_2 q_2) = 0 \qquad \text{(A-48)}$$

[12] In the above, it is clear that the class offered a lower price (class 1 when x = 0.8, for example) increases its consumption while the opposite is true of the other class. Therefore, any reduction in peak demand is not due to the class opting for lower reliability but is a result of decreased consumption of the other class. Hence, in the longer term, any reduction in fixed costs F due to reduced reserve equipment should be passed on to the class desiring a higher reliability of supply. This has the opposite effect of reliability-based pricing. A further examination of this issue is deferred to a later section where the results of the studies are examined.

$$f_2 = P_2(1 + \lambda_2 + \lambda_2/\eta_2) - C' + \frac{q_1}{q_T^2}[C + Fx](\lambda_1 - \lambda_2) - \frac{C'}{q_T}(\lambda_1 q_1 + \lambda_2 q_2) = 0 \qquad \text{(A-49)}$$

$$f_3 = P_1 q_1 - \frac{q_1}{q_T}(F\,x + C) \qquad \text{(A-50)}$$

$$f_4 = P_2 q_2 - \frac{q_2}{q_T}(F + C) - \frac{q_1}{q_T}F(1 - x)\ . \qquad \text{(A-51)}$$

The above equations (A-48) to (A-51) are solved by the Gradient or the Newton-Raphson technique using a digital computer. The procedure is to first assume a value for the unknowns, q_1, q_2, λ_1, and λ_2. The corrections (Δq, etc.) to the assumed vector are given by

$$\begin{bmatrix} \Delta q_1 \\ \Delta q_2 \\ \Delta\lambda_1 \\ \Delta\lambda_2 \end{bmatrix} = \begin{bmatrix} \\ J^{-1} \\ \\ \end{bmatrix} \begin{bmatrix} f_{10} \\ f_{20} \\ f_{30} \\ f_{40} \end{bmatrix} \qquad \text{(A-52)}$$

where J^{-1} is the inverse of the Jacobian matrix and f_{10}, f_{20}, f_{30}, and f_{40} are the values of functions (A-48) to (A-51) at the assumed values of the variables. The elements j_{mn} (m^{th} row, n^{th} column) of the Jacobian are[13]

$$j_{11} = \frac{\partial f_1}{\partial q_1} = P'D_1 - \frac{2q_2}{q_T^2}\left[\frac{C + Fx}{q_T} - C'(\lambda_2 - \lambda_1)\right] - \frac{C''}{q_T}(q_T + D_{33}) \qquad \text{(A-53)}$$

---

[13] $D_1 = (1 + \lambda_1 + \lambda_1/\eta_1)$, $D_2 = (1 + \lambda_2 + \lambda\ /\eta_2)$, $D_{11} = \lambda_2 - \lambda_1$, $D_{33} = \lambda_1 q_1 + \lambda_2 q_2$, and $D_{22} = \lambda_1 - \lambda_2$.

$$j_{12} = \frac{\partial f_1}{\partial q_2} = \frac{q_1 - q_2}{q_T^2} \quad D_{11} \left(\frac{C + Fx}{q_T} - C' \right) - \frac{C''}{q_T} (D_{33} + q_T) \tag{A-54}$$

$$j_{13} = \frac{\partial f_1}{\partial \lambda_1} = P_1 + P_1' q_1 - \frac{q_2}{q_T^2} (C + Fx) - \frac{C' q_1}{q_T} \tag{A-55}$$

$$j_{14} = \frac{\partial f_1}{\partial \lambda_2} = \frac{q_2}{q_T^2} (C + Fx) - \frac{C' q_2}{q_T} = \frac{q_2}{q_T} \left(\frac{C + Fx}{q_T} - C' \right) \tag{A-56}$$

$$j_{21} = \frac{\partial f_2}{\partial q_1} = \frac{q_2 - q_1}{q_T^2} \left[\frac{C + Fx}{q_T} - C' \right] D_{22} - \frac{C''}{q_T} (q_T + D_{33}) \tag{A-57}$$

$$j_{22} = \frac{\partial f_2}{\partial q_2} = P_2' D_2 - -P_2 \lambda_2 K_2 - \frac{2q_1}{q_T^2} D_{22} \left[\frac{C + Fx}{q_T} \right.$$

$$\left. - C' \right] - \frac{C''}{q_T} (q_T + D_{33}) \tag{A-58}$$

$$j_{23} = \frac{\partial f_2}{\partial \lambda_1} = \frac{q_1}{q_T^2} [C + Fx] - \frac{C' q_1}{q_T} \tag{A-59}$$

$$j_{24} = \frac{\partial f_2}{\partial \lambda_2} = - \frac{q_1}{q_T^2} (C + Fx) - \frac{C' q_2}{q_T} + P_2 (1 + 1/\eta_2) \tag{A-60}$$

$$j_{31} = \frac{\partial f_3}{\partial q_1} = P_1 + P_1' q_1 - \frac{q_2}{q_T^2} (Fx + C) - \frac{q_1}{q_T} C' \tag{A-61}$$

$$j_{32} = \frac{\partial f_3}{\partial q_2} = \frac{q_1}{q_T^2} (Fx + C) - \frac{q_1}{q_T} C' \tag{A-62}$$

