ORNL/CON-472

ENERGY DIVISION

MAINTAINING GENERATION ADEQUACY IN A RESTRUCTURING U.S. ELECTRICITY INDUSTRY

ERIC HIRST and STAN HADLEY

October 1999

Sponsored by
U.S. Environmental Protection Agency
401 M Street, SW
Washington, DC 20460

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
LOCKHEED MARTIN ENERGY RESEARCH CORPORATION
for the
U.S. DEPARTMENT OF ENERGY
under contract No. DE-AC05-96OR22464

CONTENTS

Page

SUMMARY

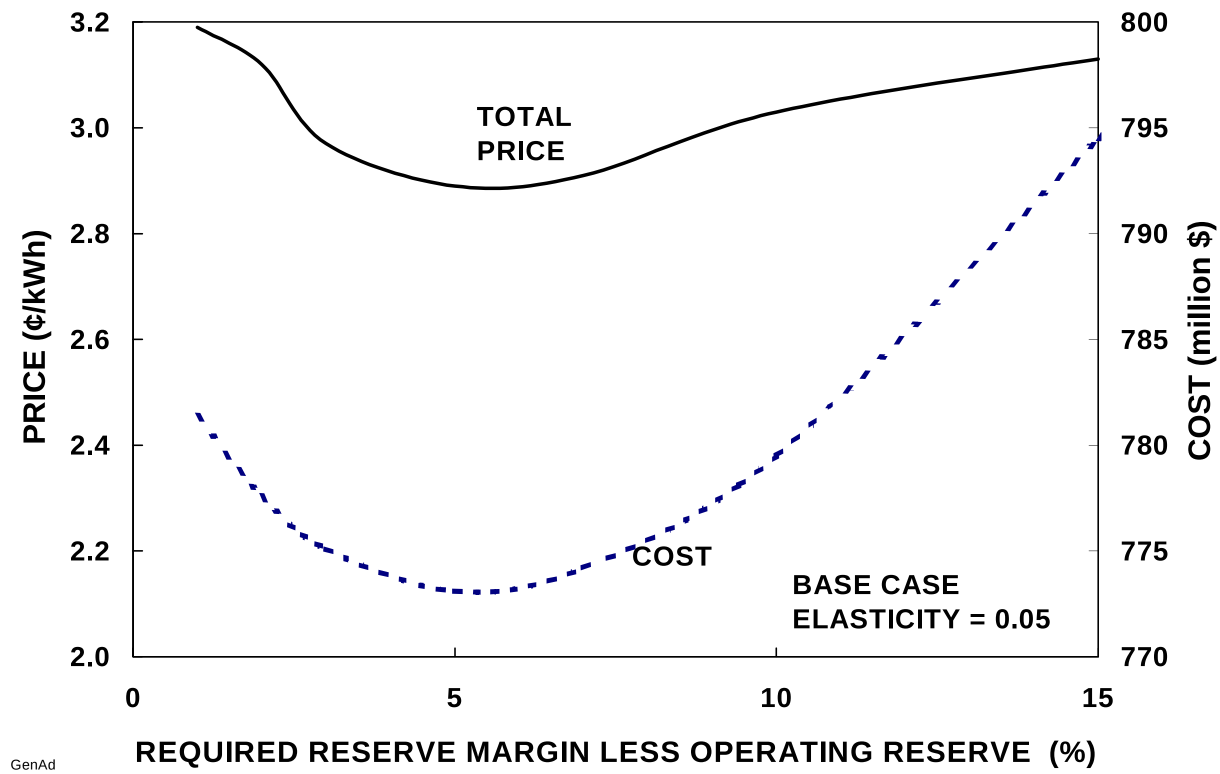
Historically, decisions on the amounts, locations, types, and timing of investments in new generation have been made by vertically integrated utilities with approval from state public utility commissions. As the U.S. electricity industry is restructured, these decisions are being fragmented and dispersed among a variety of organizations.

As generation is deregulated and becomes increasingly competitive, decisions on whether to build new generators and to retire, maintain, or repower existing units will increasingly be made by unregulated for-profit corporations. These decisions will be based largely on investor assessments of future profitability and only secondarily on regional reliability requirements. In addition, some customers will choose to face real-time (spot) prices and will respond to the occasionally very high prices by reducing electricity use at those times. Market-determined generation levels will, relative to centrally mandated reserve margins, lead to: (1) more volatile energy prices; (2) lower electricity costs and prices; and (3) a generation mix with more baseload, and less peaking, capacity.

During the transition from a vertically integrated, regulated industry to a deintegrated, competitive industry, government regulators and system operators may continue to impose minimum-installed-capacity requirements on load-serving entities. As the industry gains experience with customer responses to real-time pricing and with operation of competitive intrahour energy markets, these requirements will likely disappear.

We quantitatively analyzed these issues with the Oak Ridge Competitive Electricity Dispatch model (ORCED). Model results show that the "optimal" reserve margin depends on various factors, including fuel prices, initial mix of generation capacity, and customer response to electricity prices (load shapes and system load factor). Because the correct reserve margin depends on these generally unpredictable factors, mandated reserve margins might be too high, leading to higher electricity costs and prices (top of Fig. S-1). Absent mandated reserve margins, electricity prices and costs decline with increasing customer response to prices during high-demand periods (bottom of Fig. S-1).

The issues discussed here are primarily transitional rather than enduring. However, the transition from a highly regulated, vertically integrated industry to one dominated by competition is likely to take another five to ten years.



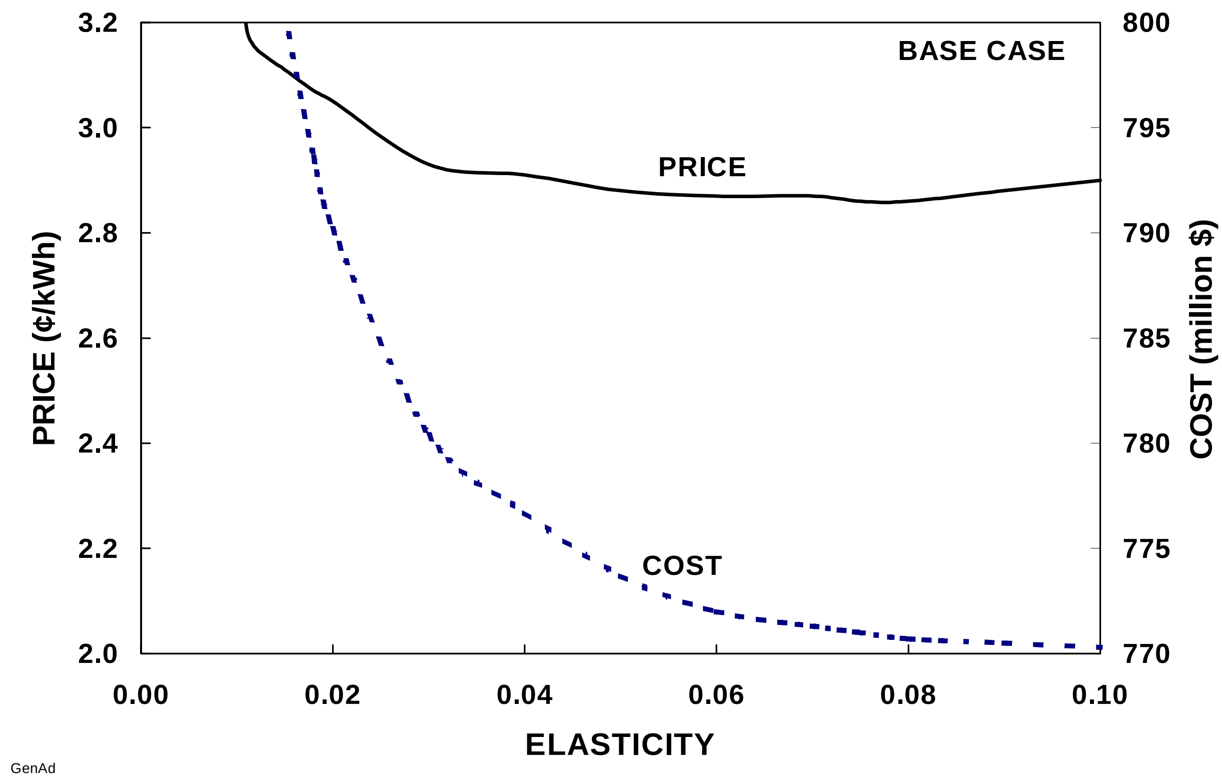


Fig. S-1. **ORCED results showing prices and costs as functions of required reserve margin (top) and unserved-energy price elasticity (bottom).**

LIST OF ACRONYMS

EIA	Energy Information Administration
EPSA	Electric Power Supply Association
ERCOT	Electric Reliability Council of Texas
FERC	U.S. Federal Energy Regulatory Commission
ISO	Independent system operator
LDC	Load-duration curve
LOLP	Loss-of-load probability
LSE	Load-serving entity
M	Million
NEPOOL	New England Power Pool
NERC	North American Electric Reliability Council
ORCED	Oak Ridge Competitive Electricity Dispatch model
O&M	Operations and maintenance
PJM	Pennsylvania-New Jersey-Maryland Interconnection
PX	Power exchange
VOLL	Value of lost load
WSCC	Western Systems Coordinating Council

INTRODUCTION

The dramatic price spikes in the Midwest in June 1998 [Federal Energy Regulatory Commission (FERC) 1998] and the occasional power shortages and price spikes during the summer of 1999 (*Electric Power Daily* 1999a) demonstrate the importance of electricity to our modern society. Do these sporadic high prices signal serious deficiencies in the reliability of bulk-power systems? Or do they illustrate the normal workings of a competitive market for electricity? At the heart of these questions is the concept of generation adequacy and how it can best be maintained in a future electricity industry very different from today's.

Restructuring the electricity industry in general, and the bulk-power sectors in particular, calls into question the entire concept of adequacy. What is adequacy? Is it purely a reliability concept, or does it also have commercial significance? Is adequacy even a relevant term for a restructured electricity industry?

Because of the dramatic changes under way in the ownership, operation, and structure of bulk-power systems and markets, we explore possible changes in generation adequacy. [This project expands on the generation-adequacy part of an earlier project conducted for the Edison Electric Institute by Hirst, Kirby, and Hadley (1999).] This project examines generation adequacy both qualitatively and quantitatively. These questions are important and difficult because the United States is unbundling the traditional, vertically integrated utilities that, historically, managed adequacy within a single organizational entity.

Resolving these issues is difficult for several reasons: (1) generation is likely to become increasingly competitive and deregulated while transmission remains regulated; (2) transmission operations are likely to be combined into large, independent, regional organizations, the scope and structure of which are far from clear; (3) because of these differences, decisions on generation adequacy might be left to competitive markets while decisions on transmission adequacy continue to be made, at least in part, by regulators and central planners; (4) adequacy and security are both complements and substitutes; and (5) generation and transmission are both complements and substitutes.

The rest of this chapter provides background on the concept and definition of adequacy, and presents historical data and projections on generation investments and capacity. Chapter 2 presents the findings from our literature review and our discussions with several industry experts. Chapter 3 explains the workings of the Oak Ridge Competitive Electricity Dispatch model (ORCED), and Chapter 4 uses the model to assess the effects on consumer and producer costs of letting markets decide on the appropriate level of generation capacity vs having central planners specify a minimum planning reserve margin. Chapter 5 presents our conclusions.

BACKGROUND

The North American Electric Reliability Council (NERC) defines reliability as "the degree to which the performance of the elements of [the electrical] system results in power being delivered to consumers within accepted standards and in the amount desired." NERC's definition of reliability encompasses two concepts, adequacy and security. Adequacy, the subject of this report, is defined as "the ability of the system to supply the aggregate electric power and energy requirements of the consumers at all times." Security is "the ability of the system to withstand sudden disturbances."

In plain language, adequacy deals with planning and investment, and security deals with short-term operations. Adequacy implies that there are sufficient generation and transmission resources available to meet projected needs plus reserves for contingencies. Security implies that the system will remain intact even after outages or other equipment failures occur. Although adequacy is a reliability concept, it has strong commercial implications; the same is true of security. Although we might like to pretend otherwise, bulk-power reliability and commerce are strongly interdependent.

Obviously, adequacy and security are complements. Without system security, the output of the generation resources, no matter how abundant, cannot be delivered to customers. Correspondingly, a high degree of security is of little value if there are insufficient resources to meet customer needs.

Adequacy and security can also be substitutes; more of one can make up for less of the other. For example, an abundance of resources makes it easier to maintain a high degree of security (i.e., reduces the need for emergency actions). That is, system operators can manage the system in real time with less data and fewer analytical tools if there are ample generation resources and redundant transmission facilities. Similarly, high-quality system operation can extract more output from a system that might otherwise be considered underbuilt. For example, the near-real-time collection and analysis of data on the current and projected states of the transmission system can allow system operators to run the system closer to its limits than would be feasible with less data collection and analysis.

Utilities divide their generation reserves into two categories, related to the differences between short-term security and long-term adequacy. For day-ahead planning and real-time operation, utilities are required by NERC and regional-reliability-council rules to maintain minimum levels of operating reserves, typically 4 to 8% of the projected daily peak.[*] These short-term reserves protect bulk-power systems from the effects of major generation and transmission outages and correct for errors in day-ahead load forecasts.

---

[*]The minimum operating-reserve requirement is typically based on the size of the largest generating unit online within the regional reliability council.

Planning reserves (of which operating reserves are a subset) are the focus of this report. These reserves provide long-term insurance against problems that might otherwise arise when units are not available (e.g., for planned maintenance) and allow for unanticipated long-term load growth. Generator outage rates are on the order of 5 to 30%; that is, units are available 70 to 95% of the time. Planning reserves provide sufficient capacity to offset these planned and sudden losses.

DATA AND PROJECTIONS

The U.S. electricity industry is currently in an awkward position—half regulated and half competitive. Many utilities are understandably reluctant to make investments until the rules and the separation between competitive and regulated activities are clear.

Figure 1 shows utility forecasts, as reported to the regional reliability councils and NERC (1998a), of generation-capacity margins from 1990 through 1998 and projections through 2007. Nationwide, reserve margins declined from 22% in 1990 to 16% in 1997 at almost 0.7 percentage points a year (top of Fig. 1). Annual capacity additions declined from 9700 MW between 1991 and 1995 to only 5200 MW between 1996 and 1998 (Terry 1999).[*] Reserve margins are expected to decline further to 10% in 2007. The bottom of Fig. 1 shows the regional trends, with a precipitous drop in reserve margin in the Electric Reliability Council of Texas (ERCOT).[#] NERC notes that the projections for the out years (2003 to 2007) are highly uncertain. This uncertainty occurs because the owners of merchant plants often do not reveal their plans early and because new generating units can often be constructed in only a few years (reducing the need for long-term projections of generating capacity).

Various organizations collect and report data on capacity additions, including the Energy Information Administration (EIA), the Electric Power Supply Association (EPSA) (Slater 1999), and NERC (1998b). EIA (1998) uses its National Energy Modeling System to forecast capacity additions required for reliability, which provides a useful reference point against which to compare capacity-addition figures. Unfortunately, the various estimates of planned capacity additions may not be consistent with each other, primarily because of differing definitions. For example, the EPSA data are for merchant plants,[§] defined as all generation not part of utility ratebase (a definition that has little to do with whether the plant output is sold under long-term

---

[*] Utilities added less than 500 MW of new capacity in 1998 and retired almost 2900 MW that year, a net loss of 2400 MW (EIA 1999). On the other hand, nonutility companies added 3000 MW in 1998.

[#] The top part of Fig. 1 shows results for the United States, while the regional numbers in the bottom of the figure are for all of North America (including Canada and a small part of Mexico) within the ten NERC regions.

[§] Some of these plants may not be built because of problems with siting, state approval, financing, or transmission access.

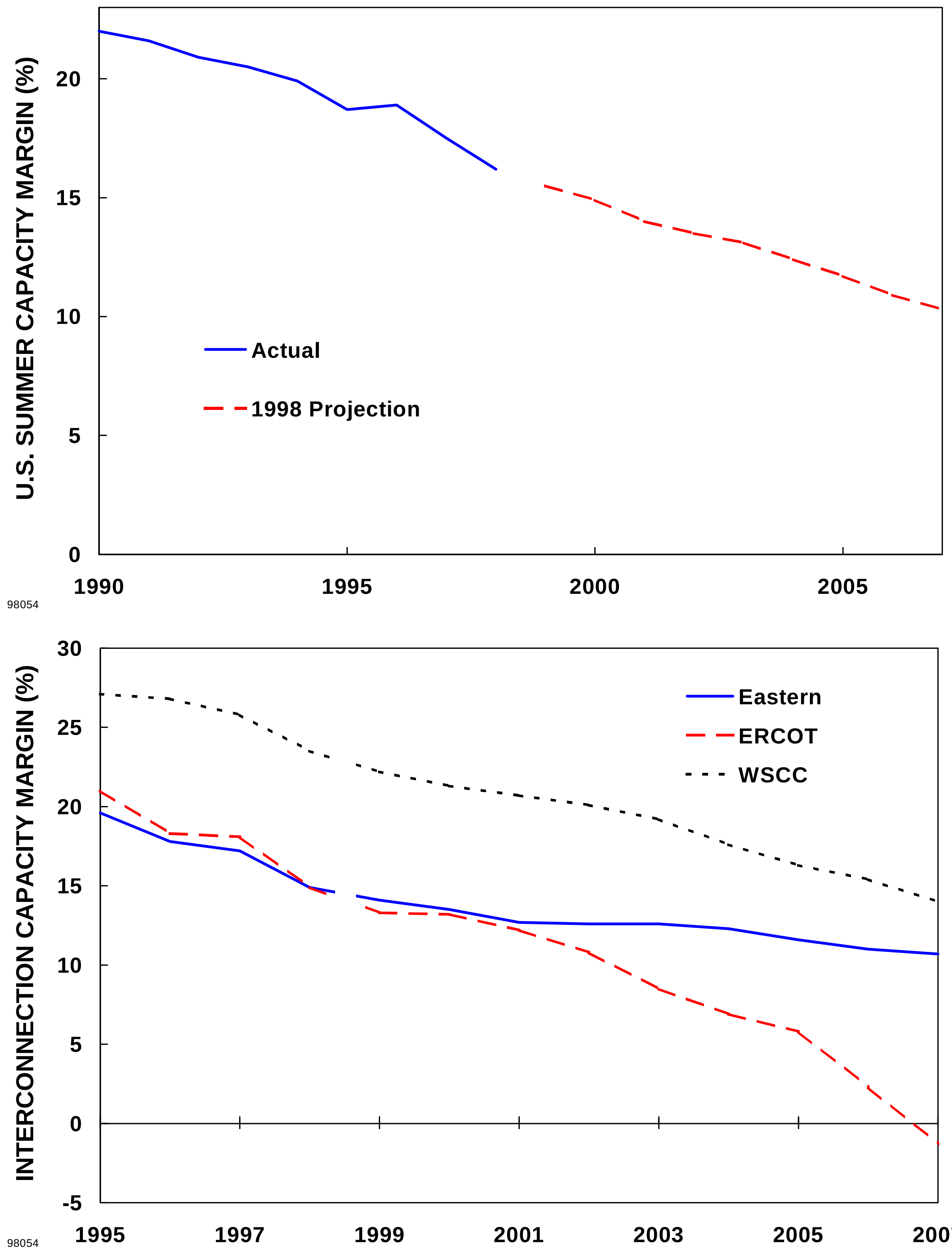


Fig. 1. **U.S. summer generation-capacity margins from 1990 through 2007 (top) and Interconnection capacity margins from 1995 through 2007 (bottom).**

contract). EIA's definition of nonutility capacity is similar to EPSA's for merchant plants in that it, too, includes utility generation that is not part of the ratebase.

