# Retail-Load Participation in Competitive Wholesale Electricity Markets

*Eric Hirst and Brendan Kirby*

January 2001

*Prepared for*

**Edison Electric Institute**
Washington, D.C.

*and*

**Project for Sustainable FERC Energy Policy**
Alexandria, VA

# Edison Electric Institute

Edison Electric Institute (EEI) is the association of U.S. shareholder-owned electric companies, international affiliates and industry associates worldwide. Our U.S. members serve over 90 percent of all customers served by the shareholder-owned segment of the industry. They generate approximately three-quarters of all the electricity generated by electric companies in the country and service about 70 percent of all ultimate customers in the nation.

Organized in 1933 and incorporated in 1970, EEI works closely with its members, representing their interests and advocating equitable policies in legislative and regulatory arenas. In its leadership role, the Institute provides authoritative analysis and critical industry data to its members, Congress, government agencies, the financial community and other influential audiences. EEI provides forums for member company representatives to discuss issues and strategies to advance the industry and to ensure a competitive position in a changing marketplace.

EEI's mission is to ensure members' success in a new competitive environment by:

- Advocating Public Policy
- Expanding Market Opportunities
- Providing Strategic Business Information

For more information on EEI programs and activities, products and services, or membership, visit our Web site at www.eei.org.

To place a publications order, call **1-800-EEI-5453**.

# The Project for Sustainable FERC Energy Policy

The Project for Sustainable FERC Energy Policy is an education and advocacy initiative of state and national public interest organizations. It addresses the regulatory policy options available to the Federal Energy Regulatory Commission (FERC) as it responds to and directs changes in the electric industry. The Project's mission is to secure FERC implementation of policies that help to assure a sustainable energy future – policies that support long-term efficiency, cost minimization, and environmental responsibility in the production and use of electric power. The Project supports policies that advance the long-term public interests in electric system planning, resource acquisition, and grid operations that assure reliability, affordability, and environmental sustainability.

# Retail-Load Participation in Competitive Wholesale Electricity Markets

*Eric Hirst and Brendan Kirby*

January 2001

*Prepared for*

**Edison Electric Institute**
Russell Tucker, Manager, Public Policy Analysis
701 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2696
202-508-5000

*and*

**Project for Sustainable FERC Energy Policy**
Terry R. Black, Director
107 Roberts Court
Alexandria, VA 22314
703-836-9547

# CONTENTS

# LIST OF ACRONYMS

| | |
|---|---|
| AMR | Automated-meter reading |
| CBL | Customer baseline load |
| EMS | Energy-management system |
| FERC | Federal Energy Regulatory Commission |
| ISO | Independent system operator |
| LSE | Load-serving entity |
| PGE | Portland General Electric |
| PJM | Pennsylvania-New Jersey-Maryland Interconnection, LLC |
| PUC | Public utility commission |
| PX | Power exchange |
| RTO | Regional transmission organization |
| RTP | Real-time pricing |
| TOU | Time-of-use |

# SUMMARY

Competitive wholesale markets today resemble the sound of one hand clapping. They are often inefficient and not fully competitive, in part because retail-customer loads do not participate in these markets.

Electricity costs vary substantially from hour to hour, often by a factor of ten within a single day. Because most customers buy electricity as they always have—under time-invariant prices that are set months or years ahead of actual use—consumers are fully insulated from the volatility of wholesale electricity markets.

Retail electricity prices reflect two components that should be priced separately:

- the electricity commodity and
- the insurance premium that protects customers from price variations.

Customers should have the *opportunity* to see electricity prices that vary from hour to hour, reflecting wholesale-market price variations. Permitting customers to face the underlying variability in electricity costs can improve economic efficiency, increase reliability, and reduce the environmental impacts of electricity production. Economic efficiency requires a range of customer choices. Offering customers a variety of pricing options is an essential component of competitive markets and a key to improving customer well-being. Customers who choose to face the volatility of electricity prices can lower their electricity bills in two ways. First, they provide their own insurance. Second, they can modify electricity usage in response to changing prices, increasing usage during low-price periods and cutting usage during high-price periods.

Retail customers who modify their usage in response to price volatility help lower the size of price spikes. This demand-induced reduction in prices is a powerful way to discipline the market power that some generators would otherwise have when demand is high and supplies tight. And these price-spike reductions benefit all retail customers, not just those who modify their consumption in response to changing prices; see Figure S-1.

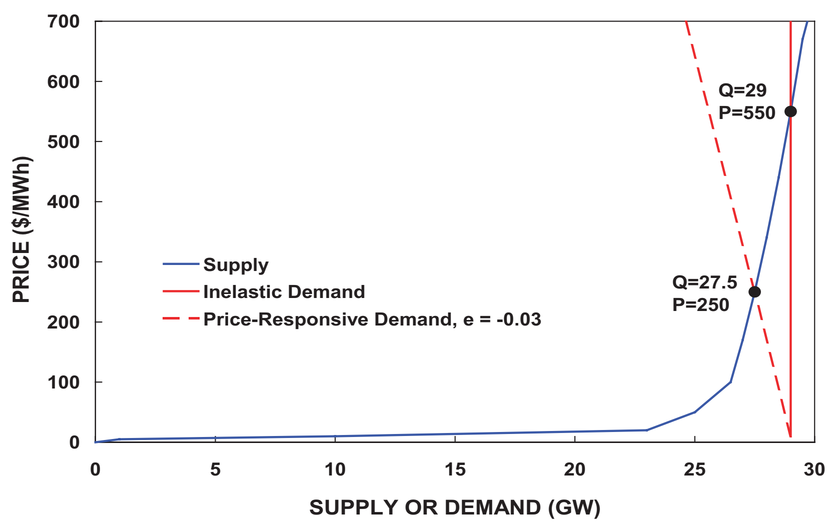



**Figure S-1.** Hypothetical wholesale supply and demand curves. The solid vertical line represents demand that is insensitive to price; the dashed line represents demand that varies with price. For this case, consumer response to price reduces demand by 5 percent and cuts the price by 55 percent. This price-spike reduction benefits all customers, not just those with price-responsive demand.

Customers who face real-time prices and respond to those prices provide valuable reliability services to the local control area. Specifically, load reductions at times of high prices (generally caused by tight supplies) provide the same reliability benefits that the same amount of additional generating capacity would.

Finally, strategically timed demand reductions decrease the need to build new generation and transmission facilities. When demand responds to price, system load factors improve, increasing the utilization of existing generation and reducing the need to build new facilities. Deferring such construction may improve environmental quality. Cutting demand at times of high prices may also encourage the retirement of aging and inefficient generating units.

Ultimately, competitive electricity markets will feature two kinds of demand-response programs. First, some customers will choose to face electricity prices that vary from hour to hour. Typically, these prices will be established in the day-ahead markets run by regional transmission organizations, such as those now operating in California, New York, and the mid-Atlantic region. Second, some customers will select fixed prices, as they have in the past, but voluntarily cut demand during periods of very high prices. In this second option, the customer and the electricity supplier will share the savings associated with such load reductions.

Although the potential benefits of dynamic pricing are large, so too are the barriers to widespread adoption. State legislatures and regulatory commissions have inadvertently blocked customer access to wholesale markets in their efforts to protect retail customers, especially residential consumers, from the vagaries of competition. State regulators need to rethink their decisions on standard-offer rates that are set so low that new suppliers are unable to compete and customers have no incentive to look elsewhere for a better deal. Although regulators should not force consumers to face dynamic pricing, neither should they make it difficult for them to do so. Ultimately, consumers will have to pay for prices that are set too low today.

Metering, communications, and computing technologies are needed to support dynamic pricing and voluntary-load-reduction programs. The cost and bother of designing and installing this infrastructure represents another important barrier to these programs, as is the unfamiliarity with the bewildering array of choices. Fortunately, these technology barriers will fall as more dynamic-pricing programs are implemented.

In summary, the convergence of retail competition, wholesale competition, and improved technologies should greatly expand the type and magnitude of price-responsive demand. Permitting and encouraging retail customers to respond to dynamic prices will improve economic efficiency, discipline market power, improve reliability, and reduce the need to build new generation and transmission facilities.

# Chapter 1: INTRODUCTION

Competitive electricity markets in the United States are under siege. The substantial summer 2000 increases in electricity prices in California, price spikes in other parts of the country, concerns about generator market power in all four operating bulk-power markets (New England, New York, mid-Atlantic [the PJM Interconnection], and California), and many reliability near-misses suggest that electricity competition is not yet working well. Many factors account for these problems, including insufficient construction of new generation and transmission facilities during the past several years, flaws in the design and operation of these new markets, inexperience on the part of many market participants, and the uneven patterns and trends in electric-industry restructuring across the country.

Recently, one potential solution to these problems has received increasing attention—demand-side participation in bulk-power markets. We now recognize, perhaps belatedly, that electricity markets can be competitive and efficient only when both the supply and demand sides participate. As the Federal Energy Regulatory Commission recently noted, "lack of price responsive demand is a major impediment to the competitiveness of electricity markets"(FERC 2000). The California Public Utilities Commission (cited in San Diego Gas & Electric 2000) was more specific:

> The revelation of the real-time price of electricity coupled with a rate alternative that allows the customer to respond intelligently will produce savings for any customer who is able to shift demand from peak to off-peak hours. The potential that many customers will respond to this opportunity to take significant control over the cost of their consumption will produce a collective benefit, in that demand will be redistributed away from the current peaks. Future generation demands will be forestalled even as existing investments in generation are made more productive. The result is a triple win, embracing the individual consumer of any class who is able to reduce costs by shifting load, the society at large which defers the demand for new generation, and investors in existing plant and equipment who see it put to more productive use.

Because the vast majority of U.S. electricity consumers are fully insulated from the price variability of wholesale markets, those markets do not—and cannot be expected to—work as intended. Almost all retail customers continue to pay for electricity at a fixed price determined long before consumption occurs.

Permitting (but not requiring) retail customers to see dynamically varying electricity prices is essential because electricity prices vary dramatically from hour to hour, between weekdays and weekends, and by season. Because of this high variability in wholesale electricity costs and prices, the price we pay for electricity includes two components: (1) the electricity commodity and (2) the insurance premium that protects customers from price variations. Most consumers and many government regulators are unaware of this second component—the risk premium—associated with traditional electricity pricing.

Why should consumers, regional transmission organizations (RTOs), government regulators, and others care whether consumers are permitted to face dynamic electricity prices? The answers fall into three categories: economic efficiency, reliability, and environmental quality. With respect to economic efficiency, the essence of competition is to expand the range of customer choices. Offering customers a variety of pricing options is an essential component of competitive markets and a key to improving customer well-being. Customers who choose to face some or all of the volatility of electricity prices can lower their electricity bills in two ways: (1) by self-insuring and (2) by shifting electricity use away from high-price periods to low-price periods. Retail customers who modify their usage in response to prices reduce price volatility by lowering the magnitudes of price spikes. This demand-induced reduction in price spikes is a powerful way to discipline the market power that some generators would otherwise have during periods of peak demand and constrained supply.[1] And these reductions in price spikes benefit all retail customers, not just those who modify their consumption in response to changing prices.

Customers who choose to face real-time prices and respond to those prices provide valuable reliability services to the local control area. The North American Electric Reliability Council (2000) noted that to "improve the reliability of electric supply, some or all electric customers will have to be exposed to market prices." Specifically, load reductions at times of high prices (generally caused by tight supplies) provide the same reliability benefits as the same amount of additional generating capacity. From the reliability perspective, a reduction in demand is equivalent to an increase in generation. Indeed, to the extent the demand reduction is spread among many (perhaps thousands) of customers, diversity enhances the reliability benefits of load reductions.[2] On the other hand, nondispatchable loads may provide less certainty to system operators than generators under their direct control.

Finally, strategically timed demand reductions decrease the need to build new generation and transmission facilities. When demand responds to price, system load factors improve, increasing the utilization of existing generation and reducing the need to build new facilities. Higher asset utilization should lower overall electricity costs. Avoiding, or at least deferring, such construction improves environmental quality. Cutting demand at times of high prices may also encourage retirement of aging and inefficient generating units.

The preceding paragraphs suggest that customer participation in wholesale markets benefits retail consumers, system operators, the environment, and society at large. These benefits can be realized in both competitive and regulated electricity markets. These programs can be implemented by regulated utilities absent retail competition if, and only if, a visible spot price for wholesale electricity exists. As discussed in the following chapter, if even a small fraction of retail load responds only slightly to time-varying electricity prices, the results would be dramatic.

This report addresses the issues discussed above. (Two important issues not covered here are (1) the direct customer and supplier-marketing costs for these programs and (2) the role of distributed generation in competitive wholesale markets.) The results are based on telephone discussions with experts from various sectors of the electricity industry, including independent system operators (ISOs), electric utilities, load-serving entities (LSEs), communication-system providers, energy-service companies, regulators, and environmental groups. We also drew on the relevant literature, including ISO filings with FERC, utility-program descriptions, and published articles and reports.

---

[1] All four of the operational independent system operators (ISOs) in the United States experience market-power problems when demand is high (typically during the summer). As a consequence, the four ISOs impose price or bid caps on the participating generators. A robust demand side that participated in bulk-power markets might obviate the need for such caps.

[2] A large generator that provides reliability services (e.g., 100 MW of 10-minute reserves) that trips offline provides no reliability benefit. It is very unlikely that hundreds or thousands of customers who, together, provide 100 MW of reserves would all fail to respond at the same time.

The remainder of this report is organized as follows. Chapter 2 presents key concepts related to price elasticity, the inherent volatility of electricity costs, and the commodity and insurance components of electricity prices. Chapter 3 discusses the various kinds of demand-side pricing programs that apply in competitive markets, and gives examples of these programs. Chapter 4 discusses the differences among customers and customer types that might affect their interest in, and ability to respond to, dynamic prices. Chapter 5 discusses the enabling metering, communications, computing, and control technologies. Chapter 6 discusses the potential roles of the providers of these pricing and program options. Chapter 7 discusses the key technical and regulatory barriers to widespread adoption of these options. Finally, Chapter 8 summarizes the results of this study and suggests future actions.

# Chapter 2: BASIC CONCEPTS

## SUPPLY, DEMAND, AND PRICE ELASTICITY

Figure 1 illustrates the current disconnect between wholesale and retail markets, using the economists' traditional supply-demand curves (Eakin 2000). The top graph shows a typical supply curve, in which increasing amounts of generating capacity are offered at higher and higher prices. (This supply curve is based on the generator offers to the California Power Exchange [PX] in June 2000. It is similar to the supply curves seen in the three northeastern ISOs.) The graph also shows consumer demand during on- and off-peak periods; both curves are vertical, indicating consumer insensitivity to changes in electricity prices. This price insensitivity is a consequence of the current status of electricity restructuring in which the vast majority of consumers, regardless of whether they have chosen another supplier, face time-invariant electricity prices. The bottom graph shows how consumers might respond to price changes during the same on- and off-peak periods if they faced time-varying prices; this graph also shows the fixed price that most consumers pay for electricity.

Figure 2, using the same supply curve shown in the top of Figure 1, illustrates the effects of price-sensitive demand during peak periods. If consumers are insulated from time-varying prices, the market clears at 29 GW at a price of $550/MWh. However, if some consumers during this high-demand hour are even modestly sensitive to prices, the market clears at 27.5 GW at a price of $250/MWh. In other words, a 5 percent reduction in peak demand reduces prices by more than 50 percent. In this example, customers as a whole, not just those that respond to price, save almost $9 million an hour. In addition, such responses reduce price volatility.

The concept presented in Figure 2 has major practical implications. On June 7, 1999, maximum demand and spot price in the PJM

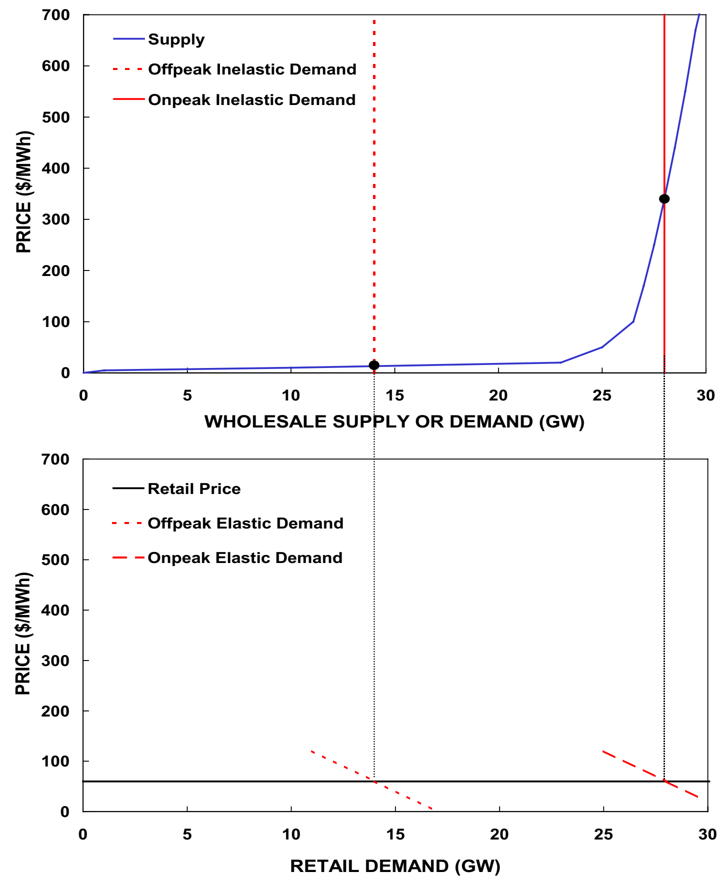



**Figure 1.** Schematic illustration of the disconnect between wholesale and retail electricity markets. The top graph shows a typical wholesale supply curve and price-insensitive consumer demand during on- and off-peak periods. The bottom graph shows how consumer demand might respond to time-varying prices as well as the fixed price that most retail consumers face.

