THE CAPACITY MARKET ENIGMA

Why haven't reliability markets developed?

By Jonathan A. Lesser, Ph.D. and Guillermo Israilevich, Ph.D.

Ask two economists to define electric reliability, and you may get three different answers. That's bad, but better than asking two engineers, who may bombard you with technical acronyms. So, the first step in discussing markets for reliability is to agree on a workable definition. Only then does it make sense to discuss whether such markets are needed, and if so, why they haven't already sprung up by themselves.

The Very Public Blackout

The genesis of providing reliability in wholesale (or bulk power) electric markets has been the evolving structure of the electric industry, including wholesale and retail deregulation efforts that have led to the creation of independent system operators (ISOs) and regional transmission organizations (RTOs), in conjunction with FERC's so-called "standard market design" efforts. The 1965 Northeast blackout prompted the formation of the North American Electric Reliability Council (NERC) in 1968, and subsequently 10 regional reliability councils and power pools, whose mission is to coordinate the operations of the many independent electric utilities, and thus reduce the risk of future blackouts. A few of those power pools, such as in New England, fully coordinate and dispatch generation; the others operate more loosely, allowing individual utilities to determine how they dispatch their generating plants, but coordinating operations to ensure the overall system functions smoothly.

The need to coordinate generation in a region provides a crucial clue to defining reliability, as well as acknowledging the challenges of designing a reliability market. Reliability, which we define as the ability to meet the demand for electricity over time, whether during the next 10 minutes or the next 10 years, is a public good. What that means is, first, that reliability for one is reliability for all, something economists call non-exclusivity and non-rivalry in consumption. Second, it means that the operating decisions made for an individual generating unit may create spillovers, that is, effects on others, both good and bad. Third, as with all public goods, individual suppliers, left on their own, won't provide enough system reliability, because they can't reap the full economic benefits of doing so and would rather "free ride" on other suppliers' investments. This is a typical characteristic of public goods: Non-exclusivity means that someone who doesn't pay still can consume the public good just as much as someone who does. As a result, no one has an incentive to invest; after all, why invest when you can "free ride?"

How do these public-good characteristics manifest themselves in transmission markets? Consider New England, which has spawned vigorous debate before the Federal Energy Regulatory Commission (FERC) over development of an installed capacity market to be run by ISO-New England (ISO-NE), the region's regional transmission organization.[1] Regardless of whether this capacity market is established—and there are strong proponents and opponents of such a market—ISO-NE almost surely will continue to establish reliability targets and operating standards. Reliability targets establish the levels of installed capacity ISO-NE deems necessary to ensure enough generation is available to meet consumers' electric demand at any time. Operating standards ensure that generators do not operate in ways that compromise the safety and integrity of the transmission system. ISO-NE also will continue to require utilities and other wholesale customers to own or purchase their "fair" share of capacity requirements, to prevent them from "free riding" on everyone else's generation capacity.

Thus, the controversy is less over the need for reliability standards (although how much reliability is enough has been a subtext of the debate) and the associated need for market intervention to eliminate free riders, as it is about whether those obligations can be met by establishing a separate, long-term market in which the price of installed capacity is set by supply-and-demand conditions. Finally, and more fundamentally, there is debate as to whether having a well-functioning energy market would ensure sufficient capacity in the long-run, eliminating the need for any separate long-run installed capacity market.[2]

Is Today's System Broken?

Opponents of establishing an installed capacity market, or paying generators anything for the inherent resource adequacy they provide, argue that an installed capacity market only will provide a windfall for generators at the expense of utilities and their customers, without providing a more reliable system. Generators, not surprisingly, have very different views, arguing that, in the current energy market, with its price caps and availability rules, many generators needed for reliability are hemorrhaging dollars.

Currently, power pools designate a number of generators as reliability-must-run (RMR) units. Such a designation prevents individual generation owners from shutting down units needed to maintain the integrity—*i.e.*, the reliability—of the entire regional power grid, and ensures they are adequately compensated for providing that regional benefit. Rather than being compensated based on the dictates of supply and demand, however, the prices paid to generators under RMR agreements are based on cost-of-service, in much the same way as regulated local utilities' cost-of-service filings are made with their state regulators. FERC has sought to replace such cost-based arrangements with market-based ones whenever possible, believing that market solutions can provide resource adequacy and security more efficiently than can traditional, cost-of-service arrangements. However, some opponents of long-term installed capacity markets have argued against even these cost-based arrangements as being unnecessary and overly compensating generators at the expense of customers.

Opponents of installed capacity markets have adopted two conflicting positions. Some opponents argue there is no need for separate capacity markets, because a well-functioning energy market will provide all the capacity needed. Others argue that separate capacity markets only provide "windfalls" to generators, and have no impact on energy markets. Taken together, these arguments represent a classic free-rider response: Capacity market opponents want to "rely on" others' generation investments to provide system reliability, and not have to pay for it themselves. But, in the tradition of public goods, such behavior ultimately will result in a system that is unreliable and harmful to customers.