In spite of these caveats, the data and projections provide useful insights. Nationwide, EIA estimates a need for 90,000 MW of new generating capacity between 1998 and 2003 to replace units that are being retired and to meet growing electricity demand. NERC's database, the basis for Fig. 1, shows capacity additions of 71,000 MW for this time period, not enough to meet EIA's projected requirements. On the other hand, announced merchant-plant capacity alone (i.e., exclusive of utility-owned generation and generation under long-term contract) totals 99,000 MW during this period, more than enough to meet the EIA requirement.[*]

The data and projections show considerable differences among regions. The amount of planned capacity is greater than what EIA forecasts is needed in the Northeast, East-Central region, Southeast, Florida, Texas, and the West. On the other hand, the amount of planned capacity may not be enough in the mid-Atlantic region, much of the Midwest and the Southwestern Power Pool.

In summary, generation adequacy has declined during the past several years. Utility reports to NERC suggest that these trends will continue for the next decade (although other entities plan to build substantial new generation). Indeed, the latest NERC (1998a) reliability assessment is more pessimistic than earlier ones, primarily because of the restructuring changes under way in the industry. This pessimism relates to a reluctance on the part of utilities to build new generation because of uncertainties about cost recovery for such investments,[#] loss of integration between generation and transmission planning, reluctance of independent power producers to reveal their generation plans much in advance of actual construction,[§] possible double-counting of some generating capacity as more suppliers rely on purchases from other entities, and uncertainty over the extent to which demand-side responses will reduce the need for new generation. (But, as noted above, independent power producers plan to build 99,000 MW of new capacity by 2003.) NERC notes that "… the level of uncertainty has increased tremendously. Purchases from undisclosed resources and the reluctance of generation developers to disclose plans for future capacity additions are making modeling for long-term transmission analysis virtually impossible."

Government agencies and reliability organizations face growing difficulties in obtaining consistent and complete data on existing and planned generation, a consequence of the increasing competitiveness of the industry and the growing diversity of entities that own and operate such facilities. EIA is redesigning its data-collection forms in an effort to deal with (1)

---

[*]On the other hand, this 99,000-MW total from EPSA is about 25,000 MW more than the EIA estimate of planned merchant capacity.

[#]Most of the generation planned by investor-owned utilities is to be built outside of the ratebase.

[§]As of 1998, nonutility entities accounted for 12% of total U.S. electric capability (Hakes 1999).

the concern of many market participants that their data should be confidential, (2) differences in the amounts and frequency of data collected from nonutilities and utilities, and (3) the need to define data elements consistently across market segments (Hakes 1999).

Finally, the time to construct new transmission facilities has increased to the point that it often takes less time to build a gas-fired generating unit than a transmission line.

GENERATION ADEQUACY

CONCEPTS

Historically, utilities maintained "extra" generating resources for short- and long-term purposes; this report focuses on long-term reserves, often called planning reserves, and does not deal with operating reserves. At least two mechanisms can be used to maintain generation adequacy:

- Rely on markets, the interactions of consumers and suppliers acting through the mechanism of volatile spot prices, to decide (1) what types of generation to build when and (2) how much electricity to consume when. California adopted this approach.

- Rely on the traditional system of having a central agency [e.g., an independent system operator (ISO) or state regulator] specify an appropriate minimum reserve margin based on analysis of loss-of-load probability (LOLP) and estimates of the value of lost load (VOLL) and other factors (e.g., forced and planned outage rates for different types of generating units).[*] This reserve margin is then imposed on all load-serving entities (LSEs). The three Northeastern ISOs (PJM, New York, and New England), all of which developed from traditional tight power pools, use this approach.[#]

The United Kingdom uses a third system. There, the National Grid Company calculates, on a day-ahead basis, the expected LOLP for each 30-minute period. This LOLP is then multiplied by the assumed VOLL of about $4/kWh to develop a capacity charge, which is added to the system marginal price (SMP). Thus, the price each online generator receives each half hour is the sum of two components:

$$\text{Pool purchase price} = \text{SMP} + \text{LOLP} \times (\text{VOLL} - \text{SMP}) \,,$$

Wolak and Patrick (1997) note that "the strategic declaration of [generator] availability [is] a very attractive way … to obtain large values of the day-ahead spot price." The nonlinear relationship between the expected reserve margin and the LOLP yields large benefits from

[*]This system of specifying a minimum planning reserve is quite different from the integrated resource planning process that many states required of their electric utilities. While IRP dealt with the technologies, fuels, and costs of generating capacity, minimum reserve margins determine only how much capacity must be installed.

[#]The traditional LOLP criterion of one day in ten years is equivalent to a VOLL of $21/kWh (based on 2.4 hours a year with unserved energy and a $50/kW-year annualized cost of a combustion turbine).

strategically withholding capacity to obtain a small reserve margin and a high LOLP and, therefore, a high capacity-charge payment.

The capacity-charge term was added to provide market signals concerning investment in new generating capacity. But the evidence to date suggests that this administratively determined factor is a source of market power rather than a useful economic incentive to build new generation.

This approach has received little attention in the United States, perhaps because the capacity charge is too easy to manipulate for companies that own large amounts of generation. In addition, administratively setting the VOLL is, at best, a rough approximation of how consumers value electricity. Finally, the day-to-day and seasonal volatility in this capacity charge may make it a poor mechanism to encourage investors to build new generating capacity.[*]

Thus, the key issue on generation adequacy is whether (1) competitive generation markets for capacity and energy will be sufficient to maintain societally desirable levels of reliability or (2) government regulators and central planners (e.g., ISOs or Transcos) will need to impose mandatory minimum-reserve obligations on LSEs to ensure that customers are not involuntarily interrupted from their electricity supplies.

These two options should produce different outcomes in:

- hourly energy prices, with reliance on real-time markets likely to yield lower average prices and costs but greater price volatility;
- customer load shapes, with reliance on real-time markets likely to yield higher load factors; and
- generation portfolios, with reliance on real-time markets likely to yield more baseload capacity and less total capacity.

DISCUSSION

Our review of the literature as well as our discussions with several market participants yielded surprising agreement. Almost everything we read and everyone we spoke with believes that—in the long run—generation adequacy will be left to markets with little involvement by

[*]Ruff (1999) believes that "The England and Wales Pool has probably come the closest to the right approach, using an explicit capacity-adder … . This procedure would be more logical and effective if it were combined with a more-or-less real-time market to price within-day effects." However, the UK is about to abandon this approach to pricing generation capacity.

government regulators. To do otherwise, most people recognize, would interfere with the workings of what are supposed to be competitive energy markets because the energy and capacity markets are closely coupled.

On the other hand, most people agreed that we may need a multiyear transition period while suppliers and, especially, retail customers learn how to respond appropriately to rapidly changing (e.g., hourly) electricity prices. We first have to permit retail customers to face these time-varying prices; in most parts of the country, customers still face prices based on embedded costs that are largely time invariant. We also need to establish intrahour (real time) balancing markets, as proposed by FERC (1999) in its notice on regional transmission organizations. During this transition period, prudence may require maintenance of mandated planning-reserve margins.

Proponents of market-based decisions on generation retirements and expansions worry, however, that electricity price spikes, such as occurred in the Midwest in June 1998, will bring forth inappropriate government price controls.[*] According to Lapson et al. (1998):

> Market fluctuations heighten regulatory risk. The jury is still out on whether policy markets (legislators and regulators—elected and appointed officials) and the public can tolerate price fluctuations in the energy market. After the [June 1998 Midwest] price spike, industrial consumers, utilities, legislators, and others called for price caps or price regulation to limit prices on the upside. (No consumers or legislators have clamored for price floors to limit producers' losses during shoulder seasons when prices are microscopic.) So far, FERC and the Congress have resisted the call for price caps. However, in the future, additional price anomalies, even for brief periods, will reduce regulators' and politicians' enthusiasm for a competitive electricity commodity market.

In support of the market option, Michaels and Ellig (1998) note that:

> Price spikes … provide market participants with important information needed for trading and capacity investment decisions. Price increases signal price-sensitive customers that it is time to conserve, and they tell producers that it may be time to expand capacity. Price increases also give producers and consumers incentives to change their behavior in ways that mitigate severe spikes; producers can profit by investing in new capacity, and consumers can make themselves better off by reducing peak period demand. … It is true that only some customers need to moderate their usage to reduce peak prices for everyone. But in the absence of a competitive market, we have no way of knowing which customers are most willing to do this.

---

[*]Prices spiked, in part, because almost all retail customers paid only traditional, embedded-cost rates and did not face these very high wholesale prices.

The Federal Trade Commission (1999) emphasizes the importance of real-time pricing to improve economic efficiency and reduce market power: “Real-time metering is important because average pricing creates a competitive disconnect that artificially decreases the price elasticity of demand faced by suppliers. Artificial demand inelasticity provides inefficient investment and consumption incentives and facilitates the exercise of market power. Both of these disadvantage customers in the long run by increasing the costs of supplying power and by preventing customers from saving money by responding to real-time price signals … .”

Figure 2 schematically illustrates the supply/demand balances with and without an explicit installed-capacity requirement. The dashed line that slopes up to the left represents consumer demand, and the stairstep line that slopes up to the right represents generating capacity. With a reserve-margin requirement of 11,500 MW, supply and demand equilibrate at a price equal to the variable cost (fuel plus variable O&M) of the last (marginal) unit online at that time (point A). If, however, there is no required reserve margin and market forces yield only 10,500 MW of available capacity, the price of electricity will rise above the variable cost of the last unit online when unconstrained demand exceeds 10,500 MW (point B). The amount of price increase (the pure capacity price in Fig. 2) is a function of the demand elasticity for electricity. The more responsive customer demand is to changing electricity prices (i.e., the flatter the demand curve is), the smaller this capacity price will be. Rose (1997) writes that “This premium [pure capacity price] emerges in peak demand hours in which the chance of a shortage of generation capacity becomes significant.”

This example makes two points. First, even if there is “insufficient” capacity from an engineering perspective, price-responsive demand and supply will equilibrate, and the bulk-power system will not necessarily crash. This equilibrium occurs because some customers would rather forego some consumption than pay the high price associated with this situation. Second, at times of high demand, spot prices will be higher if there is no required planning-reserve margin. In other words, specifying a minimum amount of installed generating capacity will suppress spot prices at certain times. Economists argue that this suppression of a valuable price signal will undercut energy and capacity markets.

Requiring a minimum reserve margin creates two markets (installed capacity and energy) with no assurance that they will be in equilibrium with each other (Graves et al. 1998). This requirement will suppress energy prices and demand-side participation in reliability, and thereby “will undermine the benefits of power industry restructuring.” On the other hand, energy-only markets “will induce efficient capacity planning—which has been the real problem in the past (not inefficient dispatch) and which is where the real opportunities for future efficiency gains lie. It will also encourage demand-side participation in peaking reserves, and forward contracting for risk protection … .”

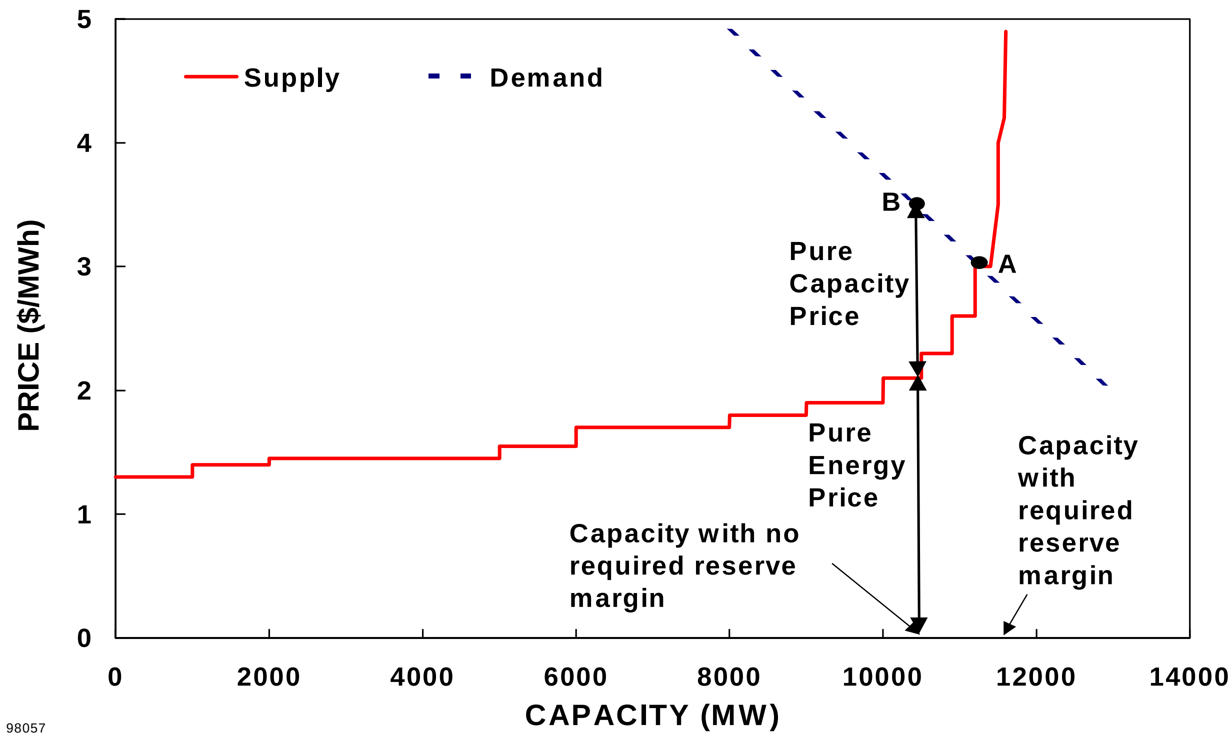


Fig. 2. **With an installed-capacity requirement of 11,500 MW, supply and demand balance at a price of 3.0¢/kWh. With no capacity requirement and only 10,500 MW online, unconstrained demand exceeds supply, and prices rise to 3.4¢/kWh.**

Graves et al. (1998) note several problems with the traditional engineering approach to maintaining generation adequacy, including:

- Setting fixed capacity requirements to deal with what is inherently a very uncertain situation. The uncertainties deal with the timing, extent, and duration of forced outages and with the tremendous variation among customers in their value of lost load.

- Static demand curves with zero price elasticity in spite of the evidence from real-time pricing programs that customers differ substantially in their willingness and ability to respond to changing electricity prices.

- The assumption that forced outages occur in a completely random fashion. In competitive markets, generation owners will work hard to assure that their plants are available during high-price periods.[*]

---

[*]Seiple (1999) concludes that "The move to an asset-based management philosophy will induce companies to improve forced outage rates and reactivate assets that can generate additional profits, thus increasing the overall supply."

- With customer choice, customers may want to choose their level of reliability, not have it specified for them by a central authority.*

Henney (1998) is similarly critical of mandating minimum-reserve requirements:

> This [PJM] proposal [for installed-capacity requirements] artificially separates the market into a capacity market and an energy market. Yet that is not what customers generally buy: they buy kilowatt hours at different times of the day and year. … splitting capacity and energy is a carryover from the regulatory revenue collection practices of monopolies and from the traditional approach to ensuring generation reliability by physical capacity planning. Fundamentally, this provision subordinates the design of a trading market to a function that the PJM-OI has wearing another hat, namely, a responsibility for keeping the lights on by ensuring generation reliability.

Ruff (1999) notes that "international experience demonstrates that market-driven electricity systems can stimulate large amounts of generation investment without long-term contracts … ." However, he also notes that if market prices do not reasonably reflect market conditions [i.e., if the system operator imposes too many restrictive operating rules on market participants], there can be either too much or too little or the wrong kind of new investment.

Hourly prices are a reasonable reflection of the value of energy for most hours of the year. However, when conditions on the grid change rapidly and capacity is scarce, hourly prices are likely to underestimate actual value. During these periods, market participants (generators and loads) have incentives to game the hourly prices by over- or under-generating or consuming. Such gaming behavior has been a problem in California, where the ISO calculates and posts spot prices every 10 minutes, but bases financial settlements on hourly quantities. For either reliability reasons or because of physical constraints on the unit (e.g., slow ramp rate), the ISO sometimes skips bids in its real-time market. The effects of this bid skipping on 10-minute prices is most pronounced during high-demand periods, when the slope of the supply curve is very steep (California ISO 1999). ISO New England has had similar problems, which has required it occasionally to modify prices after the fact (*Electric Power Daily* 1999b). Intrahour prices from two ISOs show a strong seasonal dependence with these differences averaging more than \$5/MWh (and exceeding \$10/MWh for at least 10% of the hours) during the summer months.

The example in Fig. 3 illustrates the kind of situation that can occur when intrahour markets are not carefully defined. In this example, a sudden generator outage at 20 minutes after the hour requires the system operator to call on additional generation, which, in turn, increases the short-term price of electricity. By 30 minutes after the hour, the system is once again in generation/load balance and the short-term price has begun to drop back to its

---

*Large customers may have some choice of reliability level through interruptible contracts with their local utility.

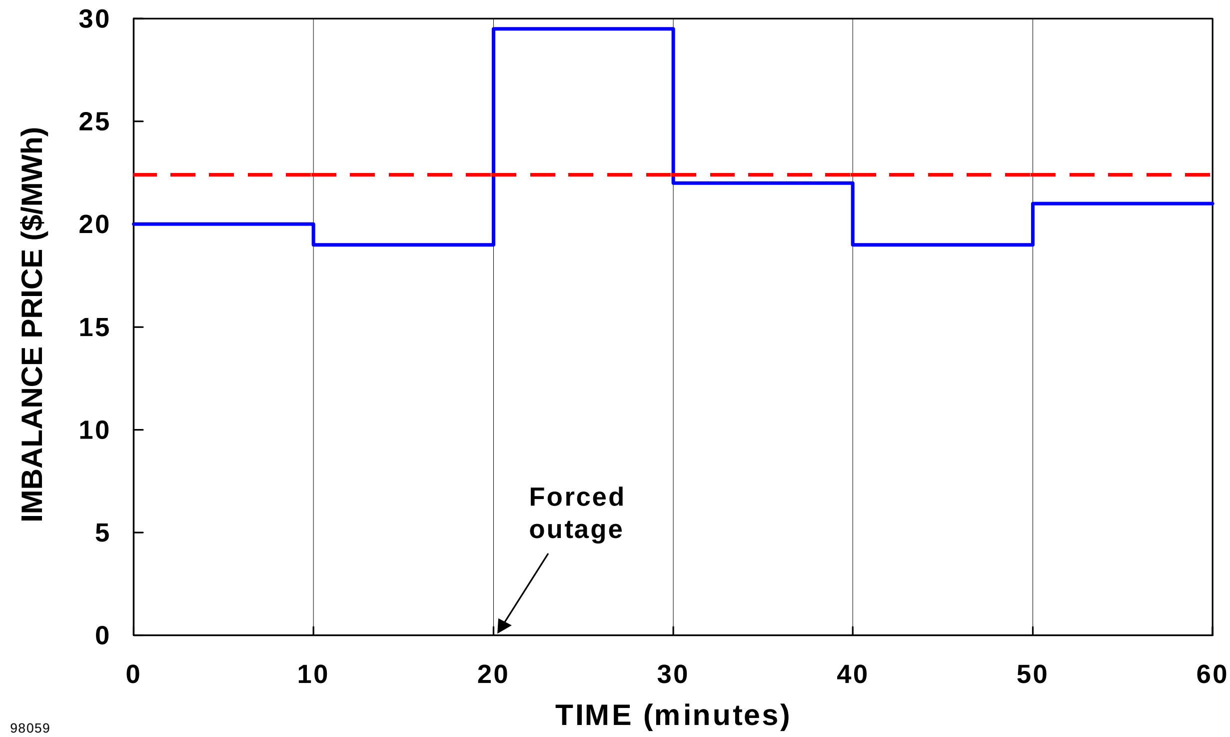


Fig. 3. **Hypothetical example showing the hourly price (dashed line) and 10-minute prices (solid line) during an hour with an outage at 20 minutes past the top of the hour.**

precontingency value. The system operator calculates and posts prices every ten minutes based on the price bids submitted by generators. If these bids do not adequately account for the costs of ramping, starting, and shutting down a generator to meet the kinds of short-term changes shown at 20 minutes, the 10-minute prices will not accurately reflect supplier costs. Therefore, the hourly prices will not accurately reflect supplier costs. This error leads to two problems. First, it encourages suppliers to game the system to profit from these inaccurate prices. Second, these prices do not reflect all the generator costs associated with producing electricity during these high-demand and rapid-change periods. As a consequence, investors may not build the economically correct amount of new generation.