Interconnection were 48,000 MW and $850/MWh. "On June 7, during the peak demand hours, an increase in supply or a decrease in demand of 500 MW would have reduced the price by about $100/MW. An increase in supply or a decrease in demand of 1,000 MW would have reduced the price by in excess of $200/MW. An increase in supply or a decrease in demand of 2,000 MW would have reduced the price by about $400/MW." Thus, a 4 percent drop in demand could have cut the hourly price by almost 50 percent (PJM Interconnection 2000a). This 4 percent reduction could occur if, say, 20 percent of the retail load faces hourly prices and, in response to changes in these prices, cuts their demand by 20 percent. This example demonstrates that substantial benefits can occur when even a small fraction of the load takes modest actions in response to time-varying prices.

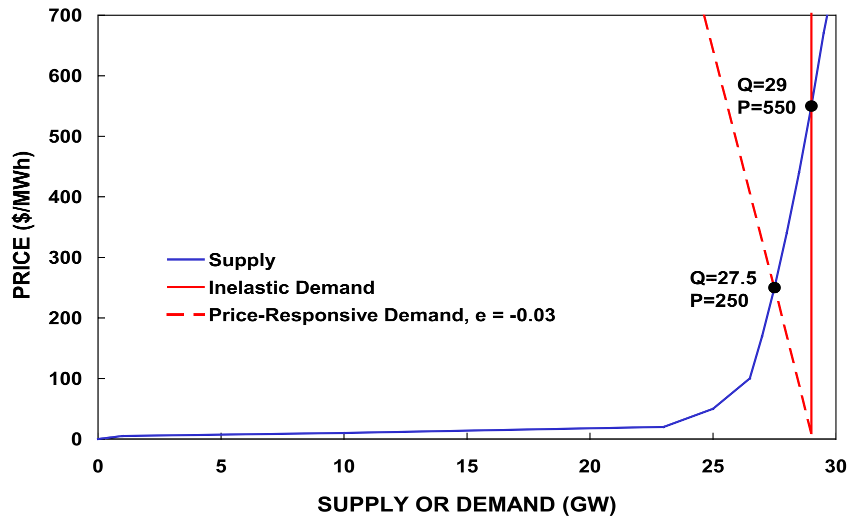



**Figure 2.** Generation-supply curve and two demand curves. The solid vertical line represents inelastic demand (as shown in the top of Figure 1). The dashed sloped line represents demand with a price elasticity of about 0.03.

The effects of a nonzero price elasticity of demand[3] during offpeak periods (say, between 10 and 20 GW on Figure 2), are very small. Because the supply curve is nearly horizontal at low and modest loads, the effects of consumer sensitivity to prices are small, on both demand and price. However, as shown in Figure 2, when supplies are tight, small changes in the amount of power demanded can have dramatic effects on market-clearing prices.

## VOLATILITY OF ELECTRICITY PRICES

Providing retail customers an opportunity to face time-varying electricity prices is important because electricity prices are volatile, perhaps more volatile than for any other commodity. Prices are so variable for several reasons:

- Generators differ substantially in their costs to produce electricity (e.g., the running costs for hydro and nuclear units are typically well below $10/MWh, while the cost for an old combustion turbine might be $100/MWh or more).[4]

[3] The price elasticity of demand is the percentage change in the demand for the product (electricity in this case) caused by a 1 percent change in its price. Mathematically, the elasticity is equal to (dQ/Q)/(dP/P), where d is a small change, Q is quantity, and P is price. For example, doubling the price would cut demand by 6.7 percent if the price elasticity is –0.1. Although price elasticities are almost always negative, reflecting the fact that consumers use less of a product when its price increases, we report elasticities here as positive numbers for simplicity.

[4] Generators also have fixed costs (e.g., fixed operations and maintenance expenses, taxes, depreciation, interest payments, and return on equity). In competitive markets, these fixed costs are recovered when the market price of power is above the generator's variable cost.

- System loads vary from hour to hour (e.g., by a factor of two to three during a single day).

- Electricity cannot easily be stored and therefore must be produced and consumed at the same time.

- Sudden generator outages, transmission outages, extreme weather conditions, and other events can trigger unexpected imbalances between generation and demand; rebalancing the electrical system can be expensive.

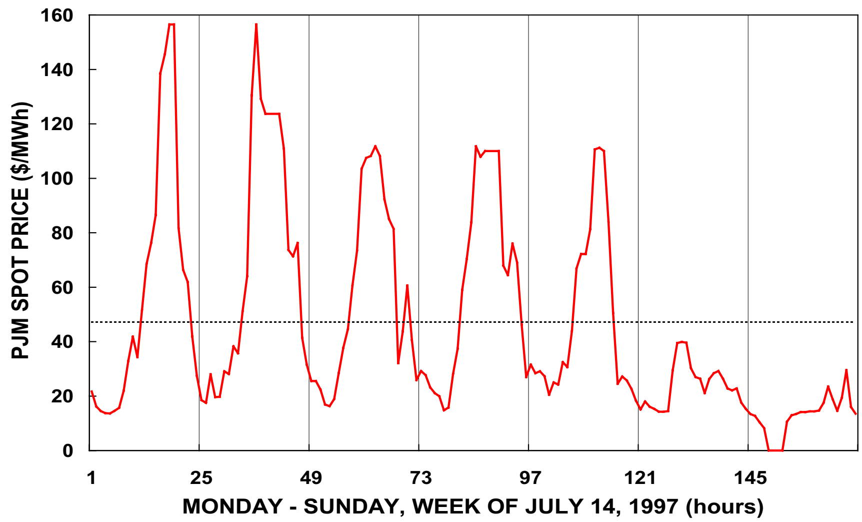



**Figure 3.** Hourly prices in the PJM Interconnection for one week. Prices ranged from $0 to $157, with an average of $47/MWh and a standard deviation of $38/MWh.

- Intertemporal constraints limit generator flexibility so that at certain low-load hours the price can be zero or negative because it costs more to turn a unit off and turn it on again later than to keep it running.

- When unconstrained demand exceeds supply, the price is set by consumer demand at a level above the running cost of the most expensive unit then online. During these few, high-load hours, generators must bid prices above their running costs to recover their startup and noload costs.

To illustrate the volatility in electricity prices, consider the one-week period for PJM shown in Figure 3. Wholesale hourly prices averaged $47/MWh. This average, however, subsumes substantial volatility, as prices ranged from a low of zero on Sunday morning to a high of $157/MWh on Monday and Tuesday afternoons. These large hour-to-hour differences in electricity prices provide substantial opportunities to make money by selling electricity at the right times (when prices are high rather than low) and by shifting consumption from one time to another (from high-priced to low-priced periods).

The statistics for hourly prices are skewed, with the mean generally much greater than the median. In the California day-ahead market in June 2000, the mean hourly price was $120/MWh, while the median was $61. This relationship between the mean and median occurs because there are many hours with low prices and a few hours with very high prices.

Table 1 summarizes variations in the day-ahead hourly electricity markets run by the California PX, New York ISO, and PJM Interconnection. Prices were more volatile and reached much higher levels in California than in the Northeast because of the mild summer 2000 weather in the Northeast. Real-time hourly prices are much more volatile than the day-ahead prices; the standard deviations of the real-time prices were 20 percent higher for PJM, 200 percent higher for New York, and 50 percent higher for California.

**Table 1. Statistical characteristics of day-ahead hourly electricity markets**

| | California (11/99–10/00) | New York (12/99–10/00) | PJM (6/00–10/00) |
|---|---|---|---|
| Average ($/MWh) | 70.1 | 36.8 | 28.6 |
| Maximum ($/MWh) | 750 | 523 | 140 |
| Minimum ($/MWh) | 6 | 0 | 0 |
| Standard deviation ($/MWh) | 81.8 | 22.3 | 18.9 |
| Fraction of hours (%) | | | |
| < $10 | 0.7 | 0.2 | 4.2 |
| < $20 | 4.0 | 15.5 | 45.2 |
| > $100 | 18.1 | 0.9 | 0.3 |
| > $200 | 5.9 | 0.2 | 0.0 |

## TWO COMPONENTS OF ELECTRICITY PRICES

The high variability of spot-market electricity prices illustrates an important, but often overlooked, feature of retail electricity markets: the traditional regulated price paid for electricity includes two components. The first is the electricity commodity, the kilowatt-hours of electricity we consume whenever we want, in whatever quantities we choose. The second is a risk premium (an insurance policy) that protects customers from the volatility of electricity prices by permitting them to buy unlimited quantities of electricity at a fixed price that is determined months (or even years) before consumption.

Traditionally, retail customers bore these risks—but not in real time. Utilities managed risk by building "extra" generating capacity. This planning reserve margin was intended to ensure the utility's ability to meet almost any foreseeable load. The capital and operating costs associated with this extra capacity is an economically inefficient way to manage risk because it assumes that all customers value electricity the same and have almost no ability to shift usage from on- to off-peak hours.

In addition, utilities collected any approved changes in power-supply costs (e.g., caused by changes in fuel costs) relative to those in base rates through monthly or quarterly fuel-adjustment clauses. If these cost changes persisted, utilities sought rate increases from the public utility commission (PUC). These delays, coupled with the absence of hourly meters, prevented customers from responding to the time-varying costs of electricity.

Competitive retail suppliers that offer electricity at a fixed price accept both quantity and price risks. The price risk (as illustrated in Figure 3) reflects the a priori unknown level and volatility of electricity prices. The quantity risk reflects the likelihood that electricity consumption will be higher when prices are higher. Electricity consumption and prices are highly correlated, as one would expect. For the week shown in Figure 3, the correlation between prices and quantities suggests that the amount of electricity consumed explained 72 percent of the variation in electricity prices. Of course, other factors, such as sudden generator or transmission outages, can also yield high prices.

A supplier selling fixed-price electricity would have to charge more than the average price to make a reasonable profit. Returning to the Figure 3 example, a customer wanting to purchase electricity during this week and willing to face spot prices would pay $47/MWh on average, assuming a time-invariant load. If, however, the customer consumed electricity using the same load shape as the PJM system load, its consumption-weighted price for the week would be $54/MWh, 15 percent higher than the unweighted average price. Altogether, a supplier might offer customers a fixed price of, say, $60/MWh for this period to account for price and quantity risks.

An analysis of California markets estimated the cost of an option to buy 1 MWh at $250/MWh during all weekdays in June and July 2001 between the hours of 11 am and 6 pm. The cost of such an option would be $15,400 (Boston Pacific 2000). Spread over all 1,464 hours in the two months, this risk premium is equivalent to $11/MWh. An option to protect against price spikes above $100/MWh would be almost three times as expensive as the $250 option.

On the other hand, customers willing to accept the quantity and price risks would, in the end, pay less for electricity. These customers would do so by buying only the electricity commodity and by providing the insurance. In addition, customers who face hourly prices can modify their loads in response to those prices and further lower their electricity costs.

As shown in Figure 4, if even a small fraction of retail load responded only slightly to time-varying electricity prices, the results would be dramatic (Caves, Eakin, and Faruqui 2000). For example, if 10 percent of the customers responded to a $10,000/MWh price spike with an elasticity of 0.1, the price would be cut almost 60 percent. (The percentage reduction in price declines as the price spike declines. At an initial price of $500/MWh, the price reduction would be 30 percent.)

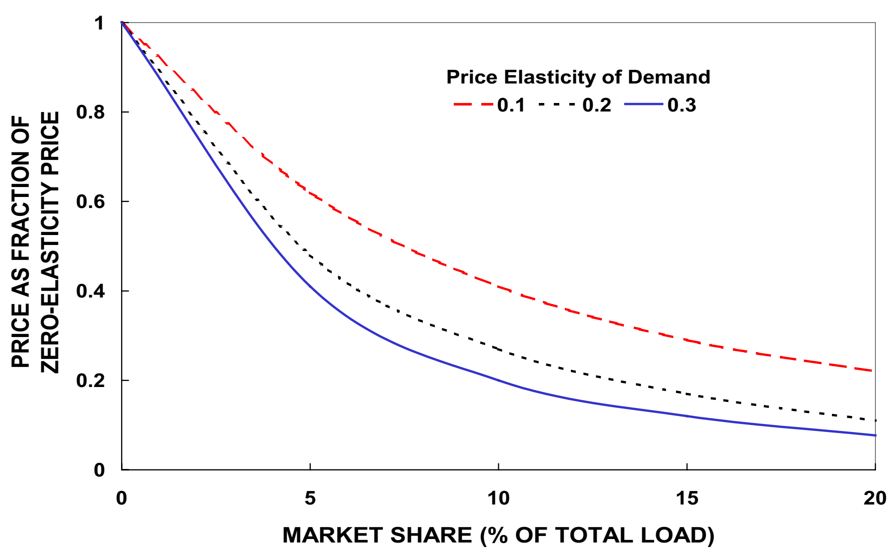



**Figure 4.** The effects of price-responsive demand on the market price of electricity as a function of the price elasticity of demand, the percentage of load that responds to time-varying prices, and the initial price of $10,000/MWh.

# Chapter 3: DYNAMIC-PRICING AND LOAD-REDUCTION PROGRAMS

## PRICING AND PROGRAM TYPES

Competitive electricity providers can offer prospective customers several options that share the financial risks between the provider and customer (Eakin and Faruqui 2000). The building blocks for these products include the hourly spot price, forward contracts (e.g., for fixed amounts for specified monthly blocks[5]), and financial instruments (e.g., put and call options[6]). Because customers have different electricity-use patterns, varying abilities to modify their electricity use in response to price signals, and different risk preferences, they will differ in their preferences among various product offerings (discussed in Chapter 4).

The ends of the risk spectrum are anchored by (1) a guaranteed fixed price that is announced well in advance and applies to all units of consumption and (2) a simple pass-through of hourly wholesale electricity prices to the customer. Seasonal or time-of-use (TOU) rates shift some risks from the supplier to the customer. (Seasonal rates vary from season to season but are time-invariant during each season. TOU rates differ across hours of the day but are predetermined; e.g., 4¢/kWh from 10 pm to 8 am, 6¢/kWh from 8 am to noon, 10¢/kWh from noon to 4 pm, and 6¢/kWh from 4 pm to 10 pm.) Additional risk can be shifted to the customer through a combination of a forward contract and a balancing contract for incremental or decremental load, in which the incremental or decremental consumption is settled at the wholesale spot price. A combination of hourly spot price plus a price cap (perhaps paid for with a price floor) also shifts risk to the customer relative to a guaranteed fixed price alone. These products represent different combinations of a forward contract and financial options.

Customers with a guaranteed fixed price could lower their electricity costs by participating in a voluntary load-reduction program and reducing their demands at times of high prices. In such programs, the LSE and customer share the savings. Alternatively, the customers could commit to reduce their loads during certain times when prices are high or reliability is threatened. Such interruptible contracts provide a call option to the supplier, the right but not the obligation to reduce those customers' demands under certain prespecified conditions.

Historically, utilities have run a variety of demand-side-management programs aimed at affecting customer electricity use levels and load shapes. These programs were driven primarily by the utility's resource planning needs rather than by competitive-market pressures and the interests of individual consumers. Such programs emphasized energy efficiency (affecting overall levels of electricity use) and load management (affecting the timing of electricity use). Typical energy-efficiency programs offered free or low-cost home energy audits,

---

[5] As an example, the California Power Exchange runs a market for monthly block forwards, each of which is for 16 hours every weekday during a particular month.

[6] A put gives the buyer the right to sell a certain asset at a certain date for a fixed price, called the strike price. A call gives the buyer the right to buy a certain asset at the strike price. For example, a call option might be written for 100 MWh of electricity during any hour of the summer months with a strike price of $150/MWh. The buyer would exercise this option only when the hourly spot price exceeded $150.

compact fluorescent lamps for commercial and industrial facilities, and other forms of advice and financial incentives. Load-management programs offered interruptible rates to large customers in exchange for the right to interrupt service to that customer under predetermined reliability conditions and direct load control (especially of water heaters and air conditioners) for residential customers.

Customer programs in competitive markets are likely to (1) focus on market, rather than regulated, prices and costs and (2) emphasize value to the customer. Competitive markets are likely to feature three broad classes of programs: dynamic pricing, voluntary load reductions at times of high prices, and customer sales of ancillary services to the system operator:

- Dynamic pricing[7] generally involves supplier provision of a set of 24 hourly prices, usually one day before real time (sometimes only an hour or two ahead). Pure spot pricing exposes all of a customer's loads to hourly price changes. Spot pricing with risk management adds a forward contract for a certain amount of load at a fixed price. In this second case, the customer pays for (is paid for) any hourly consumption above (below) the baseline at the hourly price.

- Interruption rights permit the utility (in traditional programs) to reduce participating- customer loads at certain times in return for a reduction in rates. Because customers received payments up front and the utility counted on the contracted-for resource response, these traditional interruptible-rate programs usually had penalties for noncompliance. Recently, utilities and LSEs have offered customers a shared-savings option in combination with a guaranteed price. Under such arrangements, the supplier agrees to share with the customer some of the savings that occur when the customer reduces consumption during high-priced periods. For example, the supplier and customer might agree that the customer will reduce load by 2 MW when prices exceed $200/MWh. In return, the customer and supplier will share the savings. If, for example, the fixed price of energy was $50/MWh and the load reduction occurred at a spot price of $300/MWh, the customer and supplier might share the $250/MWh savings 50:50. Table 2 compares the characteristics of the "traditional" reliability-based interruptible programs with voluntary price-based programs. The voluntary programs are typically more flexible, focus more on economics (high prices) than on reliability, and rely more on incentives to customers than penalties for nonperformance.

- Customer sale of ancillary services entails the use of load reductions as an alternative to generator provision of operating reserves (Hirst and Kirby 1998). These reserves include 10-minute contingency reserves and either 30- or 60-minute replacement reserves.

The greater flexibility of the new breed of interruptible programs relative to the traditional programs (Table 2) will likely substantially expand the number and range of customers who participate and increase economic benefits to those participants (Newcomb and Byrne 1995). Analysis of industrial customers participating in a Niagara Mohawk program showed large differences between schemes based on prices versus those based on mandated interruptions:

> When the customers were charged $0.50/kWh for 6 hours during a critical demand period, they reduced load voluntarily in ways that revealed marginal outage costs of only $0.13/kWh for the forgone

[7] We prefer the term dynamic pricing to real-time pricing (RTP) because real-time prices are determined by current conditions on the grid. Dynamic pricing, a more general term, encompasses RTP as well as forward (e.g., day-ahead or hour-ahead) prices.