$$j_{33} = \frac{\partial f_3}{\partial \lambda_1} = 0 \qquad \text{(A-63)}$$

$$j_{34} = \frac{\partial f_3}{\partial \lambda_2} = 0 \qquad \text{(A-64)}$$

$$j_{41} = \frac{\partial f_4}{\partial q_1} = \frac{q_2}{q_T} \left[\frac{C + F\chi}{q_T} - C' \right] \qquad \text{(A-65)}$$

$$j_{42} = \frac{\partial f_4}{\partial q_2} = P_2 + q_2 P_2' - \frac{q_1}{q_T^2}(C + Fx) - \frac{q_2}{q_T} C' \qquad \text{(A-66)}$$

$$j_{43} = \frac{\partial f_4}{\partial \lambda_1} = 0 \qquad \text{(A-67)}$$

$$j_{44} = \frac{\partial f_4}{\partial \lambda_2} = 0 \; , \qquad \text{(A-68)}$$

where C' and C'' are the first and second derivatives of the cost of production with respect to quantity, and P'_1 and P'_2 are the derivatives of the demand function (A-8) with respect to quantity.

Computational Procedure

The procedure is straightforward. Starting values for q_1, q_2, λ_1, and λ_2 are assumed. The values of the f_1 to f_4 [equations (A-48) to A-51)] for these q and λ are computed. The Jacobian matrix J and its inverse are computed. Then, the procedure of (A-52) gives the corrections to the starting values. The procedure is repeated until subsequent changes to q and λ change the revenue less than a small tolerance value ($10.00). To ascertain that there was one global solution to the problem (absence of local minima), the program was run assuming different starting values for the variables. For all the hours, the same solution was obtained irrespective of the starting values, thus confirming global minima.

In terms of the convergence properties, the nature of the problem is not different from the previous Ramsey pricing formulation. In this instance, there is a revenue constraint for each class. Consider class 2 for illustration. The revenue from this class, R_2 is given by

$$R_2 = q_2 P_2 = q_2 \alpha_2 K_2 e^{-K_2 q_2} \quad . \tag{A-69}$$

The variation of revenue with quantity is given by

$$\frac{\partial R_2}{\partial q_2} = \alpha_2 K_2 e^{-k_2 q_2} - \alpha_2 q_2 K_2{}^2 e^{-K_2 q_2} , \tag{A-70}$$

which can be simplified to

$$\frac{\partial R_2}{\partial q_2} = \alpha_2 K_2 e^{-K_2 q_2}(1 - K_2 q_2) \quad . \tag{A-71}$$

The revenue is a maximum when (A-71) is set equal to zero which gives $q_2 = 1/K_2$ for maximum revenue (or demand elasticity is equal to one).

Figure A-5 illustrates the relationship of (A-71). The cost curves are also indicated therein as in the case of Ramsey pricing. Note that (A-47a) for cost can be written as

$$P_2 q_2 = \frac{q_2}{q_T} \; C(q) + F - \frac{q_1}{q_T} \; Fx \; , \tag{A-72}$$

which is an increasing curve as shown in the figure. Therefore, as x decreases (that is, when class 1 subscribes increasingly for lower reliability levels) there may be no intersection and hence the solution to the problem does not exist.

For acceptable values of x, an examination of figure A-4 suggests that there are two solutions. However, this was found not to be the case during computation. Different starting values yielded the same unique solution at every hour. The reason for the unique solution is the following.