As a consequence of these intrahour dynamics during high-demand and large-imbalance periods, hourly prices may poorly signal markets on needed generation investments. According to Ruff (1999), "Hourly prices that are too low will provide too little incentive to invest in new generation capacity and particularly peaking capacity at critical times." The lesson here is that hourly markets alone may not provide the appropriate incentives for investments in new generation; *intra*hour markets may be required also (FERC 1999).[*] And these markets should

[*]Both PJM and the California ISO are considering changing their settlements to 5 and 10 minutes, respectively, from the current hourly level.

be structured so that suppliers and consumers face appropriate economic incentives. Complicated operating rules and penalties should be avoided as much as possible.

EPSA (1999) states that "… generation adequacy, which is oftentimes characterized as a reliability issue, is in fact a mismatch between purchasers' and suppliers' views of the adequate level (and cost) of supply." It opposes interventions in generation markets by regional transmission organizations, calling strategies like contracts for new generation "ominous" for long-term market operations. On the other hand, such organizations should examine their market rules to ensure that the incentives to suppliers and LSEs are appropriately aligned.

Loehr (1998) offers an opposing, cautionary view of generation adequacy:

> The argument in favor of deregulation is that investors will build new units in response to the demands of the market. Perhaps so. But there are some major concerns. For most electric power systems in the United States, the actual load exceeds 90% of the peak load only 1 to 2% of the time. In the past, utilities had an "obligation to serve" all of the load all of the time; even the last 10%. This was part of the regulatory compact. Thus they planned, built and operated as much generation as was required by the peak load. But today there is a real question as to whether, in an industry driven by competition and the marketplace, investors will be willing to commit financial resources to supply customer load which will be realized only a few hours a year. As far as actual or potential generation owners are concerned, this is a basic question of price and price signal.

Loehr appears to suggest that, in a competitive electricity industry, generation owners cannot earn a profit building plants that operate only 1 to 2% of the time. Given a choice, they would not build such plants. Therefore, society needs to make them do so through minimum-reserve requirements. Society then requires all electricity consumers, regardless of how highly they value electricity consumption at times of tight supplies, to pay for this capacity. Loehr appears not to consider the possibility that these plants will be unprofitable because consumers would rather reduce consumption than pay such high prices. The economists argue that this "extra" generating capacity should be built only if customers are willing to pay the very high spot prices associated with the very infrequent use of these units; otherwise, enough customers will reduce their demand sufficiently to yield a supply/demand balance at a lower level of generating resources and lower peak-period prices.

Jaffe and Felder (1996) believe that mandated capacity requirements are needed because such capacity benefits society at large, not just the owners of such capacity (what the economists call positive externalities). Such societal benefits are especially large for electricity because of its pivotal role in modern society, the real-time nature of electricity production and consumption (which occur within milliseconds of each other), and the difficulty of storing electricity.[*] They note that policymakers can either set minimum-reserve margins or subsidize capacity with an up-front $/kW-year payment for capacity. In principle, the two approaches should yield the same outcome.

NERC (1998a) raises concerns that "few, if any, customers understand the implications of contracting for other than firm power supplies and firm transmission services." Because of the long tradition of ample supplies and the use of interruptible rates to offer implicit discounts to large industrial customers, these customers are used to very few interruptions in service. Indeed, industrial customers, when interrupted, often are angry. Thus, it is an open question how customers will respond to real-time pricing. In addition, only a few electric utilities (e.g., Georgia Power) have much experience and a clear understanding of whether and how customers might respond to real-time pricing. On the other hand, customer loads received about one-third the total payment for operating reserves during the second half of 1998 in New Zealand (Wilson 1999).

ISO APPROACHES

Bulk-power operations in California are split between the Power Exchange (PX) and the ISO (California ISO 1999). The PX runs day-ahead and day-of energy markets for each hour. In addition, the ISO operates a real-time energy market to balance generation and load during each hour. Neither the PX nor the ISO specifies installed-capacity requirements for market participants in California. And neither entity operates an installed-capacity market.

Between April 1, 1998, and March 31, 1999, the weighted average price of electricity in the PX day-ahead market was $26.6/MWh. For 12% of the hours, prices were at or below $10/MWh. At the other extreme, prices were at or above $100/MWh for 1.1% of the hours, with prices ranging as high as $200/MWh. These prices are well above the marginal costs of the most expensive units in California and reflect the pure capacity price shown in Fig. 2.

The California ISO (1998) points to the number and size of the proposed power plants in California (16 projects with a total capacity of more than 10,000 MW as of spring 1999) as evidence that competitive markets for capacity can work.

The PJM (1998a) *Reliability Assurance Agreement* (RAA) establishes the obligations of all LSEs within the PJM control area to provide the amount of installed generating capacity that PJM determines is needed to maintain reliability. The PJM Reliability Committee determines the forecast pool requirement, the reserve margin for the PJM Control Area required as part of this agreement. The RAA "is intended to ensure that adequate Capacity Resources will be planned and made available to provide reliable service to loads within the PJM Control

[*]These societal benefits might include avoidance of the looting and violence that can erupt during a major blackout and the maintenance of electrical service to vital societal functions, such as hospitals, police and fire stations, traffic lights, and airport traffic-control systems.

Area, to assist other Parties during Emergencies, and to coordinate planning of Capacity Resources consistent with the Reliability Principles and Standards."

The PJM Reliability Committee determines the forecast pool requirement for capacity resources using "probability methods" and establishes criteria for use of capacity resources during emergencies. The forecast pool requirement is intended to "ensure a sufficient amount of capacity to meet the forecast load plus reserves adequate to provide for the unavailability of Capacity Resources, load forecasting uncertainty, and planned and maintenance outages." The focus is on the peak season, which for PJM overall is the summer.

In October 1998, PJM (1998b) established monthly Capacity Credit Markets to allow PJM market participants to buy and sell capacity credits to meet their obligations under the RAA. Any PJM member that has PJM-qualified resources or is an LSE *must* bid any excess or deficiencies into these markets. The markets for January through November 1999 cleared 6300 MW at prices that ranged from $1 to $160/MW-day, with an average of $60/MW-day (equivalent to $22/kW-year, assuming that capacity is valued equally for every day of the year). Prices were highest for June and July and lowest for September through November.

PJM began daily markets (conducted a day ahead) in installed capacity in January 1999. From January through September 1999, the daily price averaged less than $4/MW-day, far below the prices in the monthly markets. We do not know whether these price differences reflect seasonal differences, differences between daily and monthly markets, or lack of familiarity with these new products.

The New York ISO (1998) proposal for installed-capacity requirements is similar to the PJM approach. New York explains clearly the purpose of its capacity requirement: "Adequate resource capability shall exist in New York State when, after due allowance for scheduled and forced outages and scheduled and forced deratings assistance of interconnection with neighboring Control Areas and regions, and capacity and/or load relief from available operating procedures, the probability of disconnecting firm load due to resource deficiency will be, on average, no more than once in ten years."

The ISO New England (1997) approach differs from the PJM and New York approaches in that New England has two capacity components: monthly installed capability and hourly operable capability.* Operable capability refers to "any generating unit or units in any hour … which is operating or available to respond within an appropriate period to the System Operator's call to meet the Energy and/or Operating Reserve and/or AGC [automatic generation control] requirements of the NEPOOL [New England Power Pool] Control Area … ." New England market participants are *required* to bid all their operable capacity in excess of their obligations into the hourly operable capability market.

---

*The difference between installed and operable capability appears to be inoperable capability. Because no one should want to purchase inoperable capability, these two markets may be redundant. In addition, operable capacity seems to duplicate the real-power ancillary services. Indeed, ISO New England (1999) stated that its market for operable capability is "fundamentally flawed" and likely to be either replaced or eliminated.

It is unclear why New England requires two capacity markets in addition to the energy and ancillary-services markets. Our discussions with several people in the region suggest that the two markets are historical artifacts and that, within a few years, one or both will be eliminated.

ISO New England (1998) noted that "NEPOOL has had an ICAP [installed-capability] requirement since its inception. This requirement has been important in maintaining reliability in New England for over 25 years." These comments were in response to criticism from Cramton and Wilson (1998), who had been hired by ISO New England to review New England's proposed market rules and were quite critical of those rules. Specifically, they wrote about the installed-capability requirement:

> This holdover from an era of regulation is unique in the electricity industry, which is the only one that does not expect suppliers to cover fixed costs, such as capital and maintenance, from the market price of its output. … The capacity markets are a holdover from the regulated setting, when capacity decisions were not made in response to price expectations. In the transition to a competitive market, the capacity markets may serve a useful role in coordinating investments in capacity. However, once competitive electricity markets are established in New England, it would be appropriate for the capacity markets to terminate.

New England may maintain both installed- and operable-capability requirements because the installed-capability requirements are largely independent of availability. The installed-capability requirements relate primarily to "iron in the ground" without regard to the ability of that unit to operate any time soon. For example, the three large Millstone nuclear units were out of service for 18 months or longer, during which time they continued to qualify as installed capability in New England.

The clearing price for installed capacity in the New England market was zero for all months between April 1998 (when ISO New England started the market) and February 1999. Prices were positive in March and April 1999 and returned to essentially zero in May.

This discussion raises the difficulty in determining what to include as installed capacity. After all, installed but unavailable capacity does not contribute to reliability. Over what time period should generating-unit availability be measured? Should availability be determined on a daily, monthly, seasonal, or annual basis? Because the need for capacity is generally greatest during winter and summer peak periods, it is most important to measure availability during those time periods. PJM (1997) adjusts availability on the basis of maintenance outages that occur during the peak season (the 24th through 36th weeks of the year). The shorter the time period over which the capacity requirements are determined and paid for, the more accurately capacity prices will reflect their value to the grid. Monthly requirements are better than annual requirements, and daily requirements are better still.

THE ORCED MODEL

We developed ORCED to analyze various issues related to the restructuring of the U.S. electricity industry, including the generation-adequacy issues discussed here (Hadley and Hirst 1998). ORCED is a simple strategic planning model that simulates the operations of, and resultant prices and producer profits from, competitive bulk-power systems.

DESCRIPTION

ORCED can analyze the construction of new generation, retirement of existing generation, and the operation of competitive (as opposed to the traditional regulated) bulk-power markets. The model can be used to examine issues related to emissions, prices in low- vs high-cost regions, stranded costs, market power, generator profitability, and the mix of different generation technologies and fuels.

The model can simulate different bulk-power market structures. In particular, the user can specify one of three pricing schemes:

- An energy-only spot price in ¢/kWh (as implemented by the California PX and ISO). When unconstrained demand exceeds available supply, what would otherwise be unserved energy is “curtailed” because spot prices rise sufficiently to suppress demand to match the level of available generating capacity. The user simulates this situation by specifying a value for the price elasticity during these time periods. ORCED uses the amount of demand to be curtailed and the price elasticity to calculate the value of unserved energy in ¢/kWh.

- An energy-only spot price plus a capacity-reservation price (in $/kW-year), as implemented by the PJM Interconnection.

- An energy-only spot price plus the loss-of-load probability (capacity) component used in the United Kingdom. Here, the user specifies a value for unserved energy (e.g., 200¢/kWh), which the model multiplies by the loss-of-load probability. The resultant product is then added to the energy-only spot price during hours with unserved energy.

ORCED analyses and results are functions of the following factors:

- Characteristics of individual generators: capital and other fixed costs ($/kW-year), fuel and variable O&M costs (¢/kWh), dispatchability, and outage rates (%);

- Generating-resource portfolio: the mix of generating units and the relationship between generating capacity available and unconstrained peak demand;

- Customer characteristics: the load shapes and price elasticities of demand; and

- Capacity, cost, and losses in the transmission link between two regions.

Although ORCED is a two-region production-costing model, we used a single-region version of ORCED for the present analysis. We used the simpler version because our focus is on generation adequacy (not transmission issues), it runs much faster, and it includes 51 generating units in the single region (rather than 26 units in each of the two regions).

ORCED uses load-duration curves rather than chronological loads as inputs (Fig. 4).* The model is run twice for the year of simulation: once for an onpeak season and a second time for an offpeak season.

Use of load-duration curves is computationally much simpler and faster than the hour-by-hour analysis of chronological-dispatch models. This simplification, however, has a price: because it obscures the timing of system loads, production-cost analysis on the basis of load-duration curves cannot analyze the details of generator operations and costs, especially those associated with minimum and maximum loading points, incremental heat rates, startup times and costs, and minimum shutdown times. To partially remedy these problems, ORCED analyzes two user-specified seasons each year and adds a startup cost (in $/kW) for units that operate less than 10% of the hours in each season.

The model has available to it 51 generating units. All but one of the units are characterized in terms of capacity, forced- and planned-outage rates, fuel type, heat rate, variable and fixed O&M costs, and annual capital costs (based on initial construction cost, year of completion, and capitalization structure). One unit is an energy-limited hydro unit, for which the inputs include, in addition to those noted above, the plant's onpeak and offpeak capacity factors (equivalent to its maximum energy output for each season). This treatment of hydro as energy-limited ensures that hydro displaces the most expensive energy (i.e., at the top of the load-duration curves).

The model dispatches these generating units separately for the two seasons. Although the calculation process is the same for the two seasons, the results differ because of differences in the load-duration curves and because all the planned maintenance is assumed to occur in the offpeak season.

---

*A load-duration curve is created by ordering hourly system demands (in MW) from highest to lowest. The resultant curve shows the fraction of time (for the specified time period) that demand exceeds a particular value, ranging from the one-hour system peak down to the minimum demand.

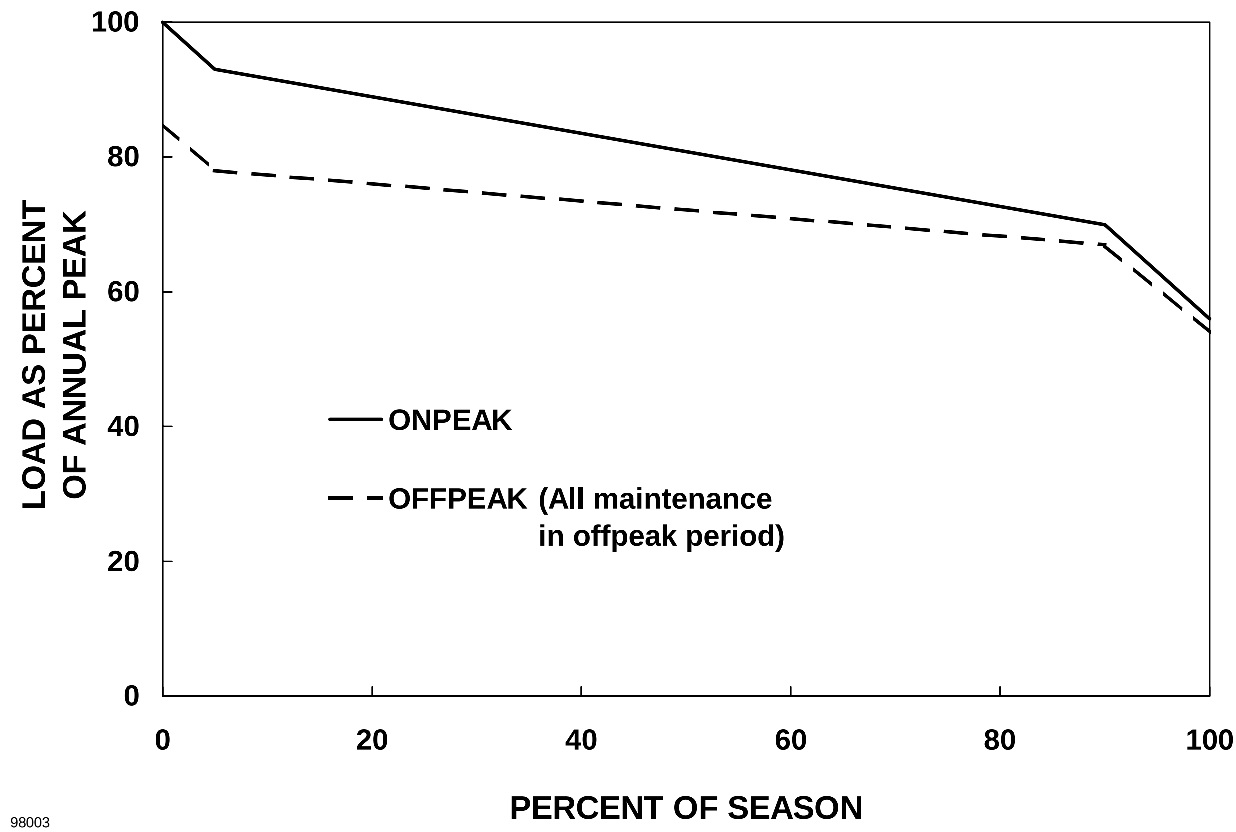


Fig. 4. **ORCED analyzes customer loads on the basis of load-duration curves for two user-specified seasons.**

The plants are first dispatched against the load-duration curve on the basis of bid price, the default for which is variable (fuel plus variable O&M) costs. Figure 5 shows a typical supply curve with marginal costs (and prices) increasing with increasing demand. (If the plant owner bids a zero price for a unit, the generator is treated as a must-run unit and is dispatched first by the model.) Because plants are not available 100% of the time, we model forced outages on a probabilistic basis.* Thus, the higher-cost plants will see not only customer loads but also "equivalent demands" based on the probability that plants lower in the dispatch order (i.e., less expensive to operate) will be undergoing a forced outage. The model creates an equivalent load-duration curve for each plant, which extends the amount of time the plant runs based on the forced-outage rates of the plants lower in the dispatch order.

ORCED calculates market prices (based on the bids from individual generators) for each time period during the two seasons. The prices also incorporate any externally imposed uplift

*The amount of computer time required for a full simulation depends strongly on the number of generators treated probabilistically. We found a reasonable tradeoff between computing time and accuracy when about 10 plants are modeled probabilistically and the other 41 are derated.