> load. When an interruptible service program was used instead—and the utility curtailed load to produce the same total load reduction as in the first case—the [revealed] outage cost to customers was $1.19/kWh of unserved load. This nearly 10-fold difference results from the fact that price changes give customers greater choice than do curtailments, allowing them to configure usage more easily.

One potential problem with load-reduction programs is the need for a baseline against which to measure and pay for load reductions. This baseline (the load shape the customer would have achieved were it not for the load-reduction program) can be based on load shapes during one or more previous days, the load during the previous hour, a statistical model of customer-specific electricity use that accounts for weather and other factors, the load of the particular end use or equipment affected, or some other method. Regardless of the method chosen, the key is to ensure that the LSE and customer agree on the baseline and are both comfortable with it. It may not be necessary, from the societal perspective, to worry about the validity of the baseline used to measure voluntary load reductions. However, if the demand reduction is required for reliability (e.g., used as contingency reserves), the RTO, as well as state and federal regulators, will want to be sure that the expected reliability benefits actually materialize when called on.

The new voluntary programs, described above, offer a low-risk way for retail customers to cut their electricity costs in exchange for limited exposure to price volatility; see Table 3. In these programs, customers enjoy the benefits of a guaranteed fixed price, and they pay the risk premium for this protection. These programs permit customers, at their discretion, to recoup part of the risk premium from the supplier by agreeing to share the cost savings associated with demand reductions at times of very high prices.

However, the load-reduction programs can be more complicated than dynamic pricing. They may contain various restrictions, including the number of times the supplier can invoke the program, the minimum advance-notification time, the minimum and maximum duration for each interruption, the load-shape basis for payment for interruptions, the minimum amount of load that can participate, the existence and magnitude of penalties for noncompliance, and so on. Of course, these restrictions apply more to the traditional interruptible programs than to the new voluntary, price-based load-reduction programs.

**Table 2. Comparison of traditional and voluntary load-reduction programs**

| | *Traditional interruptible programs* | *Voluntary load-reduction programs* |
|---|---|---|
| *Purpose* | Reliability; often used as discounts in disguise | Economics; reduce costs when prices are high |
| *Duration* | Tariffs of one or more years | Short-term (e.g., monthly or seasonal) contracts |
| *Dispatch* | Rarely; for reliability purposes only | When prices exceed specified level |
| *Payments* | Upfront payments (often bill discounts); no performance payments | Usually no reservation payments; explicit sharing of actual savings (performance based) |
| *Penalties* | Substantial | None |

**Table 3. Comparison of dynamic-pricing and voluntary load-reduction programs**

| | *Dynamic pricing* | *Voluntary load-reduction programs* |
|---|---|---|
| *Purpose* | Economics | Economics |
| *Dispatch* | Whenever customer wants | When supplier calls for response |
| *Baseline* | None needed | Either “down to” or “down by,” both of which are imperfect[a] |
| *Payments* | None | Explicit sharing of dollar savings associated with load reduction |
| *Penalties* | Never | Usually none; sometimes for failure to comply with committed reduction |
| *Type of customer response* | Do whatever makes sense | Cut demand at specified times, usually at customer option |

[a] A down-to baseline is a set MW level, while a down-by baseline is a set reduction in load from the current level.

## PRICING AND PROGRAM EXAMPLES

### Dynamic Pricing

Some electric utilities began small real-time pricing programs in the mid-1980s (Mak and Chapman 1993). Newcomb and Byrne (1995) identified 29 North American utilities with real-time pricing programs as of 1995. These programs were aimed primarily at large industrial customers and secondarily at large commercial customers. Most had a per customer minimum load of about 1 MW. Typically, these programs involved fewer than 100 customers.

The hourly pricing for these programs was complicated because the programs were run by vertically integrated utilities under cost-of-service regulation. The prices typically included demand charges, customer charges, energy charges, and adders or multipliers. The adders and multipliers were used to ensure revenue neutrality for the utility. These factors sometimes reflected extreme conditions, such as capacity shortages (marginal outage costs).

Customers in these programs were typically notified of hourly prices by 4 pm on the previous day. Communication methods typically focused on use of the telephone and faxes. (These programs were operating before the Internet became the preferred communication medium.)

Four of the nine utilities reviewed by Mak and Chapman conducted customer surveys, which generally showed considerable satisfaction with these programs. “Overall, customer satisfaction was high for all four RTP programs. Most participants planned to continue on their respective programs and expressed interest in investing in modifications of their facilities if and when RTP tariffs were to become permanent and longer contracts were made available.” The key reasons for customer satisfaction included bill savings and customer control over electricity costs. Customers also mentioned reliability in price notification, a desire for more advance notice of prices, and a limit on the number of times the utility could revise its prices upward.

Four of the utilities analyzed customer responses to time-varying prices and found “that high prices do indeed induce customers to significantly reduce loads. This general conclusion prevails despite significant differences in structural characteristics and real-time prices among these programs.” Customer responses to dynamic prices can take two forms: “(1) Intraday flexibility which reflects the customer’s ability to shift usage from one hour to other hours of the day, and (2) Interday flexibility, which reflects the customer’s ability to shift usage from one hour to hours on other days.”

Medium and large industrial and commercial customers have been buying electricity on the basis of half-hourly prices in England and Wales since 1990. Patrick and Wolak (1997) analyzed customer data for four years (1991 through 1995) for their response to these prices. Their analysis found substantial heterogeneity across industries in their price elasticity and the pattern of within-day substitutions in electricity consumption. A substantial amount of load could be shifted away from high-price periods because, even though the elasticities were small, the price changes can be substantial.

Another analysis of customer response to dynamic pricing in England and Wales found that “the amount of load shifting that occurred between days is much larger [about 10 times greater] than the load shifting that occurred within days” (King and Shatrawka 1994). The between-day elasticities ranged from 0.1 to 0.2, while the within-day elasticities were about one-tenth the between-day values. This study also found that between one-third and one-half of the customers responded to time-varying prices.

Today, Georgia Power operates the largest dynamic-pricing program in the United States. About 1,600 of its large industrial and commercial customers (representing about 5,000 MW of load) face hourly prices (Braithwait and O’Sheasy 2000; O’Sheasy 2000). Participation in this program has increased steadily since it began as a pilot program in 1992. About 85 percent of the participating load is industrial, with the commercial sector accounting for the remainder.

The utility offers day- or hour-ahead notification of firm prices. About 35 of its largest industrial customers participate in the hour-ahead program. The day-ahead participants encompass a wide range of loads, from 0.25 MW up to 100 MW or more. These customers include a variety of industries (which account for about two-thirds of the participating load) and commercial businesses (e.g., grocery stores, shopping malls, hospitals, public schools, and universities). Some of these customers, especially the large industrials, have onsite generation.

Georgia Power posts the day-ahead prices at 4 pm by e-mail and increasingly through an Internet site. The day-ahead price is equal to the utility’s expected hourly marginal cost plus $4/MWh. Because its risks are lower with hour-ahead prices, the markup over marginal cost is only $3/MWh for this option.

Georgia Power uses a two-part design, based on a predetermined customer baseline load (CBL). The CBL is intended to reflect the customer’s preparticipation hourly load shape for purposes of traditional cost-of-service revenue requirements. In other words, if the customer uses electricity according to its CBL, the utility is revenue neutral with respect to the time-invariant rate the customer would otherwise pay.[8]

[8] According to the Georgia Power RTP schedule: “The customer baseline load (CBL) is developed using one complete year of customer-specific hourly firm data that represents the electricity consumption pattern and level agreed to by the customer and Georgia Power Company as typical of this customer’s operation for billing under his conventional tariff and from which to measure changes in consumption for [RTP] billing. The CBL is the basis for achieving revenue neutrality with the appropriate non-real-time-pricing *firm* load tariff on a customer-specific basis.”

Wholesale electricity prices have been increasing in level and volatility during the past few years (O'Sheasy 2000). Historically, Georgia Power counted on about 500 MW of load reduction when prices reached $500/MWh or more. In August 1999, when prices exceeded $1,000/MWh, customer response reached 800 MW (roughly a 20 percent load reduction). Although more customers have complained about high prices during the past few summers, very few have left the RTP tariff because of this concern. Instead many customers bought "pricing-protection products" from Georgia Power: risk management packages, such as contracts for differences, caps, and collars that allow customers to protect specific loads during specific time periods.

Figure 5 shows the responses to price changes for four Georgia Power customers. The prices during these two July days ranged from $19 to about $130/MWh, with an average of about $54/MWh. The mining and chemical companies responded strongly to these time-varying prices; the correlation coefficients between their hourly consumption and hourly prices were –0.89 and –0.99, respectively. Mid-afternoon consumption for the mining customer was only 29 percent of its maximum value; the comparable value for the chemical company was 83 percent. The marble and college customers, on the other hand, showed very little response to these hourly price changes. Indeed, electricity use at the colleges increased when prices were highest, although the increase might have been less than would have occurred at lower prices. These results show that customers differ enormously in their willingness and ability to respond to dynamic prices. The results also show that some price-insensitive customers still choose dynamic pricing over the standard tariff; these customers save money by self insuring. Roughly half of the load on dynamic pricing responds to price changes; in general, customers with onsite generation are more likely than other customers to respond (Braithwait 2000b).

Georgia Power customers also respond differently to moderately and very high prices (Table 4). As expected, the percentage reductions in load are much higher at very high prices. Also, the largest industrial customers, the ones that chose the hour-ahead prices, exhibit higher elasticities than the day-ahead customers. Overall, elasticities range from 0.03 for commercial customers to 0.3 for the hour-ahead customers.

Duke Power (in North and South Carolina) has a similar dynamic-pricing program, with about 110 customers, representing 1,000 MW of load. Duke's program began as a pilot with 12 participants in 1994. Only about a dozen customers have dropped out of the program since its inception. Duke's analysis (Schwarz et al. 2000) showed the following results:

**Table 4. Georgia Power customer responses to dynamic prices by notification period and maximum daily price**

| | *Percentage reduction in hourly electricity use* | |
|---|---|---|
| *Maximum daily price* | *Hour-ahead participants* | *Day-ahead participants* |
| $300 to $350/MWh | 29% | 8% |
| $1,500 to $2,000/MWh | 60% | 20% |

Source: Braithwait (2000b).

- Customer responses to Duke's hourly prices during the four summer months show lower elasticities than other utilities report, roughly 0.04 versus 0.1. As in other studies, Duke found that only some customers respond significantly to price changes.

- Customers with self-generation respond significantly to electricity prices above a threshold point at which self-generation becomes economical. Below the price at which it makes sense to operate the onsite generation, these customers show little response to rising prices; above the price at which the onsite generation is economical, the customers show substantial price response.

- Customer responses to price changes increase with experience on the RTP rates.

- Customers with a "discrete production process that allows delay" show a large price response. For example, paper manufacturers use large grinders for making pulp; these grinders can be operated independently of the rest of the production process to avoid hours with high prices.

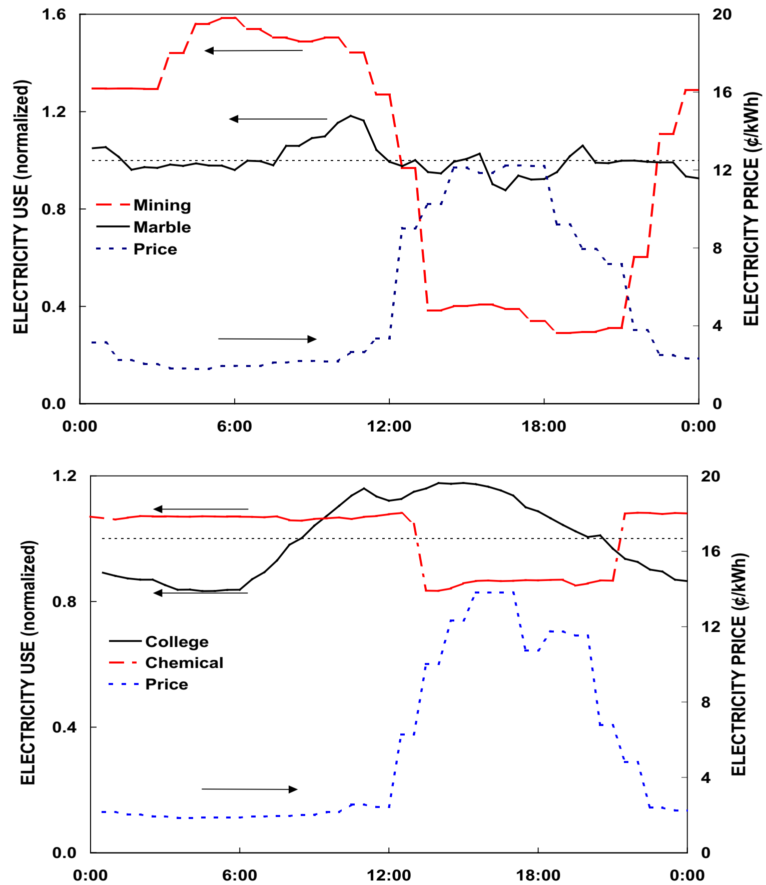



**Figure 5.** Customer responses to Georgia Power's dynamic-pricing program. The top graph shows the responses (normalized to average hourly use, the light dotted lines) from a large mining company (87 MW) and a marble manufacturer (4 MW). The bottom graph shows the responses from a chemical plant (47 MW) and several colleges (18 MW). (The dotted lines and right-hand axes show hourly electricity prices.)

While Figure 5 showed differences among customers in their responses to hourly price changes (for Georgia Power), Figure 6 shows how customers in aggregate respond to different price levels. The graph shows the daily load profile for all of Duke Power's RTP participants as a function of the maximum hourly price. As prices increase, mid-afternoon loads decline. The ratio of maximum hourly load to minimum hourly load increases from 1.04 when prices are below \$50/MWh to 1.13 when prices are between \$100 and \$150/MWh and to 1.21 when prices exceed \$250/MWh.

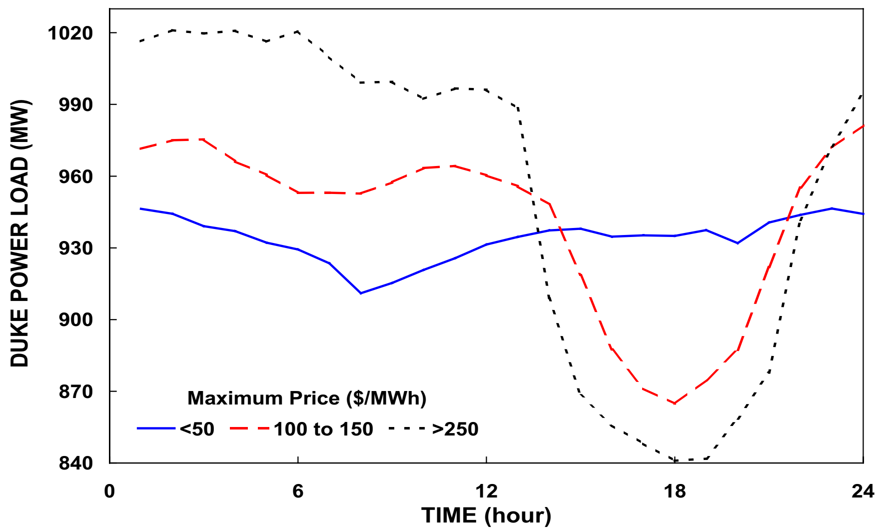



**Figure 6.** Customer response to Duke Power's day-ahead hourly pricing program as a function of the maximum hourly price.

The authors of the Duke analysis also investigated the optimal amount of advance notice of electricity prices, trading off the benefits to customers of more advance notice (e.g., day- vs. hour-ahead notification) against the risks to the supplier of having real-time prices different from the a priori price guarantee (Taylor and Schwarz 2000). Increasing the amount of advance notice allows customers to find and implement more ways to adjust their production processes to respond to these prices. On the other hand, the price forecast error increases with more advance notice. "Under reasonable assumptions, customer benefits from advance notice tend to exceed the cost of increased risk associated with forecast error. … The simulation suggests that day-ahead advance notice increases welfare for reasonable magnitudes of customer elasticity and utility forecast error."

Although most of the dynamic-pricing programs today focus on large industrial and commercial customers, some utilities run such programs for residential customers. GPU Energy ran an innovative time-of-use[9] pilot program that featured an interactive communication system (Braithwait 2000a). This system allowed the utility to communicate price signals to the home thermostat so that electricity use for heating or air conditioning could be automatically adjusted in response to price changes. The overall elasticity for these customers was about 0.30, much higher than that experienced in other residential TOU programs. During the summer months, participants cut their weekday electricity consumption by 7 percent, on average.

Several other utilities continue to run RTP programs. In general, these programs have very few participating customers. In addition, many customers have left these programs during the past couple of years as they have experienced dramatic price spikes. Interestingly, neither Georgia Power nor Duke Power experienced such customer losses.

In principle, large customers or their LSEs[10] could bid price-responsive demand into the ISO/PX day-ahead, hour-ahead, or real-time markets. However, very little such behavior is occurring. Given the need for some advance notice to plan operations, we had expected to see very little price response to real-time markets, but

[9] Recall that TOU prices differ from real-time prices in that TOU prices are fixed well in advance of consumption for set blocks of hours each day. Real-time prices, on the other hand, vary from hour to hour and are set either day- or hour-ahead, or sometimes in real time (either ex ante or ex post).

[10] For purposes of this report, LSEs can include load aggregators, energy-service providers, energy- service companies, scheduling coordinators, the marketing affiliate of a distribution utility, or a municipal distribution utility.

we had expected to see considerable such behavior in the New York, PJM, and California day-ahead energy markets. (New England does not yet operate a day-ahead energy market; it operates only a real-time energy market.) None of the people we spoke with had an explanation for why so little load was responding to prices in day-ahead markets; we return to this issue in Chapter 6.