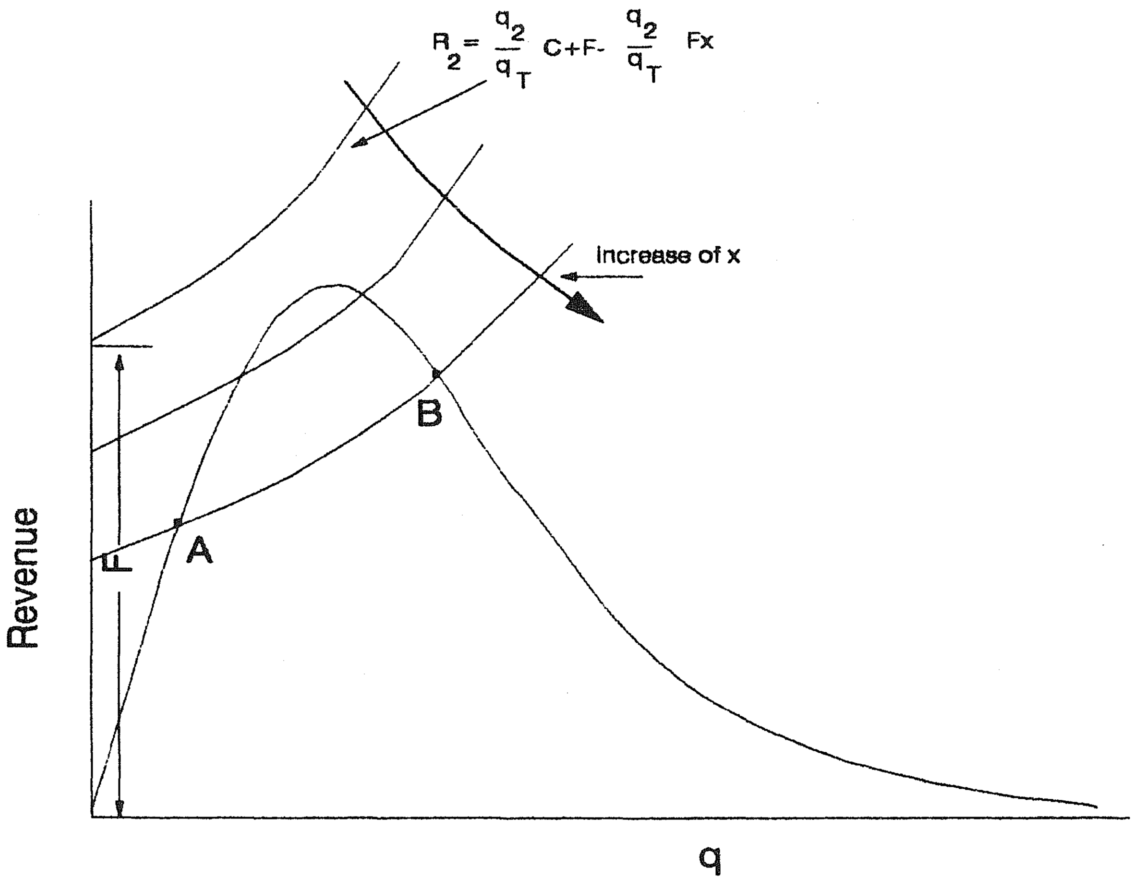


Fig. A-5. Variation of revenue and cost with x.

Suppose that in the gradient search procedure, one has obtained point A (figure A-5) as the acceptable solution for class 2. Then, from a similar reasoning for class 1, at these values of λ and q of the gradient search procedure, it is required that an acceptable solution (or intersection) be simultaneously obtained. If this is not the case, the search procedure will move the search until the revenue constraint of both classes is satisfied. Note that the revenue of each class is coupled to the consumption of another class as in (A-47a) or (A-46b). Extensive computer trials always gave unique solutions. A formal mathematical proof to prove the uniqueness of the solution is left for further investigation.

Results

As in the case of Ramsey-type pricing, it is necessary to establish the component of F for each hour such that the sum of the components is equal to F. The comments made earlier regarding the lack of a priori guidance to select the hourly values are valid here as well. However, for purposes of comparison of similar procedures, the same hourly allocation of F as in table A-7 was used in the following simulations.

Table A-8 portrays the results obtained. There are two sets of results. The first with different values of x represent the reliability subscriptions of class 1. The second with values of y relate to that of class 2. The table shows extreme values for x and y for illustrative purposes. From a practical standpoint, realistic values may be in the range of 0.9 to 1.0.

Before attempting comparison of results to draw conclusions one has to be aware of certain limitations. These are the following.

The base case has been used to establish hourly price-demand relationships. It was assumed that the consumers' reaction is to the average price of electricity implying that the price signal was as if fixed charges were recovered equally in all hours. For the Ramsey-type pricing, the hourly demand-price relations are assumed to be the same as in the base case. It is further assumed that the consumer reacts to price signals at each hour. That the recovery of fixed charges in the Ramsey-type pricing is set equal to the same percentage as in spot pricing has been made clear earlier. The reliability pricing schemes made the same assumption in regard to the hourly

TABLE A-8

RESULTS OF RELIABILITY PRICING STUDIES COMPARED TO OTHER PRICING SCHEMES

	Base Case	Spot Pricing	Ramsey Type	Reliability Pricing						
				x = 0.85	x = 0.9	x = 1.0	x = 1.2	y = .6	y = .8	y - 1.2
Welfare $	3,431,009	3,446,291	3,439,848	3,438,485	3,439,489	3,439,266	3,434,526	3,433,016	3,437,250	3,439,564
Peak load MW	2,850	2,396	2,584	2,482	2,539	2,600	2,662	2,769	2,685	2,515
Energy MWh	51,809	48,890	52,518	51,929	52,275	52,707	53,123	56,152	54,455	50,894
Prdctn. Cost $	525,230	456,940	523,069	510,887	517,870	526,602	534,903	596,770	561,310	492,439
Fixed Cost $	282,359	282,359	282,359	282,359	282,359	282,359	282,359	282,359	282,359	282,359
Revenue $	847,288	817,085	805,901	793,198	800,211	808,896	817,344	797,030	803,068	813,120
Class 1 Surplus $	2,732,895	2,734,468	2,786,898	2,811,234	2,800,482	2,780,564	2,744,112	2,732,808	2,758,475	2,799,664
Class 2 Surplus $	374,625	351,681	370,003	349,871	356,675	376,405	407,973	417,827	396,398	357,533
Producer Surplus $	322,057	360,140	282,831	282,311	282,341	282,287	282,441	282,381	282,377	382,367
Load Factor	0.757	0.851	0.847	0.871	0.858	0.844	0.831	0.844	0.845	0.850

recovery of F. Presumably, other choices for the hourly recovery of F could be made. In view of the above, it is important to recognize that the values indicated in table A-8 would be different for a different set of assumptions. Therefore, any conclusion to be drawn should not stem from the actual values in table A-8. Rather, they should emanate from an examination of the trend of changes. Thus, in the following discussion, when reference is made to the values in table A-8, the purpose is to examine the changes in the values of certain outcomes rather than the actual values of the outcomes themselves.