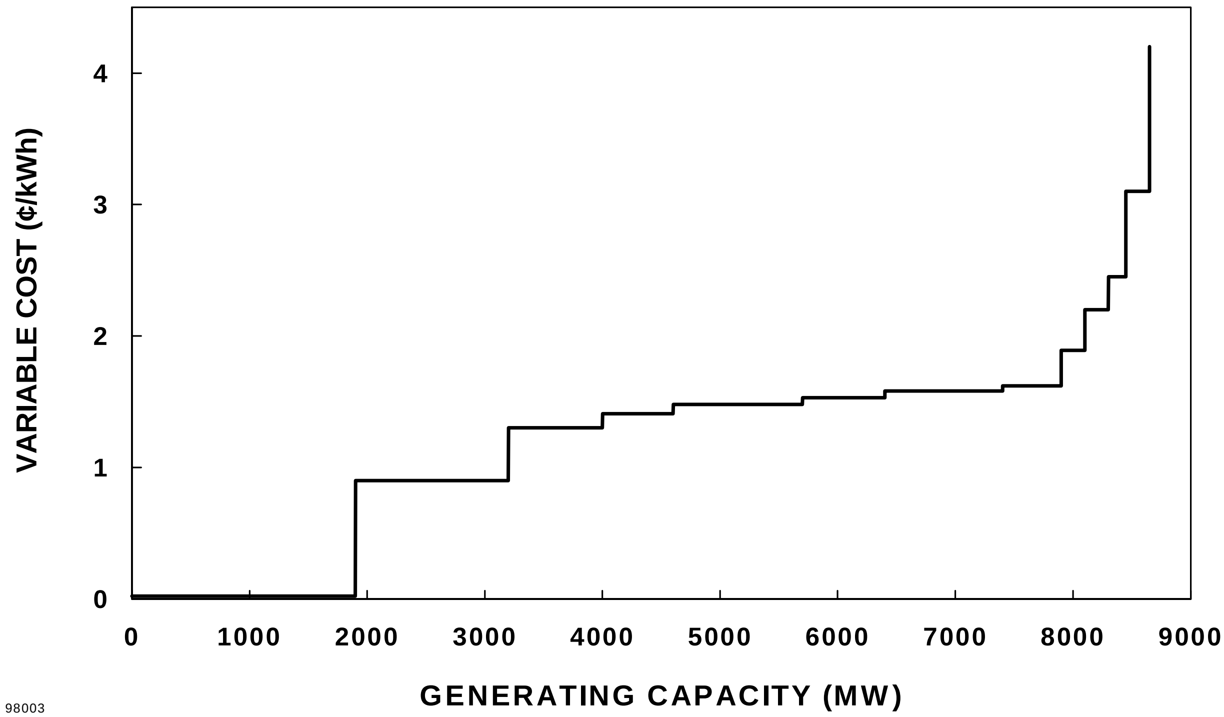


Fig. 5. User inputs specify generating-unit characteristics for 51 units. ORCED dispatches these units on the basis of either bid price or variable cost.

charge (e.g., an O&M adder to pay for capacity with low capacity factors), capacity charge (e.g., that associated with any mandated planning-reserve margin), and emission taxes. The prices during high-demand hours also reflect generator startup costs and the costs of any unserved energy for those hours during which unconstrained demand exceeds supply.

ORCED can be run iteratively to estimate customer response to changes in overall electricity-price levels and to real-time pricing. Consumer responses to changes in electricity prices are represented by three input demand elasticities:*

- Overall (annual) elasticity that adjusts annual consumption up or down on the basis of decreases or increases in overall electricity price. This elasticity is used to adjust the entire load-duration curve up (or down) in response to decreases (or increases) in the overall price of electricity.

- Time-of-use (hourly) elasticity that changes the shape of the load-duration curve in response to changes in hourly spot prices. This elasticity is used to adjust each point on the load-duration curve up (or down) based on decreases (or increases) in the price of electricity during that time period.

*The price elasticity of demand is the change in demand for a product caused by a change in its price.

- Unserved-energy (very short-term) elasticity that adjusts demand down during those periods when unconstrained demand would otherwise exceed capacity, used to calculate the market price at which supply and demand equilibrate. This elasticity is used to calculate the cost of unserved energy.

The analyses discussed in the next chapter use only the second and third elasticity factors.

ORCED can be run in either a simulation mode or an optimization mode. In the simulation mode, ORCED is a production-costing model that determines the least-cost way to meet customer electricity demands given a fixed set of generating units. In the optimization mode, ORCED is a combined capacity-optimization and production-costing model that determines the "optimal" mix of generating units available that year as well as the least-cost use of those generators to meet customer demands. The user can specify different objectives in the optimization routine, such as minimizing the total cost of producing electricity, minimizing the sum of variable plus avoidable fixed costs, minimizing electricity price, or maximizing producer profits.

The user can also impose constraints on the optimization. These constraints can apply to individual generating units or to the system as a whole. For example, maximum-capacity constraints could be imposed on existing generating units (i.e., those units constructed before the year of the simulation). Other constraints could limit the amounts of new capacity of certain types that could be constructed. System constraints could specify a planning-reserve margin or a carbon-emission cap, as examples.

FEATURES THAT AFFECT THIS ANALYSIS

This section describes the critical features of ORCED that have substantial effects on our analysis of alternative ways to ensure sufficient generation capacity. These features include the price elasticities of demand, determination of the O&M adder for the cases in which price elasticities determine the amount of installed generating capacity, determination of the annual capacity payment for the cases in which a minimum planning-reserve margin is specified, and the optimization process.

Price Elasticities of Demand

As indicated above, ORCED includes three price elasticities that affect consumer responses to price changes in the very short term, hourly, and on an annual basis. These elasticities are averages over all customer classes.

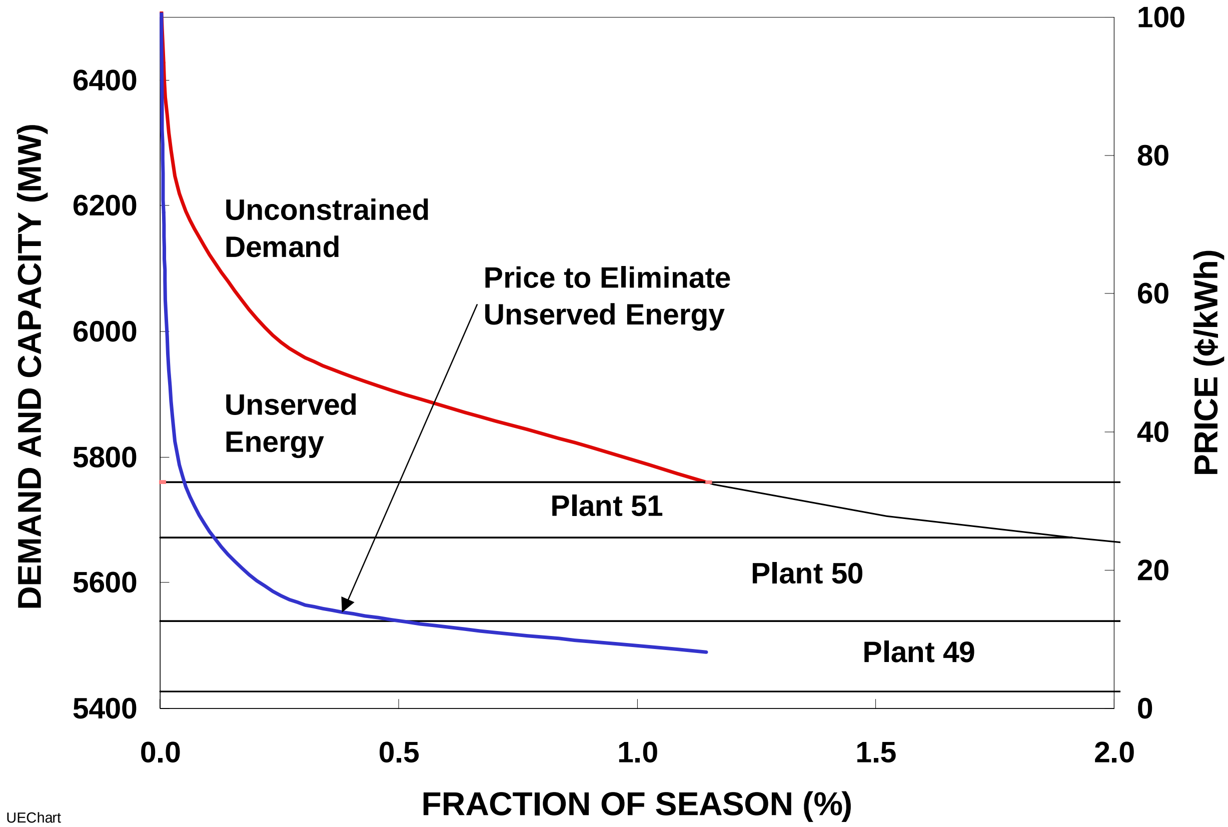


Fig. 6. Electricity supply, demand, and prices when unconstrained demand exceeds supply.

The unserved energy elasticity is used within ORCED. For a small part of each season, the combination of high customer demands and generator forced outages can cause unconstrained demand to exceed available capacity. In the example of Fig. 6, demand exceeds the capacity of all 51 plants 1% of the peak season. During the time that plant 51 (the most expensive generator) is on the margin, the market price of power is based on its bid price. As unconstrained demand exceeds the total capacity, the spot price is raised to lower demand until demand is, once again, equal to the online supply. Using the input unserved elasticity factor, prices are calculated that reduce demand to match total capacity. These prices are the market prices during the period of constrained demand.

In the Fig. 6 example, 5760 MW of supply are available to meet 6500 MW of unconstrained demand. Although the bid price of the most expensive generator online (an old oil-fired unit in this case) is 4.95¢/kWh[*], the price during this unserved-energy period reaches as high as $1.06/kWh and averages 18¢/kWh to reduce demand to the level of supply available. The input unserved-energy price elasticity (−0.05 in this case) determines the price increase

[*]This 4.95¢/kWh consists of 3.53¢/kWh for fuel plus variable O&M costs, 1.2¢/kWh for startup costs, and 0.22¢/kWh for the O&M adder described below.

needed to equilibrate demand to supply; the higher the price elasticity, the lower the price increase.

The other two elasticities available within ORCED are used to change demand over longer (hourly and annual) periods. They are used in a separate worksheet to change the LDCs for the peak and off-peak periods. From a given ORCED run, we determine the real-time price curve and corresponding customer demands (based on the original LDC) for each season. We assume that the original demands are based on a uniform price over the entire season (e.g., the cost-based price that ORCED calculates to represent regulated rates). A separate nongeneration price (to reflect the costs of transmission, distribution, and customer service) is added to the ORCED-calculated generation price to obtain the average retail price of electricity.

Given the real-time rates and customer load profile (from the LDC), an energy-weighted average market-based price is calculated. This average price is calculated by multiplying the demand during each part of the season by the associated price. For example, the minimum price is charged when demand is at its lowest; prices are highest at peak times. Consequently, the annual *energy*-weighted price may be 5% to 10% higher than the *time*-weighted price. Once the annual average price is calculated, it is used to define an adjustment factor to raise or lower the power level for the entire season with the equation:

$$R_{avg} = (\text{Annual Average Market Price}_{new}/\text{Regulated Price}_{old})^{\text{e-average}} ,$$

where e-average is the elasticity for the average price change (typically −0.5). We did not use this overall price elasticity in these analyses of generation adequacy; we assumed that the overall price elasticity of demand is zero.

Next, for each time during each season, the real-time market price is compared to the average market price. Demand is adjusted for this time with the equation:

$$\text{Demand}_{new\text{-}t} = \text{Demand}_{old\text{-}t} \times R_{avg} \times (\text{Real-Time Price}_t/\text{Average Market Price})^{\text{e-real time}} ,$$

where e-real time is the elasticity for the real-time price change. In these analyses, we set this elasticity equal to double the value of the unserved-energy elasticity (typically −0.1 and −0.05, respectively). This procedure results in a series of demands and associated times that produce a new LDC. The resulting LDC is then used in a subsequent run of ORCED (Fig. 7).

O&M Adder

Based on the costs and characteristics of the generators, ORCED calculates the dispatch of these units and the consequent marginal-cost-based prices for both seasons of the analysis year. These results are then used to calculate the revenues and operating costs for each generator; the difference between revenues and operating costs is net operating income for the year.

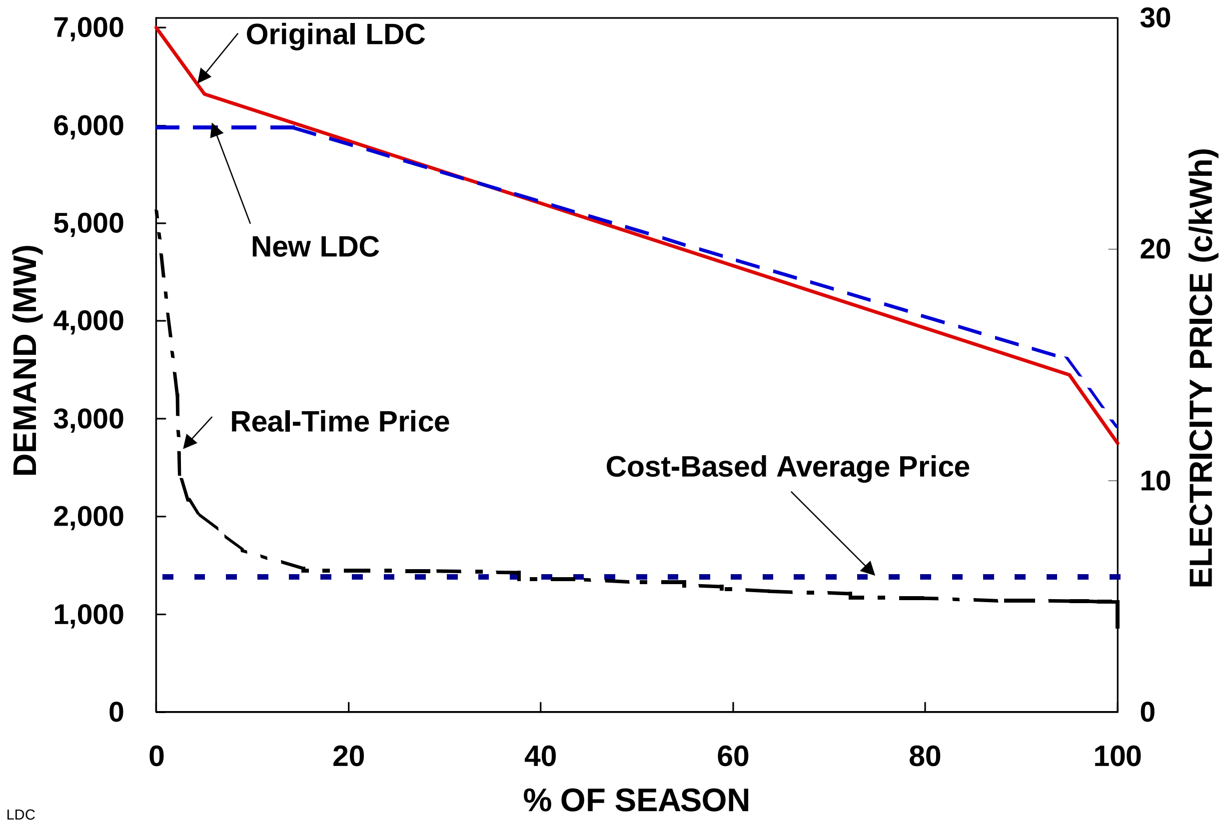


Fig. 7. **The effects of a change in the real-time price of electricity on the load-duration curve.**

If a unit bids its variable costs (as it should in an economically efficient market), it might not recover all its avoidable fixed costs or its unavoidable capital costs. The structure of competitive generation markets is such that existing generators *must* recover all their avoidable costs (which generally includes fuel plus variable and fixed O&M costs); failure to recover these costs will lead the owners of these units to shut them down. The hurdle for new units is greater; they must recover all their costs, both operating and capital, from energy revenues.

Basing market prices on the marginal variable costs of the most expensive unit online at any time works for most of the year. However, this approach may not provide sufficient revenues for those units that have high fixed operating costs and therefore are economical to run for only a few hours a year (i.e., they have low capacity factors). The owners of these units are likely to bid prices higher than their variable costs because they require additional revenues to remain profitable and because their units are likely to be the most expensive ones online and therefore the price setters.

Rather than calculate a separate O&M adder for each unit, which would greatly complicate the calculations within ORCED, we calculate a minimum adder that ensures that all generators recover their avoidable costs (i.e., they at least break even). This premium, specified in terms of \$/kW-year, is calculated only for units with a capacity factor of 10% or less. ORCED converts this factor into a ¢/kWh term and adds it to each unit's price after calculating the dispatch order for the season. The conversion factor for each generator is:

$$\text{Premium (¢/kWh)} = \text{Capacity Factor} \times \text{Adder (\$/kW-year)} \times (100¢/\$)/(8760\ \text{hr/year})\ .$$

For example, a unit with a 5% capacity factor might add to its bid price a premium of 0.5¢/kWh for its output, while another unit with a 2% capacity factor would add 1.25¢/kWh. This O&M adder is employed for those cases in which markets (rather than central planners) specify the amount of generating capacity that will be available each year.

Capacity Adder

For those cases with a mandated planning-reserve margin, the electric system must provide some mechanism that pays for any "extra" generating capacity that is installed. In this context, "extra" refers to generation that would not be economical to build or operate in markets that did not include a minimum planning reserve and would therefore lose money. To ensure that such plants are built, we developed a capacity payment in ORCED along the lines of that used by the PJM Interconnection.

The minimum capacity payment needed to provide enough capacity to meet the minimum planning reserve is determined by ordering all 51 generators in terms of increasing losses (expressed as \$/kW-year), as shown in Fig. 8. Some units with low fixed costs and high capacity factors have negative capacity costs (e.g., those on the left side of Fig. 8). But other units with high fixed costs and/or low capacity factors cannot make enough money in energy markets to cover their fixed costs.

To ensure that these units, needed to meet the planning-reserve requirements, remain available, ORCED calculates the minimum annual capacity payment based on the losses of each generator. The payment to all generators is based on the loss of the last unit (i.e., the unit with the highest loss) that ORCED pays to meet the reliability requirement. This capacity payment is adjusted for each unit's availability:

$$\text{Capacity Payment (\$)} = \text{Capacity Adder (\$/kW-year)} \times \text{Capacity} \times (1 - \text{FOR} - \text{POR}),$$

where FOR is the unit's forced outage rate and POR is the unit's planned outage rate. The factor (1 − FOR − POR) is the unit's availability (i.e., the fraction of the year that the unit is available to produce energy). The capacity payment is provided to each generator and is added to each customer's electricity bill.