**Interruptible and Voluntary Load-Reduction Programs**

Electric utilities have been running programs to encourage demand reductions at times of tight supply for at least two decades. These programs have focused on large industrial and commercial customers, typically offering a discount in the demand charge (expressed in $/kW-month) in exchange for the right to interrupt service to a portion of the customer's load (Table 2). These programs are characterized by a rigid structure that specifies far in advance the maximum number of times a year the utility can call for interruptions, the minimum amount of advance notice it must provide, the maximum time permitted for each interruption, and the penalty imposed on customers who do not meet their contractual obligation to interrupt demand when called upon to do so. A 1994 survey identified 89 such programs, covering almost 7,000 industrial and commercial customers (Plexus Research 1995).

The same survey identified 333 load-control programs, most of which focused on residential water heaters and air conditioners. The programs typically used a communication system to remotely turn off and on the end-use equipment. The 6.7 million controlled installations included nearly 3.2 million air conditioners, nearly 2.4 million water heaters, and nearly 1.0 million space-heating systems, almost all of them in residential applications. Participating customers typically received a bill credit of $20 to $30 a year for each piece of equipment under utility control. Reported load reductions ranged from 0.6 kW for water heaters to 0.9 kW for cycled air conditioners, 1.8 kW for shed air conditioners, and 1.4 kW for space-heating systems.[11]

The Energy Information Administration collects data each year on electric-utility demand-side-management programs. The results for 1998 show that these programs achieved a potential peak load reduction of 41,400 MW, down from the 48,300 MW reported for 1996 (Energy Information Administration 1999). (Two-thirds of the 41,400 MW reduction was attributed to load-management programs, with the remainder attributed to energy-efficiency programs.) This potential reduction was split roughly equally among the residential, commercial, and industrial sectors. Utilities spent about $540 million on these load-management programs in 1998.

Utilities are not the only entities operating reliability-based load-management programs. Planergy, working as an aggregator, provides a similar program with a total load reduction of about 100 MW to two utilities in Texas: Reliant Energy and CSW (now part of American Electric Power). Planergy offers a near-turnkey service to utilities that includes marketing, program operations, monitoring, and verification for settlements (Slifer 1999). By aggregating across customers and selling the aggregate load reduction to the utility, Planergy's program is available to smaller commercial and industrial customers than those that typically participate in utility interruptible programs.

Planergy uses 15-minute load data. It measures load reductions relative to the load during the hour before the utility called for the load reduction, a very simple way to define the baseline (Figure 7). The utility calls

[11] Air-conditioner *cycling* involves repeatedly turning the unit off (say, for 10 minutes every half hour) and then on (for 20 minutes). Air-conditioner *shedding* involves turning the unit off for an extended time, say four hours.

Planergy to announce a curtailment at least 75 minutes before it is to begin. Planergy then uses an automated telephone dialer to notify customers of the forthcoming interruption. Planergy uses telephone lines or satellite phones to collect 15-minute customer-load data to monitor the progress of the interruption, usually with a 15- to 20-minute delay.[12] (This delay occurs because of the inherent lags in getting data from meters that record cumulative consumption only once every 15 minutes, as well as the time it takes the Planergy system to make all the phone calls to the end-use meters and to process the data.) Planergy uses these data to follow up with customers who are not meeting their load-reduction obligations. If a customer is not able to respond, Planergy will solicit backup customers who are willing to participate from time to time but not on an ongoing basis. Because of this active followup, Planergy obtains near 100 percent compliance with the utility interruption requests. Customers who consistently fail to respond to the interruption requests are removed from the program.

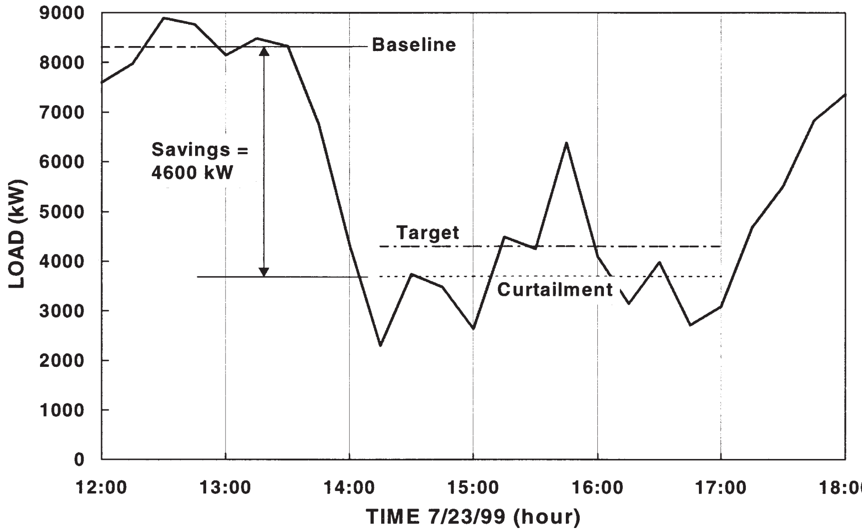



**Figure 7.** Typical load reduction achieved by Planergy in its demand cooperative program. The baseline is equal to the load of the participating customers during the hour before the interruption begins. In this example, the average load reduction exceeded the target during the 3-hour curtailment.

Advance notice is typically one hour, and the interruptions typically last three or four hours. As with the utility programs, there are limits on the number of times a year interruptions can be called and on maximum outage duration. Because customers are paid on a monthly basis in terms of $/kW-month for each kW of load reduction offered, customers are penalized for failure to respond.

In addition to these "traditional" load-management programs, ISOs and utilities operate a variety of new, primarily voluntary load-reduction programs. We first discuss the summer 2000 ISO programs and then discuss a few of the new utility programs.

ISO New England developed a Load Response Program in spring 2000 for implementation during summer 2000 (June 1 through September 30). In this program, distribution utilities and LSEs solicited retail customers and executed contracts with these customers for voluntary interruptions. ISO New England sought three 200-MW blocks of interruptible load at prices of $500, $750, and $1000/MWh. (The higher the strike price, the less often the ISO would call on these load reductions. In addition, these high-price interruptions would have shorter durations than the lower-price interruptions.) The ISO would call on these blocks of demand when a forecast of the market-clearing price of energy would exceed one of these block prices and if the ISO's

[12] These delays refer to the time it takes for Planergy to learn how customers responded to the interruption request, not to the time for the response to occur. The latter lag is a function of the time it takes Planergy to notify participating customers of an interruption and of the time it takes customers to implement control actions.

Operating Procedure 4 (Action During a Capacity Deficiency) had been implemented. In other words, this program is triggered by both reliability (expected regional capacity deficiency) and economics (market prices above the block prices). ISO New England (2000) explains the program as follows:

> During OP 4 and before Action 12 [of OP 4] has been utilized, ISO-NE may only call upon an interruptible load (in price order) if its strike price is at or below the ECP [energy clearing price]. Following activation of Action 12, ISO-NE may call upon an interruptible block regardless of whether the ECP exceeds the price of the interruptible block. In this event (post Action 12), the interruptible block would be eligible to set the ECP and would essentially act as a floor price until that hour following the restoration of the ten-minute reserve.

Customer responses to interruption requests are voluntary, with no ISO or LSE penalties imposed for noncompliance. The ISO pays the LSEs for measured load reductions. The payments to participating customers are of no concern to the ISO and are left entirely to the LSE and its customers. In a similar fashion, the ISO specifies no method for determining each customer's baseline load; instead, the LSEs are responsible for determining customer-specific baselines against which to measure load reductions.[13] The ISO imposes no minimum load size or load-reduction magnitude. The only requirements the ISO imposes on customers is that they have hourly meters and can respond to interruption requests within one hour of being asked to interrupt.

More than 900 retail customers, accounting for 253 MW, signed up for this program. Almost all of this load (240 MW) signed up at the $500/MWh price, 6 MW at $750, and 7 MW at $1,000. The program was never activated because summer temperatures were unseasonably mild.

PJM operated a similar program, called the Customer Load Reduction Pilot Program (PJM Interconnection 2000b). Like the New England program, the PJM program is reliability-based, voluntary, operated through the LSEs, and of limited duration (summer only). Participating loads must be able to reduce at least 0.1 MW of load, be able to participate for a minimum of 10 hours over the program-operation period ending September 30, 2000, be available all hours between 9 am and 10 pm all seven days of the week, be capable of achieving full reduction within one hour of PJM's request to reduce load, maintain the load reduction for at least two hours,[14] and have meters that can record consumption on an hourly basis.

The program design permitted PJM to call on these interruptible loads following the declaration of Maximum Emergency Generation and before implementation of other load-management programs. (Maximum Emergency Generation is used to increase the amount of generating capacity available to the PJM system beyond the maximum economic level during reliability emergencies.) PJM will pay the LSEs the higher of the appropriate zonal locational marginal price of energy or $500/MWh. As with the New England Program, PJM leaves it to the LSEs to determine the baseline load shape against which to measure load reductions. To the extent that participating customers are high-load-factor industrials, the baseline load shape is largely time invariant.

[13] The ISO requires only that the LSEs "calculate an estimate of the load reduction actually achieved in each hour of request. The estimate should be based on hourly meter readings adjusted to account for normal load shapes and temperature differences."

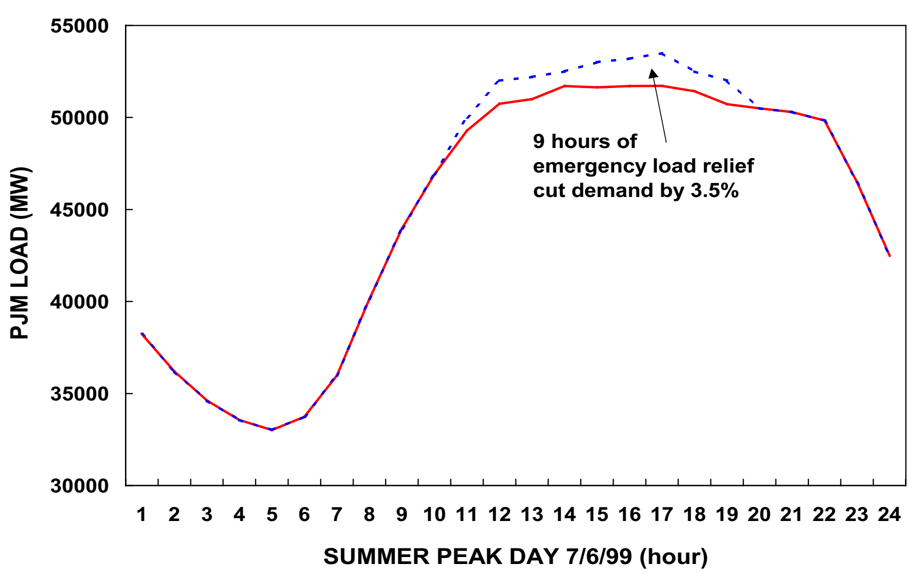
[14] PJM included this minimum load-reduction duration to ensure that participating customers could earn enough money to warrant their participation in the program (Bresler 2000).

About 30 customers representing 100 MW (ranging in size from 0.1 to 15 MW) signed up for the program. The program was never activated during summer 2000 because, as was true for New England, mild weather kept loads from reaching very high levels.

In addition to its new pilot program, PJM has also operated an Active Load Management program since 1991 (Exhibit 1). This program, operated primarily by the distribution utilities, includes direct control of residential equipment, customer load reduction to a firm level (interruptible contracts), and guaranteed load drops implemented through the use of onsite generation. In this program, PJM provides no monetary payment. Instead, participating utilities receive installed-capability credits for the load reductions, which reduce the utilities' costs of installed generating capacity. Participating loads must be available for up to ten PJM-initiated interruptions during the planning period (October through May and June through September), for interruptions lasting up to six hours between noon and 8 pm on weekdays, and within two hours of notification to the LSE by PJM. The baseline is either the customer's load one hour before the event or the customer's hourly loads on a comparable day, as determined by the LSE. Failure to perform can lead to penalty charges related to PJM's capacity deficiency charge; that is, the penalty is comparable to that which would apply for providing insufficient generating capacity to meet the required installed-capability requirement (PJM Interconnection 2000c).

About 1,700 MW of load (roughly half of which is residential and small-commercial direct-load control and half of which is industrial loads and onsite generation) qualify for installed capability in PJM. The program was called upon six times during the summer of 1999 but not at all during the summer of 2000. The program

**Exhibit 1. PJM's Active Load Management Program**




**PJM hourly loads on its peak day in 1999 with and without the effects of active load management. Absent load management, peak demand would have reached 53,500 MW.**

The situation that occurred in PJM on July 6, 1999, illustrates well the value of load management. PJM's load reached an all-time high that day and, as a consequence, it deployed its Active Load Management program to reduce demand during the mid-day hours. The program cut demand by 3.5 percent at 5 pm and by an average of 1 percent during nine peak-load hours. Because electricity prices reached $920/MWh during these hours, this demand reduction cut electricity costs by $10 million. (Absent this load reduction, prices would have been even higher!) Interestingly, the same electricity savings on the following day would have saved only $0.4 million because electricity prices on July 7 reached only $50/MWh would be called upon immediately after PJM calls on the loads participating in the previously discussed load-reduction pilot program.

The California ISO operated a similar Demand Relief Program, in which it sought 1,000 MW of participating loads, each of which must be greater than 1 MW. The program ran from June 15 through October 15, 2000, and was contracted on a monthly basis. Unlike the other ISO programs, which paid only for actual energy reductions (on a $/MWh basis), the California program paid both for capacity reserved (on a $/MW-month basis) and for energy saved (based on the difference between scheduled and actual energy use). These loads were to be called upon when the ISO declared a Stage 1 emergency, which means that operating reserves are expected to fall below the required 7 percent of daily peak demand. Unlike the two other ISOs, the California program specified the baseline against which load reductions were to be measured as the demand for the same hour averaged over the ten prior weekdays. As with the other programs, meters that record hourly consumption were sufficient for participation. Of the 180 MW of bids accepted by the ISO, contracts were signed with about 65 MW (Dozier 2000).

Unlike the programs in the Northeast, the California program was invoked 20 times during June, July, August, and September 2000 because California experienced hot weather, high demand, and limited supplies many times. As of November 2000, the ISO had published no information on the actual performance of the program.

**Table 5. Comparison of ISO summer 2000 load-reduction programs**

| | *California ISO* | *PJM Interconnection* | *ISO New England* |
|---|---|---|---|
| *Payment basis* | Monthly capacity plus energy | Energy reduction | Energy reduction |
| *Payment amounts* | As-bid capacity plus energy price | Higher of zonal price or $500/MWh | $500, $750, or $1,000/MWh |
| *Availability of load reductions* | Weekdays, noon to 8 pm | 9 am to 10 pm | Midnight to midnight |
| *Minimum capacity per customer* | 1 MW | 0.1 MW | None |
| *Penalties* | Reduced capacity payment | None | None |
| *Baseline* | 10-day rolling average | Determined by the load-serving entities | Determined by the load-serving entities |
| *Dispatch* | Stage 1 emergency | Maximum emergency generation | Operating Procedure 4 |
| *Limits* | 30 hours/month | None | None |
| *Time to respond* | 30 minutes | 1 hour | 1 hour |

Source: Siegel (2000).

Table 5 compares the characteristics of the summer 2000 load-reduction programs run by the three ISOs. All three programs represent small but important steps towards the creation of competitive markets in which loads fully participate. Because these programs are driven more by reliability requirements than by economics, their benefits are much less than they could have been. For example, the New England program received 240 MW of load-reduction offers at a price of $500/MWh. However, the ISO cannot call on these loads when the supply offers exceed $500/MWh (e.g., the price bid by the last MW of supply accepted by ISO New England is $600/MWh) unless OP 4 conditions also exist. In other words, the ISO must ignore lower-cost demand resources in favor of more expensive generation.

Several utilities also operate load-reduction programs that differ substantially from the traditional programs. Portland General Electric (PGE, in Portland, Oregon), for example, has an Electricity Exchange Rider that went into effect in July 2000 (Jonee-Guinn 2000). The rider "allows participating Customers an opportunity to voluntarily reduce their electricity usage in exchange for a payment, at times determined by the Company."

Participating loads must be greater than 5 MW with at least 1 MW of pledged load reduction. PGE announces an "event" by 4 pm a day ahead, customers have until early morning of the event day to pledge load reductions between 6 am and 11 pm, based on the hours that PGE specified for the event. Customers are paid on the basis of the northern California ISO real-time hourly price, which can differ from the day-ahead PX prices that PGE uses to decide whether to call an event. Customers receive 50 percent of the difference between the real-time ISO hourly price and the tariff energy charge; the energy charge is about $36/MWh. In no event will a customer get paid less than zero, which could theoretically occur if the real-time ISO price was lower than the tariff price.

PGE defines the baseline for each customer based on its hourly loads during the 14 consecutive weekdays before the day of the event. Participating customers have hourly meters that are read through telephone lines.

PGE imposes no penalties for noncompliance. However, if a customer pledges a load reduction for an event and then fails to deliver three times, it is dropped from the program.

Between July and November 2000, PGE called 32 events. The six participating customers achieved load reductions of up to 121 MW. PGE paid $1 million to the participating customers. To encourage customer participation, PGE did not call an event unless it expected the real-time price to exceed $250/MWh. PGE, based on discussions with its customers, believes that the customer costs for interruption are such that lower payments (e.g., at prices of $100/MWh) would greatly reduce participation.

GPU Energy (operating in New Jersey and Pennsylvania) operated a program similar to PGE's (Stathis 2000). GPU targeted retail customers with a minimum demand of 0.3 MW that had 15-minute interval meters. Altogether, GPU signed up seven companies with 15 MW of load for this pilot program. GPU found that mail solicitations did not work. Successful marketing required face-to-face meetings with customers.

Unlike the PGE program, the GPU program was based on firm day-ahead prices and commitments. GPU, based on its expectations for PJM's day-ahead locational prices, sent the customers participating in its Voluntary Load Reduction Program a set of hourly prices for the hours ending noon to 8 pm for the following day by 9 am. Customers were required to bid hourly MW reductions to GPU by noon of that day. Unlike the PGE program, which had no penalties for noncompliance, GPU penalized customers who provided less than 90 percent of their day-ahead load-reduction bids. As with the PGE program, GPU shared the benefits of load reductions 50:50 with participating customers.

