THE CALIFORNIA CRISIS

LAST WINTER, AMERICANS RECEIVED A LESSON in the fundamentals of economics as blackout after blackout rolled through California. The state, which depends on natural gas-driven turbines and hydroelectric generators to provide two-thirds of its internally produced power, suffered a devastating one-two punch of an extreme drought and soaring natural gas prices brought on by a severe imbalance between supply and demand.

The resulting scarcity in electricity increased costs for the state's three major electricity distributors, Southern California Edison, Pacific Gas & Electric, and San Diego Gas & Electric. But California law limited the distributors' ability to pass the costs on to consumers. The result was demand unchecked by the cost of supply; power was cut off to consumers while the distributors incurred enormous financial losses because of the discrepancy between producer prices and consumer rates.

The crisis has subsided recently. Natural gas supplies have increased and prices have returned nearly to pre-crisis levels. Retail electricity prices have increased for some consumers in California. In turn, supplies have increased and demand is less than expected and spot wholesale electricity prices are now 4.5¢ per kilowatt-hour (kWh), more than the 3¢ per kWh average price in 1998 and '99, but less than the 37.2¢ per kWh average price during December 2000.

Why did the California crisis occur? What policy changes would prevent its recurrence? Some economists and policy analysts have answered the first question by arguing that California's big mistake was failing to deregulate more broadly. Others have claimed that generators "gamed" California's electricity auctions to extract high prices. Still others blame the lack of long-term, price-established contracts between producers and distributors.

The answers to the second question are also varied. Some believe in the necessity of real-time retail pricing. Others call for temporary price caps. And many believe that electricity is so special that its markets must be regulated.

In this special section, three articles will consider those questions and answer them. In the first article, EPRI researchers Ahmad Faruqui, Hung-po Chao, Victor Niemeyer, Jeremy Platt, and Karl Stahlkopf examine the causes of the crisis and argue that the implementation of real-time pricing—pricing in accordance with the cost of electricity at various times of the day—would go a long way toward averting future crises.

In the second article, University of Maryland, Baltimore County economist Tim Brennan analyzes seven pieces of "conventional wisdom" that developed in the wake of the California fiasco. He argues that the California market did have some features that exacerbated the effects of the natural gas and hydro shortages. If the defects were corrected, electricity might join the list of industries in which deregulation has worked. But, he admits, electricity could be the sector in which markets "meet their match" because it is crucial to the economy and an electricity system is vulnerable to even momentary inequalities of production and consumption.

Finally, in the third article, George Mason University economists Stephen Rassenti, Vernon Smith, and Bart Wilson argue that the history of regulation has created an institutional environment in which adjustment to the daily, weekly, and seasonal variation in demand is exclusively a supply responsibility. The result is an inefficient, costly, and inflexible system that has produced the recent price shocks and involuntary disruption of energy flows in California. Demand-side bidding, coupled with interruptible-service incentive contracts, can eliminate price spikes and price increases, and reduce the need for reserve supplies of generator and transmission capacity. R

"CALIFORNICA" BY MORGAN BALLARD

GETTING OUT OF THE DARK

Market-based pricing could prevent future crises.

By Ahmad Faruqui, Hung-po Chao, Victor Niemeyer, Jeremy Platt, and Karl Stahlkopf
Electric Power Research Institute

California is the nation's largest state, with a population of 34 million, and its economy produces more that all but a handful of nations on the planet. Thus, the electricity shortage that first struck the state more than a year ago caused significant hardships for both a large group of people and a thriving economy. The most severe symptoms of the crisis have now passed, but more hardships could lie ahead because the problems that gave rise to the crisis still remain.

Ahmad Faruqui is with Charles River Associates and was with the Electric Power Research Institute (EPRI) when this article was written. He can be contacted by E-mail at afaruqui@crai.com.
Hung-po Chao, Victor Niemeyer, Jeremy Platt, and **Karl Stahlkopf** are with EPRI.

California's experience has important implications for power markets worldwide. That is because many U.S. states and foreign countries have modeled their own electricity market restructuring plans after the Golden State's design, and because of the interdependence of other economies with California's. Thus, it is important for us to examine the cause and find a possible solution to that power crisis if we want other states and countries to enjoy the benefits, and avoid the problems, of electricity market restructuring.

SEEDS OF CRISIS

The basic problem underlying the California crisis was a fundamental imbalance between the steadily growing demand for power and the limited increases in generation and transmission capacity during the 1990s. That problem was made worse by an inadequate market design in which the wholesale market was based on an hourly spot market, retail price signals were not available to moderate demand, and widely available concepts of risk management were not used. Beginning in 2000, a combination of market forces and external events — including a hot summer, an extensive drought that crippled hydroelectric power production, high natural gas prices, an above-normal number of power plant outages, and rapid economic growth in places like Silicon Valley — precipitated the crisis.

Declining investment The National Energy Policy Act (NEPA) of 1992 deregulated wholesale power markets and gave states the impetus to begin moving toward retail competition. However, the ensuing transition created considerable uncertainty in both wholesale and retail markets. The uncertainty stifled investment in the power sector because it introduced ambiguity into market incentives for building generation facilities, increasing transmission grid capacity, and providing customers with better ways of managing their electricity usage.

Ultimately, reduced investment in the electric power infrastructure for more than a decade caused an imbalance between electricity supply and demand in the Golden State and the rest of the nation. From 1988 to 1998, total U.S. electricity demand grew by nearly 30 percent but the transmission network grew by only 15 percent.

The effect of inadequate infrastructure investment was especially pronounced in California, where rapid economic growth resulted in an 18-percent increase in peak electricity demand in just six years, between 1993 and 1999. During the same time period, the state's generating capacity increased by only 0.1 percent. That lack of investment in infrastructure made the state vulnerable to a power shortage.

Weather woes What pushed California over the edge was a "perfect storm" of colliding events. The first of those events is the continuing drought that has long gripped the Pacific Northwest. That drought produced a dramatic decline in hydroelectric power production, which provides 25 percent of California's installed power production capacity.

Making matters worse, 2000 featured an unusually hot summer and cold winter over much of the western region. The anomalous weather increased the demand for natural gas for both heating and electricity production, pushing gas prices upward. The early fall explosion of a major natural gas pipeline in New Mexico further worsened the imbalance between supply and demand and produced even higher gas prices. By the end of 2000 at the Henry Hub in Louisiana, gas prices had escalated from their historical value of around $2 per million Btu (/MMBtu) to roughly $10/MMBtu.

As a result, electricity prices rose throughout the West, from an average of around $30 per megawatt-hour (/MWh) in 1999 to $100/MWh or more in 2000. With California depending on natural gas-driven turbines to provide nearly 50 percent of its internally produced power, the price spike led to skyrocketing electricity costs. By January of this year, the price of a forward contract for power to be delivered in August had risen to $500/MWh. That led the California Independent System Operator (CAISO) to conclude that the state was "facing an electricity shortage of unprecedented proportions," and forecast that the state would face a 3,700-megawatt shortage in generation capacity during peak use periods.

What pushed California over the edge was a "perfect storm" of events: a hot summer and cold winter, drought, high natural gas prices, and a maket design that prevented passing higher costs on to consumers.

RESTRUCTURING'S FLAWED DESIGN

The supply problems were compounded by the design of California's restructured power market. A fundamental weakness of the state's transition to a competitive market was that wholesale and retail markets were de-coupled, and that customers were never fully engaged in the new paradigm.

In particular, customers received no market signals to encourage demand response; that is, prices did not increase in times of electricity scarcity. Because of that, customers did not have the opportunity to take advantage of new pricing and service opportunities such as time-of-use rates and contracts to sell "negawatts" (reduced demand from unused load) back to suppliers.

When California was hit by a combination of external events, the market design showed its many vulnerabilities. First, the market transition was almost entirely dependent on an hourly spot market, in accordance with state law. Second, the market organization was fragmented between a power exchange (PX) and an independent system operator (ISO). And third, the market rules lacked incentives for either demand-side participation or provision of sufficient capacity.

Last man bidding Partly due to frozen retail rates and lack of incentives for utilities to pursue innovative pricing programs, demands were bid into the market without any price elasticity. Thus, the market saw a vertical, completely price-inelastic demand curve.

When a vertical demand curve (say, with high load) and a vertical supply curve (say, with low capacity) do not intersect or even lie close to each other, it leads to a condition known as the "last man bidding" problem. In other words, suppliers are rewarded for holding their bids off the market until the last minute, when buyers are desperate and high prices can be established. When that occurs, noncompetitive behavior would arise under any auction design, whether uniform or pay-as-bid, causing the market to break down. The combination of those features produced significant price volatility and irregularity in California's power market.

CORRECTING THE PROBLEM

Over the next two years, California will be hard pressed to add sufficient generating and transmission/distribution capacity to satisfy the electricity demands of the burgeoning digital economy. That leaves the state with only one alternative in order to avoid the rolling blackouts of last winter and spring: Engage retail customers by providing them with the correct price signals. The simplest way to accomplish that is to raise retail rates and pass on a portion of the increase in wholesale power costs.

California's public utilities commission has raised rates by up to 50 percent for most customers in the state. However, the rate increase — while large in comparison to consumer expectations — may be insufficient to bring demand and supply into balance, or to restore the state's two leading investor-owned utilities to financial solvency. Moreover, there are political limits to how much retail rates can be raised. Finally, raising rates for all hours of the day and all days of the year ignores the fact that electricity costs show considerable hourly and day-to-day variation. An alternative strategy is to pursue market-based pricing programs that lower peak usage.

California has only one alternative to future rolling blackouts: Engage retail customers by providing them with the correct price signals.

Market-based programs There are a variety of market-based pricing programs that make customer demand responsive to price changes. Such programs can reduce demand rapidly, at low cost, and without adverse environmental effects. Customers who are willing to pay for the high cost of power may continue to use it at their "normal" levels, while those who are willing to lower demand or shift it to lower-cost periods benefit from lower bills.