An alternative approach would pay each generator that would otherwise lose money a unit-specific capacity charge. These charges would be set so that each such unit would just break even. On the surface, such a system would appear to save money for consumers because

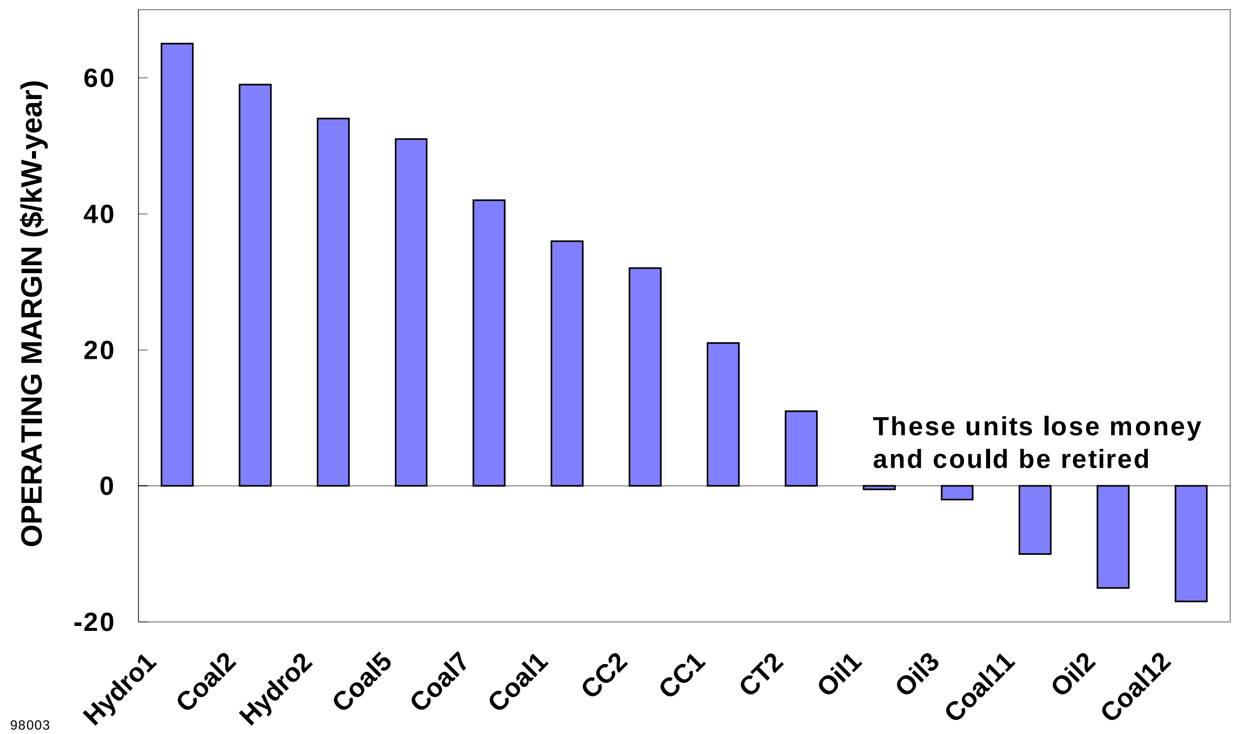


Fig. 8. ORCED results showing the contribution to margin for a sample of the generating units.

it would not pay all generators the same (higher) annual capacity charge. However, this alternative approach invites gaming by generators; each generator would raise its bid to the level that it guesses will match that of the most expensive winning bid. Such incentives to bid dishonestly (i.e., to submit bids different from marginal costs) would reduce economic efficiency. Therefore, this approach was not considered further.

Optimization Process

ORCED operates as a simulation model in that, given a set of generating units and demand curves, the units are dispatched and the consequent finances calculated. The Excel Solver is used to add an optimization feature to the model.

To run Solver, one must specify an objective function, variables, and constraints. Typically, the *objective* is to minimize the avoidable costs, those costs that are not sunk. This factor includes all fuel costs, fixed and variable O&M costs, unserved energy costs, and the capital costs of plants not yet built. The *variables* refer to those factors that ORCED can change as it searches for an optimal solution. They include the capacities of a subset of the 51 generators.* *Constraints* set limits on the values of the variables (i.e., the capacities of the

---

*We could include all plants but only at the expense of greatly increased calculation times. Experience with ORCED showed that some plants are never modified by the optimization routine and can be cut from the variable list.

generators). The capacity of existing plants can only be reduced while the capacity of new plants can be reduced or increased. We set upper limits on the capacity of individual new plants so that they do not become too large a share of total capacity, typically about 5% of demand.

After some experimentation with different objective functions, we decided to minimize the avoided cost of electricity production, taking into account both the construction costs for new generation and the operation of existing generation. Running ORCED to minimize the price of electricity or to maximize generator profitability led to results that were, in our view, unreasonable. For example, minimizing the price of electricity led to the construction of many baseload units, which, in turn, led to major earnings losses for the owners of these units.

Solver works best when the model is linear and without discontinuities in the objective function. However, ORCED has inherent discontinuities and nonlinearities. The unserved-energy prices depend on the most expensive generator then online. If the most expensive plant is retired, then the unserved energy price instantly drops. The probabilistic treatment of forced outages creates nonlinear solutions as well. As a consequence, Solver does not always find the global optimum, but instead stops at a local optimum. The user then must either run Solver with a range of initial conditions or manipulate the Solver results to see if costs can be further reduced. One method to further optimize the solution is to retire those plants that are the least profitable, checking to see if the avoidable costs continue to decline. At some point, the increase in unserved energy raises costs with further retirements, and the optimization process stops. Alternatively, we achieved considerable success by running each case twice, first with Solver used to minimize avoidable cost and second with Solver used to minimize price, constraining avoidable cost to be no more than its “optimized” value.

QUANTITATIVE ANALYSIS

STRUCTURE OF THE ANALYSIS

We ran two sets of cases with ORCED for this project, one with mandated planning reserves and one with market-determined reserves.

- **Specified planning-reserve margins.** We fixed the unserved-energy elasticity at 0.05* and ran several cases with different values of reserve margin. [Because ORCED deals with energy and not with ancillary services, these reserve margins should be increased by at least 5 percentage points to reflect the need for generating capacity for regulation, spinning reserve, and supplemental reserve (Hirst and Kirby 1998).] We set up these model runs to minimize the avoided cost of electricity production. We then added an annual capacity payment (in $/kW-year) to ensure that the most unprofitable unit needed to meet the minimum reserve requirement just broke even. This capacity payment was determined by dividing the monetary losses for each generator (for those generators that lost money) by the availability-adjusted capacity of each generator. Given the required amount of installed capacity, the payment was set equal to the highest dollar-per-kW loss to ensure that no generator lost money.# This capacity payment was then added to the price of electricity that consumers pay.§

- **Market-determined reserves.** We varied the unserved-energy price elasticity and let the model determine the "optimal" reserve margin. Here, too, ORCED selected generating units to retire and build to minimize the avoidable cost of electricity production, taking into account both the construction of new generators and the

---

*We have no empirical basis for choosing a value of 0.05 for this elasticity. Initially, this elasticity will be low because consumers will not yet have installed the technologies allowing a full response to real-time pricing. In the long run, as consumers install such technologies, elasticities will increase. An elasticity of 0.05 (−0.05, to be precise) means that a 1% increase in the price of electricity cuts demand by 0.05%. Doubling the unserved electricity price cuts demand by 3.4% ($1 - 2^{-0.05}$).

#This no-loss constraint is essential in competitive electricity markets. Were a unit to lose money continuously, it would go out of business, which would reduce the amount of installed capacity below the minimum specified. This problem is not solved by a change of ownership because it is a function only of operating, not capital, costs.

§At low values of specified reserve margins, all the generators are profitable in the energy market, and there is no need for a capacity payment (i.e., the total and energy prices are the same). (See the left side of the graph at the top of Fig. 9 and subsequent figures in this chapter.) In these situations, investors, reacting to these earnings opportunities, would build more generating capacity. To distinguish clearly between the two sets of cases, one regulated and one market dominated, we did not allow ORCED to add such generation in these low-reserve-margin cases.

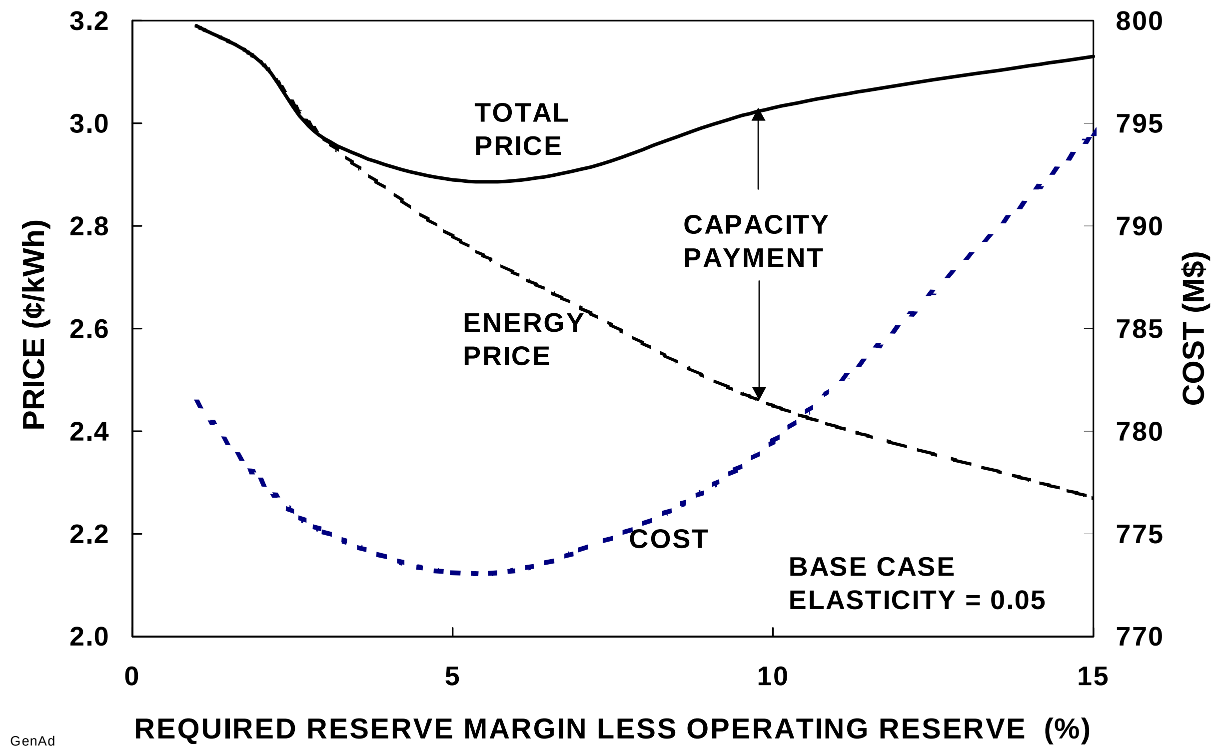
operation of the existing generators. We added an O&M adder (expressed in $/kW-year) for those units that operate less than 10% of the hours. We added this factor (which ranged between 0 and about $2/kW-year) to ensure that plants operating for only a few hours a year would recover their avoidable fixed costs. (ORCED converts the adder to an energy-price premium paid to those units; the premium increases as capacity factor declines below 10%.) The rationale for including this O&M adder is the same as that used to justify the capacity adder in the fixed-reserve margin cases—to guarantee that no generator loses money.

We ran these two sets of cases for five scenarios: a base case, natural gas prices 100% higher than in the base case, a system load factor 16% lower than in the base case, customer responses to time-of-use pricing as well as to unserved-energy prices, and an unserved-energy elasticity of 0.02 instead of 0.05.

BASE CASE

Our base case is an electric system with peak demand of 7000 MW and annual energy consumption of 38,600 GWh, yielding a system load factor of 63%. With an unserved-energy elasticity of 0.05 and a reserve margin of 5%, the shares of generating capacity and energy are: coal (39 and 50%), gas (33 and 25%), nuclear (13 and 17%), hydro (11 and 8%), and oil (4 and 0%). The price of natural gas is about double that of coal ($2.25 and $1.39/MBtu, respectively). All the new generating units (960 MW) are gas-fired combined-cycle units. The annual average price of electricity is 2.89¢/kWh (with hourly prices ranging from 0.5 to 22.8¢/kWh throughout the year), and the avoidable cost of electricity is $773 million. This base-case system is roughly consistent with EIA's (1998) *Annual Energy Outlook 1999* Reference Case Forecast for the years 2000 and 2005.

We ran cases with reserve margins ranging from 1 to 15% (Fig. 9). Model results show, as expected, that the total cost of electricity production increases at very low and very high levels of reserve margins. For the assumed unserved-energy elasticity of 0.05, the cost curve has a broad minimum at around a 5% reserve margin. As the required reserve margin increases beyond the optimum value, the total price of electricity (i.e., the spot price of energy plus the capacity payment) increases. However, as the required reserve margin increases, the spot price of energy declines, and the capacity charge increases. Thus, setting the reserve margin too high drives a substantial wedge between hourly energy prices and the total price that consumers see, undercutting the operation of the energy market. In addition, higher specified reserve margins lower peak-period prices and the volatility of prices during the year. As the required reserve margin is increased from 1% to 5% to 10%, the maximum spot (unserved-energy) price declines from 28¢/kWh to 23¢/kWh to 19¢/kWh.



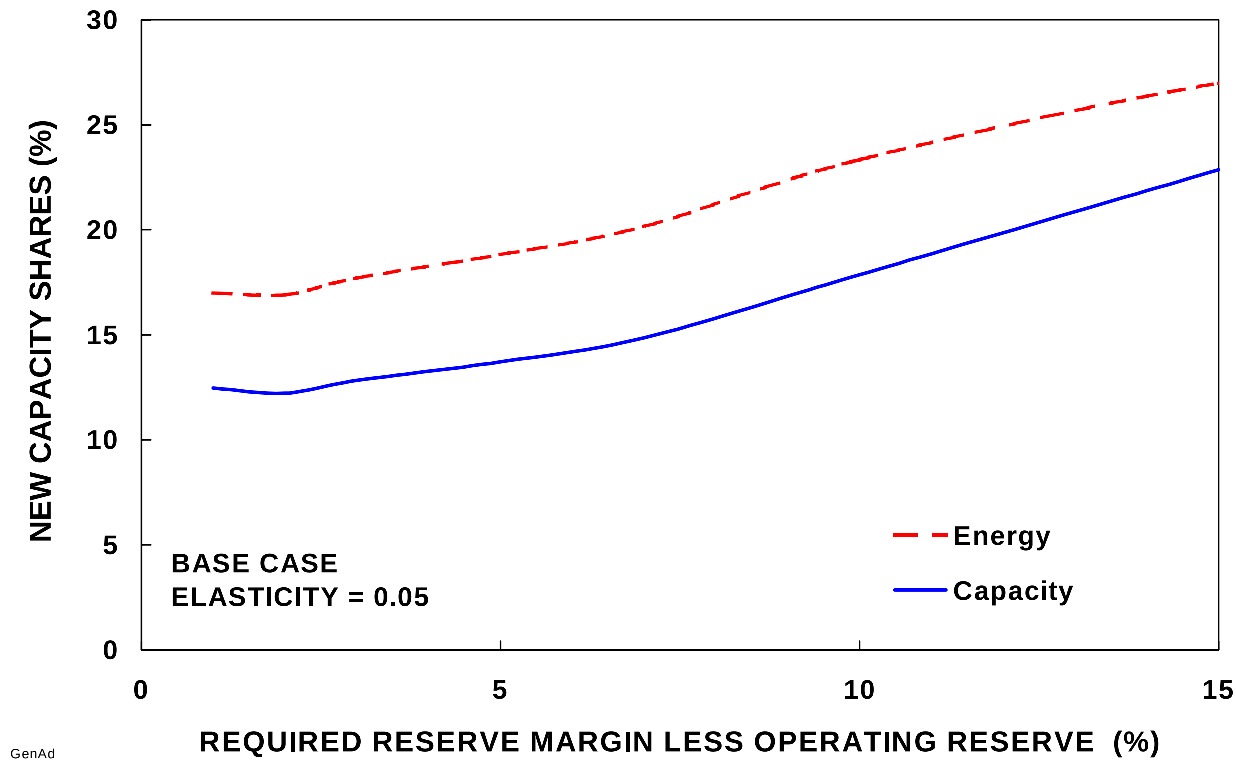


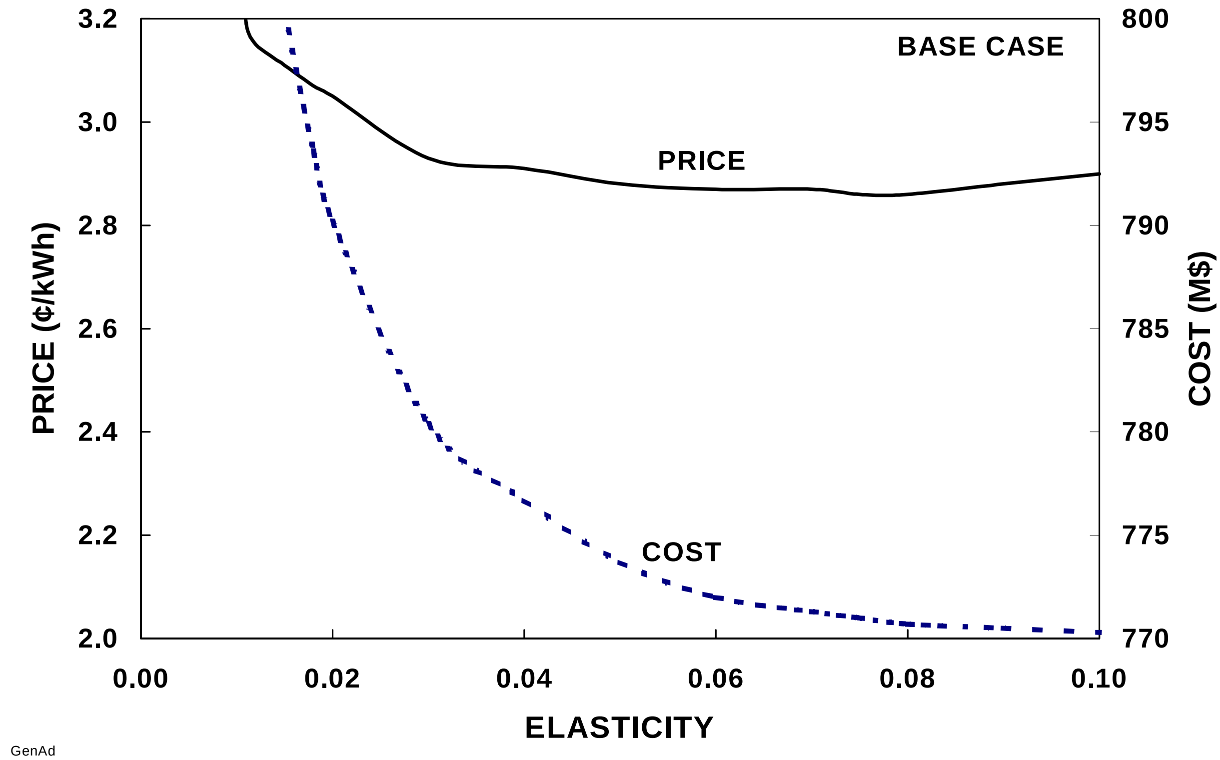
Fig. 9. **ORCED results showing electricity prices and costs (top) and percentages of total capacity and energy from new generation (bottom) as functions of required reserve margins.**

Electricity costs and prices are high at very low values of reserve margin because the amount and cost of unserved energy are high. At the other end of the spectrum, prices and costs are high because of all the "excess" capacity that must be supported through the capacity payment.

As the specified reserve margin is increased, the amount of new generating capacity brought online increases (bottom of Fig. 9). For these cases, all the new capacity is gas fired. Because this new capacity is very efficient, it accounts for larger shares of energy production than of generating capacity. For example, at the "optimal" 5% reserve margin, new gas-fired generation accounts for 14% of generating capacity and 19% of energy production. [Once again, these results are roughly consistent with EIA's (1998) forecasts, which show that about 95% of new generation capacity additions are gas fired.] For reserve margins of 10% or below, all the new capacity consists of combined-cycle units; at higher values of reserve margins, some of the new capacity is combustion turbines. At high values, the capacity factor for generation declines. For low-capacity-factor operation, combustion turbines are a more economical choice than combined-cycle units because of their low capital costs (offset by their higher operating costs). Thus, mandating high reserve margins increases the share of peaking units (and lowers the share of baseload units) in the mix of generating capacity.