Because of the mild summer, GPU invoked the program only five times. Like PGE, GPU did not offer the program unless the day-ahead price was sufficiently high, about $200/MWh for GPU. On average, about 4 MW of load reduction was obtained.

The PGE and GPU programs used the same Internet-based communication system, developed and sold by Apogee Interactive, to send price information to participating customers and to receive customer responses. The Apogee system also calculated the actual load reductions, based on its proprietary analysis of customer-specific baseline load shapes.

Wisconsin Electric developed two programs for implementation in summer 2000 (Sodemann 2000).[15] The company was motivated in part by negative customer reaction to its traditional interruptible-rate programs. After years of enjoying the rate discount and not having been called upon to interrupt loads, customers were angry that they were forced to reduce loads during the summer of 1998. Wisconsin Electric wanted to develop programs that gave more control to customers. The goal was to develop programs that paid customers for voluntary load reductions when regional electricity prices were high.

The Dollars for Power program was aimed at commercial and industrial customers and used fixed bid prices of $400, $800, and $1,250/MWh for load reductions, with participation by about 4 MW, 4 MW, and 7 MW, respectively, of load relief. Participation was limited to customers who could provide reductions greater than 0.05 MW and who had hourly meters. Customers signed up for load reductions at one of the three prices, but were free to switch from one price to another on a monthly basis.

The utility used a combination of its website plus e-mails and pagers to post prices and announce the time deadline for customer MW-reduction bids. Customers who provided load reductions were paid according to the fixed price; no penalties were applied for failure to reduce load. Unlike the PGE and GPU programs, this was a day-of program, in which Wisconsin Electric announced prices and solicited MW-reduction bids in the morning for use later that day.

The Power Market Incentives program was aimed at Wisconsin Electric's largest and most sophisticated industrial customers. In this program, the utility established the price at which it would buy back load reductions and then solicited MW-reduction bids from the customers. The minimum load reduction for this program was 0.5 MW, ten times that for the Dollars for Power program. Altogether, 27 customers with 85 MW of load relief participated in this program. Unlike the Dollars for Power program, this program included penalties for nonperformance, with the amount of the penalty based on the actual wholesale price of power.

In both programs, the utility and customer negotiated a baseline load shape prior to program initiation. Because many of the very large customers had essentially flat load shapes, agreement on a baseline was not difficult.

---

[15] In addition, the utility paid a large mining company (with a load of 90 MW) $82,000 to shut down on one day when prices were especially high.

Utilities are not the only entities that offer such load-reduction programs. Several power marketers and energy-service providers offer similar programs. We were not able to obtain details on any of these programs, however.

**Ancillary Services**

For both reliability and commerce, bulk-power systems require certain services beyond the basics of energy, generating capacity, and power delivery (Hirst 1999). Some of these ancillary services (such as regulation and reactive power) are required during normal operations to maintain the necessary balance between generation and load in real time and maintain voltages within the required ranges. Other ancillary services (such as contingency reserves) provide insurance to prevent minor problems from becoming catastrophes. Finally, some services (such as system blackstart) are required to restore the bulk-power system to normal operations after a major outage occurs.

In response to a rapidly changing structure for the U.S. electricity industry and the requirements of FERC's (1996 and 1999) Orders 888 and 2000, the industry is unbundling these services and pricing them separately. In particular, certain services are being bought and sold competitively, primarily through markets operated by ISOs. Some of these services, such as contingency reserves and load following, can be provided by loads as well as by generators:

- Load following is the use of online generation equipment to track the interhour changes in customer loads. These hourly and diurnal changes in customer loads are generally correlated with each other and are predictable.

- Spinning reserve is the use of generating equipment that is online and synchronized to the grid that can begin to increase output immediately in response to changes in interconnection frequency and that can be fully available within ten minutes[16] to correct for generation/load imbalances caused by generation or transmission outages. In principle, loads under the control of the system operator could help provide this service.

- Supplemental reserve is the use of generating equipment and interruptible load that can be fully available within ten minutes to correct for generation/load imbalances caused by generation or transmission outages. Supplemental reserve differs from spinning reserve only in that supplemental reserve need not begin responding to an outage immediately. This service may also include the provision of additional generating capacity that must be fully available within 30 or 60 minutes (the exact time depends on the rules of the regional reliability council) and can then be maintained until commercial arrangements can be made (e.g., for two hours) to "back up" the normal supply for the load.

- Backup supply is a service customers could purchase to protect against forced outages at the generating units that provide their energy or against loss of transmission between their normal supply and their load. This service is used to restore operating reserves to normal levels and therefore must be available within 30 to 120 minutes. Unlike spinning and supplemental reserves, which are system services required for reliabil-

[16] In November 1999, the North American Electric Reliability Council changed the recovery period for its Disturbance Control Standard from ten minutes to fifteen minutes. The ISOs are using the extra five minutes to determine whether an outage has occurred that warrants use of contingency reserves and, if so, which reserves to use.

ity, backup supply is a commercial service that supports individual transactions. Customers who did not purchase backup supply would be required to disconnect their loads from the grid when their power supply failed or to buy power on the spot market.

Municipal water pumping accounts for 3 to 4 percent of U.S. electricity use (Kueck 2000). Because water systems have extensive storage in pipelines and water towers, the pump motors could be turned off immediately in response to bulk-power-system contingencies. This pumping load could provide half the nation's spinning-reserve requirements. Unfortunately, current reliability rules prohibit the use of such rapid-response loads as spinning reserve.

The three Northeastern ISOs as well as the California ISO run markets for ten-minute spinning and supplemental reserves and either 30-minute (Northeastern ISOs) or 60-minute (California) replacement reserves. None of the ISOs runs a market for load following; implicitly, this service is captured in their intrahour energy markets (five minutes for the Northeastern ISOs and ten minutes for California). The ISOs do not run markets for backup supply because the longer-term nature of this service removes it from the ISOs' reliability responsibilities.

Although loads can, in principle, provide any of these services, in general they are not doing so. Only in California are some loads participating in ancillary-service markets: about one-third of supplemental reserve and 1 percent of replacement reserve comes from loads. Specifically, a couple of large government water agencies provide about 300 MW, on average, of supplemental reserve.

In February 2000, the California ISO issued a request for bids inviting loads greater than 1 MW to participate in its ten-minute nonspinning-reserve and 60-minute reserve markets as well as its supplemental-energy (real-time) market (California ISO 2000a). The ISO was hoping to obtain about 400 MW in each of the two reserve markets plus 1,000 MW in the supplemental-energy market. The ISO received six bids for 400 to 500 MW in all three markets. Ultimately, the ISO signed Participating Load Agreements with about 230 MW of this load.

Load responses were limited because of the ISO's stringent requirements for real-time data. (The ISOs' metering requirements are more relaxed for the voluntary load-reduction programs, presumably because the load-response time is an hour instead of ten minutes for nonspinning reserve.) The ISO standards for metering and telemetry of load data from the customer's meter to the ISO are based on its historical practice with generation, which calls for data transmissions every four seconds. The ISO believes that it must have such frequent data to comply with reliability requirements.[17] To accommodate loads that have meters that are read less often, the ISO adopted the concept of an Aggregating Load Meter Data Server, a data-acquisition and processing system that collects data from individual loads and passes the aggregate data to the ISO's computer system. Although the data server would be required to send data to the ISO every four seconds for nonspinning reserve (once a minute for replacement reserve), the individual loads could report data to the data server at one-minute intervals (once every five minutes for replacement reserve).

Loads wanting to sell nonspinning reserve are required to be able to reach their load-reduction commitment within ten minutes and maintain that reduction for at least two hours. Loads wanting to sell replacement

[17] The California ISO's technical standard for loads notes its need for "near real-time [data] … to continuously monitor the status, location, and amount of reserves available to meet reliability criteria set by the Western Systems Coordinating Council and the North American Electric Reliability Council" (California ISO 2000a).

reserve are required to reach their load-reduction commitment in no more than 60 minutes and, once again, maintain that reduction for at least two hours.

Most of the load that planned to participate in the ISO's ancillary-service markets was already participating in existing utility interruptible programs and, therefore, subject to approval from the California PUC. (Some parties were concerned that loads might get paid twice for the same interruption, while others felt there would be no overlap because the utility programs and ancillary-service requests would be called at different times and for different reasons.) Because the PUC had not ruled on such participation by November 2000, there was essentially no new participation by loads in the ISO markets (California ISO 2000b).

# Chapter 4: CUSTOMER CHARACTERISTICS

This chapter discusses various factors that might affect a customer's interest in and ability to participate in programs that modify the timing of electricity use. First, however, we examine the distribution of customers by energy use and demand. The infrastructure costs of metering, communications, and control are, to some extent, independent of the magnitude of load. Therefore, it is useful to know, roughly at least, how large customer-specific load reductions must be to justify the infrastructure costs.

Nationwide, the 538,000 industrial customers represent 0.4 percent of all customers but 30 percent of total demand (Edison Electric Institute 1999), equivalent to about 200 GW of demand at the time of summer peak. As illustrated in Figure 8, a very small percentage of customers (1 percent in the case of this utility) account for a substantial share of total electricity sales and demand (about 50 percent in this case). A mere 0.01 percent of this utility's customers account for 16 percent of its total load; 0.1 percent of its customers account for about one-third of its total load.[18]

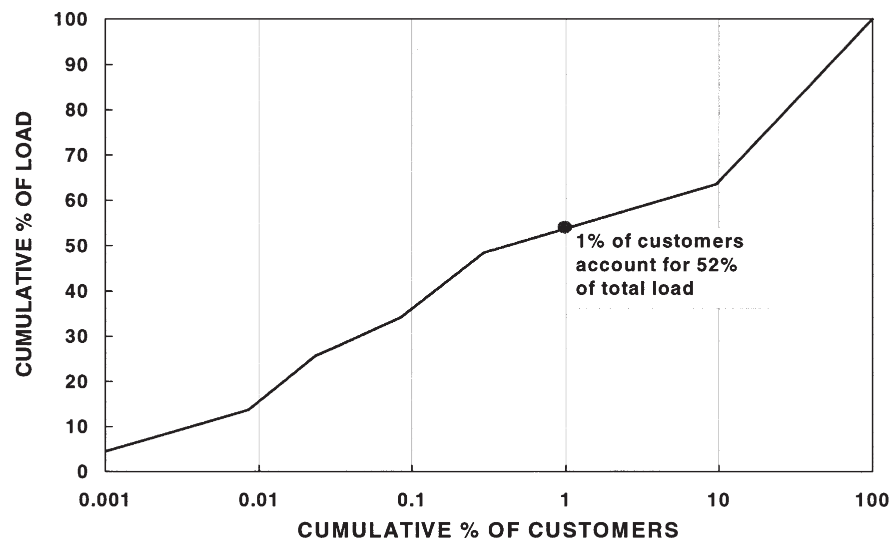



**Figure 8.** Cumulative load as a function of the cumulative number of customers for an electric utility. Note that the horizontal axis is logarithmic.

These results show that large benefits can likely be achieved by offering dynamic pricing and related programs to a tiny fraction of the nation's electricity consumers. In other words, your grandmother need not worry about responding to dynamic pricing.

In addition to size, customers differ in their hour-to-hour load shapes and energy intensity, as shown in Figure 9 (Rabl 2000). The top part of the figure shows the typical weekday load shapes for four types of commercial facilities. Overall, grocery stores and restaurants are more electricity intensive than office and retail facilities. However, the latter two building types have much lower load factors (about 60 percent relative to the roughly 80 percent for the two other building types).

[18] We know of one utility with about 500,000 customers for which the largest 600 customers (all of which have five-minute interval meters) account for 60 percent of total load. We know of another utility for which the largest 15 customers represent 20 percent of total load.

The bottom part of Figure 9 shows how various commercial building types differ in overall energy intensity for lighting and air conditioning. Facilities on the left of the graph may be good candidates for load management because they use more electricity for these end uses.

The value of certain end uses, such as air conditioning and space heating, is very time sensitive. Customers want to cool their buildings during hot summer days and heat them during cold winter days. The thermal storage implicit in these end uses make them good candidates for short-term (less than an hour or so) interruptions. On the other hand, some end uses, such as clothes washing and drying, can be shifted from one hour to another with little loss in customer service. In California, residential clothes dryers use almost 1,000 MW of load at the time of system peak (Reed 2000). Agricultural water pumping for irrigation accounts for an additional 2,200 MW. If residential and agricultural customers faced dynamic pricing, they might willingly shift these loads to lower-cost hours.

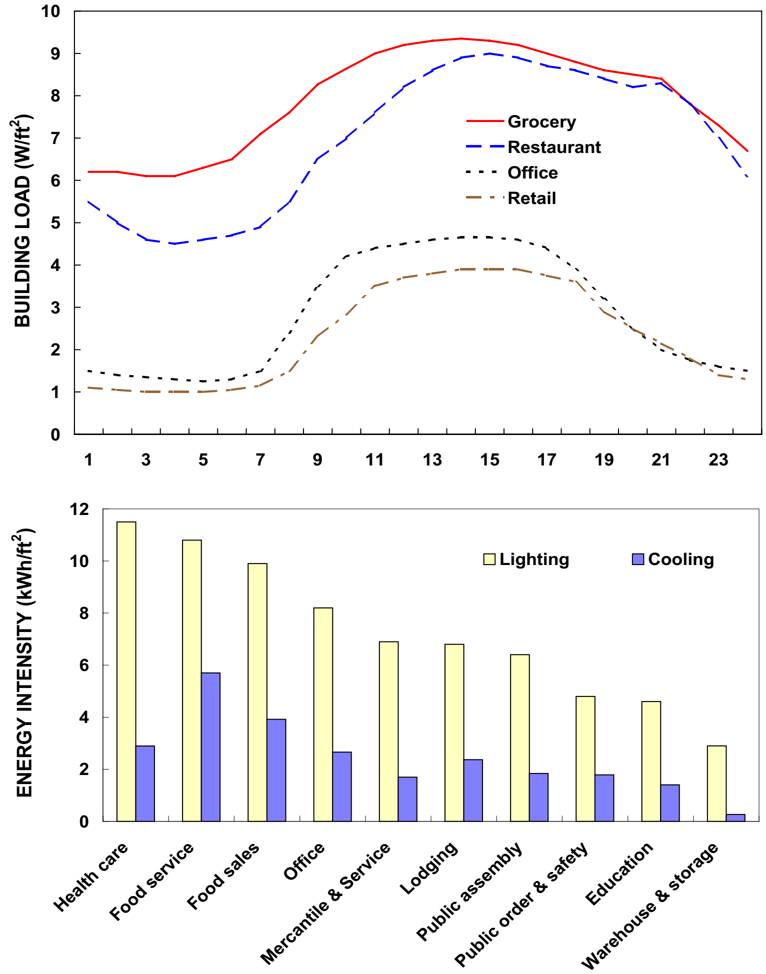



**Figure 9.** Customers differ in their ability to modify their loads, in part as a function of their daily load shape (top) and annual intensity of use (bottom).

Customers differ in the flexibility of their operations, including the speed and ease with which they can reduce load and then later increase load. Some customers may be able to interrupt loads for short periods but not for several hours. On the other hand, some customers may be able to interrupt loads economically only if they are paid for the interruption over several hours (e.g., a full shift). As shown in Figure 10, the distribution of high-price hours can be very nonuniform. In this case—the California PX prices from July 1999 through June 2000—prices exceeded \$100/MWh for 260 hours. On eight occasions, the price was above \$100/MWh for two consecutive hours. At the other extreme, the price remained above this level for 11 or more hours 11 times (including one time for 17 hours).

Finally, customers differ in whether they have automated controls for some of their electricity-using equipment. For example, many commercial facilities have building automation systems, which automatically adjust ventilation levels, chiller operations, and lighting levels as functions of outdoor conditions and facility use. Providing these control systems with information on hourly electricity prices permits automatic adjustment of equipment to save money on electricity bills and maintain occupant comfort, health, and safety.

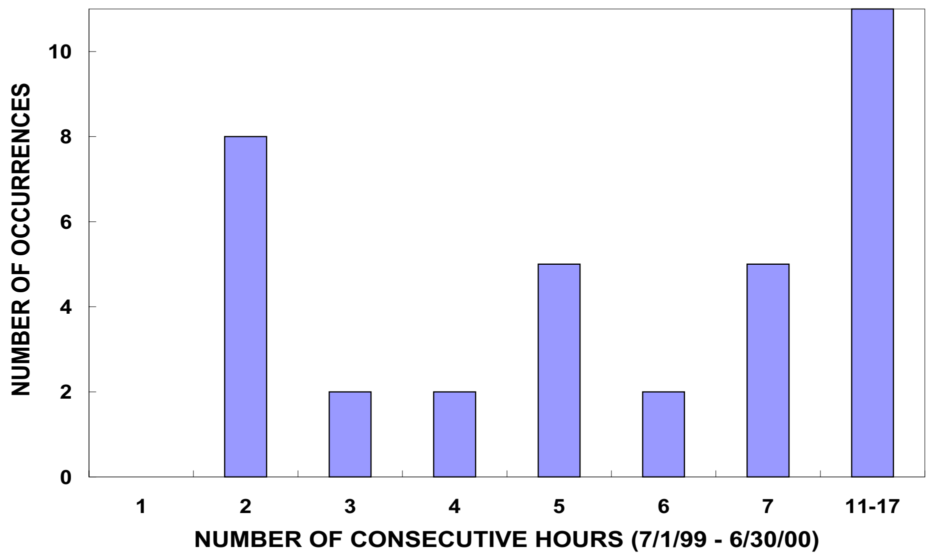



**Figure 10.** The number of times prices exceeded $100/MWh in the California Power Exchange day-ahead market as a function of the duration of these high prices.

The Marriott Marquis (with a peak demand of 6 MW) in New York City offers an interesting case study on the benefits of automatic control. This project, which involved the hotel, Consolidated Edison (the local utility), and EPRI, linked real-time price information from the utility with the hotel's energy-management system (EMS). The EMS controls hundreds of parameters to manage heating, ventilation, air conditioning, lighting, and other loads. To optimize building energy use in response to hourly prices would require either operators to make manual adjustments several times a day on set points and start/stop times for various pieces of equipment or an electronic link between the price signals and the EMS. Implementing this linkage was not trivial in the mid-1990s when this project took place (Hoffman and Renner 1997): "Software was needed at the utility to prepare the real-time prices, a communication system was needed so that the utility could 'talk RTP' with the customer, and a controller was needed at the customer site to decide how to manage loads."