Examples of such pricing programs include the following:

Real-time pricing (RTP), in which the customer is usually given some advance warning (generally day-ahead) of the hourly prices for a future time (typically a 24-hour period), and then he adjusts his consumption accordingly.

Coincident peak pricing, in which the hourly prices for the projected high-cost hours for a year are averaged, and the average price is applied to those peak hours (probably 100 to 300 hours). Prices for all other projected low-cost hours are similarly averaged and applied to the appropriate hours. The customer then pays the low-cost-hour price unless the energy provider notifies the customer that certain hours (say the following day) will be high-cost hours.

Time-of-use (TOU) rates, which differentiate prices by sets of hours in a day, between weekdays and weekends, and between seasons. The rates are pre-set, compared to the constantly fluctuating prices of RTP.

The theme across all market-based pricing approaches is to encourage the customer to modify operating practices and/or invest in new technologies that reduce demand during expensive peak periods. All of the approaches require the installation of some type of smart meter, often called an electronic (interval) meter, that can measure consumption during the different periods when different prices are in effect. Such a meter needs to be distinguished from the traditional spinning disk meter that has been the mainstay of the electric industry for the past half-century. Electronic meters are commercially available today but are expensive; however, their cost will decrease if enough are installed that they go into mass production.

REAL-TIME PRICING

From many vantage points, the preferred market-based pricing option is RTP of electricity. Under that arrangement, customers pay a variable price for power that moves in proportion to its price in a wholesale spot market. Often, the pricing scheme focuses on large power customers who have the necessary interval metering systems already in place, or customers in locations where it can be put in place at relatively low cost.

For example, the state of California has approved a budget of $35 million for installing or upgrading interval metering systems. Those meters will go to the state's 21,000 largest customers with demands in excess of 200 kW, representing 30 percent of peak demand and 38 percent of energy consumption. The implementation of the meters is currently underway, albeit at a slow pace because of numerous administrative difficulties.

Consumer response While unresponsive demand is often cited as a contributing factor to high prices in times of capacity constraints, skeptics have questioned whether customers will actually respond to changes in hourly prices. Fortunately, ample evidence is available of the effects of recently implemented RTP programs. That evidence confirms that several, but not all, commercial and industrial (C&I) customers respond in a consistent and predictable manner to changing hourly prices, particularly at extremely high levels.

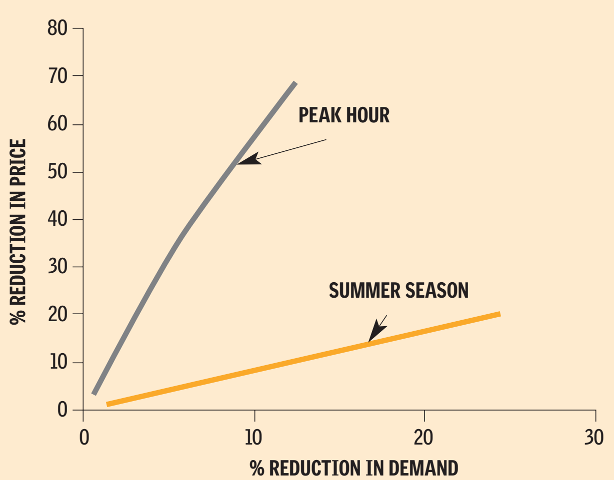
For example, Georgia Power Company operates the largest RTP program in the United States, with more than 1,600 C&I customers accounting for as much as 5,000 MW of demand. Georgia Power estimates that it achieved load reductions ranging from 400 to 750 MW on moderate- to very high-price days in 1999. One group of the most responsive Georgia Power customers reduced load by 30 percent during periods of moderately high prices of about 30¢/kWh, and 60 percent in the few hours in which prices exceeded $1.00/kWh. On average, customers in Georgia can be expected to reduce loads by about 17 percent because of the RTP program.

RTP lets customers respond by reducing their usage during expensive periods and increasing their usage during inexpensive periods, thereby lowering their total electricity costs. That would remedy a key deficiency in California's market design: the disconnection between wholesale and retail markets. In the current design, regardless of how high wholesale prices climb, customers see the same fixed retail price and therefore have no incentive to reduce consumption at peak hours. If some portion of the retail customers faced hourly prices that reflected wholesale market prices, then their demand response during periods of tight capacity and high prices would help relieve resource constraints and hold down market prices.

Figure 1

Prices and Demand

A decrease in consumption produces a decrease in price.



Simulation To show the potential benefits of RTP in California, we conducted a counterfactual simulation to determine what would have happened in the year 2000 had California implemented RTP. The simulation was performed in several steps:

First, we specified the share of the market that would be eligible for hourly pricing. Two scenarios were defined: one that only allowed the large C&I customers to participate in the program and another one that allowed both medium and large C&I customers to participate. We assumed that the large C&I customers represented 50 percent of the total C&I load, and that medium C&I customers represented another 25 percent.

Second, we specified the share of the eligible market that would choose to participate in hourly pricing. Both voluntary and mandatory designs were considered. In the voluntary design, we considered a low participation scenario in which 25 percent of the eligible customers chose to participate in RTP. We also considered a second voluntary scenario in which the participation rate rose to 50 percent. In the mandatory scenario, participation was set at 100 percent of the eligible market.

Third, we specified the likely level of customer price responsiveness. Three levels were considered, with the first being moderate responsiveness, where 50 percent of the customers are price responsive and display an average hourly elasticity of substitution (HES) of 0.053 — meaning that there would be a 5.3-percent decrease in the ratio of consumption during two hourly periods if there is a 100-percent increase in the ratio of prices during the same hourly periods. (We derive our HES values from StatsBank, compiled by EPRI, using the results of econometric studies carried out on metered hourly load shape data from 1,000 customers in the United States and the United Kingdom.) In the second case of high responsiveness, 100 percent of the customers are price responsive, and the HES is 0.135. In the third case of ultra high responsiveness, 100 percent of the customers are price responsive, and the existence of enabling technologies doubles the HES to 0.25.

Fourth, we simulated customer demand response at actual prices from the year 2000, using a nested constant elasticity of substitution (CES) model of hourly customer demand for electricity. The reduced level of demand would move the market down the cost curve, yielding reductions in wholesale market prices during the peak hour. We also estimated the slope of the cost curve by plotting pricing data from the Power Exchange against loads from the CAISO. Finally, we calculated the economic benefits that would flow to all customers from lower wholesale prices, which arise from the reduction in peak demand caused by implementation of RTP.

Results To illustrate the effect of demand response, let us consider a prototypical high-price hour, such as an hour in which prices hit the June 2000 price cap of $750/MWh. Our analysis, carried out across all scenarios and cases within scenarios, shows that customer response to hourly, market-based retail prices could reduce peak loads by 193 MW (to 5,199 MW), thereby reducing prices from peak hourly levels of $750 per MWh to $517 per MWh. For the summer 2000 season as a whole, energy costs would have been reduced on high priced days by $81 million (to $1,494 million). Thus, demand response would have resolved much of the problem in the short-term, giving policy makers and market participants a chance to put in place other measures that would solve the longer-term financial and resource issues.

Figure 1 displays a graphical summary of the results. The line entitled "peak hour" traces the relationship between progressively higher demand reductions (brought about by expanding customer eligibility, participation rate, and HES) and progressively higher price reductions. It has a slope of about five, indicating that a one-percentage point reduction in demand during the peak hour is likely to bring about a five-percentage point reduction in peak hourly prices.

The line entitled "summer season" shows the corresponding relationship between progressively higher demand reductions and progressively higher reductions in total energy cost. It has a slope of about 0.9, indicating that a one-percentage point reduction in demand over the summer season is likely to reduce costs by about 0.9 percent. As expected, demand reductions during peak hours have a greater impact on energy prices and costs than demand reductions during high-priced hours over the entire season. The results illustrate the potential benefits of hourly pricing in a tight power market.

Technology Market-based interruptible programs, as well as RTP programs, will benefit from the application of new monitoring, energy control, and communication technologies. Technologies that improve the ability of customers to respond to hourly prices in an automated fashion — with predetermined strategies that cause minimal disruption — will enhance demand response and customer benefits.

The Automated Energy Control System (AECS) is an example of the effect of enabling technology. The program allows a commercial building to modulate its energy usage in response to a real time price signal. All major circuits inside the building are connected to receive instructions from AECS. It uses modern digital technology to bring the benefits of RTP to customers.

By adopting RTP or other market-based pricing programs, California would offer consumers considerable incentive to make more efficient use of electricity.

OTHER PRICING INCENTIVES

There is a potential incentive for many more customers to respond as if they face hourly prices, even if they do not do so contractually. The incentive, which has not been appreciated to date, arises as more customers have their energy usage metered on an hourly basis. When hourly metered data are available, suppliers will be able to calculate the cost to serve customers for a recent historical period, and adjust their price offers accordingly (i.e., charge higher prices to those customers whose usage tends to be high during high-cost hours).

When customers understand the process, they will have an incentive to reduce load during high-price periods even if they do not face hourly prices explicitly. In a recent step in that direction, Puget Sound Energy has begun to install hourly meters for nearly half of its customers and is providing information on hourly energy costs. However, the company is not yet billing customers at hourly prices.

So why not? Clearly, it appears that there are significant benefits to be gained from hourly pricing, and nearly everyone agrees that price-responsive demand is a key part of the process of mending California's broken energy market. That should lead us to ask, Why has hourly pricing not been adopted to any degree? A number of barriers and misconceptions appear to be delaying the move toward implementing hourly pricing.

The most notable problem arises from lack of experience with hourly pricing. Operating an hourly pricing program requires hardware and software for communicating prices to customers (e.g. on a day-ahead basis), metering customers' energy consumption on an hourly basis, and billing customers based on hourly prices and usage values. The major California utilities ran pilot RTP programs about two decades ago, but then discontinued them. However, some of the key features are in place (e.g., most large C&I customers already have hourly meters installed), and expertise on implementing RTP is available from other states.

