Conference on Competition in Wholesale Power Markets

Docket No. AD07-7-000

Federal Energy Regulatory Commission

Acting in Time: Regulating Wholesale Electricity Markets

Comments of William W. Hogan[i]

May 8, 2007

Introduction

The Federal Energy Regulatory Commission is now completing a series of technical conferences addressing questions of performance, market design, seams between market regions, market monitoring, and the challenges facing regulation of organized wholesale electricity markets.[1] The related inquiries are important and overdue. These discussions stand in the context of the larger consideration of regulation of the transmission grid under principles for open access without undue discrimination, applied to both the organized markets in Regional Transmission Organizations (RTOs) and regions outside of RTOs.

The present technical conference emphasizes a subset of the issues dealing with demand response, long-term contracting and governance of RTOs. I have been asked to address primarily the first of these issues, and in particular the role of operating reserve demand curves in facilitating demand response. In pursuing this charge, it is important to place the discussion within a larger framework. This larger discussion addresses issues of market design, regulatory policy, and the pressure to act in time.

Acting in Time

We are all familiar with "…the concern that governments, communities, and nations often seem unable or unwilling to act expeditiously to solve problems, even when prompt action will almost certainly be less costly and more effective than delay."[2] There could be many explanations of the causes of inaction in any particular instance.

In regulating wholesale electricity markets there are at least three conditions that appear necessary to guide and precipitate timely action. First, we need to decide who is in a position to act, or we must create someone if there is a vacuum. Second, we need a

---

[1] FERC Technical Conferences: Wholesale Power Market Competition, February 27, 2007; Seams Issues in the Eastern Interconnection, March 29, 2007; Policies Regarding Market Monitoring, April 5, 2007; Wholesale Power Market Competition, May 8, 2007.

[2] David Ellwood, "Acting in Time" initiative, John F. Kennedy School of Government, Cambridge. MA. For further details on this research effort covering topics as diverse as global warming and the health care system, see http://www.ksg.harvard.edu/dean/acting_in_time/ ..

framework for integrating what will be necessarily a set of related decisions and actions rather than shooting a single silver bullet. Third, there has to be a well-defined problem statement with understandable actions available that could address the issue; diagnosis without prescription presents only a circumstance, not a problem. A list of sufficient conditions for effective action would be longer, to include cost benefit analysis of alternatives, but these three necessary conditions will serve for the moment.

In the case of wholesale electricity markets, the Commission should be a locus of action. There is no doubt that the Commission has both the jurisdiction and the responsibility. As Chairman Kelliher observed at the technical conference in February, "[t]he Commission's policy is not and has never been deregulation," so there has been no abdication of responsibility. Chairman Kelliher went on to say that the nature of regulation has changed, with the task being to find the best possible mix of competition and regulation.[3] Furthermore, the prior technical conferences reinforced the straightforward opening statement supporting competition in wholesale markets as a clear and continuing national policy:

> "While competitive markets face challenges, we should acknowledge that competition in wholesale power markets is national policy. The Energy Policy Act of 2005 Federal Energy Regulatory Commission embraced wholesale competition as national policy for this country. It represented the third major federal law enacted in the last 25 years to embrace wholesale competition. To my mind, the question before the Commission is not whether competition is the correct national policy. That question has been asked and answered three times by Congress.
>
> If we accept the Commission has a duty to guard the consumer, and that competition is national policy, our duty is clear. It is to make existing wholesale markets more competitive. That is the heart of this review: to not only identify the challenges facing competitive wholesale markets but also identify and assess solutions."[4]

Although there are many actors in the complex political process of developing regulatory policies to support wholesale market competition, leadership must come from the Commission. Absent this leadership, even the status quo is in jeopardy.

This leadership fundamentally involves market design, and the design of compatible regulations to support that market design. In the case of electricity, with its complex and large interactions across the transmission grid, the market can't solve the problem of market design.