The Marriott project demonstrated all these systems and their successful integration. The communication links between the hotel and utility used ordinary phone lines for two-way communication. The Marriott system included indoor air-quality sensors used to adjust ventilation levels; the system also precooled conference rooms and ballrooms as a function of hourly prices. This hotel saved $500,000 a year from this combined system of RTP and automated controls, primarily through reductions in electricity costs for air-handling units (60 percent of total), chillers for air conditioning (20 percent), and exhaust fans (16 percent). Overall, integration of RTP data with the hotel's EMS increased the load reductions from about 3 percent of total load (with manual responses) to almost 20 percent.

Automatic control (e.g., of chiller shutdown and restart) permits a facility to achieve load reductions and cost savings on all days, not just on those when prices are very high. Manually controlling equipment to respond to dynamic prices makes sense primarily when high prices occur infrequently and predictably. Automatically controlling equipment permits the building owner to save energy and money whenever prices are above their average values. Although the per-hour savings are smaller on average, the number of hours when savings occur is so much higher that automatic controls are often justified.

Customers who have either product- or thermal-storage capabilities (e.g., water-heater tanks and air-conditioned spaces) might easily provide load reductions for the 30 to 60 minutes required for contingency reserves. Such customers might find it more difficult to participate in load-reduction programs that required curtailments for a few hours.

# Chapter 5: METERING, COMMUNICATIONS, AND COMPUTING

Implementing the kinds of programs discussed here requires various technologies (Figure 11). To begin, customers must have meters that record, store, and communicate data on their electricity use at the hourly or subhourly (e.g., once every five minutes) level. Second, the LSE must be able to provide price and interruption-request information to the customer in a timely, usually automated, fashion. Similarly, the customer's interval kWh-consumption data must be transmitted periodically to the LSE. The LSE may provide additional data and analysis on the customer's electricity use and bill-saving options to the customer. Finally, based on these data and analysis, the customer needs to implement control actions, either automatically or manually, to manage electricity use and costs. This chapter discusses these technologies and how they enable customer participation in wholesale markets.

The type of program and kind of customer response determine how quickly information must be transmitted from the LSE to the customer and from the customer to the LSE. For example, if the customer is participating in a day-ahead dynamic-pricing program, the LSE needs to communicate with the customer only once a day. Indeed, the LSE can post the next day's 24 hourly prices on its website; customers can then access these data at their convenience. On the other hand, if the customer is providing nonspinning reserve, the RTO (perhaps acting through the LSE) may issue customer-specific control signals on a minute-by-minute basis in near-real time (e.g., reduce your load by 2 MW within ten minutes).

The frequency with which the customer's electricity-use data must be transmitted to the LSE also varies. Although all these programs require that the customer's load data be measured and stored hourly or subhourly, these data need not be reported to the LSE more often than once a month. For example, a customer participating in a dynamic-pricing program would need a meter that records and stores hourly consumption data. But the LSE needs that data on only a monthly basis for billing purposes. On the other hand, if the customer is participating in a load-reduction program, the LSE might want to know the load reduction as it occurs, in which case the communication from the customer's meter to the LSE needs to be either much more frequent or based on a system that permits the LSE to poll the customer's meter at any time.

If the customer participates in a program that provides it with current information on its electricity use, ways to reduce that use, and current prices, the LSE will need frequent access to the meter data so that it can analyze these data and provide the customer with the results of its analysis. Alternatively, the customer's electric meter could communicate directly with its computer or EMS. This second approach requires a meter with multiple communication ports that can send

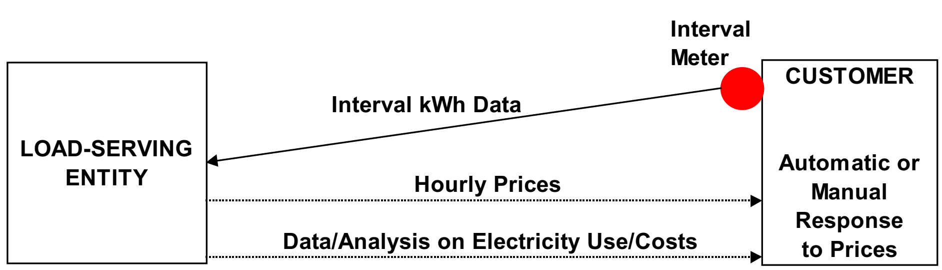



**Figure 11.** Dynamic-pricing and load-reduction programs require interval meters, communication systems to move data and instructions between the customer and its LSE, and perhaps automatic-control systems that respond to time-varying prices.

data to both the customer and the LSE and customer-side software that converts raw meter data into useful information.

Because of the factors discussed above, it is difficult to estimate the costs of these enabling technologies. Table 6 offers a broad range, suggesting an annual cost of $30 to $250 per residential or small commercial customer. For large commercial and industrial customers (which use about 100 times more electricity than residential and small commercial customers), the range in annual cost is roughly $100 to $800.

## METERS

Meters differ in what they measure and how frequently they do so (Otero-Goodwin 1999). Traditionally, utilities have deployed two kinds of meters, one for operations and the other for billing. Operational meters collect and transmit data to the control center once every few seconds but are not very accurate. Revenue meters, on the other hand, take longer to measure and record electricity use but are sufficiently accurate to use for billing. Use of revenue meters to make real-time load-control decisions is difficult because these meters report consumption only at the end of each interval.

Meters for residential and small commercial customers usually measure single-phase real energy use only. Meters for large commercial and industrial customers measure three-phase demand and energy and may also measure reactive-power consumption and power-quality characteristics. The traditional residential meter records cumulative electricity use, typically read once a month by a meter reader who walks from house to house reading the dials on each meter. Advanced meters record and store (within the meter) electricity consumption at much finer intervals of five, ten, 15, 30, or 60 clock-synchronized minutes.[19]

For residential and small commercial customers, it may be cost-effective to retrofit their existing meters. These upgrade packages include a pulse initiator (that generates an electrical pulse for every revolution of the meter disc), a data recorder (that records the number of pulses), and a communication interface. These packages permit the capture of electricity-use data at 15-minute intervals with communication of that data to the LSE on a daily or weekly basis. The cost of such upgrade packages is roughly $50 to $200.

**Table 6. Approximate costs for metering, communication, and decision-support technologies**

| | *Residential and small commercial* | *Large commercial and industrial* |
|---|---|---|
| *Interval meter* | $50 (retrofit) to $200 (new meter) | $250 to $1,000 (new meter) |
| *Communication systems (from LSE to customer and from meter to LSE) plus analysis software* | $2 to $20/metermonth | -$5 to $50/metermonth |

[19] The frequency with which the meter stores the consumption data depends on the kinds of programs the customer might participate in. A meter that records hourly data is sufficient for a customer who plans to respond to hourly prices. A customer who wants to sell spinning reserve to the system operator, on the other hand, might need data recorded once every five minutes.

For new installations, larger customers, or more sophisticated applications, a new electronic solid-state meter may be installed. Such meters range in cost from $200 to more than $3,000, depending on their capabilities. The price range is so large because these meters differ in the number of channels of data they record, the amount of data that can be stored within the meter, the number of communication ports, and the communication medium. Four manufacturers dominate this market: ABB, General Electric, Schlumberger, and Siemens.

Expanding meter functionality can spread these costs over more purposes. Some advanced meters can accept data from gas and water meters and can transmit that data to the appropriate utility. Some meters can accept inputs from customer alarms, such as those that provide notification of flooding, freezing, or meter tampering.

## COMMUNICATION SYSTEMS

### Price Communication from LSE to Customers

As with all products, customers have to know the price (or have a reasonable expectation of what the price will be) before they can make consumption decisions. Historically, fixed prices were communicated through published tariffs. This type of communication is adequate for prices that do not change or for prices that change with a fixed pattern, such as TOU rates.

Automatic price communication is required to communicate time-varying market-based prices. Because customers are generally price takers, price signals can be *broadcast* to all the applicable loads simultaneously. Therefore, the LSE need not send a unique message to each consumer.

The frequency with which prices need to be communicated depends on the market the consumer participates in. It may be acceptable to take several hours to deliver day-ahead energy prices. Real-time prices (e.g., those that are produced by the ISOs' five- or ten-minute markets) must be delivered within minutes. The communication system for transmitting such real-time prices must be automated and integrated with the ISO that operates the real-time market.

Price communications need to be reasonably secure. If the prices are public information (as with a clearing-price power exchange), eavesdropping is not a concern, but injection of a bogus price could be very disruptive.

Several technologies are used today to communicate real-time price information, including fax, telephone, e-mail, pagers, PC-to-PC communication, and the Internet. E-mail is currently the most popular choice because it is cheap and easy to automate. In some cases, the customer determines when it wants to obtain current price data, possibly downloading from a Web site. In other cases, the supplier initiates the communication either when prices change or on a regularly scheduled basis, usually by e-mail or possibly by telephone. In either case, the communication is with the customer's staff, computer, EMS, or end-use equipment (e.g., thermostat), not with the electricity meter.

Communicating price signals over the Internet (i.e., posting prices on a Web site) is cheap, easy to use, and provides broad access to customers. It is much simpler for an LSE to host a Web site than to establish telephone communications with each customer in a large group. Communications with customer personnel by telephone, pager, e-mail, or Web site is easier than communicating with customer equipment. Standards will be required on the data to be transmitted as well as the communication protocols before machine-to-machine communication becomes cheap and widespread.

### Consumption Communication from Customer Meter to LSE

The kilowatt-hour consumption data recorded by revenue meters must be transmitted to the LSE and to the customer to guide load-control actions. For the vast majority of electricity consumers, meters are read manually once a month by meter readers. Although this process is slow, it is cost effective for customers with time-invariant prices, typically costing less than $1/meter-month (Taub 2000).

Automated-meter reading (AMR), now in use for 5 to 10 percent of the nation's electricity meters, is required to record and communicate electricity-use data at the hourly or subhourly level (Nesbit 2000).[20] Two technologies dominate the AMR field: telephone and fixed radio networks (Skog 2000).[21]

Telephone lines are often used to communicate with meters, usually through a telephone modem. The data transmission can either be initiated by the LSE (called outbound communication) or by the meter (inbound communication). In both cases, the communication can be conducted with the same phone line the customer uses; in some cases, a separate phone line is installed, which increases the cost of this option. Another option is to use a dedicated cellular or satellite phone, which would substantially increase costs. This option is useful for loads that do not have standard telephone access, such as billboards, street lights, and irrigation pumps.

For inbound communication, the meter is programmed to call a phone number (either a local number or a 1-800 number) at certain times (e.g., at 2:30 am every night). The meter modems are smart enough to recognize when the phone is in use and will then place their phone call later. These modems can also recognize when the telephone is picked up while they are transmitting data to the LSE; they immediately terminate their use of the phone to permit the customer to use the phone and then resume data transmission later.

Outbound communication provides more flexibility for the LSE, permitting it to poll the meters whenever it wants to. On the other hand, the LSE now has to manage all these phone numbers and needs a ringless technology (so the customer does not hear the phone ring when the LSE calls the meter).

Typically, the phone calls last from 20 seconds to connect and transfer 24 hourly consumption values to 90 seconds to connect and transfer hourly data for a full month. These times are short if the system needs to read only a few meters, but can cause problems if the system must read thousands of meters.

Phone systems can be used for customers who are geographically dispersed. Where the loads are densely packed, fixed radio systems may be preferred.[22] In such systems, each meter has a transmitter/receiver attached to it.[23] Radio transmitters/receivers are mounted on utility poles or outside walls of buildings, situated so they can "see" the meter-mounted radio systems. The pole- or wall-mounted systems then communicate with central transmitters/receivers, which in turn send the electricity-consumption data to the LSE's central computer.

---

[20] AMR offers benefits beyond those applicable to dynamic pricing, such as more accurate meter readings, shorter lag times between meter reading and bill preparation, reduction in the need for estimated readings, and elimination of the personal hazards associated with manual reading in some locations (e.g., a dog in the backyard).

[21] Power-line carrier (PLC) communications have also been used, but most new systems use telephone or radio networks. PLC systems are expensive because they require shunts aro~êd distribution transformers, and their data-transfer rates are slow.

[22] Some AMR systems use *mobile* radio networks, in which a person walks or drives past each meter and uses a hand-held transmitter/receiver to read meter data. Such mobile systems are relatively inexpensive, but their ability to support dynamic pricing and load-reduction programs is very limited.

[23] The meter needs a transmitter so it can send kWh-consumption data to the pole-mounted unit. It needs a receiver so the system can tell the meter when to report data.

Because radio systems do not rely on phone lines, they can communicate with the meters at any time. Communications are faster with fixed radio networks than with telephone systems. For low volumes, readings can take less than ten seconds per meter. Such systems are used today to serve commercial and industrial customers. For high volumes, up to 100,000 meters can be read per hour, but this increases the processing time greatly. In most high-volume applications, the systems collect and deliver a daily read file for a million or more end points by 8 am next business day. As the market moves to more dynamic pricing scenarios, CellNet, a major provider of fixed radio networks, envisions having to deliver data within the next hour of use (Eskew 2000).

The costs of fixed radio systems depend on the number of customers connected to the system, the geographic density of the customers, the topology of the area (which affects the propagation of the radio waves), and the location of meters (indoors vs. outdoors). Fixed-radio networks are lowest in cost when all (or almost all) the customers in a particular area are served by the same communication system. These radio systems are typically owned and operated by third-party vendors, which sell the service to utilities on a dollar-per-meter-month basis. The costs vary from about $1 to $5/meter-month, depending on the frequency of meter reading and the amount of data transferred (Eskew 2000).

## ANALYSIS, DECISION, AND RESPONSE SYSTEMS

Communicating price signals from the LSE to the customer and consumption data from the customer to the LSE is, technically speaking, all that is required for customer participation in dynamic-pricing programs. But the benefits of such participation are much greater if the customer responds intelligently to these price and consumption data. Customers, based on this information might decide to:

- Do nothing and continue to consume electricity as before,
- Respond to price changes by manually adjusting consumption,
- Respond to price changes by automatically adjusting consumption, or
- Invest in new equipment and controls to expand the range of possible responses.

In some cases, consumers might like to see data and analyses of their energy bills on a monthly or quarterly basis. They may use this after-the-fact data to decide on appropriate operations and maintenance changes or capital investments that, on a long-term basis, will cost-effectively reduce their electricity bills. In such cases, the meter resolution is hourly but the communications can occur infrequently.

Other customers will choose to respond closer to real time. For these customers, LSEs may provide data and analysis tools that help customers decide what to do when. Such systems can be especially helpful to customers who decide to take manual-control actions only during the few hundred highest-priced hours during the year. Systems such as Silicon Energy's Enerscape software (and similar packages from iTron and Active Energy Management) can facilitate such decision making.

If the customer plans to respond routinely to price changes, automatic response is more practical. Johnson Controls and Honeywell market systems that integrate electricity-price data with their automated EMSs. In general, these systems use the Internet to communicate with customers, although pagers are also used. The

Honeywell system was tested at a few large facilities and demonstrated savings of 10 to 20 percent through automated response to dynamic prices (Kiernan 1999).

Carrier Electronics developed a thermostat for residential application that, upon receipt of a control or price signal from the LSE, can adjust air-conditioning or space-heating temperature settings up or down. Puget Sound Energy conducted a small Home Comfort Control pilot program during the 1999/2000 winter season (Boice 2000). The program featured a two-way radio-communication system from Schlumberger, programmable thermostats from Carrier Electronics, and energy management and Internet services from Silicon Energy. The program tested customer comfort and acceptance of 2° and 4° F decreases in temperature setting for two hours at a time, periodically invoked by the utility. Customers could, at a cost of $2 per event, override the utility setting. The pilot program demonstrated the reliable operation and integration of the technologies and very high customer satisfaction with the program.

Cannon Technologies (2000) has more than 250 load-control systems installed for all kinds of electric utilities. Its Internet-based system can implement various strategies to cycle equipment on and off, control generators and irrigation pumps, and manage other electricity-using equipment based on price changes. Cannon reduces the infrastructure burden for the LSE by creating a private load-control master station on its server. Communications from its server to the loads is via 900-MHz FLEX™ national paging. By integrating the system with real-time meter data, Cannon provides a complete system for monitoring and controlling demand anywhere in the country. Utility Data offers similar capabilities using telephone lines for communications.

Apogee Interactive operates an Internet-based demand exchange that permits customers to bid load reductions to their LSE. The PGE and GPU Energy programs discussed in Chapter 3 use the Apogee system, as do programs run by the Bonneville Power Administration, Entergy, and other utilities (Gilbert 2000). Apogee developed software to calculate customer-specific baseline load shapes (based on prior consumption, weather data, and other factors) that can be used to determine the extent of actual load reductions. Apogee has about 2,800 MW of load reduction under contract, involving more than 500 customers of all types. For the utility, this is a free call option in return for which the electricity-bill savings are shared (often 50:50) between the customer and the utility.

Future technology developments might lower the cost of communication to the end user and, especially, communications within the end-user's facility. Developments in wireless communications over short distances may make it possible to easily interconnect meters to the telephone system, the Internet, and individual electricity-using equipment throughout the facility. For example, a "smart" refrigerator might receive data on current electricity prices and, based on the contents of the refrigerator and the temperatures in the freezer and refrigerator compartments, turn off the compressor for an hour or two. Such distributed intelligence connected by a fast and inexpensive communication system should permit broader, easier, and more effective responses to time-varying electricity prices. Such systems have applications far beyond electricity, affecting natural gas use, water use, and appliance repair.

In addition, standardization can help lower the costs of advanced meters and communication systems. The tasks required of these systems are no more complicated than those performed by devices like cell phones, pagers, and calculators. To realize cost reductions requires mass production, which should not be a problem, given the existence of 125 million electric meters in the United States. Such mass production, however, requires standardization of the functions, communication protocols, and markets so that inexpensive devices can be designed and manufactured.