CONCLUSION

A number of factors contributed to California's recent energy crisis. But the design of the market, which prohibited the passing of appropriate price signals on to consumers, made the crisis considerably worse than it needed to be. By adopting RTP or other market-based pricing programs, California would offer consumers considerable incentive to make more efficient use of electricity. That would enable the state and its residents to take fuller advantage of the benefits, and avoid the difficulties, of electricity market restructuring. R

QUESTIONING THE CONVENTIONAL "WISDOM"

The causes and solutions to the California crisis are not as simple as some say.

By Tim Brennan
University of Maryland, Baltimore County

Deregulating an industry, like flying an airplane, makes headlines only when there has been a crash. Opening electricity markets, formerly a topic of interest to only a few aficionados, has become a major news story thanks to the California experience.

Ideally, the problems with high prices, rolling blackouts, bankrupted utilities, state bailouts, and allegations of anticompetitive conduct would be educational, providing lessons that other states could use to realize the promise that opening electricity markets may bring. The fear is that a bungled experience, at least viewed in hindsight, might dissuade us from opening other electricity markets, even when that restructuring promises significant benefits.

There is little dispute that the California crisis began with demand outstripping supply in the western United States in the summer of 2000. That problem was quickly followed and overshadowed by financial catastrophe, as the utilities responsible for selling electricity were forced to buy all of their power at wholesale prices many times higher than what California law permitted them to charge at retail. Political paralysis followed as lawmakers were unable to decide how to distribute the mounting, multibillion-dollar debt among utility stockholders and creditors (through bankruptcy), customers (through higher rates), taxpayers (through assorted subsidies and bailouts), and generators (through federally-ordered rebates).

Tim Brennan is a professor of policy sciences and economics at the University of Maryland, Baltimore County and a senior fellow with Resources for the Future. He is co-author with Karen Palmer and Salvador Martinez of the forthcoming book *Alternating Currents: Electricity and Public Policy*. Brennan can be contacted by E-mail at brennan@umbc.edu.

CONVENTIONAL WISDOM

Many tellers of the above tale have added some embellishments to the story. Among the most frequently heard are:

- California's big mistake was failing to deregulate more broadly.
- Generators "gamed" California's electricity auctions to extract high prices.
- The lack of "real-time" meters precluded consumers from seeing high peak prices, removing an incentive to conserve power and reschedule uses when electricity is more plentiful.
- The state's requirement that distribution utilities buy power from the California Power Exchange (PX) discouraged them from insuring against high wholesale prices via long-term contracts.
- Through either collective action or unilateral conduct, generators exercised market power, reflected by prices substantially above the competitive level.
- Temporary imposition of wholesale price caps would smooth out transitional bumps toward a competitive electricity industry.
- But for the mistakes made by California, electricity markets will work.

Those assertions, part of the increasing body of "conventional wisdom" regarding electricity policy, may have some merit. But, as we will see, that merit is not without qualification.

"CALIFORNIA'S DEREGULATION DID NOT GO FAR ENOUGH"

Without doubt, the primary factor that turned a tight electricity market in California into a financial and political disaster was the continuation of retail price regulation while wholesale prices were free to fluctuate. It prevented rates from going up during the peak summer demand season in 2000, thus sustaining inelastic demands that overtaxed the system and drove up wholesale prices.

But the most significant effect of retail price controls was in conjunction with the steep increase in wholesale prices. In June of 2000 when wholesale rates rose to five or more times their prior levels, the distribution utilities started spending more money than they were taking in because they were forced to buy power through the PX at extremely high rates.

Partial deregulation Admittedly, in California, "electricity deregulation" was something of an oxymoron. Even without the financial carnage brought by forcing retailers to sell low when they had to buy high, the state ought to lift retail price controls as well as wholesale price controls. But whether reality meets the ideal depends on a host of structural policies. That is because, at this stage of the electricity game, flash-cut deregulation of the entire sector is probably not a good idea.

Among the industries deregulated in recent years, electricity resembles telecommunications more than others in that only some sectors of the telecommunications and electricity industries are amenable to ready deregulation. Telecommunications deregulation focused on competitive markets in long distance service, customer premises equipment, and information services, but local telephony remained a regulated monopoly. In electricity, only the generation and marketing sectors are ripe for deregulation. In those sectors, scale economies appear sufficiently small relative to the size of the market, allowing multiple vendors to compete.

Wires The same cannot be said of the "wires" sectors — local distribution and long-distance transmission. Local distribution is a monopoly service largely because one set of lines, poles, and conduits can supply the electricity that consumers are likely to demand. Overbuilding a competitive local grid would be wasteful.

Were it simply a matter of scale economies, competition in long-distance electricity transmission could match that which we have in long-distance telephone service. But long-distance electricity transmission differs from the telephone model. Transmission lines of different companies are interconnected so that power can be sent in any direction to alleviate regional shortages. However, it is prohibitively costly to "route" electricity onto selected interconnected lines. Interconnection and inability to route electricity combine to create "loop flow" or "parallel flow" — that is, electricity going from Generator A to Distributor B travels on every open transmission path between the two points. One utility's ability to transmit electricity depends on the capacity of lines owned by others. Despite having separately owned components, the transmission grid is, functionally, a single economic unit.

Until "distributed generation," i.e., producing electricity on the users' premises, becomes more economical, local distribution and long-distance transmission are likely to be regulated for the foreseeable future. The electricity industry, then, will be only partially deregulated. As we saw in telecommunications, partial deregulation in an industry creates special problems in managing the relationship between the regulated and unregulated sectors, particularly if utilities remain vertically integrated across the boundary.

Quarantining monopoly One concern in full deregulation of an industry like telecommunications or electricity is cross-subsidization — that is, a firm shifting some of its costs from unregulated production to regulated operations. That would raise the firm's regulated rates and create an artificial competitive advantage in its competitive markets. A second is discrimination, in which the firm gives its competitors in the unregulated markets delayed or inferior access to its regulated service. In electricity, discrimination could appear as a transmission grid owner providing better and timelier line capacity to its own generators than it gives to generators owned by its competitors. The result can be monopolization of the nominally unregulated market, in which the firm captures from consumers the profits that regulation of its natural monopoly market was intended to suppress.

The 1984 divestiture by AT&T of its local telephone companies was predicated on the virtue of separating control of the monopoly service from ownership of competitive enterprises to prevent discrimination and cross-subsidization. The analogue in electricity has been the less draconian formation of independent system operators (ISOs) and regional transmission operators (RTOs), which allow generation companies to keep transmission facilities while vesting control in nominally independent nonprofit boards.

In California, and in most other jurisdictions that have opened electricity markets to competition, the incumbent local distribution utility has retained a near-monopoly in retailing electric power. The cleanest solution to extend deregulation through to retail would be to get the distribution utility out of the retailing business, and instead have it provide local distribution service to independent retailers. The retailers could then compete on price, after purchasing generation in the wholesale market and paying regulated rates for transmission and distribution. Whether the retailers procure electricity on spot markets or long-term contracts, or are themselves generators, would then be a matter for markets, rather than regulators, to decide.

"CALIFORNIA'S MARKET INSTITUTIONS WERE INHERENTLY FLAWED"

Any analysis of California's problems has to recognize that, from the standpoint of prices and reliability, restructuring in the Golden State seemed to work well for over two years. Data from the California ISO indicate that, until June of 2000, electricity prices remained fairly low. Wholesale prices ranged roughly between 1¢ and 3.5¢ per kWh off-peak, with peak prices being roughly a penny higher.

Any inferences regarding the causes of the California crisis have to recognize that opening wholesale markets to competition in and of itself, and California's process in particular, cannot be the culprit. Other regions that opened electricity markets, such as the mid-Atlantic "PJM" market, have had good performance.

Gaming the auction While the data do not establish that deregulation in general and California's rules specifically led to high prices, the state's market mechanism may have made a bad situation worse. The California PX employed an auction in which generators could bid in "supply curves" with up to 16 price-quantity pairs. Everyone would be paid the amount needed to get supply sufficient to meet demand. In theory, generators would have little reason to act strategically because each gets the market-clearing price.

But in practice, each generator might have an incentive to bid in a little bit of power at a very high price. If such a bid is not taken, the generator does not lose much — only the profits from the small amount of sales. If the bid is taken, however, the generator could reap a windfall in that it receives a high price on all of its output, not just the amount bid in at the high price.

To see how that differs from ordinary markets, imagine a car dealer considering putting a very high price on one of his cars, on the chance that a desperate customer would be willing to pay for it. The tactic seems unprofitable: The car would almost certainly remain unsold. But suppose that, because of some rule governing the automobile market, if the car dealer makes the high-price sale, the dealer could then charge the same high price for every other car it has already sold. Posting the high sticker price for one car would look more promising. The car might go unsold, but the expected benefits from setting the high price are considerably greater. With inelastic demand and the ability to charge a high price to all buyers, the California electricity market may resemble the second scenario.

The incentive to "game the system" via low-volume high bids need not pan out. Each generator may prefer to free ride, letting someone else set the high bid — so no one does it. Getting high bids requires some assumption that the benefit-cost ratio of the small high bid is so great that it is worth doing unilaterally even if no one else goes along. In addition, to get gaming, the auction may have to require a minimum amount a generator must offer at any price it bids, yet not so large as to render significant the losses if that electricity goes unsold.

Redesigning the auction could mitigate gaming. Paying generators only their bid price rather than the market-clearing price would eliminate the gains from gaming. But it would also substitute incentives to raise the bid price to some extent on all power, which would likely raise prices overall. (See "Electricity Markets: Should the Rest of the World Adopt the United Kingdom's Reforms?" *Regulation*, Vol. 22, No. 4, Fall 1999.) A better response would be to increase minimum bid quantities, to raise the cost of putting in a high bid if that quantity goes unsold most of the time. In the extreme, one could force generators to offer all of their power at the same price.