Certain trends are evident and are to be expected. They are the following.

The producer surplus is approximately the same in all cases ($282,359), for that is the recoverable amount by the regulated monopoly. The exception is the case of spot pricing permitting overrecovery. The welfare is a maximum when P = MC of production. Ramsey-type pricing gives the second best solution. The reliability pricing schemes with x or y = 1 (no reliability differentiation) produces almost the same benefits as the Ramsey-type pricing.

Additional observations from an examination of table A-8 are as follows. The subscription of a lower reliability by class 1 (residential plus commercial in our example) increases welfare and reduces peak demand compared to $x = 1$ for some cases (see $x = 0.9$). Note that the welfare and peak reduction are inferior to the Ramsey-type pricing scheme.

It was observed that the total welfare increased in some hours and decreased in some other hours. In all hours, however, the surplus of the class subscribing for lower reliability increased with the opposite being true for the other class. The net surplus in each hour (class 1 and class 2 together) was either a loss or gain compared to the case with $x = 1$. The value of net surplus depends on the elasticities of the two classes. Table 8 shows the total net surplus over the twenty-four hours of simulation. It so happens that the total welfare slightly increases (compared to $x = 1$) at $x = 0.9$. With this exception, in all other cases the total welfare decreases.

The table also shows the peak demand. The reduction in peak demand is a result of reduced consumption by class 2 due to a higher price arising from the subscription of class 1 to lower reliability. Note that the load factor also improves with decreasing x.

The subscription of class 2 has similar effects. Lower the reliability desired by class 2 ($y < 1$), the lesser is the welfare. Also, the peak demand increases with lower subscription of reliability by class 2.

In practice, it is not possible to predict the level of reliability subscription by customers. The choice may not necessarily be related to reliability worth in view of the fact that the customer may not make proper choices. Even if the subscriptions are proper there is no guarantee that customers' option for reliability will be optimally related to elasticities. That not being the case, the uncertainty associated with the benefits does not warrant reliability pricing.

A subsidiary outcome of the pricing schemes is that there would be less energy available for noncore customers. In chapter 4, a method of pricing service to noncore or interruptible customers is outlined. It is clear from the above that an improved load factor with a corresponding reserve reduction resulting from certain subscriptions would leave less energy and capacity available for sale to noncore customers.

Improved load factors could also result in lower overall reliability due to the same reasons shown by the example in chapter 4. The operation of the system should be such that the reliability of the class not subscribing to any reliability level would not be less than under the base case scenario. Such operation may or may not be possible due to the complex nature and interaction between transmission and generation outages.

In addition to the above difficulties, the method of relating the reliability of supply to rates precipitates an anomaly. Strictly, the reduction in the rate ($x < 1$ for class 1) should reflect the operational benefit and the longer-term gains. While the operational gains may be achieved by tripping off the customer desiring lower reliability earlier than the other classes, the latter concern of longer-term gains poses a contradiction.

Under the above pricing scheme, the consumption of the class desiring lower reliability is higher at all times including the time of peak. The opposite is true of the class not opting for a lower reliability of service. Therefore, the benefits due to peak reduction arise from the consumption pattern of the class not opting for lower reliability. To examine the second order effects, one must run the above simulation again, reducing F to a lower value to the class not desiring reduced reliability to account for the longer term reduction in reserve. Clearly, such a rerun of the simulation will reveal that the consumption of the class not desiring lower reliability also increases. Therefore, it is to be expected that the reduction in the peak is somewhat less than what is shown in table A-8. In addition, the passing of the resource expansion benefits arising from reserve reduction and improved load factor to the class not desiring a lower reliability counteracts the reliability pricing.

As said earlier, the class desiring lower reliability would increases consumption. Then, the fact that the class contributing to a higher consumption at peak would be charged a lower price is anomalous and contradicts ratemaking principles. Furthermore, it would be impossible to implement the above rates practically because of the difficulty in determining exactly how much peak consumption was increased or reduced by the respective classes in the future. It would be impossible to substantiate which consumption was influenced by which price signal from a posterior examination of consumption. A theoretical conjecture arising from a simulation can hardly take the place of a practical determination of such relations. Therefore, the ratemaking principles under such circumstances can be contested.

The above shows the uncertainties associated with the welfare gains from reliability pricing schemes. Assumptions have been made regarding the price-demand relationships (and therefore, the price elasticity of demand). Any change in these relationships could change the results substantially. In view of the sensitivity of the results to these relationships, it would be necessary to determine the relationships from well designed tests. Furthermore, the consumers' behavior is dynamic in the sense that the relationships could change due to price signals or due to prevailing market conditions. Therefore, unless one is certain about the consumer behavior, the benefits due to reliability pricing appear uncertain.

Effect of Interrelated Demand Functions

In the above analysis, it was assumed that all the hourly demands are independent. It is shown in chapter 2 that certain types of consumption might demand constant or near constant energy over a period of time and, therefore, the demand not materializing at the time of peak would show up during nonpeak hours. This "demand payback" has the effect of introducing a coupling between the hourly inverse demand functions.