As for the larger framework integrating the many components and decisions that arise in designing wholesale electricity markets, there is a clear answer that follows from both analysis and the by now extensive experience in organized wholesale markets: bid-based-security-constrained-economic-dispatch-with-locational-prices-and-financial-

---

[3] Endorsing comments attributed to Fred Kahn.

[4] Joseph T. Kelliher, "Statement of Chairman Joseph T. Kelliher," Federal Energy Regulatory Commission, Conference on Competition on Wholesale Power Markets AD07-7-000. February 27, 2007.

transmission-rights.[5] As the Commission knows well, there has been an enormous amount of analysis and experimentation, and this basic design should by now be uncontroversial. This is not one market design framework among many possible alternatives; for wholesale market competition under the principles of open access without undue discrimination, it is the only framework known that works in both theory and practice. This framework captures the core of the designs in place or soon to be implemented in every organized market in the United States.

Although this basic market design should not be controversial, it requires continuous repetition. It is a simple fact that the nature and complexity of the interactions in the transmission grid are not well known to those who have not been taught by the engineers, and the results often seem counterintuitive. With a constant influx of new market participants, and new regulators, it is necessary to remind that there is a real problem here—electricity is different—and the proven market design contains many complementary pieces that fit together to provide a solution. Furthermore, for those who have not been taught by the economists, it is easy to forget that the prices and resulting incentives flowing from the market design are also complex, and are just as important to get right and consistent with the engineering details.

The fundamental problem centers on getting market pricing in place to provide the proper operating and investment incentives while creating the associated property rights to allow market-based investments to go forward. Ironically, for reasons that are peculiar to electricity, the critical pricing rules and conditions arise in the wholesale spot market. Electricity is different, and this difference makes a necessity of the apparent contradiction of a regulator as market designer.

The regulator can choose one of two paths. Consistent with the goal outlined by the Commission, the regulator can pursue little "r" regulation through designing rules and policies that are the "best mixture" to support competitive wholesale electricity markets. In pursuing the little "r" approach, a key requirement is to relate any proposed solution to the larger framework and to ask for alternatives that better support or are complementary to the market design. Many seemingly innocuous decisions appear isolated and sui generis, but on closer inspection are fundamentally incompatible with and undermine the larger framework.

The other path is to frame every problem in its own terms—inadequate demand response, insufficient infrastructure investment, or market power—and design ad hoc regulatory fixes that accumulate to undermine market incentives. This creates a slippery slope problem, where one ad hoc solution creates another problem, and regulators are driven more and more to intervene in ever more ad hoc ways. This leads to big "R" regulatory micro-management. For example, socialized costs for preferred infrastructure investment can easily reduce the incentives for other market-based investments, thereby

---

[5] For a similar succinct statement, review the comments of former Commission chair Besty Moler at the February 27, 2007, technical conference. For a more discursive summary, see John Chandley and William Hogan, "A Path To Preventing Undue Discrimination And Preference In Transmission Services," (August 2, 2006) Submitted to the Federal Energy Regulatory Commission, Docket No. RM05-25-000, August 25, 2006.

increasing the need for regulators to select among additional appropriate investments and socialize even more costs.

A core idea of an electricity market that relies on market incentives for investment is that these incentives appear through the largely voluntary interactions of the participants in the market. A main feature of the market would be prices determined without either price caps or other interventions that would depress prices below high opportunity costs and leave money missing. The real-time prices of electric energy, and participant actions, including contracting and other hedging strategies in anticipation of these prices, would be the primary drivers of decisions in the market. The principal investment decisions would be made by market participants rather than the regulators, and this decentralized process would improve innovation and efficiency. A goal would be to avoid repeating the problem of leaving customers with stranded costs arising from decisions in which the customers had no choice. This change in the investment decision process and the associated reallocation of risks would arguably be the most important benefit that could justify greater reliance on markets and the costs of electricity restructuring. If this were not true, and if it would be easy for planners and regulators to lay out the trajectory of investment for the best portfolio of generation, transmission and demand alternatives, then electricity restructuring would not be needed.