We then ran ORCED with no required reserve margin but with different values for the consumer response to price changes when unconstrained demand would otherwise exceed supply (the unserved-energy elasticity discussed above). The results show substantial benefits from encouraging at least a minimal level of consumer response to high prices (Fig. 10). An increase in elasticity from 0.01 to 0.04 cuts costs and prices by 7% and 15%, respectively. The results also show that, beyond an elasticity of about 0.04, there is little additional benefit to greater consumer response to price changes. (A price elasticity of 0.04 could occur if, for example, 20% of the total load had an unserved-energy elasticity of 0.20, which shows that only a small fraction of load needs to be responsive to real-time pricing.) At higher levels of elasticity, costs and prices decline very slowly. As demand elasticity increases, the maximum spot price and price volatility decrease. As the elasticity increases from 0.02 to 0.05 to 0.10, the price of unserved energy (the market price of electricity at times when supplies are not sufficient to meet *unconstrained* demand) declines from 339¢/kWh to 38¢/kWh to 13¢/kWh (bottom of Fig. 10). At elasticity levels above 0.04, the costs and prices are lower in these cases than for any of the required-reserve-margin cases except for the optimal value of 5% (at which point, the two methods of setting reserves yield the same results).

The reserve margins chosen by ORCED decrease as elasticity increases (bottom of Fig. 10). The reserve margin is about 5% at an elasticity of 0.05, consistent with the results shown in Fig. 9.



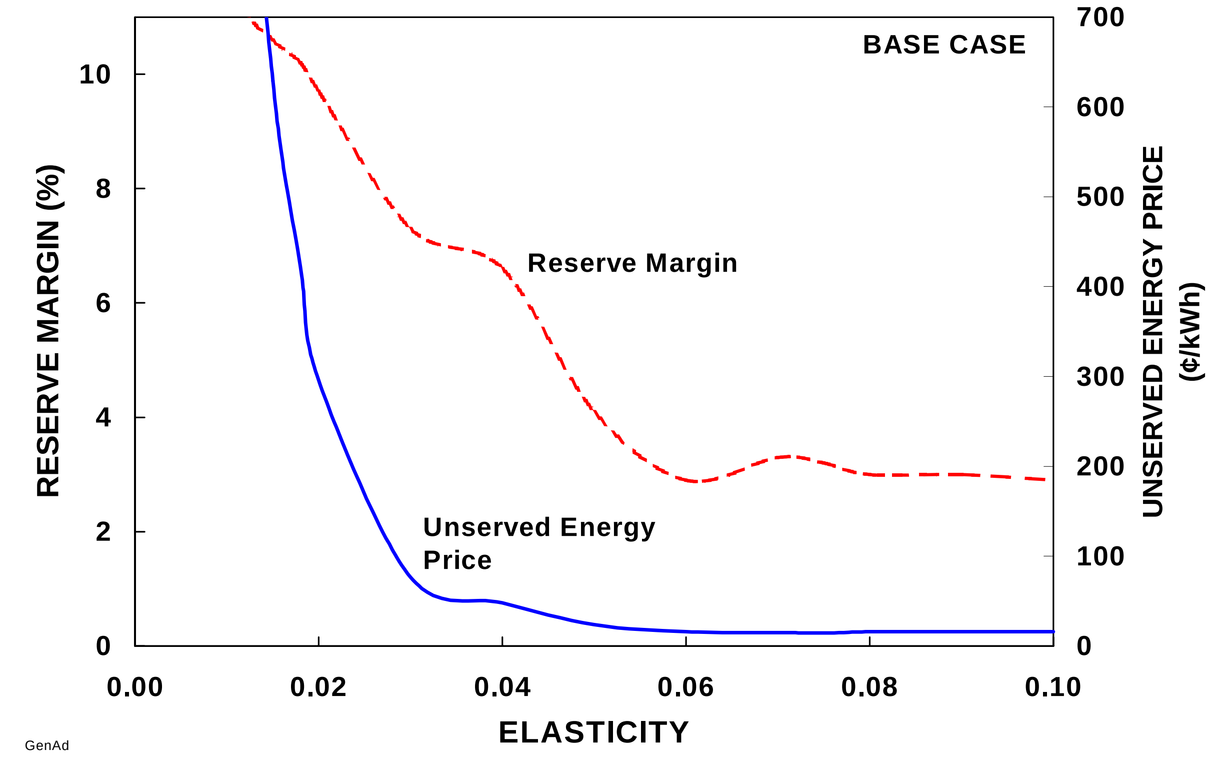


Fig. 10. **ORCED results showing electricity prices and costs (top) and market-determined reserve margins and unserved-energy prices as functions of the unserved-energy price elasticity.**

HIGHER NATURAL GAS PRICES

The base-case scenario used coal and natural gas prices of \$1.39 and \$2.55/MBtu, respectively. In the present scenario, we doubled the price of natural gas, to \$5.10/MBtu. We focus on natural gas prices, rather than coal prices, because 100% of the new generation in the base case is fueled by natural gas.

Once again, we ran several cases with required reserve margins ranging from 1 to 15% (Fig. 11). Qualitatively, the results are quite similar to those discussed above. The total cost and price of electricity are lowest over a range of required reserve margins around 7%, slightly higher than in the base case. As the reserve margin either decreases or increases away from the optimum, costs and prices increase.

These results differ from the base-case results quantitatively. Both electricity costs and prices are substantially higher because gas prices are doubled. Electricity costs are about 15% higher, and prices are almost 50% higher. The much higher percentage increase in price than cost is a consequence of the assumed marginal-cost pricing.

As the required reserve margin increases, the amount of new generating capacity that is constructed also increases. However, unlike the base cases, these cases show a substantial amount of new coal-fired capacity. The amount of new coal capacity remains the same across all the reserve-margin cases analyzed here. Any additional new capacity (beyond that provided by the coal units) comes from gas-fired units. At reserve margins below 15%, all the new gas-fired capacity consists of combined-cycle units; some combustion-turbine capacity is installed when the reserve margin reaches 15% (unlike the base case, in which combustion turbines were constructed when reserve margins reached 10%). At a required 5% reserve margin, new coal accounts for about 50% of the new capacity and almost 60% of the new energy.

The cases with no required reserve margin but with different values of the unserved-energy elasticity are also similar qualitatively to those presented above for the base case (Fig. 12). Once again, slight increases in the unserved-energy elasticity of demand yield substantial cost and price reductions. However, as above, increases in elasticity beyond about 0.03 provide little additional benefit. Once again, electricity prices and costs are lower for all cases with an elasticity greater than 0.03 than for any of the specified-reserve-margin cases, except for the optimal value of 7.5% (where the two methods of setting reserves yield the same results).

Figure 12 also shows the reserve margins chosen by ORCED as a function of the unserved-energy demand elasticity. As in the base case, the reserve margin drops rapidly as elasticity increases from 0.01 to 0.03 and then decreases more slowly. In a similar fashion, the price of unserved energy also drops very rapidly as the elasticity increases from 0.01 to 0.03 and then declines very slowly thereafter. Once again, these results are quite similar to those obtained in the base case.

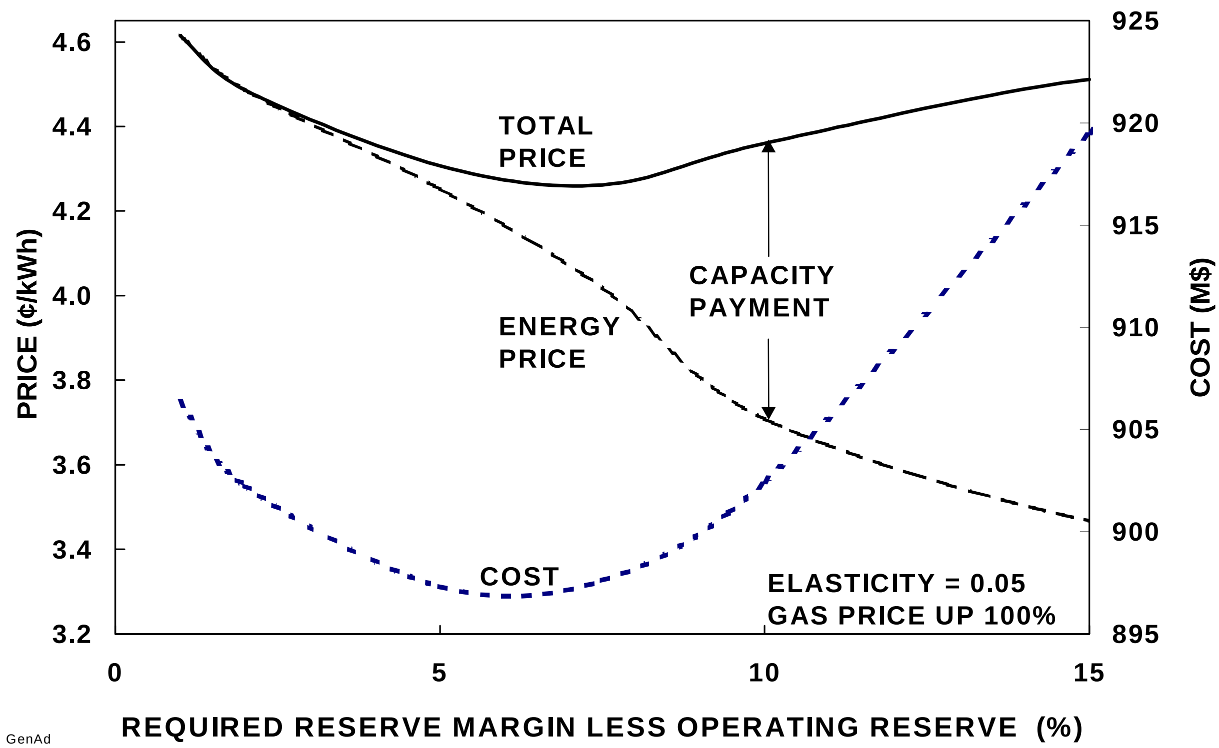


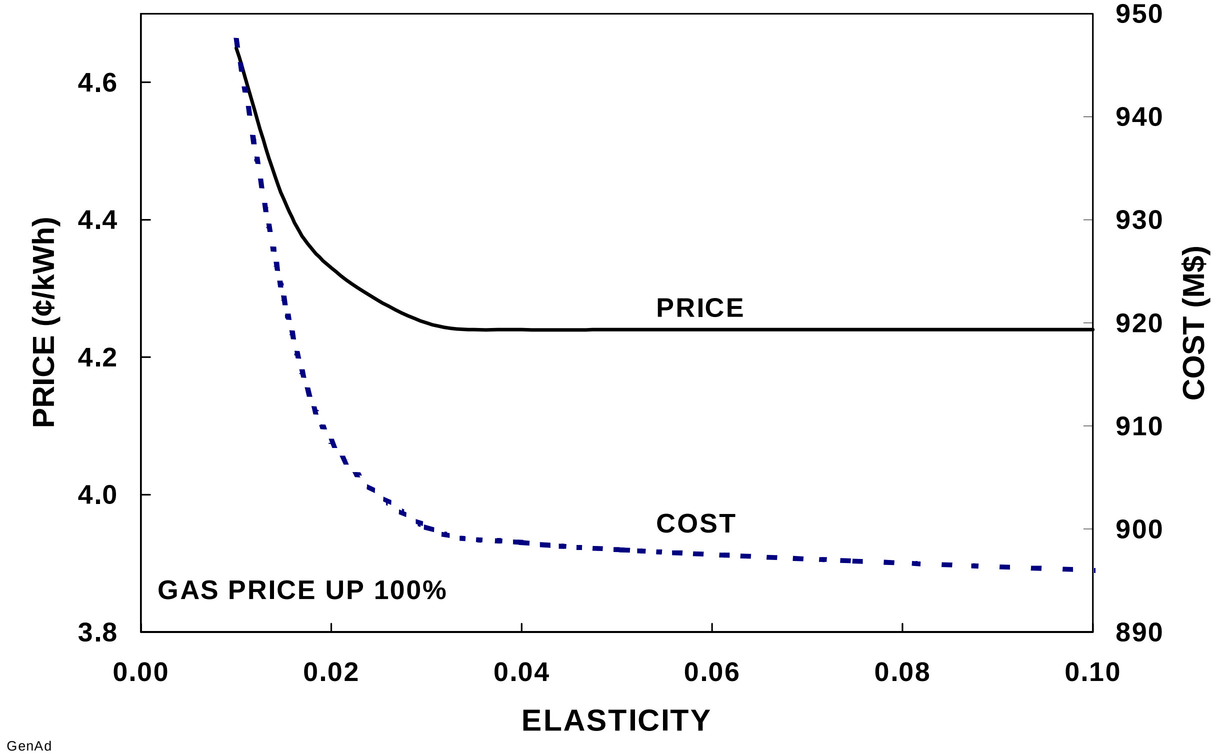
Fig. 11. **Electricity prices and costs as functions of required reserve margins when natural gas prices are double those in the base case.**

Overall, the scenario with higher natural gas prices yields results very similar to those obtained in the base-case scenario. This finding holds for both sets of cases, those with specified reserve margins and those with reserve margins determined by the unserved-energy elasticities.

LOWER SYSTEM LOAD FACTOR

The two scenarios discussed above had a system load factor of 63% (7000-MW peak demand and 38,600-GWh energy). This scenario has a load factor of 55% (the same 7000 MW peak but a 16% decline in energy use to 32,300 GWh). We tested a different load factor because the economics of more or less generating capacity should depend strongly on the amount of energy that can be sold from that base of installed capacity.

As before, we ran two sets of cases, one with a required reserve margin ranging from -10% to +10% and the second with an unserved-energy elasticity ranging from 0.01 to 0.10. Qualitatively, the results are similar to those obtained for the two preceding sets of cases. But quantitatively, the optimum required reserve margin is much lower—indeed, negative—when the system load factor is reduced. The much lower load factor yields an optimum reserve margin of about −5% (Fig. 13). The negative load factor occurs because peak demands happen



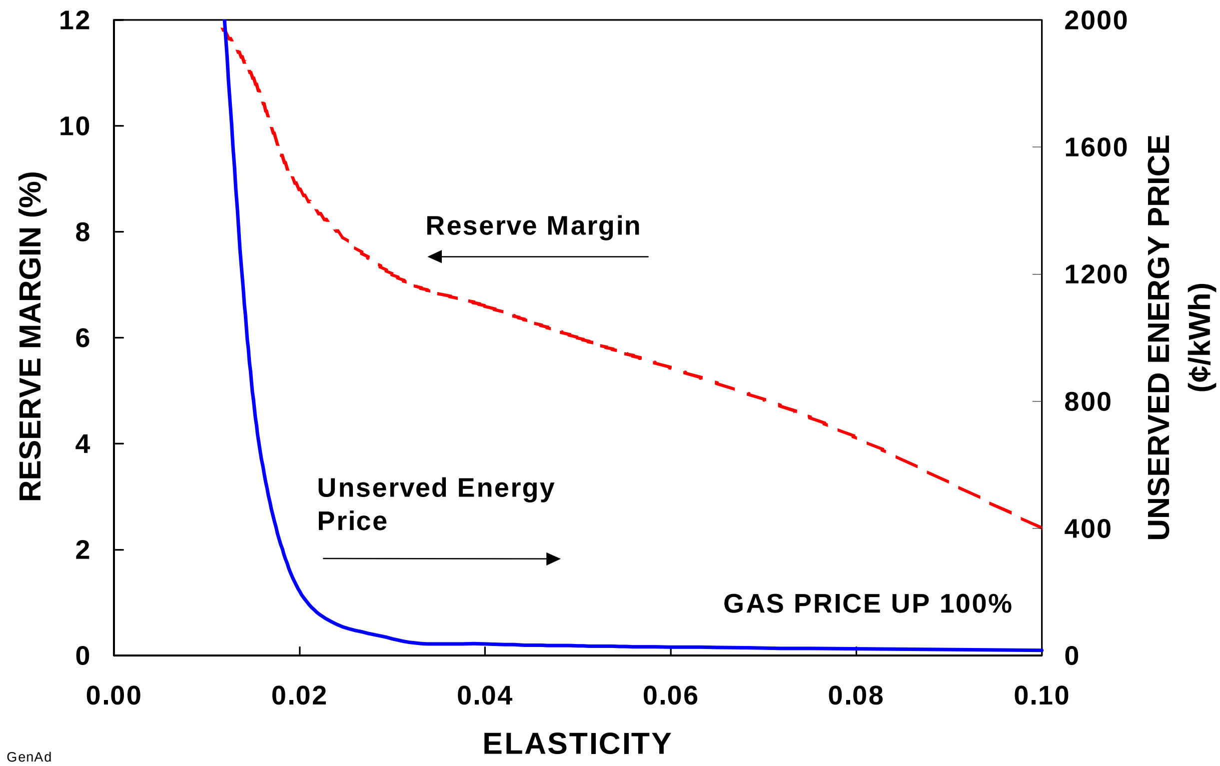


Fig. 12. **Electricity prices and costs (top) and market-determined reserve margins and unserved-energy prices as functions of the unserved-energy elasticity when natural gas prices are double those in the base case.**

for so few hours that it is cheaper to pay unserved-energy charges than to pay capacity charges for generating units that are rarely used.

Total electricity costs are lower than in the base case because electricity consumption is down 16%, but prices are slightly higher because more installed capacity is required per unit of energy use.

As in the base case, the new capacity is all gas fired. And, as in the base case, until the required reserve margin reaches a minimum value (7% here and 10% in the base case), all the new capacity consists of combined-cycle units. At higher required reserve margins, some of the new capacity is combustion turbines.

The cases with different values of unserved-energy elasticity are also qualitatively similar to those presented above (Fig. 14). As the elasticity increases from 0.01 to about 0.05, costs and prices drop sharply; thereafter, costs and prices decrease only slightly with increasing elasticity. As noted above, prices are higher than in the base case because the same amount of generating capacity is supported by 16% less energy. Once again, the costs and prices are lower for all cases with elasticity of 0.05 and above than for any except the optimal required-reserve-margin cases.