Since we are not privy to generators' account books, we cannot determine their financial status. But we can examine the structure of today's wholesale electric markets, or even an ideal wholesale energy market, to determine whether such markets, by themselves, would provide sufficient incentives to generators to provide sufficient system reliability. They won't.

Too Little Reliability, Too Late

Today's wholesale electric markets aren't fully deregulated. After early prices of $5,000 or more per megawatt-hour (MWh), breaches of contract obligations to supply promised generation, and rolling blackouts, transmission system operators entrusted with overseeing those bid-based wholesale markets and maintaining reliability reacted to regulatory and political pressure to establish price caps in markets. In New England, for example, electric prices today are capped at $1,000/MWh.

Whether or not one argues that $1,000/MWh, or any other binding price cap, is good public policy, the economic effects of price caps are clear: Price caps retard investment and adversely affect the mix of generating resources.[3]

Consider peaking units as an example. Peaking units serve a distinct purpose: They trade low capital costs for high operating costs, and thus are designed to run infrequently. Suppose an investor is considering building a new peaking unit. Without any type of separate capacity payment, the owner of a peaking unit must rely on energy market price spikes to recover that capital investment. Now, such price spikes may occur only during a few hours in a year, or may not even happen for several years. Therefore, when the price spikes do occur, they must be high enough for the investor to recover capital cost over time and earn a return on that investment high enough to compensate for the financial risk.

In the presence of price caps, however, a generation developer will be reluctant to invest because the price spikes counted on will be reduced (increasing the financial risk of the project). More important, banks that are asked to finance such investments will be less likely to provide the necessary capital. If they do agree to provide financing, they will charge higher interest rates. As a result, if new peaking units aren't built at all, overall prices paid by retail customers in the energy market will increase.[4] Moreover, other changes to wholesale market structures can increase uncertainty and retard new investment. That is why regulatory certainty is critical when establishing any new market, whether for generating capacity or air pollution permits. Investors need to know the market rules that will apply and be confident that those rules won't change in a way that increases their downside financial risk.

The Prospect of Market Intervention

When energy market prices are capped, generation owners need to be compensated for the additional financial risks such caps impose. But what if wholesale energy market prices weren't capped? Would that obviate the need for a separate long-term installed capacity market, as some claim? For a number of reasons, the answer still is no. On top of the general problem of underinvestment in public goods like reliability, generation developers also would have to contend with numerous financial risks, which would be especially troublesome for peaking and intermediate units that rely on high energy market prices to recover their costs.

First, when price spikes make headlines, it is too tempting politically not to intervene, or threaten to intervene, so as to protect defenseless customers from "selfish generators" who "take advantage of an energy crisis." One only has to look at the political reaction to high gasoline prices in the wake of Hurricane Katrina to see that political demagoguery is alive and well.

The difficulty, of course, is that high energy market prices may be associated with some form of anticompetitive supplier behavior. Although anticompetitive behavior clearly requires intervention, high prices don't necessarily mean anticompetitive behavior exists; instead, they may reflect volatile supply-and-demand conditions. The best response to anticompetitive behavior is not to use so blunt a policy instrument as price caps.

As evinced by the California meltdown, it can take regulators years of expensive litigation to determine whether high prices were the result of supply shortages, weather conditions, anticompetitive behavior, or some combination of all of those

> Short of a guaranteed schedule of sabotage, generation developers are likely to avoid such a constrained region, even though, from a broader market perspective, that is precisely where they ought to build.

factors. If generation owners have no other source of market compensation, especially owners of peaking and intermediate units that are dependent on revenues earned over a limited number of hours, they will be less likely to invest.

Taking Your Lumps

Another risk to generators arises from revenue uncertainty caused by the "lumpy" nature of capacity investments, coupled with the lack of demand responsiveness. Financing for generators with high revenue uncertainty simply may be unavailable, even when such investments may be valuable for the system as a whole. Unless the full value of reliability is captured separately, its public good status will lead to under-investment in capacity.

This is particularly true as wholesale energy markets become more focused at the zonal and even individual generator level. Although these focused markets can send more appropriate price signals to investors, one ironic consequence is a higher concentration of suppliers, with the resulting increased potential for market power and greater price volatility. As a result, individual investment decisions, as well as individual operational decisions, will have more profound impacts on wholesale energy prices in these smaller markets.

Consider a transmission-constrained region where new generation investment is valuable from the standpoint of both reliability and energy. Suppose one or more large generating units are in that region, built to exploit economies of scale. If the region is small and transmission constrained, these generating units are likely to have a disproportionate impact on market clearing prices and overall system reliability. These generators may find themselves in an odd position: As long as they operate, energy prices will be set relatively low and the generator will not recover replacement costs. But, if one of the generators suffers a forced outage, market prices may spike and reliability drop, perhaps dangerously so. This is what we call the "endogeneity" problem.