Concerns about gaming suggest a more extreme remedy — eliminate central auctions for all but the ancillary power services and emergency procurement necessary to maintain reliability. As we discuss below, some grid management is likely to be necessary. But running all power transactions through a central market seems a holdover of a regulatory mindset that refuses to trust electricity markets to dispatch the least-cost generators first.

"WE NEED MORE REAL-TIME PRICING"

Many commentators have noted that a contributing factor to the California electricity crisis was that consumers have too little incentive to conserve power during peak periods. That is because they do not see a "real-time price" equal to the cost of electricity at the time they use it. As long as consumers pay the same price for electricity regardless of when they use it, they will have too little incentive to conserve at times of the day when power plants are straining to keep up with demand.

Imagine that the average electricity price is $50 per MWh, but the peak period cost is $300 per MWh. Further, suppose that turning up the thermostat a few degrees will reduce use of a five-kilowatt air conditioner by two hours a day during peak-demand August afternoons. A consumer paying $300/MWh during those hours would reduce her bill by about $90 that month, but she would save only $15 if she were charged the standard price. One can multiply those figures many times over to estimate the savings to a commercial landlord.

Metering Retail regulation need not be the impediment. In principle, regulators could set real-time prices equal to wholesale prices plus rates for transmission and distribution. Whether or not one regulates retail rates, one cannot have real-time pricing without some method for measuring how much power consumers use at any given time rather than total use over the month, as with conventional meters.

Advocates of conservation and improving market performance thus look to real-time meters as a way to make consumers more responsive to electricity prices. But to say that real-time meters are a good thing is not to say that we need more of them. Their benefits may be positive, but that could be counterbalanced by the cost of producing and installing them. To justify increasing their use, we would want to know what is the externality, i.e., why the market currently provides too little incentive to install real-time meters. One would think that if a power company is paying $300/MWh for electricity it can sell for only a fifth of its wholesale price, the company would already be willing to pay the customer to meter use and promote conservation.

For the economy at large, real-time pricing is not a prime policy objective. As anyone who has waited in line to get into a popular restaurant knows, electricity is not the only market lacking real-time pricing. Should the government encourage restaurants to offer "early bird" or "pre-theater" discounts for dinners at 5:30 p.m., to cut down on the lines at 7:30 p.m.? One would not think so — restaurant owners presumably can balance for themselves the benefits of minimizing peak-period congestion against the costs of offering different menus.

Real-time pricing externalities depend on the benefits of setting prices at, rather than below, market-clearing levels. If sellers have to meet all demand at the too-low peak price, real-time pricing would reduce inefficient overproduction. But there is no externality. Sellers who have to supply electricity at prices below marginal cost have the correct incentive to pay buyers to reduce purchases at peak periods, including covering the cost of installing real-time meters to measure such use.

But suppose low prices result in rationing, when the amount demanded at that low price exceeds the amount supplied. In that case, there can be a positive externality from installing real-time meters when the underlying product is being rationed inefficiently, in that some who get the product have a lower willingness to pay for it than some who do not. With inefficient rationing, the economic cost in not getting prices right is compounded by a misallocation of the goods supplied. If a low-valuing user and a generator cut a deal in which the former agrees to adopt real-time pricing, positive benefits will accrue to a high-valuing user able to get more electricity as a result.

To continue the restaurant analogy, suppose I really want to get into the restaurant but am back in the queue with no way to pay to be in front. I get positive benefits if those in front of me accept a technology that gets them to adopt real-time restaurant pricing, if some of them leave and I can move up. Note that the externality does not hold with efficient rationing. If those with the highest willingness to pay are at the front of the line, there is no additional inefficiency created by rationing, apart from the reduction in output itself.

Government involvement Policies to promote real-time metering through subsidized installation or as a requirement for opening electricity markets may be warranted if the alternative is blackouts. Those who benefit include not just the power company and customer who agree to install a meter, but others for whom conservation reduces the likelihood of a blackout. In addition, because blackouts cannot be targeted on a consumer-by-consumer basis very easily, an individual consumer cannot easily buy his way out of avoiding blackouts by agreeing to install a real-time meter and pay a high peak price in exchange for not being rationed. Hence, the interest of an individual generator and its customers in adopting real-time pricing may be too small because they can do too little to guarantee a steady flow of power.

Similar arguments apply to substitutes for real-time pricing that reduce electricity use during peak periods, such as interruptible service contracts or demand-side management programs. Whether the size of the externality warrants significant subsidies for real-time meters, or more radical measures such as delaying open markets altogether, remains to be seen. The latter may be particularly dubious if the alternative is to maintain time-independent regulated rates.

"LONG-TERM CONTRACTING WOULD BE A BIG HELP"

Another idiosyncrasy of the California experience was the requirement that distribution companies purchase wholesale power exclusively through the PX. With more flexibility, the distributors might have been better able to hedge against high wholesale prices. If so, the distributors and the state of California would not be suffering the financial and political strains associated with bankruptcies and bailouts.

But it is important not to let hindsight exaggerate the potential benefits of long-term contracts going forward. Such contracts are equivalent to purchasing insurance from generators against high prices in the future. Looking at the absence of such contracts after the fact differs little from observing that someone whose house just burned down should have bought fire insurance beforehand.

Increasing entry The salient question is whether the absence of long-term contracts discourages electricity production. If fire insurance were impossible, people would buy fewer houses. Inability to spread risk efficiently in electricity markets could be important, as wholesale price volatility in California has undoubtedly increased the demand for more flexibility in sharing risk. But it is hard to see how the absence of such contracts discouraged significant supply in California. Buyers had protection: retail rate regulation. The utilities had an obligation to serve them, short of — and probably regardless of — bankruptcy. Generators would be deterred by the prospect of low prices, not high prices.

The bankruptcy disaster misleadingly can suggest that long-term contracts lead to lower prices. But, except for the benefits of avoiding the risk of low prices, a generator will not sell power on a long-term basis for substantially less than it would expect to receive in "spot" sales. The huge difference between what a long-term wholesale electricity contract might have gone for two years ago compared to current prices is unlikely to be replicated.

The only solution to high prices is more entry. But were such entry forthcoming, prices would fall absent contracts. New generators would enter if the expected profits (adjusted for risk) earned during the peak period would cover their capital costs. One would expect that off-peak power prices would be depressed, perhaps down to operating costs with little to no capital recovery. The situation is similar to that of resort hotels, which cover operating costs during the off-season and recover their capital costs during the peak season.

Other dangers Moreover, long-term contracting is not without downsides. As with other forms of insurance, it creates a potential for moral hazard. Here, it would arise as greater consumption of electricity at the lower contract price below the marginal cost of generating electricity. Just as fire insurance can lead to more fires, long-term contracts could encourage electricity consumption when we want to discourage it. In addition, long-term contracts between distribution utilities and generation companies may establish undesirable linkages between the competitive generation sector and regulated monopoly distribution. As noted above, restricting distribution companies to the passive business of carrying power to consumers would mitigate those concerns, leaving independent, competitive retail companies free to balance the costs and benefits of long-term contracts or vertical integration with generators.

Of course, we should not prohibit long-term contracting between retailers and generators. They should be free to spread risk as they see fit. But we ought not rely on long-term contracting to alleviate the supply and demand imbalances underlying the California crisis.

"GENERATORS EXERCISED MARKET POWER"

Perhaps the most controversial claim regarding the California situation is that it was the result of the exercise of market power: Generators intentionally reduced electricity supply to drive up the price.

Market power accusations are not new. In March of 2000, the Market Surveillance Committee of the California ISO found that California's electricity markets had not been "workably competitive" during the summers of 1998 and 1999, when prices were estimated to have exceeded competitive levels by less than 20 percent. Ironically, if generators exercised market power prior to June 2000, something else was responsible for causing prices to increase by factors of five or more above prior prevailing levels. That does not mean that market power played no role in wholesale electricity price increases from June 2000 onward. It is not a defense against antitrust accusations that the alleged violation did not happen earlier. But it does call into question any view that deregulation of power markets ipso facto led to the exercise of market power.

Withholding supply Market power can be exercised either collusively or unilaterally. The unilateral exercise of market power is not illegal, but antitrust law prohibits collusion.

As commodities go, electricity might be a relatively good candidate for collusion; it lacks differentiating characteristics that would make it hard to fix its price. However, the number of competitors in the California wholesale market would make collusion unlikely. In addition, only a careless cartel goes beyond raising prices to inviting headlines and state and federal investigations brought about by blackouts and bankruptcies.

A more likely concern is that generators unilaterally might have found it worthwhile to withhold supplies to raise price. The demand for electricity is quite inelastic — especially if regulation or the absence or real-time metering insulates consumers from wholesale prices. At peak demand periods (when supply is inelastic as well), oligopoly and dominant firm models suggest one could observe prices double the competitive level, even if the market looks competitive by conventional standards, e.g., in having several substantial independent suppliers.

What do studies show? Numerous empirical studies have found very high markups of price over cost in electricity prices, even taking into account increased natural gas costs and environmental regulations. But such studies are not yet conclusive. On the price side of the equation, one has to keep in mind that it is not the price one sets, but the price one gets. Utility bankruptcy and court requirements forcing generators to sell are reminders that prices may have been inflated to take the risk of nonpayment into account.

More important differences lie on the cost side. The prevailing method for measuring price-cost margins or markups is to compare price with a measure of the average variable cost of the highest-cost producer in the market. For markets with excess capacity, that approach is reasonable. However, prices during peak periods have to be sufficiently high so that generators who come on line to meet that demand cover not only their operating costs but their capital costs as well. Again, consider resort hotels, where room rates are far above operating costs during the peak tourist season to supply the revenue necessary for covering the cost of the hotel. A plant that is online only one percent of the time (80-90 hours per year) to meet extreme peak demand has to recover all of its capital costs a hundred times more quickly than a baseload plant running all of the time. Opening markets may be only revealing the very high cost of peak power that we never saw when it was averaged into regulated rates.