One method of modelling such a phenomenon is as follows. Let the on-peak hours be defined as the hours h, h = 16, 17, 18, and 19. The energy payback or make-up hours are identified as h = 12, 13, 14, 15, 20, 21, and 22.[14] Then, for the i^{th} class of consumer, let

$$\sum_h q_i^6 = \tilde{Q}_i \ , \quad h = 16,\ 17,\ 18,\ \text{and } 19, \qquad \text{(A-73)}$$

and

$$\sum_h q_i^b = Q_i^b \ , \quad h = 16,\ 17,\ 18,\ \text{and } 19, \qquad \text{(A-74)}$$

where q_i^b represents the consumption of the i^{th} class at hour h under base case price assumptions.
Define

$$\Delta Q_i \triangleq Q_i^b - \tilde{Q}_i \ . \qquad \text{(A-75)}$$

Let the payback in energy demand in the off-peak hours be a fraction β of ΔQ_i. Evidently $\beta \leq 1.0$. If the payback is assumed to be equal in all designated off-peak hours, the additional energy ϵ demanded during the seven hours 12 to 15 and 20 to 22 is given by

$$\epsilon = (\beta \Delta Q_i)/7.0 \ . \qquad \text{(A-76)}$$

Therefore, the assumed demand functions of (A-8) are modified for the hours 12 to 15 and 20 to 22 to be

[14] These hours are arbitrarily chosen for illustrative purposes. The actual hours of energy payback depend on the characteristics of the load.

$$U_i'(q, t) = \alpha_i(t)K_i(t)[e^{-(K_i q_i(t) + \epsilon)}] . \quad \text{(A-77)}$$

Equation of (A-77) is nothing but the previous inverse demand relationship right shifted by ϵ.

The application of the mathematical development for any of the pricing schemes mentioned earlier will give results identical to those previously outlined. The only difference is that the demand during the designated off-peak hours will be enhanced by ϵ.

It is apparent that the above assumptions may be rather simplistic. One could simulate energy invariance by assuming a functional approximation for the load, as in chapter 4. In any event, it is important to recognize that the relation between hourly demands must be accounted for in a proper simulation. Further work on better methods of such simulation appears warranted.

Graphical Depiction

Figures A-6 to A-10 show hourly demands. These are the results shown previously in tables portrayed in a graphical fashion.

Figures A-6 and A-7 show the total system load under different pricing schemes. It can be seen from A-6 and A-7 that welfare maximization results in the "flattest" load and hence obtains the highest reduction in peak load. Recall that in our simulation under welfare maximization, the producer has revenue in excess of costs.

Observe in figures A-6 and A-7 that the peak load reductions (compared to base case) and changes to load shape are smaller in other pricing situations.

Figures A-8, A-9, and A-10 show the effect of different pricing schemes on the relative consumption by the two classes. Note that the consumptions of the two classes decrease under welfare maximization compared to the base case. However, under reliability pricing, consumption of class 1 increases while that of class 2 decreases (for most hours) compared to the base case. Figure A-10 shows that the same is true of consumptions in comparison to the welfare maximization case.

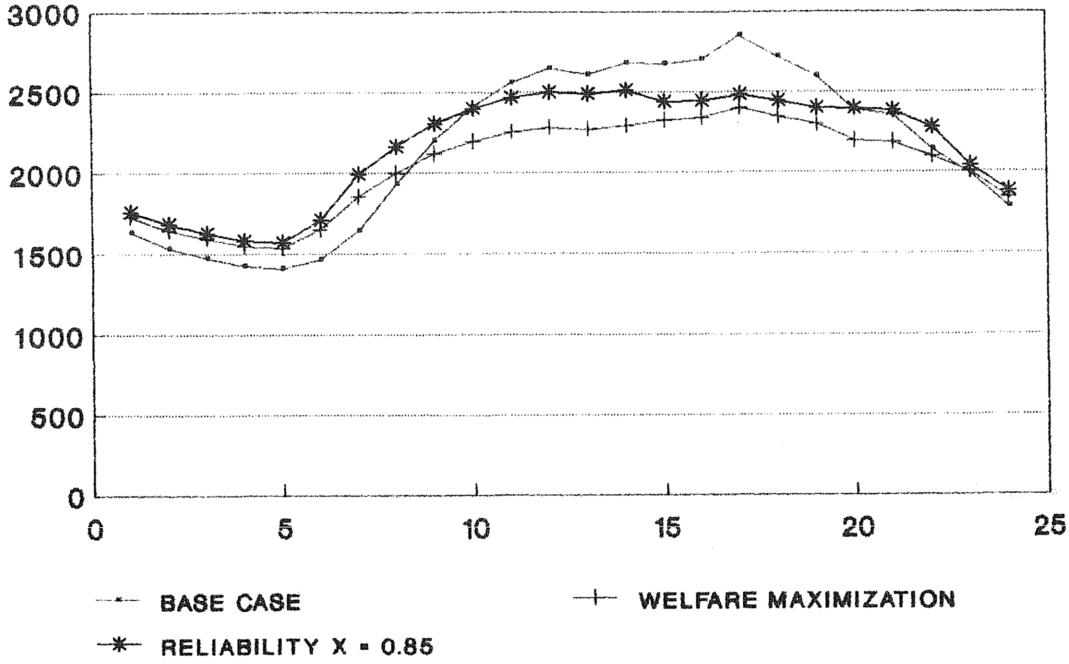


Fig. A-6. Total demand under welfare and Ramsey pricing scenarios.