A challenge for the Commission and market participants is to fulfill the joint responsibilities of regulation and support of competition. The general framework by itself does not provide all the answers, and regulatory intervention is required and ubiquitous. But the general market design framework does provide a powerful test bed for evaluating the degree to which proposed regulatory mandates address problems in a little "r" manner that is consciously supportive of market incentives and flexibility, or the degree to which big "R" decisions have some hope of avoiding the slippery slope.

The need to act in time to address the half empty glass of electricity restructuring seems greater everyday. The reverberations of the western energy crisis, extended debate over EPAct05, political tensions with state regulators over market design, and the competing attention of regulatory mandates for infrastructure expansion have delayed some of the most pressing market reforms in and out of the RTOs. The welcome move to mandatory reliability standards presents the Commission and NERC with a demanding new task to implement these standards in a manner that supports both reliability and wholesale competition. And external pressures pose real threats to the viability of markets. In an ironic twist, on December 19, 2006, the same day the Commission announced plans for this series of conferences, the Ninth Circuit promulgated its decisions in the Snohomish and CPUC cases.[6] At worst these decisions fundamentally undermine markets by making it much more difficult to develop viable long-term contracts. At best these decisions increase the pressure for the Commission to take up critical matters of market design.

Action is needed to identify and implement the best mix of regulation and competition. Crafting and evaluating initiatives in the spirit of little "r" regulation would

---

[6] Pub. Util. Dist. No. 1 v. FERC, 471 F.3d 1053 (9th Cir. 2006) (Snohomish); Pub. Util. Commission of the State of Cal. v. FERC, 474 F.3d 587 (9th Cir. 2006) (CPUC).

apply the test of compatibility with the basic market framework. Supporting better demand response and infrastructure investment provides a timely illustration of the general argument.

Demand Response, Price Incentives and Infrastructure Investment

The early sessions in this series of technical conferences provided extensive discussion about the electricity restructuring glass as being half full or half empty. There have been impressive accomplishments through the organization and operation of RTOs. The qualitative evidence is sometimes dramatic. Many have forgotten, for example, the powerful difference in the performance in PJM in 1997 without locational pricing versus 1998 when locational pricing was implemented.[7] Furthermore, the systematic quantitative evidence is accumulating, as reported by others in the first technical conference in this series.

However, all would agree that the glass is half empty, and it would be hard to justify all the costs and turmoil of the transition of electricity restructuring based on the results to date. The biggest open question is the degree to which markets can operate to improve the risk allocation and performance of major infrastructure investment decisions. And among the biggest disappointments has been the (very) limited success in eliciting greater demand side participation in the market. As suggested by the agenda in this conference, these problems are related.

There is general agreement that efficient demand-side participation could have a dramatic impact on market performance. Not the least of which would be the impact on changing the magnitude and structure of needed infrastructure investment. Although there are a few steps in the chain of logic and actions needed to provide appropriate incentives for greater demand side participation, including metering and state regulation, there is one step that falls squarely in the domain of the Commission. This step would be to improve the price signals in the wholesale market. Despite recent headlines to the contrary, the basic fact is that wholesale electricity prices have been too low to support either infrastructure investment or adequate demand side participation in the crucial spot markets.