Questions about the empirical basis for claims of market power do not prove that it was not exercised. Theory suggests that, at least in peak periods, it would not be surprising if generators unilaterally found it profitable to reduce output. But those issues do suggest that one might be cautious in intervening in wholesale power markets to deal with alleged market power prices.

"WHOLESALE PRICE CAPS CAN RESOLVE A TEMPORARY PREDICAMENT"

Few are likely to be surprised at the call for wholesale electricity price caps for California, and not many more are likely to be surprised that the Federal Energy Regulatory Commission (FERC) — with statutory authority over wholesale electricity prices — adopted so-called "soft caps" in response to the crisis. Under the caps, when reserve generation capacity is less than seven percent of supplies, generators would have to sell power below a calculated "market clearing" price based on FERC's estimate of the "marginal cost of the last unit dispatched, or justify prices charged above this level." When reserve capacity is greater than seven percent, prices would be capped at 85 percent of the price FERC calculated during the most recent preceding supply crunch.

The effects of caps In competitive markets, price caps that bind will predictably reduce supply and create excess demand. On the other hand, if generators either are unilaterally exploiting market power or gaming the auction, a price cap can increase supply by reducing the incentive to raise prices by withholding electricity. (Antitrust laws already prohibit collusion.)

The main impetus behind price caps, however, is not to improve market efficiency but to reduce the transfer of wealth from California's taxpayers, distribution utilities, and ratepayers to the generation companies. Some of the transfer is the result of deregulation. Under traditional cost-of-service regulation, additional revenues paid by users for peak power cover just the costs of the marginal plant.

By contrast, in competitive markets when marginal costs are high, everyone in the market — not only the marginal firm — gets to charge a high price. In open electricity markets, substantial wealth will flow during peak periods from consumers to producers, because the baseload producers can charge prices considerably above their operating costs.

Political effects Having all suppliers charge prices sufficient to induce marginal supply is efficient. The marginal opportunity cost of a megawatt from a low-cost baseload plant equals the cost of replacing that megawatt from a high-cost peak-load plant. But the transfer is likely to be politically upsetting. The initial symptom of the California crisis was not blackouts or bankruptcy but the political turmoil associated with higher retail rates in San Diego in the summer of 2000, during a three-month window in which retail rates were not regulated.

Over time, free entry would fix the problem. Added generation would come in, depressing prices off-peak until overall expected revenues covered capital and operating costs. But the political process may not exhibit sufficient patience. If redistribution through windfall profits taxes is not feasible, it may be politically appealing to trade lower prices today for reduced supplies now and in the future. That is particularly true if, as some suggest, the caps are sufficient to encourage most of the entry we would have gotten at genuine market-clearing prices. Moreover, price caps could be efficient if most consumers would prefer lower prices (and whatever increases in expected blackouts that accompany them) to the degree of reliability set by regulators and the higher prices required to achieve it.

Problems with caps Even if price caps theoretically discourage the exercise of market power and auction gaming, or enable a politically or economically preferable trade-off between price and reliability, their practical implementation creates difficulties. The first is a price standard based on operating costs. As noted in the previous section, such a standard is inappropriate because the marginal entrant needs to expect to recover fixed and variable costs. FERC illustrates that mistake in its price-cap order, in which it claims that generation companies can use "the amounts earned on the more efficient plants [to] cover the investment in the marginal plant." FERC's error is that no generation company will have an incentive to build a marginal plant if the money to pay for it has to come from profits from more efficient plants.

The second apprehension with price cap proposals is the argument that they are temporary. Without a very clear indication of what makes California different, those calls are tantamount to saying that competition works only in the easy case, when multiple sellers have substantial excess capacity. Some possible temporary aspects of the California situation that might be relevant are getting more long-term contracts (reducing incentives to raise peak period spot prices) or real-time pricing (reducing demand and profits when prices go up). But, as we saw above, neither is a guaranteed or incontrovertible fix.

If caps are imposed whenever electricity becomes scarce, we essentially have re-regulated the industry. Theoretical and empirical analyses indicating that even relatively unconcentrated electricity markets may lead to very high prices may suggest that some regulation of wholesale prices may be warranted. If so, price caps should not be endorsed or adopted under the pretense of being "temporary." If the California experience teaches us that intolerable levels of market power are inevitable — as scale economies in generation set lower bounds on market concentration — it would be hard to avoid the conclusion that completely unfettered competition is untenable.

"BUT FOR CALIFORNIA'S MISTAKES, ELECTRICITY DEREGULATION WORKS"

The most important cautionary note in emphasizing the severity of the California crisis is ironic or perhaps paradoxical — that it may give a false sense of security regarding the merits of opening retail markets to competition. Most of the problems in California are among those we know how to solve or prevent. Supply and demand imbalances may remain, perhaps even more intensely, were regulation to continue. Mistakes in the design of residual regulation and centralized auctions can be avoided. Antitrust laws, with additional deconcentration policies and backstop price ceilings if necessary, can deal with market power.

Thus it seems that, if only California had gotten it right, electricity competition would bring the same benefits that competition has brought to banking, transportation, and telecommunications in the last 20 years. However, even with the best institutions and intentions, "getting it right" is not so easy.

Reliability Profound difficulties in deregulation arise because of electricity's unique combination of three crucial properties:

- It is crucial to the economy.
- An electricity system is vulnerable to even momentary inequalities of production and consumption because electricity cannot be stored economically.
- The market is so interrelated that, if one supplier fails to produce enough electricity to meet its customers' demands, all customers on the grid may be blacked out.

Hence, some degree of coordination (either through explicit cooperation, regulation, or legal incentives) or centralized management is necessary to ensure that one supplier's imbalance does not bring down the entire system.

Compatibility with competition A first question is whether competition is compatible with the centralized control necessary to prevent systemic breakdowns. Excepting the California crisis, system failures in the United States have almost always been local, e.g., when lightning hits a utility pole or local substation. Wide-area failures, such as the 1965 blackout that struck New York City and other parts of the northeast, or the western U.S. problems in 1996, are exceptional.

Will that record of reliability continue with greater competition in electricity? Or is it an artifact of an era when the major utilities were not competing with each other, but each held geographically distinct franchised monopolies? If the newly competing firms cooperate to manage reliability, can fixed prices, reduced output, and divided markets be far behind?

Coordination through regulation Other means for promoting reliability are regulations and legal incentives to hold individual generators responsible for system-wide losses incurred when they fail to meet the demands of their customers. Those means might include reserve requirements and liability penalties when shortfalls in production relative to demand from one's customers lead to a blackout. Such rules may be ineffective if a generation company can declare bankruptcy rather than cover losses due to breakdowns. The time it typically takes the legal system to resolve liability disputes could be inadequate for the electricity market, in which supply and demand must be kept equal without interruption.

Central planning If the need to ensure reliability requires central planning, the compatibility question becomes whether the role of the planner—a regional reliability council, RTO, ISO, distribution utility, or some combination of those entities — will leave sufficient scope for competition to be meaningful. If the central coordinator can limit its activity to relatively small and occasional purchases of ancillary services, the rest of the generation and marketing sectors likely will remain large enough to make competition worthwhile. The more the planner has to extend its reach into managing transactions, purchasing electricity, and owning generation, the more the scope of competition will shrink.

CALIFORNIA'S LEGACY

Whether we can reap significant benefits of competition while retaining the central coordination necessary to maintain system reliability remains the most significant test that restructuring has to pass. Before the California crisis, deregulation proponents knew that the most likely threat to restructuring as a policy movement was a large-scale technical systemic breakdown followed by finger pointing, because no one would claim responsibility for ensuring reliability. Whether a true technical breakdown will happen in a deregulated electricity market remains to be seen. However, the list of institutional flaws in the California experiment imply that it was not a fair test.

In debates about deregulation, advocates and opponents generally treat it as a matter of theory at best or ideology at worst. The electricity industry, more than most (if not all) others, asks whether the answer to "Markets or not?" turns on facts as well as ideas and values. Advocating markets may be more effective if we concede that the issue is empirical rather than preordained. Electricity may turn out to join the list of other industries in which deregulation has worked. But electricity could be the sector in which markets may have met their match. R

READINGS

- "A Quantitative Analysis of Pricing Behavior in California's Wholesale Market during Summer 2000," by P. Joskow and E. Kahn. AEI-Brookings Joint Center for Regulatory Studies, Working Paper 01-01, 2001.
- *A Shock to the System*, by T. Brennan, K. Palmer, and others. Washington, D.C.: Resources for the Future, 1996.
- "Diagnosing Market Power in California's Deregulated Wholesale Electricity Market," by S. Borenstein, J. Bushnell, and F. Wolak. POWER Working Paper PWP-064, August 2000.
- "Strategic Bidding in a Multi-Unit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales," by C. Wolfram. *Rand Journal of Economics*, Vol. 29 (1998).
- "The Bay Area — A Knowledge Economy Needs Power," produced by McKinsey and Company. Bay Area Economic Forum, April 2001.
- "Why Regulated Firms Should Be Kept Out of Unregulated Markets: Understanding the Divestiture in *U.S. v. AT&T*," by T. Brennan. *Antitrust Bulletin*, Vol. 32 (1987).

TURNING OFF THE LIGHTS

Consumer-allowed service interruptions could control market power and decrease prices.

By Stephen J. Rassenti, Vernon L. Smith, and Bart J. Wilson
George Mason University

EARLY LAST DECADE, CONGRESS PASSED legislation that allows the deregulation of wholesale electricity production and prices in the United States. Under the legislation, states or regions that implement deregulation must develop restructuring plans for the power industry. The plans, in part, must specify the auction market rules for determining the hourly wholesale price of energy that wholesale producers sell to retail distributors, who in turn sell to their customers.