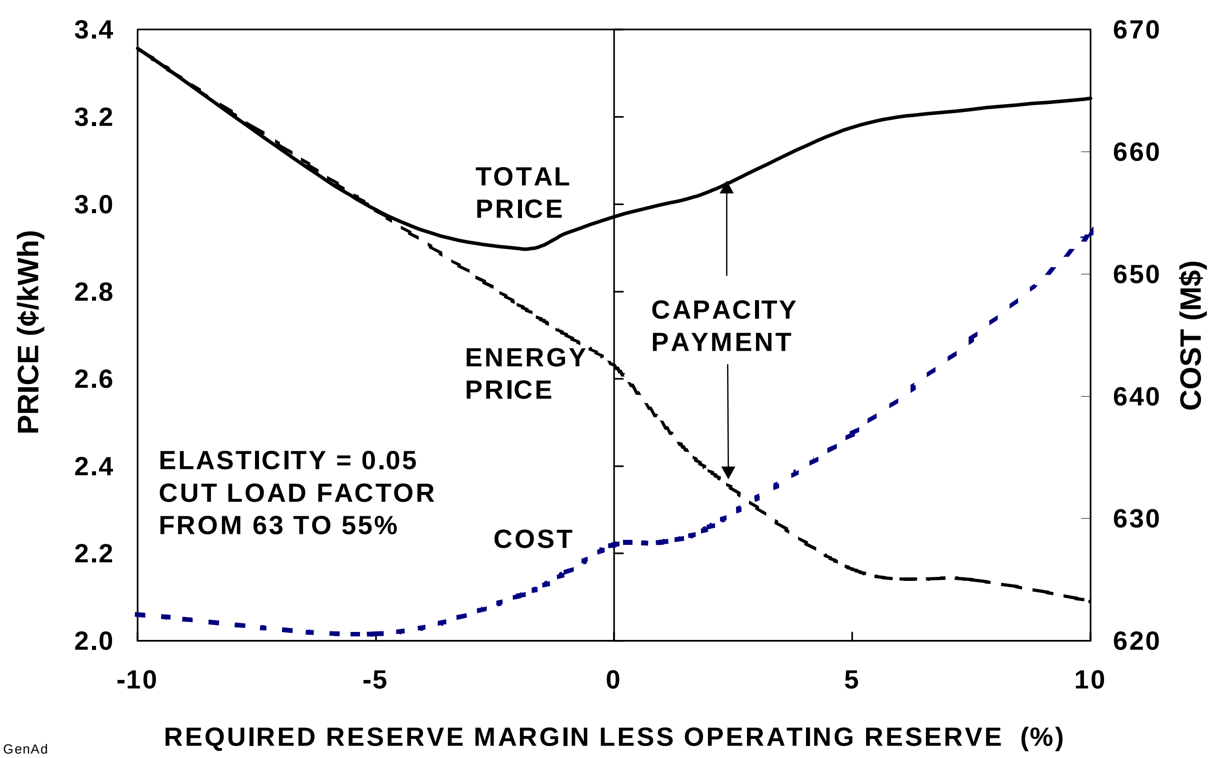
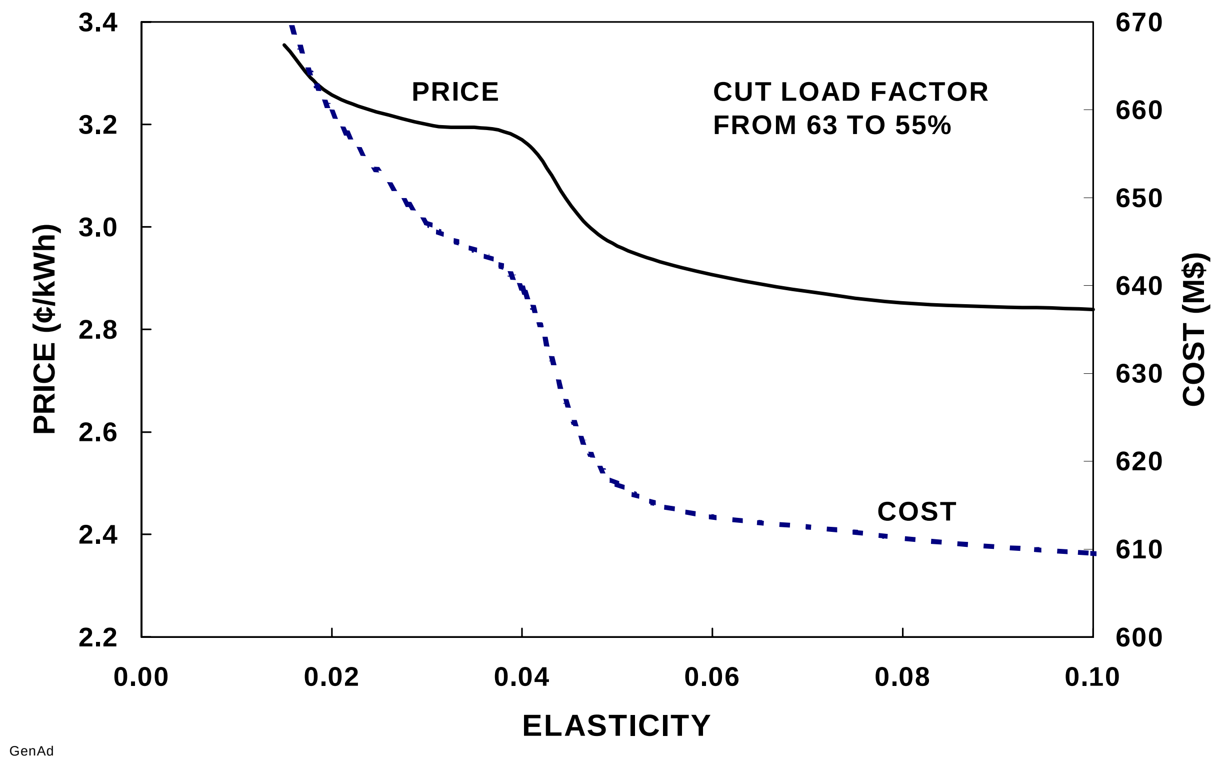


Fig. 13. **Electricity prices and costs as functions of required reserve margins when the system load factor is cut 16%.**



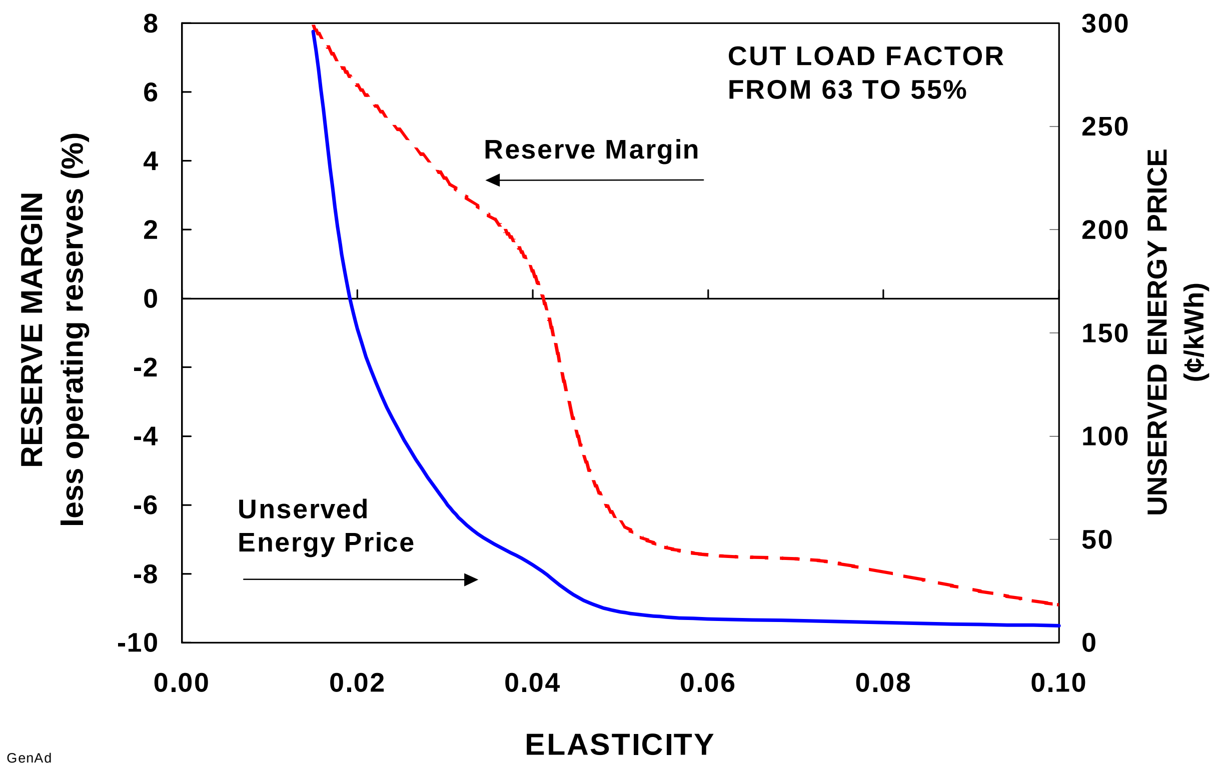


Fig. 14. **Electricity costs and prices (top) and market-determined reserve margins and unserved-energy prices (bottom) as functions of unserved-energy elasticity with the system load factor cut 16%.**

Figure 14 shows the reserve margins chosen by ORCED as a function of the unserved-energy elasticity. Again, the shape of the curve is similar to those shown above. However, consistent with the results in Fig. 13, the "optimal" reserve margin is negative for elasticities greater than 0.04. Finally, the unserved-energy price drops very sharply as the unserved-energy elasticity increases, from almost $3/kWh at an elasticity of 0.015 to 8¢/kWh at an elasticity of 0.10.

CUSTOMER RESPONSE TO REAL-TIME PRICING

The three scenarios considered so far limited customer response to changing prices only when unconstrained demand was higher than available supply (operating through the unserved-energy elasticity factor). Given the importance of customer response to real-time pricing, we developed this scenario to expand the range of customer responses to price changes. In addition to the unserved energy response analyzed above, we allow customers to respond to time-of-use pricing (the second elasticity discussed in Chapter 3). In these cases, customers adjust their hour-to-hour demands in response to changes in hourly prices. To simplify these cases, we set the time-of-use elasticity equal to double the unserved-energy elasticity.[*] Figure 15 shows how increasing the time-of-use elasticity flattens the load-duration curve. In this example, the peak demand drops 11% from 7000 MW (the original value when this elasticity factor equals zero) to 6260 MW with an elasticity of 0.12) with no change in energy consumption.

As before, we ran two sets of cases. But here we ran each case twice, once with the original LDC and a second time with the LDC modified on the basis of changes in hourly electricity prices. ORCED reoptimized the mix of generating capacity on the basis of the modified LDC. The modified LDCs lowered peak demands by 5 to 10% for the cases with a specified planning-reserve margin and by up to 15% for the cases with different price elasticities. Because peak demands are lower, the amount of extra capacity installed to meet a set reserve margin is less than in the base case. For example, a 5% reserve margin translated into a total of 7350 MW of installed capacity in the base case but only 6920 MW with a time-of-use elasticity of 0.10.

For the cases with a specified planning-reserve margin, the pattern of results is similar to that obtained in the base case (compare Figs. 9 and 16). Compared with the results in the base case, costs and prices are lower at reserve margins equal to and greater than the "optimum." These differences are consistent with the notion that additional customer response to real-time pricing should lower costs and prices. The difference between the total and energy-only price (i.e., the installed capacity payment) is much smaller in these cases because the flatter LDC leads to greater use of generators with high costs. This greater use permits them to recover more of their fixed costs from energy charges and, therefore, requires a smaller capacity

---

[*]The price elasticity of demand should increase as the time for adjustment grows. We do not know whether the ratio of elasticities is two to one (as assumed here), higher, or lower, but we believe it is greater than one.

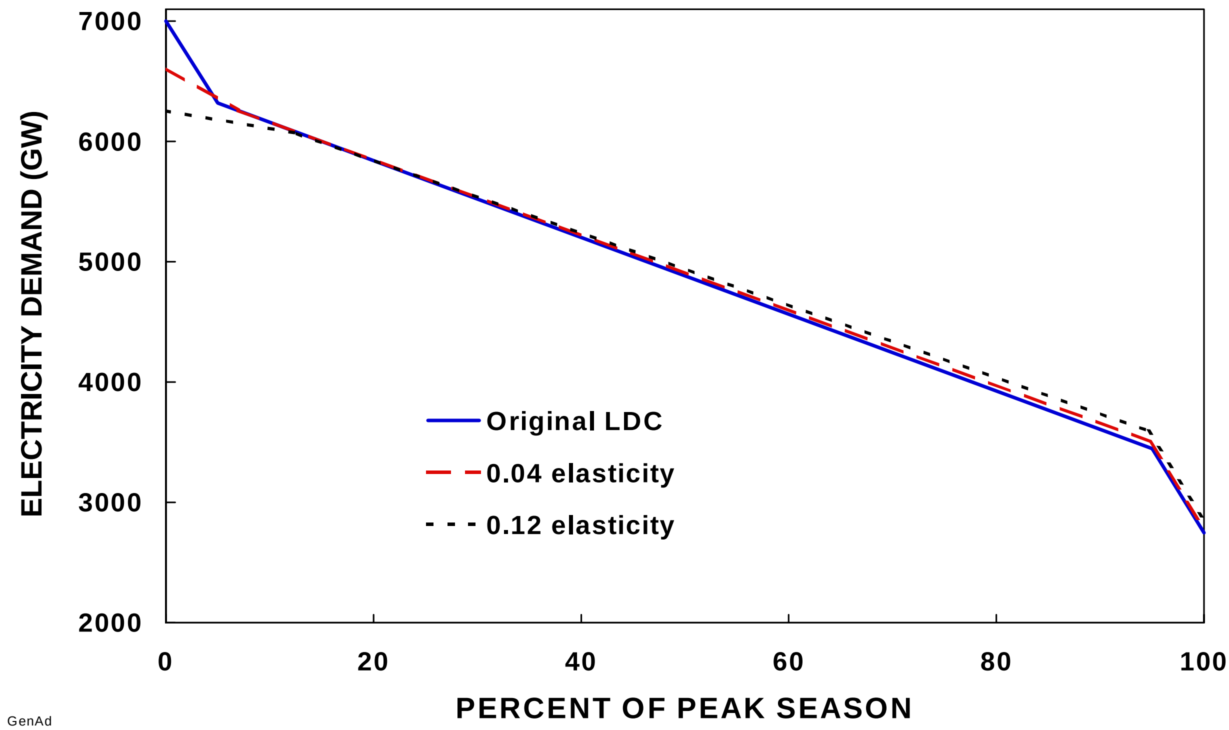


Fig. 15. **Load-duration curves for the peak season as functions of the time-of-use elasticity.**

payment for them to break even. For example, at a 7% reserve margin, the required capacity payment is $4/kW-year compared with $17 in the base-case scenario.

The cases with no required reserve margin and varying time of use and unserved-energy elasticities yielded results very similar to those in the base case; compare Figs. 10 and 17. Costs are consistently lower than in the base-case scenario; prices are very close to the base-case values. Once again, comparisons are complicated by the changes in peak demand. As the unserved-energy elasticity increases, peak demand decreases, from 7000 MW at zero elasticity to 6610 MW at an elasticity of 0.02, to 6310 MW at an elasticity of 0.05 and to 5950 MW at an elasticity of 0.10. Once again, the flatter LDCs lead to retirement of very expensive plants and greater use of the remaining plants. For example, in the base case, with an unserved-energy elasticity of 0.06, the O&M adder is $0.54/kW-year, equivalent to 0.63¢/kWh (with this unit running 86 hours during the year). In the present scenario, the same case yields an O&M adder only half the base-case-scenario value ($0.26 vs $0.54/kW-year). Because this unit runs for 117 vs 86 hours, the energy equivalence is only 0.22¢/kWh, one-third the base-case value.

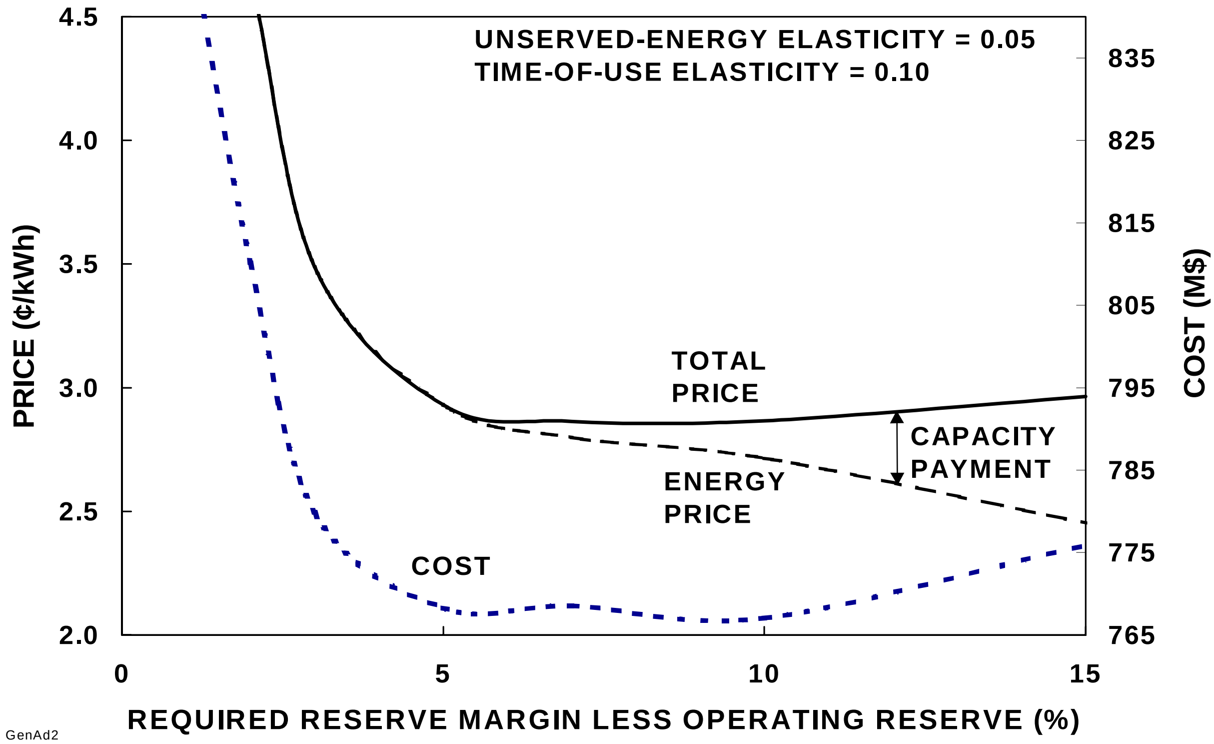
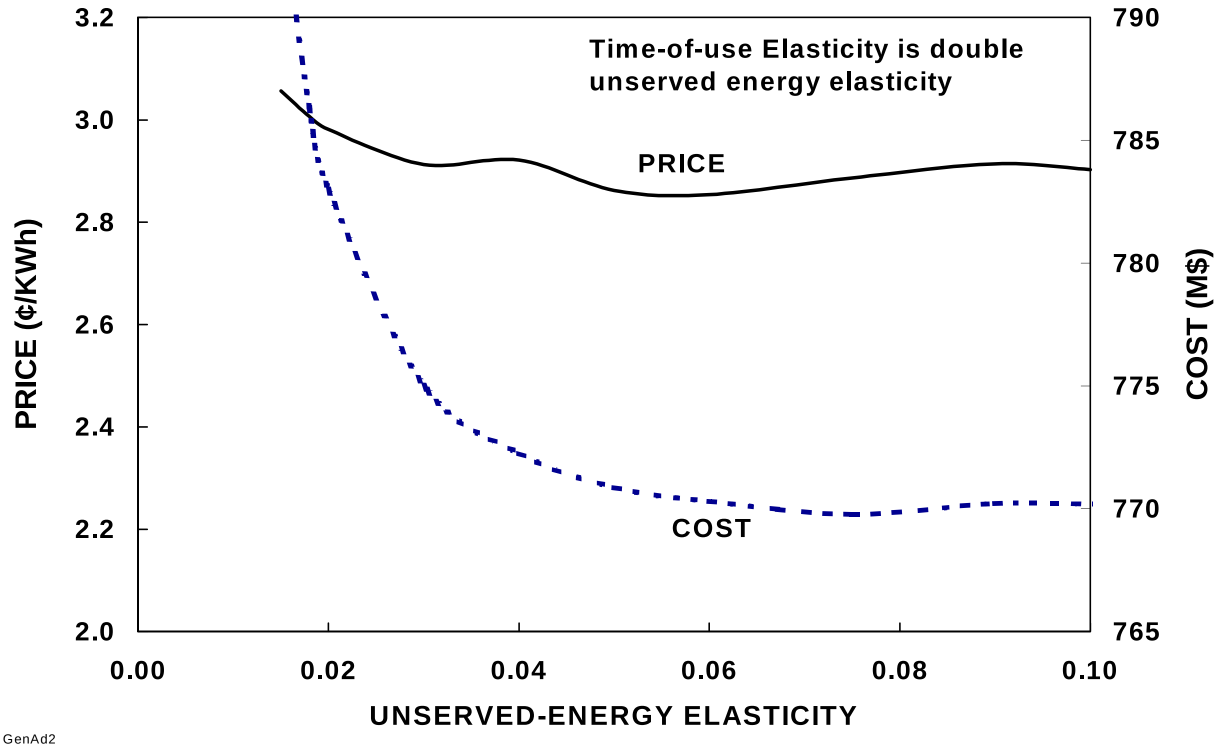


Fig. 16. **Electricity prices and costs as functions of required reserve margins with the time-of-use elasticity equal to 0.10.**

As with the earlier cases, the optimal reserve margin declines as the elasticity increases (bottom of Fig. 17). However, in these cases, the optimal reserve margin is roughly 10% for all values of unserved-energy elasticity above about 0.05. Of course, the amount of installed capacity continues to decline as the elasticity increases because, as discussed above, peak demand declines with increasing elasticity.