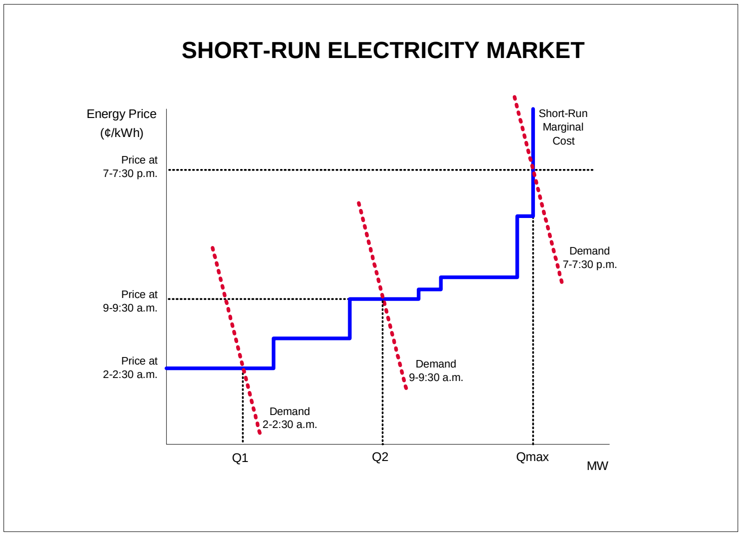
In particular, prices in organized markets tend to be too low during conditions of generation capacity scarcity, exactly the time when the unexploited demand side resource would be most valuable. But without the signal and the reward through prices, there is insufficient market incentive for demand side action or for adequate infrastructure investment. There are many reasons for this inadequate scarcity pricing that relate to both mistakes in market design and practices of system operators.[8]

---

7 William W. Hogan, "Getting the Prices Right in PJM: Analysis and Summary: April 1998 through March 1999, The First Anniversary of Full Locational Pricing," Harvard University (available at www.whogan.com), April 2, 1999.

8 Paul Joskow, "Competitive Electricity Markets and Investment In New Generating Capacity", MIT, June 12, 2006, http://econ-www.mit.edu/faculty/download_pdf.php?id=1348 .

A mistake goes back to the early discussion of the simplified story of electricity markets. A figure from the early discussions illustrated the principle of pricing in a spot market.



The basic idea was that during most periods the price would be equal to the marginal costs of the most expensive generator in the economic dispatch, as illustrated by the two lower price levels depicted in the figure. This idea was well-understood and was easy to implement because it did not require any knowledge of the shape the demand curve. The marginal cost price could be determined directly from the total level of generation and the aggregated supply offers of the generators.

However, as shown in the figure, during peak hours, when all the capacity was in use, the efficient price would be determined by the intersection of the demand bids with a vertical section of the supply offers. This would be easy to implement if there were enough demand bids. But in the absence of demand bids there is no guidance as to how to determine the appropriate scarcity price. In practice, the practice has been to apply the same pricing rules and set the spot price at the marginal cost determined by the supply offer of the most expensive plant running.

This pricing rule is both conceptually wrong and presents a major problem. The conceptual error is obvious from the figure. On a vertical segment of the supply curve the marginal cost of the most expensive plant running is too low to set the appropriate scarcity price. The major problem is that this failure to capture the proper scarcity prices in equilibrium eliminates all of the energy revenues needed to cover the capital costs of the peaking generator, and a major fraction of the revenue needed to cover the capital costs of all other mid-range and base-load generation. The same applies to investments for demand-side alternatives and incentives for demand-side participation in the spot market.

This result is known as the "missing money" problem.[9] It is important to recognize that this is not a second-order problem. As documented by many analyses and summarized in the Joskow overview, compared to the efficient equilibrium, a large fraction of the money has been missing.

The big "R" solution to this infrastructure investment and demand-aide participation problem has been to construct increasingly expansive regulatory mandates to require investment, long-term contracting, and demand-side programs that must overcome market incentives without adequate scarcity pricing. There is so much effort being devoted to these fixes that we all should hope that they will work. However, these regulatory mandates do nothing to address the little "r" problem of revising the market design to provide better scarcity pricing.

Operating Reserve Demand Curve Theory and Scarcity Pricing

A problem with identifying the little "r" approach of revising the wholesale electricity market design to provide better scarcity pricing is the almost universal judgment that this would be politically infeasible, even if it does work in Australia. This argument has been powerful and has produced an immediate segue into a variety of big "R" regulatory mandates to deal with the symptoms without further consideration of treatment directed at the fundamental problem.