In areas that have implemented deregulation, the process has resulted in market designs hammered out by regulators, consultants, industry representatives, and various power-marketing intermediaries. The resulting plans employ supply-side bidding mechanisms for the hourly spot market, coupled with varying degrees of freedom for producers and distributors to arrange bilateral long-term contracts.

In California, producers submitted price-conditional offers of energy to the spot market, known as the California Power Exchange (California PX). That way, producers supplied the instantaneous demand at the price that balanced production with consumption.

However, the widely fluctuating rates paid by retail distributors to producers were not reflected in the rates that retailers charged most of their customers. Until the recent demise of the PX, most of California's wholesale power was resold to consumers at fixed rates per kilowatt-hour, after payment of a fixed monthly access charge. That meant that any consumer would be guaranteed that his instantaneous demand would always be satisfied at the regulated delivery price. Thus, deregulation at the wholesale level was not coupled with any significant change in the retail technology for delivering, metering, and charging the end-use consumer.

Stephen J. Rassenti, Vernon L. Smith, and **Bart J. Wilson** are affiliated with the Interdisciplinary Center for Economic Science at George Mason University. They can be contacted by E-mail at rassenti@cybernomicsinc.com, vsmith2@gmu.edu, and bwilson3@gmu.edu.

THE MUST-SERVE MARKET

In order to appreciate the implications of such a system, imagine what would happen if, say, the airline industry operated under similar rules. Airlines would be required to charge all passengers an identical, regulated monthly access fee and a fixed price per mile traveled, regardless of the flyer's destination, time and day of flight, and even his willingness to pay. Also, as part of the service, airline seats would have to be available to any passenger whenever he wants to fly, even if it is on the busiest day of the year. That would mean that, in order to operate in accordance with regulation, the airlines would have to purchase and maintain enough airplanes to satisfy peak demand, yet always charge the same fare regardless of how many planes are in use at a given time.

Fortunately, that is not the way the airline market works. For airlines, efficient pricing is very sensitive to peak versus off-peak demand; tickets are less expensive at times when fewer planes are needed to satisfy demand and there is plenty of capacity at the airports involved, and more expensive when more planes are needed to satisfy demand or airport capacity has reached its limit. Competitive markets will tend to yield higher prices for peak users because it is their travel demands that require the airlines to buy more airplanes and restrict the number of passengers that they can service at an airport. (The airline market did not always work that way; among the first casualties of airline deregulation were the "fair" pricing policies under regulation that tended to be insensitive to time and destination of travel.)

However, the electricity industry operates in a "must-serve" market similar to our hypothetical airline market. That must-serve structure was inherited from America's traditional, rigidly regulated power system that, because of political pressures, placed reliability of electricity flows above all other goals, regardless of cost. Under regulation, the cost was collectivized by averaging it across all users, regardless of an individual consumer's willingness to pay to keep the lights on. The local utility was expected to maintain service, or restore it quickly (even in inclement weather), and spread the cost of super-reliability over all customers. That cost included the maintenance of substantial reserves in generation and transmission capacity, with regulators establishing standard electricity rates high enough to ensure that utilities received a return on the capital so invested.

Thus, system reliability and the capacity to satisfy all retail demand were exclusively a supply-side adjustment problem. The consequence of that mindset was uncontrolled cost creep that increased to a gallop as utilities invested large sums of money in expensive plants in an effort to satisfy peak demand, and then were granted rate adjustments to cover the expenses. Reaction to that pricing became part of the political outcry for deregulation.

Under the old regulatory scheme, utilities invested large sums of money in expensive plants and received government-approved rate increases in an effort to keep up with peak demand.

DEREGULATION BEGINS

Implicitly, the process of deregulation assumed that the built-in supply-side bias did not require fundamental rethinking when it came time to design spot markets for the new world of competition. In the formation of market institutions, however, the devil is always in the details.

Beginning three years ago in Midwestern and Eastern markets, peak prices hit short-run levels of 10- to 100-times the normal price level of \$20-\$30 per megawatt hour. That was the predictable direct consequence of a market design in which completely unresponsive spot demand impinges on a responsive discretionary supply.

California's crisis Last fall and winter, the California PX was plagued by exorbitant increases in spot prices. Investor-owned local distribution utilities, which were required to obtain all their power through the PX, were pushed to the edge of bankruptcy because they were forced to buy all of the electricity their customers demanded — even if the demand resulted in the radical bidding up of prices — yet the distribution utilities then had to resell the power to their customers at low fixed retail rates (with a temporary exception in San Diego).

The crisis led to political action that imposed price caps on the market. But, of course, the caps only reduced power producers' response to the shortages. Political action also forced unwilling out-of-state vendors to continue selling energy to the financially troubled distributors for fear that, if they cut off the flow of electricity, the distributors would go bankrupt and the suppliers would never see the money already owed them.

Changing demand But what would have happened if, instead of following a must-serve philosophy, distributors and their customers had cut back — or even cut out — electricity use during peak hours? Would such a decrease in peak demand lead to lower peak wholesale prices and more supply? To answer those questions, we carried out a series of economic experiments that examined consumer response in such a market. We found that, if as little as 16 percent of buyers during peak demand voluntarily accepted an interruption of power (in exchange for financial compensation), the high level of prices and the tendency for upward price spikes to occur when electricity supplies are tight would have been completely avoided.

In such a system, the distributors would submit bids to customers, offering to purchase the ability to interrupt service. The interruptions would be voluntary, perhaps implemented by radio-control signals sent by the distributor to selectively turn off some of the customers' appliances. Or perhaps the distributor could adopt real-time prices that would provide customers with financial incentive to "self-interrupt" their service by cutting use during peak hours.

In California, such voluntary acceptance of interruption certainly seems preferable to last winter's involuntary area-wide blackouts that affected all consumers alike, including those operating elevators, energy-dependent production lines, and computers. What is more, in such a decentralized market, no price controls would need to be imposed, no involuntary penalties would need to be inflicted for high consumption, and no action would be taken against sellers for "unjust enrichment."

OUR EXPERIMENTS

In testing our models to see if a voluntary interruption scheme would be effective, we sought to answer two questions:

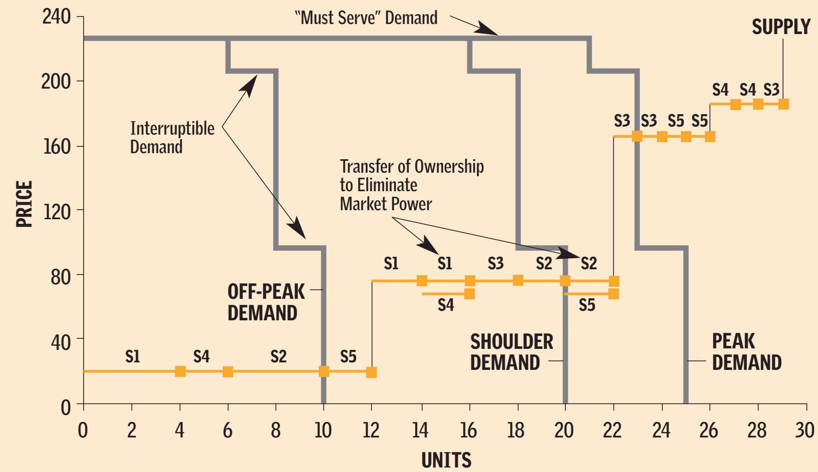
- What is the effect of supply-side market power on competition and prices? Specifically, what would be the difference between a market in which there is concentrated ownership of certain generator types vis-à-vis a market in which there is no such concentration of ownership?

- What is the effect on competition and prices when there is demand-side bidding by the distributor, who profits from buying in the spot exchange and reselling to retail customers at a given price schedule?

Figure 1

The Shape of Demand

Demand and price at peak, shoulder, and off-peak periods, overlaid by generating units that are profitable at various levels of price and demand.



Sx: the cost of operating various generator units operated by various suppliers.

We compared data from experiments that implement our theoretical demand-side bidding in a both-sides active, sealed bid-ask auction, with data from a one-sided auction in which only generator owners are submitting price conditional offers of power to the market. The latter type of auction is the one that characterizes most closely the exchanges found in California, the Midwest, and on the East Coast.

In summary, we compared data on prices for four different treatment conditions: supplier market power or not, coupled with demand-side bidding or not. Our experiments used profit-motivated subjects who were paid their earnings in cash at the end of each experiment. Buyers profited by buying low, sellers by selling high, with all participants using their own devices to work out profitable actions in light of experience.

Representing supply The supply schedule underlying all the experiments is represented by arraying generator marginal costs from lowest to highest. That arrangement is appropriate because power companies typically run their more efficient units for longer periods of time and only operate their least-efficient units at peak times when they are straining to keep up with demand. In our representation, each generator type has a constant marginal cost up to its capacity — a common approximation in power systems.

The resulting supply schedule is shown in Figure 1 as the line marked by the coming-on-line of various generators operated by suppliers S1, S2, S3, S4, and S5. In that representation, we distinguish three standard types of generators:

- Low-cost, base-load generators that are normally running and supplying power continuously at all times of the day. In Figure 1, generators with a running cost of 20 per unit of output (e.g. $20 per megawatt-hour) are shown on the lowest supply step. The generators' total supply capacity is 12 units (e.g. 1,200 megawatts) of power. At any price above 20, all of those generators are profitable.

- Intermediate cost ("load follower") generators that are normally running at all times except during the low-consumption off-peak hours. Those are shown in Figure 1 as incurring a cost of 76 per unit of output, with a total capacity of 10. The generators are profitable at any price above 76, as are the base-load generators. So, at prices above 76 but below when the high-cost units come on, the total energy supplied is 12 + 10 = 22.

- High per-unit cost generators that are normally used only during peak demand. Four of those generators are shown in Figure 1 with per-unit cost of 166, and three more with unit costs of 186. Hence, the supply moves up to 26 units when the going price is above 166, and up to 29 units when the price is above 186.