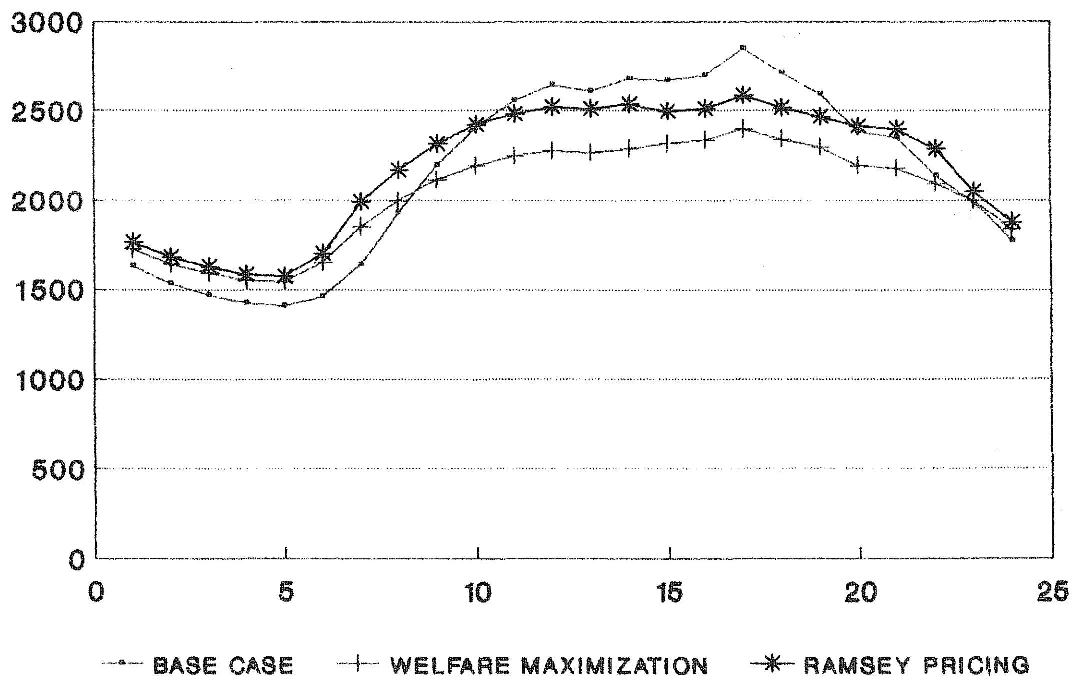


Fig. A-7. Total demand under welfare and reliability pricing scenarios.

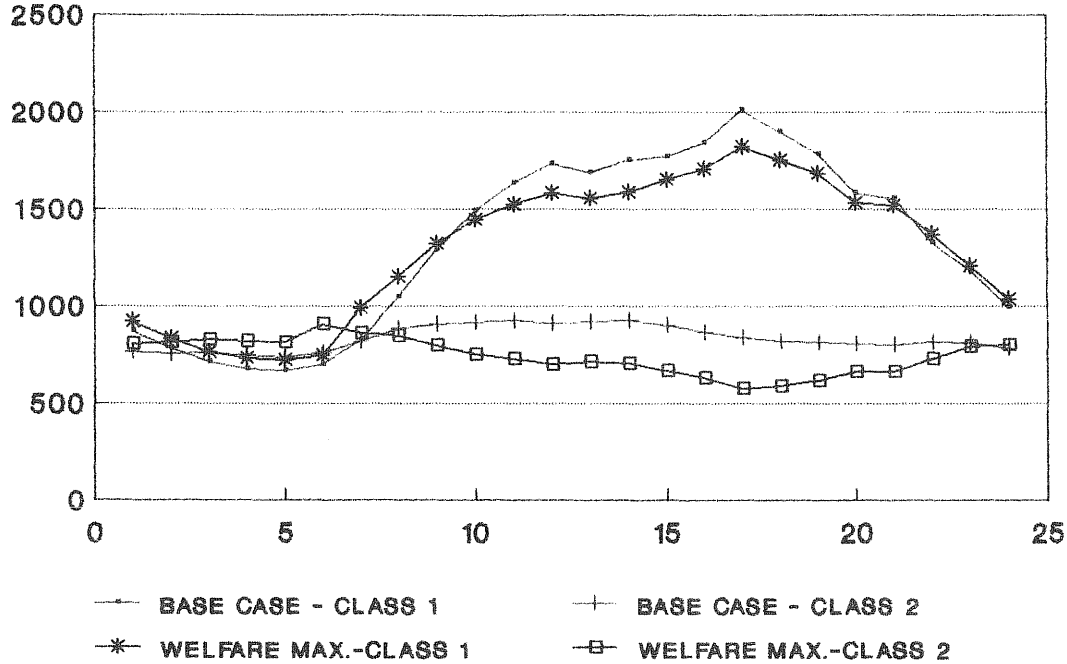


Fig. A-8. Comparison of class consumptions under welfare maximization and base case pricing scenarios.