# Chapter 6: PROGRAM AND PRICING PROVIDERS

Chapter 3 discussed a sample of current programs, most of which are offered by the traditional utilities. This chapter discusses the potential roles of various entities, including RTOs, nonutility LSEs, and the incumbent utilities. We also discuss the actions FERC and state regulators could take to encourage greater customer participation in bulk-power markets.

## REGIONAL TRANSMISSION ORGANIZATIONS

In some respects, the RTOs are logical candidates to promote demand-side participation in their markets. They run many of the day-ahead, hour-ahead, and real-time markets for energy and ancillary services. Because they are responsible for making these markets efficient and competitive, they are interested in expanding the range and diversity of resources participating in these markets. On the other hand, the RTO markets are *wholesale*, and the issues this report discusses concern *retail* customers. The RTOs are structured to deal with a few market participants, each of which is responsible for hundreds (or at least tens) of megawatts of demand or supply. The RTOs are not set up to interact with thousands (or millions) of customers. Perhaps, the issue is whether the manager of wholesale markets has a responsibility to ensure easy access for retail customers and their providers.

As a practical matter, the existing ISOs are becoming more active in soliciting demand-side participation in their markets. The ISOs operated various small scale, experimental demand-side programs for summer 2000 (Chapter 3) and are planning expanded and improved programs for 2001. On the other hand, none of the ISOs (or the California PX) has yet investigated the reasons for the very limited participation in their hourly energy markets by price-responsive load or done much to encourage such participation.

At a minimum, the RTOs must ensure that customer loads are in no way prevented from participating in RTO markets. Just as the ISOs have developed many rules and procedures to deal with the idiosyncracies of different types of generation, so must the RTOs develop rules that accommodate differences between the loads and generators (as well as among different types of loads) that might participate in their markets. Figure 12 illustrates some of the complexities that occur in dispatching generators to

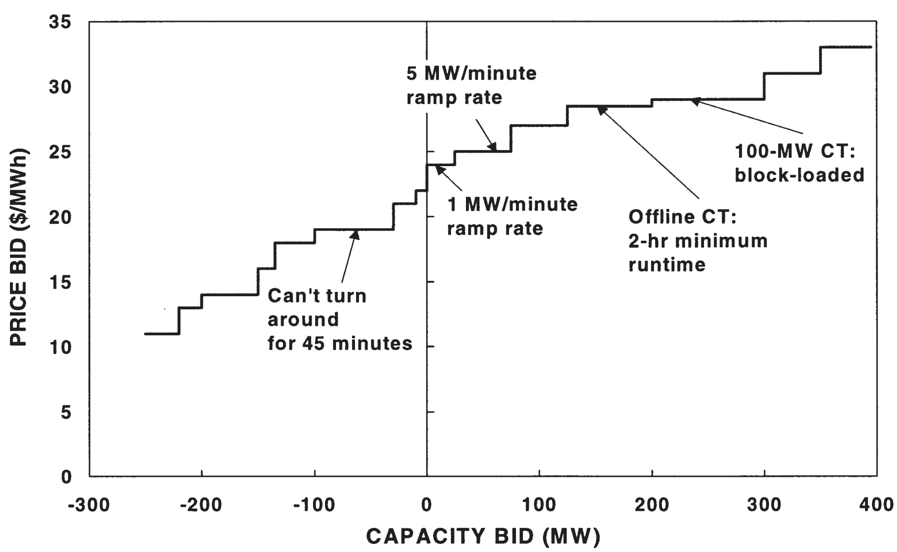



**Figure 12.** Supply bid stack for an RTO's real-time balancing market, illustrating some of the complexities caused by the operating limits of some generators.

maintain the necessary near-real-time balance between generation and load. Although it is convenient to think of a simple "stack" of resources that differ only in price, the intertemporal constraints of many generators make the optimal scheduling and dispatch solution difficult to achieve. In this example, the generators in the bid stack differ in their ramp rate, speed with which they can change direction, minimum runtime, and ability to operate at less than full output. This graph does not show the complexities associated with energy-limited hydro units, which may have daily water budgets and other environmental restrictions on their operation.

Consistent with the many rules that reflect and respect differences among generators, RTOs must accept differences among the loads that might participate in their markets. The RTOs cannot simply require that loads conform to the existing supply-side rules because those rules were developed with generators—and only generators—in mind. For example, some loads may be able to reduce their output very quickly and can therefore provide contingency reserves. However, some such loads may not be able to return to their precontingency consumption level within the RTO-prescribed time, say ten minutes. Rather than prohibiting such loads from participating in reserve markets, the RTOs should examine the basis for the ten-minute restoration rule and decide whether a longer period of reduced load might be acceptable. (Resources must be restored to their precontingency levels when directed to do so by the RTO so they can, once again, be ready to respond to the next contingency.) Perhaps a load that cannot increase consumption for at least two hours after reducing its load in response to a contingency should be permitted to participate in contingency-reserve markets but get paid a lower price than resources that can return to their precontingency state within ten minutes.[24]

The RTOs can play important roles in two other ways. First, they must ensure that their resource scheduling and dispatching software can accept bids from customer loads as well as from supply resources. Not all of the software used by today's ISOs can appropriately handle price-responsive load. Second, the RTOs should review their metering and telecommunication requirements. The typical requirement is for metering and telecommunication of generator output to the RTO control center every four to six seconds. While obtaining data from large generating units so frequently may make sense, it may not be necessary for small loads. System operators monitor the output from large generators so frequently because the failure of any one unit must be compensated for immediately. Because almost all loads are tiny compared to generators, the statistical averaging across loads reduces greatly the need to monitor the consumption of individual loads. Also, it is inherently more reliable to turn something off (e.g., interrupt a load) than to turn something on (e.g., start a combustion turbine to provide contingency reserves).

The RTOs need to be sure that the metering and telecommunications requirements they impose on participating loads are consistent with the RTO's reliability responsibilities. For example, the California ISO, in its effort to encourage load participation in its ancillary-service markets, relaxed the frequency with which individual loads must record and report their consumption from once every four seconds to once a minute (California ISO 2000a). Because these loads are required to respond within ten minutes of an ISO request, it is hard to see why the ISO would need to know the consumption levels of these loads more often than once a minute. Although the RTO needs to be able to measure the performance of loads in delivering services to the RTO, it may not need to make these measurements in real time.

[24] For example, the California Department of Water Resources has a pumping load that ranges from 1,000 to 2,600 MW (Radford 2000). The Department can turn these pumps off almost instantaneously, but it cannot meet the ancillary-service requirements of the California ISO to restore the load within ten minutes of being directed to do so. The pumps must remain off for at least 30 minutes (analogous to the minimum runtimes of some generators) and can then be turned back on at the rate of 20 MW/minute (e.g., an hour to restore 1,200 MW of pumping load).

Finally, RTOs can encourage dynamic pricing and load-reduction programs by providing appropriate accounting, billing, and settlement services. Such functions are needed to accurately track the benefits of these load reductions and assign them to the correct LSEs.

## LOAD-SERVING ENTITIES

LSEs may be the best candidates, in competitive markets, to offer a variety of risk-management options to their retail customers. They have direct contact with retail customers, they have direct contact with wholesale markets, and they have the appropriate economic incentives to manage retail loads. Because these entities bridge wholesale and retail markets, they should be in the best position to offer and profit from a full range of pricing options. Indeed, much of the customer-choice benefit of retail competition should derive from the ability of LSEs to identify and act upon customer preferences for risk management and other services. In addition, the LSEs' aggregation functions may obviate the need for the RTOs to consider the requirements and costs of communicating with thousands of small loads.

Our conversations with representatives of LSEs suggest that few are offering such pricing options today. The focus of LSE marketing efforts seems to be on building market share through the offer of price discounts. The LSEs see several obstacles to their provision of risk-management services. First and foremost, the LSEs emphasize the tremendous regulatory uncertainty about standard-offer service, provider-of-last-resort requirements, the status of incumbent utilities, and stranded-cost recovery. If, for example, the state regulator requires the local utility to provide a discounted, guaranteed rate as part of the standard offer, the LSE will have a very difficult time selling dynamic pricing to potential customers.

Second, LSEs are not sure who owns and who has rights to the metering and communication infrastructure required to implement such programs, as well as to the data these systems produce. In many states, the PUC has not yet determined whether these functions are to be provided exclusively by the distribution utility, subject to competition, or provided by any entity other than the distribution utility. Until the rights and responsibilities for meters, meter reading, communications, and data are resolved, LSEs are reluctant to invest in the equipment.

## DISTRIBUTION UTILITIES

Whether distribution utilities should offer such programs depends on whether the utility is a wires-only company or is permitted (or required) to remain in the retail-service business. Again, the role of the state regulator is paramount here. At the moment, most utilities retain retail responsibilities, such as the standard offer for those customers who do not choose another supplier, for those customers who leave another supplier, and for all customers where retail markets have not yet been opened to competition. It is for these reasons that most of the programs discussed above are operated by utilities.

In addition, distribution utilities may lose money if they encourage customers to reduce loads. This earnings loss can occur if utility revenues, but not costs, depend primarily on energy and demand charges (in ¢/kWh and $/kW-month) rather than customer charges (in $/month).

## Federal Energy Regulatory Commission

On the regulatory side, it is unclear whether FERC has any direct jurisdictional authority over demand-side participation in wholesale markets. However, FERC can ensure that RTO market rules and tariffs present no unfair obstacles to such participation. Indeed, FERC can require RTOs to fairly and fully accommodate load participation in such markets.

In addition, FERC should recognize that its approval of price caps in ISO markets can thwart demand-side participation in these markets. The programs discussed in Chapter 3 suggest that customers are unlikely to participate in voluntary load-reduction programs unless the price for such reductions exceeds about $250/MWh. (Clearly, customers who face dynamic prices do respond to prices below $250/MWh.) Thus, the California price cap of $250/MWh may effectively eliminate demand-side participation in such programs; the $1,000/MWh caps in the Northeastern markets may not inhibit demand participation.

Given the critical importance of price-responsive demand to the successful operation of competitive wholesale markets, FERC might encourage such participation. It could do so through its authority to permit generators to sell power at market-based (vs. regulated) rates, its approval of RTOs under Order 2000, its mitigation of market power, and its approval of mergers and acquisitions.

## Public Utility Commissions

PUCs can play powerful roles in encouraging demand-side participation.[25] To begin, PUCs need to decide what functions the incumbent utility may and may not perform. Specifically, is the utility a wires-only company that delivers a product from a customer's supplier to the customer's meter? Or is the utility also in the retail-service business?[26] If the latter, is the utility required to provide energy supplies to some or all of the customers in its service area? If so, what are the terms, conditions, and prices for such standard-offer service?

If PUCs require local utilities to stand ready to serve any and all customers at a standard-offer rate and if that rate includes a discount, the opportunity, both for other suppliers and consumers, to have retail loads participate in wholesale markets is greatly blunted. Any such discount, relative to the market price of generation, will ultimately be paid for, either by utility shareholders or, more likely, by customers at a later time; there is no free lunch! Other suppliers will have great difficulty competing with an unfairly discounted price. Equally important, customers will have no incentive to accept the risks associated with dynamic pricing.

Several LSEs in the northeast, where standard-offer rates are low, offer September-to-May products and then "return" retail customers to the incumbent utility for the high-price summer months. This is not how competitive markets should work and is not sustainable.

---

[25] PUCs might require all large consumers (e.g., with demands greater than 20 kW) to have interval meters. Such meters permit large customers to respond to dynamic pricing and load-reduction programs and provide a strong incentive to their LSE to manage the timing of their electricity use.

[26] If PUCs permit utilities to offer competitive retail energy services, they may require the company to be split into a regulated delivery entity (wires only) and an unregulated retail provider. As part of this separation process, PUCs need to decide whether meters and related services should remain with the wires company or be considered contestable.

If, however, the PUC establishes a standard offer that explicitly recognizes the two components of electricity—the kWh product and the risk premium—then the entity providing that standard offer can earn a reasonable return, and customers will face a reasonable price, neither too high nor too low. (If the standard-offer price is too high, customers will find better deals elsewhere; if the price is too low, customers will almost surely be required to repay these costs at a later time.) In such situations, customers will have appropriate incentives to accept some of the risk premium in exchange for a lower electricity price, and alternative suppliers will have appropriate incentives to offer such price/risk combinations.

If the utility is required to provide standard-offer service and that utility has a fuel- and purchase-power-adjustment clause, the utility has little incentive to manage price or quantity risks. (In such a situation, the utility might sign forward contracts for some of its electricity supply, but only if it believes the PUC will not second-guess its decisions through subsequent prudency reviews of those contracts.) Any increase in electricity costs is passed through to customers, although with some delay. In such situations, customers face strong incentives to manage their electricity use; however, faced with time-invariant prices and monthly bills, customers have virtually no ability to respond to these temporal changes. This is the situation that electricity consumers in San Diego faced during summer 2000.

On the other hand, if the utility is not able to recover fully its wholesale power costs and is required to provide standard-offer service, it temporarily faces all the quantity and price risks discussed above. In such cases, the utility will likely request a rate increase from the PUC, as occurred recently in California, and customers will eventually pay these costs. Correspondingly, retail customers face none of these risks in the short term. In this case, PUCs should recognize the risk-management role that the utility plays and compensate it accordingly.

Metering, communications, and billing represent another area where PUC authority is important. PUCs need to decide whether these functions should remain monopolies with the incumbent utility or be made competitive (Gromer 1999). Some metering and communication technologies offer large economies of scale (e.g., fixed radio communications), which argues for monopoly provision of these services. On the other hand, competitive markets may spur greater innovation and attention to customer service, which is especially important for rapidly evolving technologies like these. PUC indecision on this set of issues is delaying greater use of real-time metering, communications, and analysis—the infrastructure needed to support retail-customer participation in bulk-power markets.

If the PUC permits most customers to be billed on the basis of a representative load profile (i.e., with meters that read only their monthly consumption), incentives to respond to real-time prices are further dampened. "Without a retail market that has a large fraction of final demand facing an hourly price for each unit of consumption during that hour that is tied directly to the wholesale price of electricity during that hour, it is impossible for electricity suppliers to credibly bid into the day-ahead and real-time markets to reduce their demand in response to high wholesale prices" (Wolak 1999). The use of predetermined load profiles instead of interval meters eliminates all financial incentives the LSE or its customers might otherwise have to modify loads in response to price changes.

Finally, PUCs can help educate customers about the time-varying costs of electricity and the benefits to customers of adjusting to these dynamics. In the long term, LSEs will be responsible for marketing dynamic-pricing and load-reduction programs to customers, but during the transition period PUCs can play important roles.

# Chapter 7: BARRIERS

The previous chapters suggest substantial potential for price-responsive demand and large benefits, both to participating customers and to electricity consumers in general, from such behavior. Given these large benefits, why is so little happening today? And what can be done to realize more of that potential? We see three sets of obstacles: regulatory, technical, and institutional.[27]

## REGULATORY OBSTACLES

In our view, the rules that state regulators have established governing retail competition are, perhaps inadvertently, limiting customer participation in dynamic-pricing programs. Even where retail competition is not in place, PUCs can take steps to encourage consumer response to time-varying prices.

The key problem, as we see it, is PUC indecision on how to create truly competitive markets without inappropriate "protections" for retail customers. As William Weaver, the CEO of Puget Sound Energy, noted (Boschee 2000):

> If the wholesale electricity markets are ever going to rationalize themselves, it's the end users who ultimately make the buying decisions and they're the ones that must receive market price signals. Under the partial deregulatory scheme in the U.S. today, utilities insulate their customers from market price signals in most parts of the country. In effect, these utilities are providing their end use customers with a valuable hedge … .

In many states, PUCs have mandated rate discounts, imposed rate freezes, and established predetermined load profiles as part of the transition to retail competition. Perhaps uncertain about the benefits of competition to retail customers, PUCs have required the distribution utilities to offer electricity at a reduced price. In some areas, the generation component of the standard-offer retail rate is below the wholesale price, which blocks all competition. Such discounts will almost surely be paid by consumers later on. As the California ISO noted, "The retail rate freeze mutes the incentives for loads to reduce demand at times of high system load and high prices"(California ISO 2000b).

For example, the standard-offer and default rates in Massachusetts were initially set below the market price of power. As a consequence, very few retail customers selected new, competitive suppliers. The incumbent utilities wound up deferring more than $100 million, to be recovered from customers later on. More important, customers were not given the opportunity to face dynamic, market-based prices. In recognition of the lack of competition for retail customers, the Massachusetts Department of Telecommunications and Energy, in June 2000, ordered the utilities to begin charging market-based rates for default service in January 2001 (Electric Power Supply Association 2000).

Key to resolving this problem is explicit PUC recognition that the provision of fixed-price electricity includes an insurance policy as well as the electricity commodity. As mentioned several times in this report, PUCs, in establishing policies, rates, and earnings for standard-offer service, must account for these risk-management

[27] The lack of publicly available data on the potential size of various markets and on the costs and benefits of various programs are not obstacles. In competitive markets, suppliers should—and have the incentive to—determine what works and what does not.

costs. Such costs must be included in the rates customers face, and such costs must be reflected in the earnings that the provider of the standard-offer service receives for this service. Requiring the incumbent utility to offer this service at an artificially discounted price will block potential competitors from entering the retail market and will remove all incentives retail customers might have for other pricing options. PUCs should be sure the default provider service is a fair deal, for both customers and the provider, but not too good a deal for customers. A deal that is "too good to believe" today will have to be paid for later!

If PUCs impose rate caps on the local utilities, the utilities lose money if they run innovative load-reduction programs and pay for the associated metering and communication infrastructure. PUCs may want to modify their policies to encourage such programs.

PUC decisions on metering and related services (billing and access to meter data) are also critical to expanding price-responsive demand. As Gromer noted, metering is "the platform for innovative pricing and service packages"(Gromer 1999). Causey wrote: "There is tremendous regulatory uncertainty about who eventually will own the meters and related equipment and the data generated from meter reading. Thus, utilities are justifiably fearful of making an investment today that might become a stranded asset tomorrow"(Causey 1999). We believe that PUC indecision on these issues is slowing the adoption of the infrastructure technologies discussed in Chapter 5.