Next, consider the decisions faced by potential investors in new generation, especially those wanting to build peaking or intermediate units. Those investors may look askance at this constrained market, because they may not be able to recover investment costs through high energy prices (which take place only when their units are not operating). It may also be that their profits would be too little to justify building in the region, unless the existing large generators can be "relied on" to fail, and thus drive up the market price of energy during a few, high-priced hours.

ISO-NE is facing this "lumpy" investment problem in load pockets. The overall downward impact on energy prices from large baseload and intermediate units precludes them from recovering investment costs, and has forced increasing numbers of generators to apply for RMR agreements. New generation investment, sized large enough to exploit economies of scale, will drive market prices even lower.

It is possible that the new generator will be efficient enough to operate profitably. But what if the drop in energy market price causes financial distress for the existing generator? If the existing generator shuts down, prices will spike and reliability will drop too far.[5]

Thus, short of a guaranteed schedule of sabotage, generation developers likely are to avoid such a constrained region, even though, from a broader market perspective, that is pre-

> Consumers don't directly purchase reliability. But reliability, like clean air, is a public good, so harnessing the power of market forces to provide reliability requires that a market be established.

cisely where they ought to build. The answer to this paradox lies in the inability of generators to rely on high energy prices to finance investments. In other words, you cannot rely on a lack of reliability. It is just too risky, because of the inherent volatility of electric markets, the often-occurring phenomena of regulatory and political intervention, as well as other market imperfections.

Reliability Requires Capacity

The remedy for these dilemmas is to provide distinct installed capacity payments to generators. Capacity and energy, while clearly related, are not identical products. Installed capacity provides both short-term and long-term reliability, a public good. And, like other public goods, market forces alone will provide too little reliability. Moreover, if we want to encourage a diverse mix of generating resources—baseload, intermediate, and peakers—investors must expect to earn sufficient revenues to cover their risks for each generation technology. That's true regardless of whether generation markets are fully deregulated or fully regulated. Ignoring the public good component of installed capacity sooner or later will force system operators to intervene by designating a number of generators as RMR units.

Should we return to fully regulated wholesale energy markets, based strictly on generators' cost of service, and abandon creating markets for reliability and other ancillary services? FERC does not think so, and, there's no economic basis to conclude that competitive wholesale energy markets aren't working well, price caps excepted.

Unlike the market for energy, installed capacity markets always will have an administrative aspect. Just as consumers don't directly purchase clean air, they don't directly purchase reliability. And, because reliability, like clean air, is a public good, harnessing the power of market forces to provide reliability requires that a market be established.

Thus, if we are to continue to rely on competitive market forces to provide new generation supplies, we need separate, long-term, installed-capacity markets. Not only will that market reduce the risks from volatile energy markets, but it will provide a hedge against the inevitable regulatory and political pressures for energy market intervention when prices are high. New generation investment, spurred on by a separate capacity market, not only will reduce capacity market prices but will increase competition in the energy markets. The result will be lower energy prices and improved reliability.

Jonathan Lesser is a partner and Guillermo Israilevich is a manager with Bates White LLC, an economic and litigation consulting firm in Washington, D.C. Although Dr. Lesser has testified in Devon Power LLC, et al. (Docket No. ER03-563-030), the opinions in this article are not sponsored by any market participant. Contact Dr. Lesser at jonathan.lesser@bateswhite.com and Dr. Israilevich at guillermo.israilevich@bateswhite.com.

Endnotes

1. *Devon Power LLC, et. al., Docket No. ER03-563-030.* The case is awaiting an order by FERC. The New York Independent System Operator has overseen a similar capacity market since May 2003.
2. Generation capacity provides two separate but related products: security and adequacy. Security is the ability of the electric system to respond to instantaneous or short-term changes in demand, such as providing spinning and non-spinning reserves. Adequacy is the ability of the electric system to respond to long-term changes in demand. We are unaware of anyone who argues that system security products, such as spinning and non-spinning reserves, are not needed, or that a generation-only market would provide those products. The controversy focuses on whether a separate, long-term resource adequacy market is required.
3. Although we simply refer to price caps, modifications to operating-reserves requirements may have similar effects on generators' revenues, as the system operator might refrain from purchasing all the required reserves (and prevent further price increases) during capacity shortages.
4. Of course, price caps aren't the only risks facing potential generation developers. Many state regulators have become increasingly concerned about who will build new baseload generation, as continued industry restructuring, market uncertainties, and changing environmental regulations have increased risk, creating an investment climate where developers and banks want signed, long-term contracts with utilities before breaking ground, while utilities and other retail providers are reluctant to enter into long-term contracts because of the regulatory and market risks such contracts pose. Addressing those uncertainties is best left to another article.
5. The flip side of lumpy investments is the potential for market power. In California, for example, RMR contracts were used extensively during 2000 and 2001 to prevent capacity withholding in transmission-constrained regions. See Scott Harvey and William Hogan (2001), "Further Analysis of the Exercise of Market Power in the California Electricity Market." Working paper, Center for Business and Government, John F. Kennedy School of Government, Harvard University.