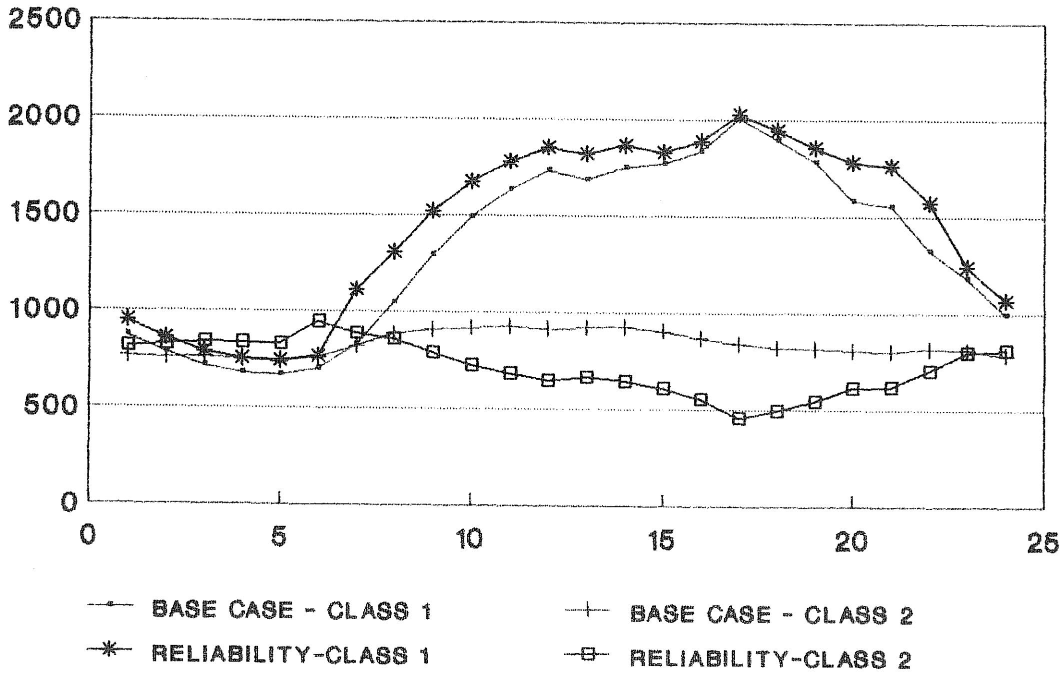


Fig. A-9. Comparison of class consumption under reliability pricing scenario and base case.

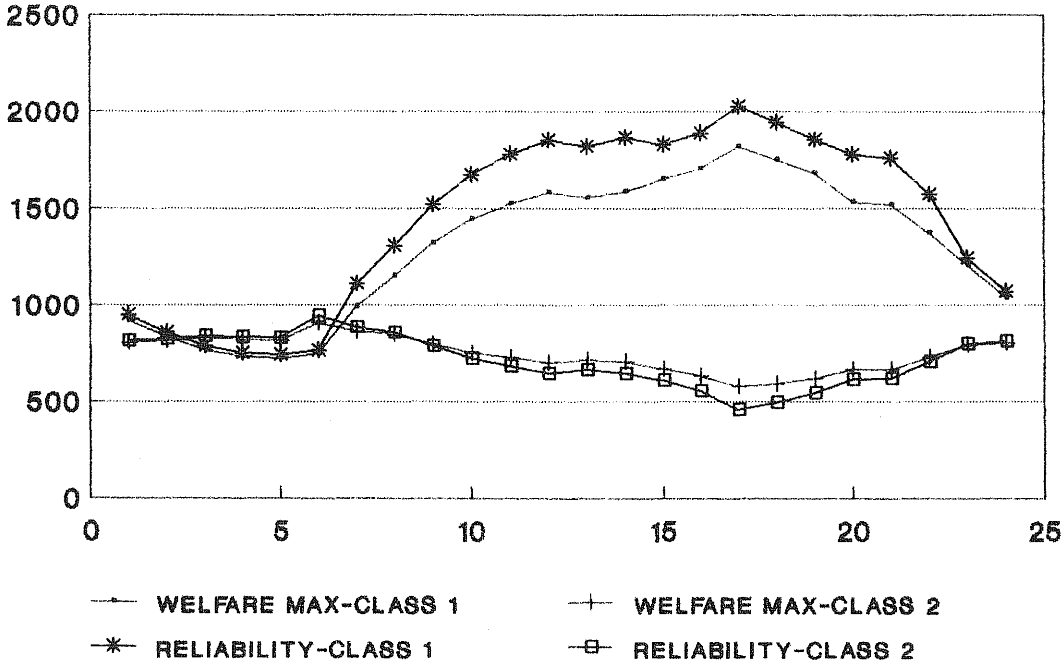


Fig. A-10. Comparison of class consumptions under reliability and welfare maximization pricing scenarios.

The above graphical portrayals are taken from the appropriate tables discussed earlier.