There are two immediate arguments against simply assuming that better scarcity pricing is impossible. First, improved scarcity pricing should not be done in isolation. It could and should be seen as a complement to improvements in long term contracting, or as an adaptation to systems like the New Jersey Basic Generation Service auction.[10] The impact of better scarcity pricing would not change the need for regulatory interest in contracts as part of long-term hedging programs, especially for smaller customers. However, it would have a major impact on the nature of such contracts and could greatly simplify matters such as dealing with the deliverability of generation capacity.

Second, improved scarcity pricing and better long-term hedges need not be thought of as mutually exclusive of the more direct big "R" mandates for infrastructure investment and demand side programs. To the contrary, in most cases well-designed mandates would be easier to implement with better scarcity pricing in the spot market. In addition, better scarcity pricing provides about the only hope as an insurance policy in the event that the carefully planned regulatory mandates don't quite deliver on the actual needs in the future spot market. If there is a commitment to the big "R" regulatory mandates for resource adequacy and demand side programs, this should not preclude

---

[9] The term "missing money" describes the condition in which prices in the markets for energy and ancillary services are kept below market-clearing levels, especially in hours of scarcity, with one result being that the prices fail to cover the fixed costs of generators. The characterization as "missing money" comes from Roy Shanker. For example, see Roy J. Shanker, "Comments on Standard Market Design: Resource Adequacy Requirement," Federal Energy Regulatory Commission, Docket RM01-12-000, January 10, 2003.

[10] William W. Hogan, "On An 'Energy Only' Electricity Market Design For Resource Adequacy," Center for Business and Government, John F. Kennedy School of Government, Harvard University, September 23, 2005, pp. 27-33, (available at www.whogan.com).

attention to better scarcity pricing. If we have to choose, better scarcity pricing should be the priority. But we do not have to choose. The obvious answer is to do both.[11]

If we were to seek better scarcity pricing, how would this be done beyond simply hoping for more bidding by dispatchable demand in the spot market? An answer appears through inspecting another "mistake" in the early discussion and implementation of market design for wholesale electricity markets. The stylized spot market figure illustrating equilibrium pricing neglects the role of operating reserves. In the presence of adequate demand side bids, the simplification of treating operating reserves as a fixed added capacity requirement is a small complication and makes no material difference in the analysis.

In the absence of adequate demand side bids, however, treating operating reserves as a fixed capacity adder is wrong both in the conceptual implications for equilibrium pricing and as a practical description of what actually occurs in system operations in any electricity system, including in the organized markets.[12]

Here the term operating reserves refers to many things including spinning reserves that are synchronized to the system and available to provide immediate energy production, quick start units that might be available in ten minutes, standby reserves that might be available in twenty minutes, voltage reductions, and so on. Dealing with the range of tools is not trivial but is doable. However, for the present discussion we can think of operating reserves generically as dispatchable supply and demand options that are immediately available but being held in reserve.

These operating reserves are inherently short-term and are quite distinct from the installed capacity reserves more commonly discussed. However, installed capacity mandates are a long-term concept, distinct from the necessary and essential operating reserve requirements.

Operating reserves are needed to meet two objectives. One is to reduce the probability that the system operator will turn to involuntary load curtailments over the time frame when there might be unexpected outages or surges in demand. Another is to ensure that there is enough immediately available capacity to protect the system in the event of a contingency that could otherwise bring down the whole system. The former involves probabilities and tradeoffs. The latter acts as a constraint given the list of monitored contingencies.

The simplifying assumption that there is a fixed requirement for operating reserves is consistent with the contingency constraint idea, but it is not compatible with the probabilistic analysis of reducing the expected but limited involuntary curtailments of load.