This lack of resolution inhibits utilities from installing these systems, for fear that regulators will either disallow these costs or that they might become stranded. LSEs are unsure whether they are permitted to install such systems. If they do install these systems, how will they recover costs if customers switch to a different energy supplier? In the meantime, what entities have access to customer-meter data? Advanced metering can occur with either a regulated monopoly or a competitive market, but it will likely not occur until regulators decide on the framework for such metering and infrastructure issues.

PUCs can encourage adoption of these dynamic-pricing and voluntary load-reduction programs regardless of whether retail markets remain regulated or are open to competition. Indeed, many of the programs described in this report are run by utilities under state regulation. A necessary prerequisite for these programs, however, is a competitive wholesale market with visible hourly prices.

At the federal level, FERC's acceptance and imposition of low price caps in the ISO and PX markets it regulates will suppress customer participation in voluntary load-reduction programs. The limited program experience to date (Chapter 3) suggests that consumers will not participate in such programs unless the price exceeds about $250/MWh. Indeed, the California ISO proposed a performance payment of $500/MWh plus a reservation payment of $20,000/MW-month for its summer 2001 Demand Relief Program, well above the current cap of $250/MWh. More generally, the greater the risk and magnitude of price spikes, the greater the incentives for price-responsive demand. As the California ISO's Market Surveillance Committee noted "Price spikes provide the economic signals for retail customers to make the investments necessary to shift their demand in response to high prices" (Wolak, Nordhaus, and Shapiro 2000).

## TECHNICAL OBSTACLES

As discussed in Chapter 5, all the technical components necessary for dynamic-pricing and voluntary load-reduction programs exist and have been applied in various settings. Unfortunately, the industry has not evolved to the point that standardized (off-the-shelf) equipment and communication packages are readily available. It seems that every program has to custom design its own enabling infrastructure. To the extent that complete systems involve components from various manufacturers (e.g., meters, communication systems, and data-analysis software), the industry may need to develop standards to ensure that the various components can work well with each other, regardless of who manufactures what.

Although the evidence we obtained on the capital and ongoing costs of these systems is sketchy, we believe there is substantial opportunity for cost reductions. In particular, as more utilities and LSEs offer such programs and the number of installations increases, the cost per customer should decline. Such cost reductions will permit the cost-effective application of these programs to smaller and smaller electricity consumers. The metering, communications, and analysis tasks are not particularly challenging. Cell phones, pagers, and calculators perform more demanding duties. If the requirements for these services can be standardized, mass-produced electronics can likely dramatically reduce the cost and increase the performance of advanced metering. This would facilitate real-time market response for even the smallest load.

## INSTITUTIONAL OBSTACLES

A variety of traditional practices are limiting widespread adoption of dynamic pricing and related competitive options. Perhaps the greatest barrier here is the widespread belief that electricity costs and prices can and should remain time invariant. At the risk of yet another repetition, we emphasize our view that consumers, suppliers, and regulators need to recognize the risk-management component of electricity pricing. A related barrier is the lack of experience with market-based prices that vary from hour to hour and with retail-customer programs that encourage economic responses to these price changes.

In addition, system operators (today's vertically integrated utilities and ISOs and tomorrow's RTOs) have traditionally focused on the supply side and ignored the demand side of the equation (by assuming, in essence, that demand is completely price inelastic). That is, they maintain reliability by managing generation and transmission assets to meet fixed customer demands. However, customer loads, as discussed throughout this report, can participate in bulk-power operations to maintain reliability and participate in commercial markets. System operators (Chapter 6) need to broaden their thinking to accommodate the unique characteristics of customer loads, just as they have done for the unique characteristics of individual generating units. Ancillary services need to be defined in terms of their functions, not with reference to the generators that traditionally provided the services. System operators should recognize the reliability benefits of using large numbers of small loads that can respond quickly.

The LSEs need to do additional market research to understand what customers want from their electricity supplier and how customers might respond to different products and services. More important, LSEs need to offer a variety of price and risk options. As discussed in Chapter 4, customers differ substantially in the magnitude of their electricity consumption, load shape, storage capabilities, automation of electricity-using processes, and other factors that affect their interest and ability to participate in dynamic-pricing programs. The LSEs and PUCs need to educate consumers about electricity, its production, costs, and alternative pricing and risk-management strategies.

Finally, not everyone benefits when loads respond to prices. In particular, generators lose money when customer demand drops in the face of high prices. Because generator earnings are very sensitive to price spikes, generator owners might object to RTO efforts to accommodate price-sensitive loads.

# Chapter 8: CONCLUSIONS

After the summer 1998 Midwest price spikes, the FERC staff wrote:

> The fact that retail customers had no incentive to adjust their usage based on price contributed to the price spike. Retail competition, coupled with the ability to respond in real time, could allow customers to see the price of the power they use and react accordingly.(U.S. FERC 1998)

Clearly, we must have customer participation in bulk-power markets; otherwise these markets will not and cannot be truly competitive, and the expected benefits of competition will not be realized.

Fortunately, not much participation is needed to realize large benefits. If even a small fraction (say 20 percent) of retail load participates and even if this load exhibits a very low price elasticity (say 0.1), the effects on price spikes, price volatility, market-power abuses, mix of new generation constructed, and overall electricity costs will be substantial.

The new breed of customer programs differs dramatically from traditional demand-side programs. Traditional programs focused on the resource needs of vertically integrated utilities and were closely overseen by PUCs. Current and emerging programs focus on customer benefits and, therefore, emphasize the timing, much more than the magnitude, of electricity use. These programs will be largely independent of regulatory oversight, market driven, and focused on customer service and benefits—helping customers cut their energy costs. Fundamentally, the new programs recognize the two-part nature of electricity, the commodity and the insurance service. Consumers can purchase insurance to protect themselves from the volatility of electricity prices. Or they can self-insure, avoid the insurance premium, and also manage their electricity use in response to time-varying prices.

There is little need to identify resource potential (in MW and cost-effectiveness) because that is the role of competitive markets. If customers face dynamic prices or pay appropriate amounts for insurance that protects them from time-varying prices, society need not be concerned with whether they respond to time-varying costs.

The notion that consumers are indifferent to electricity prices or have no interest in responding to hourly prices may be true today, given the price discounts and insulation from price volatility they enjoy thanks to PUC mandates for standard-offer service. As customers recognize the inherent volatility of electricity prices and what they pay for risk management, some will likely choose to face prices that vary from hour to hour. And some of the customers who do not want to deal with dynamic pricing might participate in voluntary load-reduction programs, in which they can, entirely at their option, cut back during occasional, very high-price hours.

The reliability benefits of customer response to market prices will become increasingly important. In the long run, the distinction between reliability and markets will diminish as system operators rely more and more on markets to obtain reliability services. Specifically, high prices will signal impending reliability problems. A focus on prices rather than on strict reliability rules will provide much more flexibility and incentive to customers to save money, automatically improving system reliability.

Although the potential benefits of dynamic pricing are large, so too are the barriers to widespread adoption. The greatest barriers are legislative and regulatory, deriving from state efforts to protect retail customers from the vagaries of competitive markets. PUCs must rethink their decisions on standard-offer requirements and competition for metering and related services. If PUCs require distribution utilities to bear the risks of price volatility through rate freezes, fixed prices, and standard-offer service with prices set too low, it will be impossible for competitive retail suppliers to offer innovative risk-management services to potential customers. Although PUCs should not force consumers to face dynamic prices, neither should they make it difficult for consumers to do so. Ultimately, PUCs need to balance their interest in promoting competitive electricity markets with their interest in "protecting" customers, especially residential customers, from the vagaries of such markets. Artificial discounts offered customers today will almost surely be repaid by those customers later on.

Another important barrier is the limited use to date of the metering, communication, computing, and control technologies needed to realize these benefits. These technology barriers will likely fall as more dynamic-pricing programs are offered throughout the country and as these functions are standardized, allowing equipment and software manufacturers to harness the benefits of mass-produced electronic systems.

The RTOs should work with LSEs and other load aggregators to simplify their data and communication requirements so that loads can fully participate in RTO markets. The RTOs should specify clearly the minimum metering and telemetry requirements consistent with reliability standards. The RTOs should review their existing market rules, in particular the special modifications that recognize the unique nature of many generators, and implement comparable rules that recognize the unique nature of many loads that might participate in their markets. Their focus should be on system reliability, not on their traditional use of generators to maintain reliability.

In summary, the convergence of retail competition, wholesale competition, and improved technologies should greatly expand the type and magnitude of price-responsive demand. Permitting and encouraging retail customers to respond to dynamic prices will improve economic efficiency, discipline market power, improve reliability, and reduce the need to build new generation and transmission facilities.

# ACKNOWLEDGMENTS

We thank Terry Black and Russell Tucker for their advice and encouragement throughout the course of this project. We thank Barbara Barkovich, Denis Bergeron, Charles Black, Terry Black, Eric Blank, Joseph Bowring, Steven Braithwait, Stu Bresler, Robert Burke, James Caldwell, Keith Casey, Richard Cowart, Laurence DeWitt, Michael Dozier, Bill Eastlake, Timothy Eskew, Ahmad Faruqui, Joel Gilbert, Charles Goldman, Daniel Griffiths, Cheryl Harrington, Steven Huntoon, Daniel Johnson, Meghan Jonee-Guinn, John Kueck, David Lawrence, Ronald Liberty, Kenneth Linder, Mathew Morey, David Moskovitz, Steven Nadel, Michael O'Sheasy, Paul Peterson, Irwin Popowsky, John Reynolds, Steven Rosenstock, Amir Shalaby, John Skog, Linda Sodemann, Rick Starr, Dean Stathis, Thomas Taylor, Tracy Terry, Susan Tierney, and Russell Tucker for providing valuable information for this project, reviewing the draft of this report, or both. Finally, we thank Fred O'Hara for editing the final report.

# REFERENCES


C. Boice 2000, personal communication, Boice Dunham Group, New York, NY, August 10.

P. Boschee 2000, "Puget Sound Energy Scores High in Service Quality While Positioning for Solid Growth," *Electric Light & Power* 78(9), 18–20, September.

Boston Pacific Co. 2000, *An Initial Analysis of Recent Wholesale Prices, Price Caps, and Their Effects on Competitive Bulk Power Markets*, Electric Power Supply Association, Washington, DC, October.

S. Braithwait 2000a, "Residential TOU Price Response in the Presence of Interactive Communication Equipment," Chapter 20 in *Pricing in Competitive Electricity Markets*, edited by A. Faruqui and K. Eakin, Kluwer Academic Publishers, Boston, MA.

S. Braithwait 2000b, "Customer Response to Market Prices — Evidence from Georgia Power RTP and Other Sources," Demand-Responsiveness Workshop sponsored by the California Power Exchange, San Francisco, CA, March 24.

S. Braithwait and M. O'Sheasy 2000, "Customer Response to Market Prices — How Much Can You Get When You Need it Most?"*EPRI International Energy Pricing Conference 2000*, Washington, DC, July 2000.

S. Bresler 2000, personal communication, PJM Interconnection, Norristown, PA, August.

California Independent System Operator 2000a, *Market Participating Load Technical Standard for Summer 2000*, Folsom, CA, February 24.

California Independent System Operator 2000b, *Request to Extend Price Caps*, Department of Market Analysis, Folsom, CA, August 10.

Cannon Technologies 2000, www.cannontech.com, Wayzata, MN, accessed September 20.

W. B. Causey 1999, "Waiting for the Call," *Electric Perspectives* 24(1), 23–36, January/February.

D. Caves, K. Eakin, and A. Faruqui 2000, "Mitigating Price Spikes in Wholesale Markets Through Market-Based Pricing in Retail Markets," *The Electricity Journal* 13(3), 13–23, April.

M. Dozier 2000, personal communication, Manager of Contracts, California Independent System Operator, Folsom, CA, August.

K. Eakin 2000, "Connecting Retail and Wholesale Electricity Markets," Edison Electric Institute Conference on RTOs and Market Design, Washington, DC, September 29.

K. Eakin and A. Faruqui 2000, "Pricing Retail Electricity: Making Money Selling a Commodity," Chapter 2 in *Pricing in Competitive Electricity Markets*, edited by A. Faruqui and K. Eakin, Kluwer Academic Publishers, Boston, MA.

Edison Electric Institute 1999, *Statistical Yearbook of the Electric Utility Industry, 1999 Edition*, Washington, DC.

Electric Power Supply Association 2000, *Retail Electric Competition: Getting it Right!*, Washington, DC, October.

Energy Information Administration 1999, *Electric Power Annual 1998, Volume II*, DOE/EIA- 0348(98)/2, U.S. Department of Energy, Washington, DC, December.

T. Eskew 2000, personal communication, Schlumberger, Norcross, GA, October 23.

J. Gilbert 2000, personal communication, Apogee Interactive, Tucker, GA, August 28.

P. Gromer 1999, *Competition Comes to Metering: Regulatory Debates and Decisions*, EIC-5, E Source, Boulder, CO, May.

E. Hirst 1999, *Bulk-Power Ancillary Services for Industrial Customers*, Electricity Consumers Resource Council, Washington, DC, June.

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

S. Hoffman and R. Renner 1997, "Taking Advantage of Real-Time Pricing," *EPRI Journal*, 22(2), 18–23, March/April.

ISO New England 2000, *Load Response Program and Changes to Market Rule 20*, Docket No. ER00-2424-000, submitted to the Federal Energy Regulatory Commission, Holyoke, MA, May 5.

M. Jonee-Guinn 2000, personal communication, Portland General Electric, Retail Energy Products and Services Department, Portland, OR, October 4.

B. Kiernan 1999, *From RTP to Dynamic Buying: Communication, Analysis, and Control Tools for Managing Risk*, EIC-7, E Source, Boulder, CO, September.

K. King and P. Shatrawka 1994, "Customer Response to Real-Time Pricing in Great Britain," 2.195–2.203, *Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, Washington, DC, August.

J. Kueck 2000, personal communication, Oak Ridge National Laboratory, Oak Ridge, TN, December 1.

J. C. Mak and B. R. Chapman 1993, "A Survey of Current Real-Time Pricing Programs," *The Electricity Journal* 6(7), 54–65, August/September.

B. Nesbit 2000, "Utilities Install Mega-Meter Systems," *Electrical World* 214(4), 28–29, July/August.

J. Newcomb and W. Byrne 1995, *Real-Time Pricing and Electric Utility Industry Restructuring: Is the Future "Out of Control?"* SIP V, E Source, Boulder, CO, April.

North American Electric Reliability Council 2000, *Reliability Assessment 2000–2009*, Princeton, NJ, October.

M. O'Sheasy 2000, personal communication, Georgia Power Company, Atlanta, GA, August 29.

C. Otero-Goodwin 1999, *When You've Gotta Have It: Options for Collecting Interval Data*, EIC-8, E Source, Boulder, CO, November.

R. H. Patrick and F. A. Wolak 1997, *Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices*, draft, Rutgers University, Newark, NJ, and Stanford University, CA, August.

PJM Interconnection 2000a, *PJM Interconnection State of the Market Report 1999*, Norristown, PA, June.

PJM Interconnection 2000b, *PJM's Customer Load Reduction Pilot Program*, Norristown, PA, June.

PJM Interconnection 2000c, *Load Data Systems, Manual M-19*, Revision 01, Norristown, PA, June 1.

Plexus Research 1995, *1994 Survey of Utility Demand-Side Programs and Services*, EPRI TR-105685, EPRI, Palo Alto, CA, November.

V. Rabl 2000, “Reinventing Load Management,” Peak Load Management Alliance Conference, New Orleans, LA, June 5–6.

B. Radford 2000, “Price Cap Follies,” *Public Utilities Fortnightly* 138(15), 4–6, August.

W. L. Reed 2000, “The Transformation of Competitive Electricity Markets by Innovative Technologies,” Sempra Energy, San Diego, CA, October.

San Diego Gas & Electric Company 2000, *Application of San Diego Gas & Electric Company (U 902-E) for Authority to Provide Customers with Real-Time Energy Meters*, Docket No. A.00-07-055, before the California Public Utilities Commission, San Diego, CA, July 31.

P. M. Schwarz, T. N. Taylor, M. Birmingham, and S. L Darden 2000, *Industrial Response to Real-Time Prices for Electricity: Short-Run and Long-Run*, University of North Carolina and Duke Power, Charlotte, NC.

T. Siegel 2000, “Comparison of Curtailable Loads,” California ISO, Folsom, CA, September 20.

J. Skog 2000, personal communication, Maintenance & Test Engineering Company, Olympia, WA, October 4.

D. Slifer 1999, “Planergy Demand Cooperative Description,” Planergy Services, Inc., Richmond, CA, October.

L. Sodemann 2000, “Wisconsin Electric’s Voluntary Load Reduction Programs,” Edison Electric Institute Conference on RTOs and Market Design, Washington, DC, September 29.

D. Stathis 2000, personal communication, Manager of Natural Gas Transactions, GPU Energy, Reading, PA, October.

S. Taub 2000, “The Information Engine,” *Electric Perspectives* 25(5), 56–71, September- October.

T. N. Taylor and P. M. Schwarz 2000, “Advance Notice of Real-Time Electricity Prices,” *Atlantic Economic Journal* 28(4), 478–488, December.

U.S. Federal Energy Regulatory Commission 1996, *Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities*; *Recovery of Stranded Costs by Public Utilities and Transmitting Utilities, Final Rule*, Docket Nos. RM95-8-000 and RM94-7-001, Order No. 888, Washington, DC, April 24.

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*, Order No. 2000, Docket No. RM99-2-000, Washington, DC, December 20.

U.S. Federal Energy Regulatory Commission 2000, *NStar Services Company, Order on Complaint and Conditionally Accepting Market Rule Revisions*, Docket No. EL00-83-000 et al., Washington, DC, July 26.

F. A. Wolak 1999, *Report on Redesign of California Real-Time Energy and Ancillary Services Markets*, Market Surveillance Committee of the California Independent System Operator, Folsom, CA, October 18.

F. A. Wolak, R. Nordhaus, and C. Shapiro 2000, *An Analysis of the June 2000 Price Spikes in the California ISO’s Energy and Ancillary Services Markets*, Market Surveillance Committee, California Independent System Operator, Folsom, CA, September 6.