Representing demand For all of our experiments, demand was represented by a three-step resale value schedule. As shown in Figure 1, the first step — a large spike of demand at price 226 — is the so-called "must serve" demand that depicts the fixed retail price for all customers requiring uninterruptible service. Here, we represent the U.S. industry in its current state in which most of the demand is not responsive to price. (Of course, under full deregulation we would expect that condition to change over time, depending on prices and incentives to conserve usage.) With full retail pass-through of wholesale energy cost-based prices, we would expect a rapid adoption of both the culture and the technology for implementing price responsiveness.

The first block of inelastic demand varies over a daily cycle that begins with the "shoulder" demand (16 units), increases to the "peak" demand (21 units), returns to the shoulder demand, then decreases to "off-peak" demand (6 units). The cycle is repeated each "day" in our experiments, for a total of 14 cycles. Thus, the relatively steady peak and off-peak hourly price sequences for typical wholesale demand are each combined into one simplified block (representing several hourly markets) for the peak and another for the off-peak periods of the day.

We also simplified the transition demand sequence from off-peak to peak and vice versa by representing the transitions with a single steady block of "shoulder" demand. Our small simplifications enabled us to focus on the key issues we wanted to study while still capturing the essence of the daily natural cycle in demand in all electrical delivery systems.

The market at work Electricity distributors profit by reselling wholesale power purchased in each auction period during the day at their fixed three-step resale prices. Of course, in our theoretical market that includes voluntary interruptions, distributors and their customers will not arrange to cut service during the must-serve first step. However, the second and third steps can be interrupted, but the customer has agreed to continue accepting service if the price does not eclipse the step-two price (206) and step-three price (96). The lower resale prices for power represent discounts by the wholesaler to various retail customers in return for their willingness to have part or all of their deliveries interrupted at the discretion of the wholesaler. Those on the second step, however, never pay more than 206, and those on the third more than 96.

Market power is about the ownership of different generators classed by marginal cost, given a fixed and unresponsive demand. Changing that distribution would disrupt market power.

The resale values remain identical for two treatments: *No Demand-side Bidding* and *Demand-side Bidding*. In the absence of demand-side bidding, all of the listed resale values and associated quantities demanded are completely revealed as bids to the market by a robot bidder. That eliminates strategic bidding behavior by buyers, while strategic behavior on the seller side is fully active.

With demand-side bidding, four profit-motivated human subjects function as wholesale buyers, and are free to bid their respective demand each period at their discretion. They can reveal (bid their true demand as in Figure 1), under-reveal, or withhold demand strategically (by bidding at prices below their demand) in the same way that the five sellers are free to reveal, under-reveal, or withhold supply from the market by submitting asking prices higher than the true marginal cost.

The spot market pricing mechanism Where there is no active demand-side bidding in each spot pricing period, sellers privately submit a schedule of asking prices for their generation capacity. The aggregate of all generator offer schedules is obtained by arraying all of the individual offer price steps from lowest to highest. The offer schedule is then "crossed" with the actual demand (resale value) schedule to determine a single uniform market-clearing price where the two schedules intersect. All generators receive, and all wholesale buyers pay, the same price. That treatment parallels the energy markets in most regions of the United States that have instituted hourly spot markets, except that in our markets we make no provision for bilateral contracts secretly priced outside the exchange; all energy transfers pass through the spot market.

In the second experimental price mechanism treatment, there is active demand-side bidding by wholesale buyers. The buyers, in addition to the sellers, must privately submit schedules of bid prices for the purchase of electricity in each spot pricing period. The aggregate of the bid array, ordered from highest to lowest, is crossed with the offer schedule of sellers. Where the bid-ask (reported demand and supply) schedules intersect determines a single uniform price applicable to all buyers and sellers. The two-sided auction market allows buyers potentially to neutralize the expression of seller market power by under-revealing their resale values or withholding some of their demand for interruptible electricity.

Table 1

Hypotheses of Treatment Effects by Level of Demand

Level of Demand	Effect of no demand side bidding and market power	Effect of demand side bidding and market power	Effect of demand side bidding and no market power
Shoulder	Higher prices and volatility	Lower prices and volatility	?
Peak	?	Lower prices and volatility	?
Off-peak	?	Lower prices and volatility	?

UNILATERAL MARKET POWER

How do we create market power in generator ownership? To answer that question, look at the second step of the supply schedule, and the shoulder demand that intersects that step. Above each generator segment in the step there is a list of the generator companies that own capacity on the step. Seller *S1* owns two generators that operate profitably at prices set at the intermediate level of demand, and *S2* similarly owns two units.

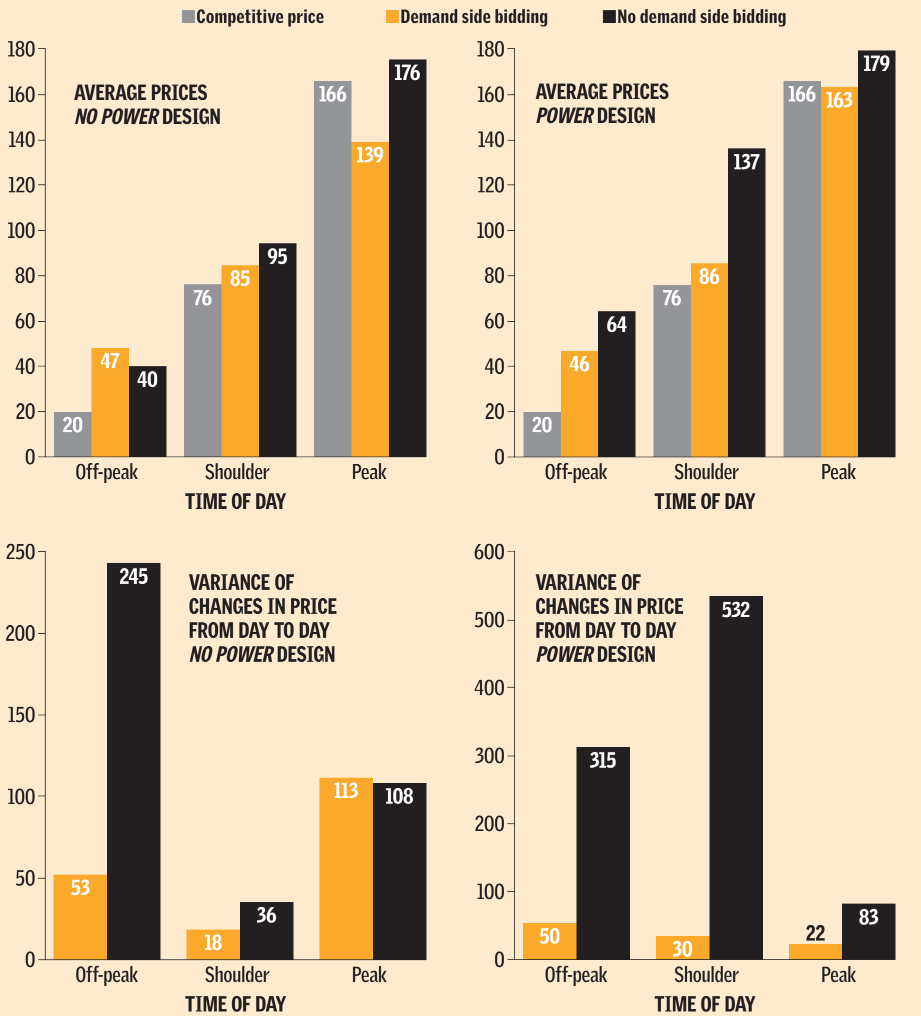
Gaming Given that ownership pattern, either *S1* or *S2* can increase its profits unilaterally by withholding some of its capacity on the intermediate cost step. During the shoulder periods, the competitive price is equal to the marginal cost (76) of the intermediate generators. However, either *S1* or *S2* can unilaterally withdraw (not submit offers for) four units of production entirely so that the price rises to the third step of the supply curve (166), where four units of peaking generation capacity contest any further attempted increase in price.

Alternatively, either *S1* or *S2* can increase the offer price for its intermediate cost generator capacity so that the offer sets the market price. It is important to note that it requires only one of the two sellers, *S1* or *S2*, to undertake that profitable action that reduces its volume sold but also benefits all other sellers. Either one of the producers will be even better off by not having reduced its sales volume if the other seller withholds supply to raise the price. Unless they tacitly coordinate their offers, each has an incentive to free-ride on the increased offer price of the other.

Figure 2

Market Power and Price

Experiment results for average prices and variances by treatment.



At the competitive price of 76, *S1* and *S2* both earn a profit of 224 [(76 – 20) x 4 units]. If either *S1* or *S2* raises the offer on its intermediate units to 166, the price-setter's profit rises to 584 [(166 – 20) x 4 units]. That unilateral deviation is even profitable at a price of 96 — the third shoulder demand step — where *S1* and *S2*'s profit would be 384.

Changing the system How can we change the ownership distribution of generators to eliminate market power during the shoulder demand periods? We can eliminate the market power incentives simply by transferring ownership of one of *S1*'s and one of *S2*'s intermediate cost generators to *S4* and *S5* respectively. We will call this the *No Power* treatment.

With that ostensibly minor reallocation of capacity at Nodes 1 and 3, not a single supplier can increase profit unilaterally by offering units at supra-marginal price levels in the shoulder period. If a single supplier raises its offer above 96, that supplier surely will not sell its intermediate units of capacity and, furthermore, will not increase the price received for its base load units. In that case, it is not profitable for any supplier to deviate unilaterally from revealing its marginal cost. Only if two suppliers — through tacit coordination — decided to raise their offers on the intermediate-cost (second-step) capacity could a supra-competitive price emerge.

Notice that, in both the *Power* and *No Power* treatments, no supplier can exercise market power during peak demands; all unilateral deviations are unprofitable. Even in the *Power* treatment, unilateral increases in offers by *S1* and *S2* to raise the price from the competitive level of 166 to the peak production costs of 186 result in a loss of profit of 360 [(166 – 76) x 4 units] from the intermediate units of production and yield a gain of only 80 [(186 – 166) x 4 units] on the base load units. That design, with generation competitive at peak demand levels but not at intermediate levels, illustrates the important principle that market power need not be associated only with peak demand conditions. Market power is about the ownership distribution of different generators classed by marginal cost, given a fixed and unresponsive demand.