REQUIRED RESERVE MARGIN WITH A LOWER ELASTICITY

Because the magnitude of price elasticity affects the costs and prices associated with different levels of required reserve margins, we ran one additional scenario. Here we set the unserved-energy elasticity equal to 0.02, substantially lower than the 0.05 used in all the other scenarios with required planning-reserve margins.



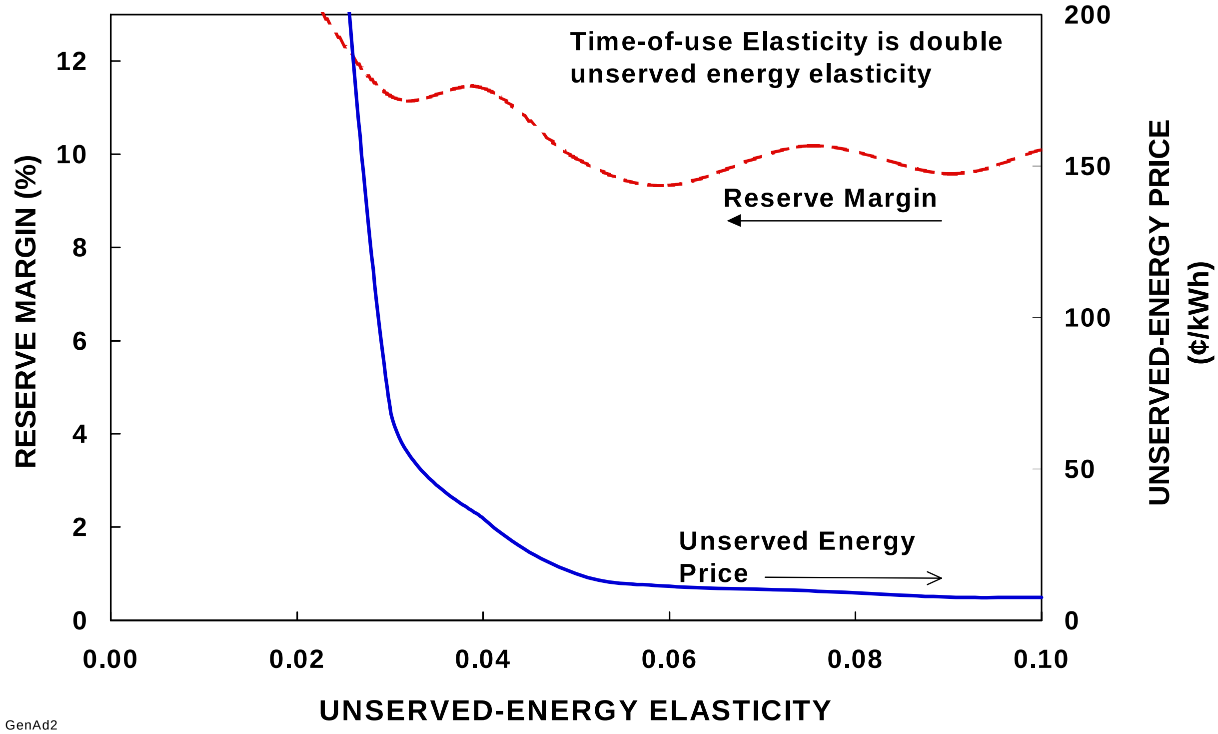


Fig. 17. **Electricity prices and costs (top) and market-determined reserve margins and unserved-energy prices (bottom) as functions of unserved-energy elasticity when the time-of-use elasticity is double the unserved-energy elasticity.**

As expected, both prices and costs are higher than in the base case. In addition, the optimal reserve margin is almost 10% instead of the 5% found in the base case (compare Figs. 18 and 9). These results suggest, once again, that the "correct" value for installed capacity depends strongly on consumer response to time-varying prices. Consistent with results presented above, unserved-energy prices are much higher with an unserved-energy elasticity of 0.02 than an elasticity of 0.05.

SUMMARY

A review of Figs. 9 through 18 shows how consistent results are across the five scenarios discussed in this chapter. Depending on the particular scenario, the optimal value of the mandated planning-reserve margin ranged from −5% to +9%. For any given scenario, costs and prices increase as the mandated reserve margin increases; Table 1 shows these costs for a 5-percentage-point increase in required reserves. The table also shows the percentage increase in costs and prices for the cases with unserved-energy elasticity equal to 0.02, relative to those at 0.05, to illustrate the value of consumer response to real-time prices.

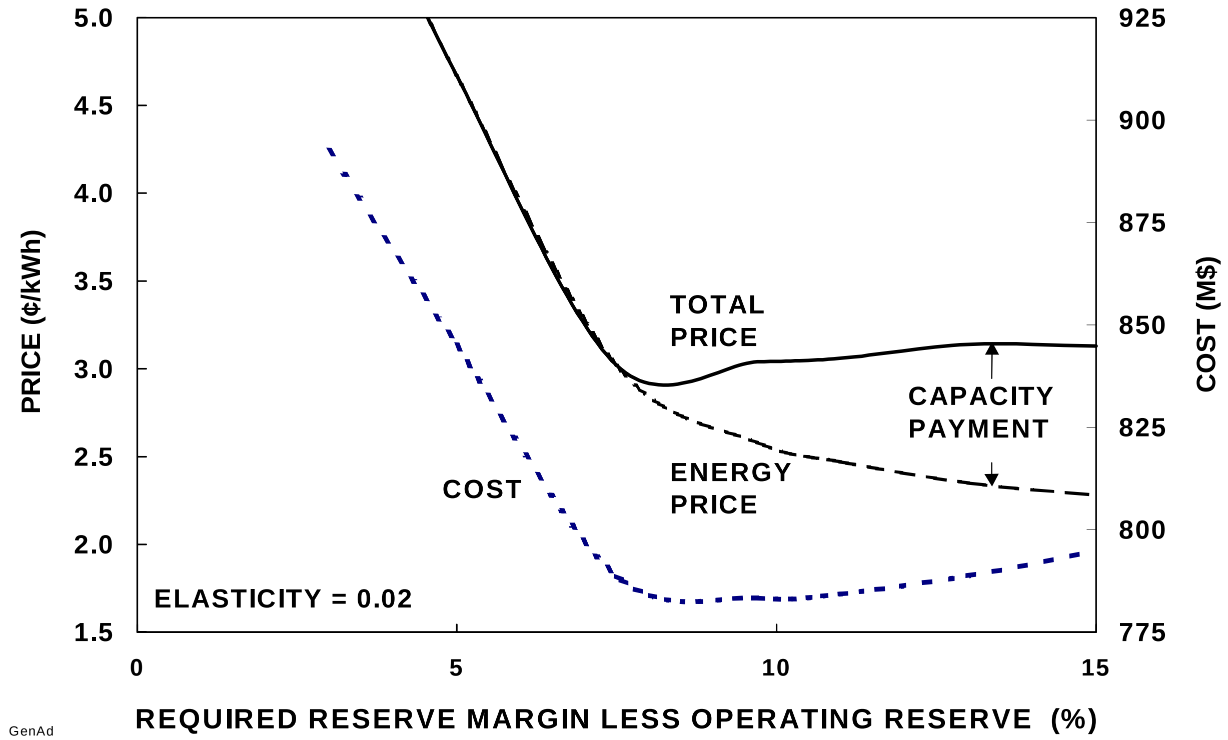


Fig. 18. Electricity prices and costs as functions of required reserve margins when the unserved-energy elasticity is 0.02.

Table 1. Percentage change in electricity costs and prices for four scenarios

	Percentage increase for reserve margin +5% points relative to optimal reserve margin		Percentage increase for elasticity of 0.02 relative to 0.05	
	Cost	Price	Cost	Price
Base case	1	5	2	6
Gas prices up 100%	1	1	1	2
Load factor down 16%	1	0	7	10
Add time-of-use elasticity	0	-2	2	4
Unserved-energy elasticity = 0.02	2	3	–	–

CONCLUSIONS

During the lengthy, awkward, and difficult transition from a highly regulated, retail-monopoly-franchise structure to a competitive and deintegrated structure, maintaining appropriate levels of installed generating capacity may be difficult.

Perhaps the key generation-adequacy problem is the absence of a demand-side response to real-time pricing. Economic theory suggests that consumers and suppliers, in response to real-time prices, will take appropriate steps to ensure generation adequacy. But, if most retail consumers continue to face traditional tariff prices that have little or no temporal variation, this approach will be short-circuited. In addition, customers must have the technical ability (including metering, communications, and computing systems), as well as the economic incentive, to respond quickly to changes in energy prices. Real-time pricing should stimulate the use of distributed supply resources as well as customer modification of loads. Until real-time pricing is available to at least some retail customers, traditional approaches to maintaining generation adequacy may be needed.

A second critical factor is creation of efficient, competitive spot markets for energy. These markets need to be integrated with those for ancillary services and transmission. And they need to accurately reflect the intrahour costs of energy (including startup, ramping, and shutdown costs) when system conditions are changing rapidly from minute to minute.

We discussed two primary approaches to ensuring that enough generating capacity is available so that customers will not be *in*voluntarily disconnected from the grid. One approach stresses reliability and calls for continuation of required minimum planning reserve margins. The other focuses on economic efficiency and the use of competitive markets to balance demand and supply. Requiring minimum planning reserves (1) ensures that "enough" generation will exist; (2) uses an approach that worked well in the past; (3) reduces the volatility of electricity prices; (4) protects customers from such volatility; and (5) ensures that the positive externalities associated with extra generating capacity are maintained. On the other hand (1) the amount of generating capacity that is "enough" depends on customer response to prices, which varies across customers and customer classes; (2) because an approach worked well in a vertically integrated, monopoly-franchise industry is no assurance that it will work well in a deintegrated and competitive industry; (3) price volatility sends important economic signals to consumers and producers concerning when and how much electricity to consume and produce; (4) only a small fraction of loads needs to be price-sensitive to equilibrate demand and supply and to eliminate the need for mandated planning reserves; and (5) no public benefits are associated with generation adequacy beyond the private benefits.

In addition, the market proponents favor market decisions on generation investment because it (1) lets markets make both capital and operating decisions, which is why we are creating competitive electricity markets in the first place; (2) encourages customer participation in the provision of reliability services; and (3) will yield lower average electricity prices and costs to consumers. The reliability proponents respond that (1) reliability is in part a public good that cannot be left entirely to the self-interests of market participants; (2) there is too much uncertainty about whether, when, and by how much customers will reduce demand in response to high spot prices to consider demand-side responses a resource for reliability; and (3) the costs of major outages are so high that they wipe out any savings associated with lower electricity prices.

The ORCED analyses suggest that centralized decisions concerning the amount of generating capacity to build and maintain may often be wrong, yielding either too much capacity or not enough. So many factors affect the "optimal" amount of generating capacity (e.g., the prices of fossil fuels, the amount and types of generating capacity already online, and customer load shapes and price elasticities) that it is difficult for central planners to make the right choice. For example, the optimal reserve margin ranged from −5% to +10% across the scenarios analyzed in Chapter 4. Imposing a required reserve margin above what markets would voluntarily produce increases electricity costs and prices. However, the ORCED results show broad optima for each of the five scenarios, with a range of several percentage points around the true optimum.

On the other hand, relying on the actions of consumers and suppliers in response to time-varying spot prices works well only if (1) consumers can and do respond to high prices and (2) market administrators ensure that intrahour prices accurately reflect costs. Therefore, electricity policymakers should encourage demand-side experiments and investments to ensure that, when prices rise, customers will be able to respond. And FERC should include in its final rule on regional transmission organizations additional detail on its proposed requirement for real-time balancing markets.

Our key findings and conclusions are:

- Generation-capacity margins have been declining for at least a decade, and utility plans show continued declines.

- Whether these declines in generation adequacy reflect increased productivity or shortfalls in reliability is unclear. It is clear, however, that the transitional state of the U.S. electricity industry (half competitive and half regulated) leads to tremendous uncertainty, which may limit investments in long-lived assets, such as generating units.

- Independent power producers plan to build large amounts of new generation capacity throughout the country during the next few years.

- Generation adequacy could be maintained in competitive electricity markets in one of two ways: (1) sole reliance on markets acting through time-varying spot prices or (2) continuation of the historical practice of setting minimum requirements on installed capacity that must be met by all load-serving entities.

- Market-based methods for generation expansion seem, both to us and to most of the people we talked with, the preferred long-term approach.

- Only a very small fraction of loads needs to respond to real-time prices for this approach to work well in maintaining generation adequacy.

ACKNOWLEDGMENTS

We thank our EPA project managers, James Turnure (now with the Federal Energy Regulatory Commission) and Tracy Terry (now with the U.S. Department of Energy) for their ongoing encouragement and guidance. We thank Bill Eastlake, Steve Huntoon, Pierre Landrieu, Kenneth Linder, Mark Reeder, Tracy Terry, and James Turnure for their very helpful comments on the draft of this report. We thank Fred O'Hara for editing this final report.

REFERENCES

California Independent System Operator 1998, *1998 Transmission Reliability Report*, Folsom, CA, July.

California Independent System Operator 1999, *Annual Report on Market Issues and Performance*, Folsom, CA, June.

P. Cramton and R. Wilson 1998, *A Review of ISO New England's Proposed Market Rules*, ISO New England, Holyoke, MA, September 9.

Electric Power Daily 1999a, "Marketers Bristle at Cinergy Use of *Force Majeure*, See Damage to Market Integrity," McGraw-Hill, August 4.

Electric Power Daily 1999b, "FERC Asked to Extend ISO-New England's Authority to Alter Prices, Add New Caps," McGraw-Hill, August 12.

Electric Power Supply Association 1999, *Comments on Regional Transmission Organizations*, Docket No. RM99-2-000, submitted to the Federal Energy Regulatory Commission, Washington, DC, July 27.

Energy Information Administration 1998, *Annual Energy Outlook 1999, With Projections to 2020*, DOE/EIA-0383(99), U.S. Department of Energy, Washington, DC, December.

Energy Information Administration 1999, *Electric Power Annual 1998, Volume 1*, DOE/EIA-0348(98)/1, U.S. Department of Energy, Washington, DC, April.

F. C. Graves, E. G. Read, P. Q. Hanser, and R. L. Earle 1998, *One-Part Markets for Electric Power: Ensuring the Benefits of Competition*, The Brattle Group, Cambridge, MA, October.

S. Hadley and E. Hirst 1998, *ORCED: A Model to Simulate the Operations and Costs of Bulk-Power Markets*, ORNL/CON-464, Oak Ridge National Laboratory, Oak Ridge, TN, June.

J. E. Hakes 1999, "Electric Power Industry Restructuring: Data Collection, Disclosure and Associated Issues," U.S. Department of Energy, presentation to the Electric Power Supply Association, July 13.

A. Henney 1998, "Contrasts in Restructuring Wholesale Electric Markets: England/Wales, California, and the PJM," *The Electricity Journal* **11**(7), 24–42, August/September.

E. Hirst and B. Kirby 1998, *Unbundling Generation and Transmission Services for Competitive Electricity Markets: Examining Ancillary Services*, NRRI 98-05, The National Regulatory Research Institute, Columbus, OH, January.

E. Hirst, B. Kirby, and S. Hadley 1999, *Generation and Transmission Adequacy in a Restructuring U.S. Electricity Industry*, Edison Electric Institute, Washington, DC, June.

ISO New England 1997, *NEPOOL Market Rules & Procedures*, Holyoke, MA, August 5.

ISO New England 1998, *ISO New England Assessment of the Competitiveness and Effectiveness of the NEPOOL Markets, Exhibit A*, Holyoke, MA, September.

ISO New England 1999, *Request for Expedited Approval of Revisions to NEPOOL Market Rules 1 and 10*, Docket No. ER99-4002-000, submitted to the Federal Energy Regulatory Commission, Holyoke, MA, August 5.

A. B. Jaffe and F. A. Felder 1996, "Should Electricity Markets Have a Capacity Requirement: If So, How Should It Be Priced?" *Proceedings: 1996 EPRI Conference on Innovative Approaches to Electricity Pricing*, EPRI TR-106232, Electric Power Research Institute, Palo Alto, CA, March.

E. Lapson, R. Hornick, S. M. Fetter, and R. Pellecchia 1998, *Electricity Price Spike: Lessons Learned*, Fitch IBCA, New York, NY, October 29.

G. Loehr 1998, "Transmission Reliability: 'Brave New World' or 'Brave New World?'" EUCI Conference on Reliability and Competition, Denver, CO, November 3.

R. J. Michaels and J. Ellig 1998, *Electricity Passes the Market Test: Price Spikes in the Summer of 1998*, Mercatus Center, George Mason University, Fairfax, VA, October.

New York Independent System Operator 1998, *NYISO Manual for Installed Capacity Requirements*, draft, Schenectady, NY, October 16.

North American Electric Reliability Council 1998a, *Reliability Assessment: 1998–2007*, Princeton, NJ, September.

North American Electric Reliability Council 1998b, *Electricity Supply & Demand Database and Software*, Princeton, NJ, October.

PJM Interconnection 1997, *PJM Manual for Accounted-for-Obligation*, Norristown, PA, April 28.

PJM Interconnection 1998a, *Reliability Assurance Agreement among Load Serving Entities in the PJM Control Area*, Norristown, PA, September 15.

PJM Interconnection 1998b, *Schedule 11: PJM Capacity Credit Markets*, Norristown, PA, October 14.

J. L. Rose 1997, "Last Summer's 'Pure' Capacity Prices, A Harbinger of Things to Come," *Public Utilities Fortnightly* **135**(22), 36–41, December.

L. E. Ruff 1999, *Competitive Electricity Markets: Why They Are Working and How to Improve Them*, n/e/r/a, Los Angeles, CA, May 12.

C. Seiple 1999, "The Magic of Greed," *Public Utilities Fortnightly* **137**(4), 10–11, February 15.

S. Slater 1999, personal communication, Electric Power Supply Association, Washington, DC, August.

T. Terry 1999, personal communication, U.S. Department of Energy, Policy Office, Washington, DC, August.

U.S. Federal Energy Regulatory Commission 1998, *Staff Report to the Federal Energy Regulatory Commission on the Causes of Wholesale Electric Pricing Abnormalities During June 1998*, Washington, DC, September 22.

U.S. Federal Energy Regulatory Commission 1999, *Regional Transmission Organizations, Notice of Proposed Rulemaking*, Docket No. RM99-2-000, Washington, DC, May 13.

U.S. Federal Trade Commission 1999, *Comments of the Bureau of Economics, Regional Transmission Organizations*, Docket No. RM99-2-000, submitted to the Federal Energy Regulatory Commission, Washington, DC, August 16.

J. Wilson 1999, Navigant Consulting, personal communication, Washington, DC, September.

F. A. Wolak and R. H. Patrick 1997, *The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market*, draft, Stanford University, CA, and Rutgers University, Newark, NJ, February.

D:\Wpd\Rel\EPAGenAd.wpd