REFERENCES

1. Telson, M. L. "The Economics of Alternative Levels of Reliability for Electric Power Generation Systems." *Bell Journal of Economics* 6, no. 2 (Autumn 1975): 679-94.

2. Munasinghe, M. "Principles of Modern Electricity Pricing." *Proceedings of IEEE* 69 no. 3 (March 1981): 332-48.

3. *Ontario Hydro Survey on Power System Reliability*. "Viewpoint of Customers in Retail Trade and Service." Ontario Hydro Report No. R&U (1977):79-7. Obtainable from Ontario Hydro.

4. Woo, Chi-Keung, and B. M. Gray. "An Econometric Analysis of Industrial Firms' Power Outage Costs in Northern California." Pacific Gas and Electric working paper, 1987.

5. "Customer Demand For Service Reliability: A Synthesis of Outage Cost Literature." EPRI Report P-6510, Project 2801-1, September 1989.

6. David, A. K., D. J. Nutt, C. S. Chang, and Y. C. Lee. "The Variation of Electricity Prices in Response to Supply-Demand Conditions and Devices for Consumer Interaction." *Electric Power and Energy Systems* 8, no. 2 (April 1986): 101-14.

7. David A. K., and Y. C. Lee. "Dynamic Tariffs: Theory of Utility-Consumer Interaction." *IEEE Transactions on Power Systems* 4 (August 1989): 904-11.

8. Billinton, R., G. Walker, and R. K. Subramuniam. "Factors Affecting the Development of a Residential Customer Damage Function." *IEEE Transactions on Power Systems* PWRS2, no.1 (February 1987): 204-208.

9. _______. "Factors Affecting the Development of a Commercial Customer Damage Function." *IEEE Transactions on Power Systems* PWRS1, no. 5 (November 1986): 28-33.

10. Schweppe, F. C., M. C. Caramanis, and R. D. Tabors. "Evaluation of Spot Price-Based Electricity Rates." *IEEE Transactions on Power Apparatus and Systems* PAS-104 no. 7 (1985): 1644-55.

11. Tokue, Ishikawa. "Customer Behavior Model Under Deregulated Power System." *Electrical Engineering in Japan* 105, no. 3 (1985): 96-103.

12. Ramsey, F. P. "A Contribution to the Theory of Taxation." *Economic Journal* 37 (March 1927): 47-61

13. Baumol, W. J., and A. K. Klevorick. "Input Choices and Rate of Return Regulation: An Overview of the Discussion." *Bell Journal of Economics* 1 (Autumn 1970): 162-90.

14. Marchand, M. S. "Pricing Power Supplied on an Interruptible Basis." *European Economic Review* 5 (1974): 263-74.

15. Smith, Stephen A. "Efficient Menu Structure for Pricing Interruptible Electric Power Service." Santa Clara University: Working Paper Series, (March 1989): 89-007.

16. Boiteux, M. "La tariffication des demand on point: application de al theorie de la vente au cout marginal." *Revue Generale de l'electricite* 58 (August 1949): 321-40.

17. Ischirhard, John, and Frank Jen. "Behavior of a Monopoly Offering Interruptible Service." *The Bell Journal of Economics* 10, no. 1 (Spring 1979): 244-57.

18. Mussa, M., and S. Rosen. "Monopoly and Product Quality." *Journal of Economic Theory* 18 (1978): 301-37.

19. "IEEE Reliability Test System Report of Task Force on APM Subcommittee," *IEEE Transactions on Power Apparatus & Systems* PAS-98, no. 6 (November/December 1979): 2047-54.

20. "Bulk System Reliability--Measurement & Indices." Measurement Indices Working Group Report, *IEEE Transaction on Power Systems* 4, no. 3 (August 1980).

21. Endrenyi, J., *Reliability Modeling in Electric Power Systems.* (Chichester, England: John Wiley & Sons, 1980).

22. IEEE Committee report, R. Billinton, Chairman. "Bibliography on the Application of Probability Methods in Power System Reliability Evaluation." *IEEE Transactions on Power Apparatus and Systems* 91 (April 1972): 649, 660.

23. Baleriaux, H., E. Jamoulle, and Fr. Linard de Quertechin. "Simulation de l'exploitation d'un parc de machines thermiques de production d'électricité couplés à des stations de pompage." *Revue E* (édition SRBE) 5, no. 7 (1967): 3-24.

24. Rau, N. S., and K. F. Schenk. "Application of Fourier Methods for the Evaluation of Capacity Outage Probabilities." A79 1023-3, IEEE Power Engineering Society Winter Meeting, New York, 1979.

25. Crew, M. A. and P. R. Kleindorfer, "The Economics of Public Utility Regulation," first edition. (The MIT Press, Cambridge, Massachusetts, 1986).

26. Noell, S. K., J. B. Wharton, and L. E. Sweeney, presentations in a session entitled "Reliability Differentiation in New Electric Services." Conference on New Dimensions in Pricing Electricity, sponsored by EPRI and Niagara Mohawk Power Corporation, Syracuse, New York, September 1986.

27. "Priority Service Methods" (or PRISM), obtainable from EPRI, Palo Alto, California, 1989.