S1 and *S2* can exert some market power during off-peak demands by raising the offer prices on two units of base-load capacity, regardless of the allocation of intermediate capacity generators. The theoretical upper bound on the price during off-peak demand is 76, where price is contested by the marginal cost of intermediate generating capacity. We included some market power incentives in the off-peak demand as a common control providing such incentives across sessions in all treatments.

DESIGN AND HYPOTHESES

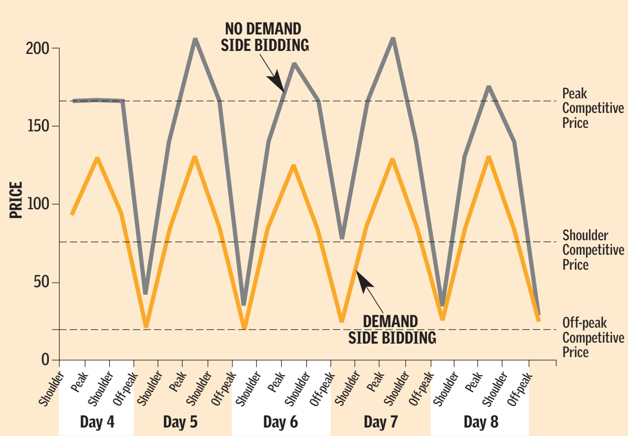
To test how active demand-side bidding and the ownership pattern of generators in electricity markets contribute singly and in tandem to the exercise of market power, we conducted 16 market experiments using students at the University of Arizona. Each session lasted approximately 90 minutes. We conducted four sessions in each of the four combinations of treatments: {*No Power, Power*} x {*No Demand-side Bidding, Demand-side Bidding*}. Each session was comprised of 14 market days that sequence through the different levels of demand as follows: shoulder, peak, shoulder, and off-peak.

Table 1 presents our hypotheses concerning the treatments effects of *Demand-side Bidding* and *Power* relative to baselines of *No Demand-side Bidding* and *No Power*. We hypothesized that active demand-side bidding will lower prices and decrease volatility for every level of demand when there is market power. We also hypothesized that the *Power* treatment will increase prices and volatility in the shoulder demand but not affect peak and off-peak prices when there is no market power.

Figure 3

Lowering the Peeks

Two examples of prices, with and without demand side bidding, in the *Power* Treatment.



Results Using the last 10 days of trading in each session, Figure 2 displays the average price and variances for each of the four combinations of treatments. Notice in the two upper panels that demand-side bidding reduces prices in the shoulder and peak periods for both the *Power* and *No Power* designs, and also reduces prices in the off-peak periods in the *No Power* treatment. The effect of demand-side bidding is particularly striking in the shoulder periods in the *Power* design. *Demand-side Bidding* completely neutralizes the exercise of market power during shoulder periods by reducing the average price from 137 to 86. Without demand-side bidding, the *No Power* suppliers push the price up to the value of the interruptible unit of demand, 96.

As hypothesized, we also find that shoulder prices are highly volatile in the Power design without demand-side bidding. The variance of price changes is 532 in the *Power* with *No Demand-side Bidding* treatment, but only 36 in the *No Power* with *No Demand-side Bidding* treatment. However, when *Demand-side Bidding* interacts with the *Power* design, the variance in shoulder periods drops from 532 to 30. In the *Power* design, demand-side bidding consistently and dramatically reduces the volatility of prices. In the *No Power* design, demand-side bidding reduces volatility in the off-peak and shoulder periods. *Demand-side bidding* effectively limits price volatility whether or not generators have market power.

PRICES WITH AND WITHOUT DEMAND-SIDE BIDDING

Figure 3 illustrates two time series paths of five experimental days for two sessions in the *Power* treatment, one with and the other without demand-side bidding. It is obvious that the session with demand-side bidding leads to lower prices in shoulder and peak periods. Furthermore, for the same time of day, prices are very stable across the days with demand-side bidding. Without demand-side bidding, the shoulder and peak prices vary rather noticeably from day to day.

Figure 4 illustrates why prices are lower and less volatile with demand-side bidding. The left panel illustrates the revealed bids and offers from both sides of the market in one *Power* session. The right panel illustrates the sellers' offers and the robotic bids at full resale value from another *Power* session. Notice that, in the *No Demand-side Bidding*, session, the offers are not very competitive because the sellers freely push the price up against the unresponsive buyers. In striking contrast, with *Demand-side Bidding* the lower bids by the buyers force the sellers to submit competitive offers resulting in lower prices in a 100-percent efficient allocation.

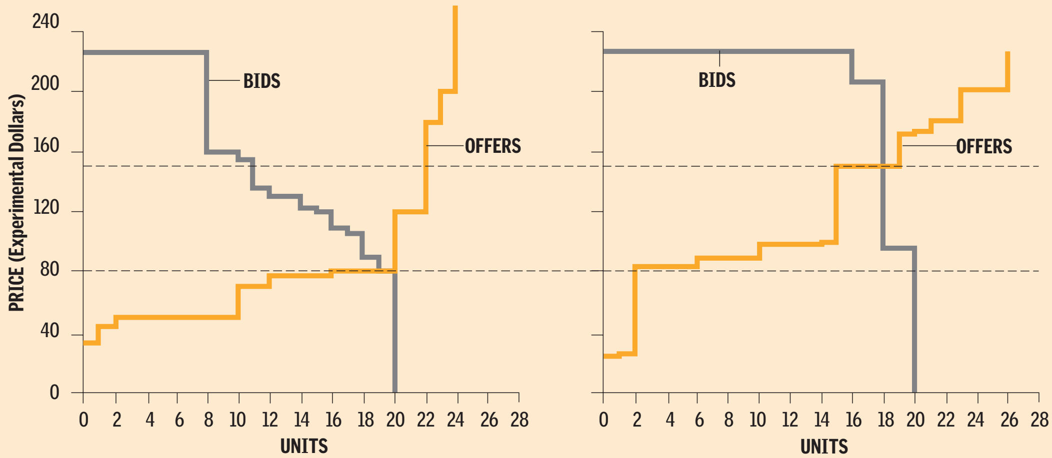
CONCLUSIONS

In this article, we have reported on laboratory experiments that used profit-motivated subjects to examine the effect of market power on the level and volatility of prices in a supply-side auction market with fully revealed demand. Using

Figure 4

More Effects of Demand-Side Bidding

Two examples of the impact of demand side bidding in the *Power* treatment.



Power vs. *No Power* baseline experiments, we extended the study to include measuring the effect of demand-side bidding, which was our primary treatment of interest.

While holding constant the cost and structure of supply, and the resale value and structure of wholesale demand, we inquired as to how effective is the introduction of profit-motivated demand-side bidders in restraining supply-side market power. Specifically, we measured the effect of demand-side bidding on both the level and volatility of wholesale prices. We did the comparisons in a technologically conservative environment in which most — 84 percent of peak — demand was what the industry and regulators call "must serve." In that sense, we studied a transition environment still influenced by regulatory rigidities likely to change and adapt over time to decentralized management by markets.

We found that, in the absence of demand-side bidding, suppliers who have market power because of the ownership distribution of generators are able to push up the general level of prices in all phases of the daily demand cycle — peak, shoulder, and off-peak. The high average level of prices peppered with frequent upward price spikes parallels the recent experience in regional electricity spot markets such as California and the Midwestern and Eastern sectors. The effect of demand-side bidding is to reduce both the level and volatility of prices that emerge from the exercise of market power by generators. When there is no generator market power, the effect of demand-side bidding is to reduce the shoulder and peak levels of prices and to reduce price volatility at off-peak and shoulder demand periods.

Interruption's payoff The public policy implications of our experiments are evident: Wholesale spot markets need to be institutionally restructured to make explicit provision for demand-side bidding. Distributors need to incentivize more of their customers to accept contracts for voluntary power interruptions. Distributors stand to gain by interrupting demand sufficiently to avoid paying higher peak and shoulder spot prices, and the savings can be used to pass on incentive discounts to customers whose demand, or portions of it, can be reduced or delayed to off-peak periods when supply capacity is ample. The technology and capacity for implementing such a policy have long existed and can be expanded, but incentives for introducing it have been inadequate.

Our policy proposal recognizes that adjustment to the daily, weekly, and seasonal variation in demand, and to the need to provide adequate security reserves, is as much a demand-side problem as it is a supply-side problem. The history of regulation has created an institutional environment that sees such adjustment as exclusively a supply responsibility. The result is an inefficient, costly, and inflexible system that has produced the recent price shocks and involuntary disruption of energy flows in California.

Demand-side bidding, coupled with the supporting interruptible-service incentive contracts, can eliminate price spikes and price increases, and reduce the need for reserve supplies of generator capacity and transmission capacity. R

READINGS

- "Controlling Market Power and Price Spikes in Electricity Networks: Demand-side Bidding," by S. Rassenti, V. Smith, and B. Wilson. Interdisciplinary Center for Economic Science Working Paper, George Mason University, 2001. Available online at http://gunston.gmu.edu/bwilson3/working_papers.htm.
- "Efficiency and Income Shares in High Demand Energy Network: Who Receives the Congestion Rends When a Line is Constrained?" by S. Backerman, S. Rassenti, and V. Smith. *Pacific Economic Review*, 1997.
- "The Exercise of Market Power in Laboratory Experiments," by C. Holt. *Journal of Law and Economics*, Vol. 32 (1989).
- "Market Design and Motivated Human Trading Behavior in Electricity Markets," by M. Olson, S. Rassenti, V. Smith, and M. Rigdon. Published in the *Proceedings of the 32nd Hawaii International Conference on Systems Sciences*, 1999.