---

11 William W. Hogan, "Resource Adequacy Mandates and Scarcity Pricing: Belts and Suspenders," Harvard University, February 23, 2006, (available at www.whogan.com) .

12 William W. Hogan, "On An 'Energy Only' Electricity Market Design For Resource Adequacy," Center for Business and Government, John F. Kennedy School of Government, Harvard University, September 23, 2005, pp. 11-14, (available at www.whogan.com). For a related discussion of the importance of an operating reserve demand curve, see ISO New England, "2006 Wholesale Markets Plan," September 2005, pp. 16-17.

For the contingency constraint, it is true that there is a fixed requirement for operating reserves (adjusted for particular momentary conditions). Below this level there is a very high value for incremental operating reserves equal to the value of loss load (VOLL), because the system operator will incur that cost by curtailing load in advance in order to restore the minimum contingency requirement for operating reserves. However, once the contingency constraint is satisfied, the value of additional operating reserves drops to zero. For the contingency constraint there is a vertical demand curve, just as in the stylized model.

By contrast, when considering the tradeoff of the probability of getting into a circumstance that requires involuntary load shedding, more operating reserves should be better. With increasing availability of operating reserves the marginal value would decline, but in the nature of such probabilistic analysis the value would never go to zero. More operating reserves would be better. For this reason, a vertical operating reserve demand curve is incorrect as a conceptual matter.

In the presence of active demand side bidding, the vertical operating reserve demand curve would not be a serious quantitative problem and would have little impact on scarcity pricing. But without active demand-side bidding, the conceptual mistake has real practical significance.

The theoretical problem of the vertical operating reserve demand curve is compounded by the practice in all markets. In practice, system operators do not adhere to a fixed operating reserve requirement. As capacity becomes shorter, the operator takes a number of steps to use some of the existing reserves, reduce voltage, or implement various emergency actions. Only as a last resort in this sequence of steps will the operators turn to involuntary load curtailments in rolling blackouts, and then only to maintain the inviolate constraints of enough reserves to meet the contingency constraints protecting against a system-wide failure.

These operating practices are in general a good thing, and have been developed over many years to provide the requisite high reliability on the grid. What is not a good thing is that these many operating practices have not been integrated with the pricing provisions in the organized markets. Perversely, for the most part the net effect of all these practices is to reduce the marginal cost of the most expensive generator running and, coupled with the pricing mistake described above, these practices interact with the pricing flaws and result in lower not higher prices during scarcity conditions.[13]

The scarcity pricing problem does not arise from the operating practices but from the conceptual failure of the simple market design to incorporate the operating reserve demand curve. The little "r" solution to this failure of market design is simply to replace the flawed concept of the vertical demand curve for operating reserves with the more realistic model that allows for different values (prices) for operating reserves above the absolute minimum level required to meet the contingency constraints.

---

[13] Paul Joskow, "Competitive Electricity Markets And Investment In New Generating Capacity", MIT, June 12, 2006, http://econ-www.mit.edu/faculty/download_pdf.php?id=1348 , p. 35.

This is not a new idea, and it is not simply a conceptual proposal, For example, the New York Independent System Operator (NYISO) adopted the concept and implemented an operating reserve demand curve integrated with the energy market design in the spot market. In this system, when capacity is constrained and operating reserves are reduced, the value of marginal reserves rises, increasing both the price of energy and the related opportunity costs of reserves.

Operating Reserve Demand Curve Implementation

There are two problems in using the actual NYISO operating demand curve to illustrate a real implementation of the concept. First, the demand curve as published is not the actual demand curve. Rather the published values are the shadow prices on various constraints, and the operating reserve demand curve is defined only implicitly through the interaction of these constraints. Second, the values of the constraints were obtained from good engineering judgment, but provide little insight into how the concept could be translated to other settings.

Using a formal albeit simplified model and representative data from the NYISO, it is possible to outline how to obtain a reasonable operating reserve demand curve and to determine the quantitative implications of its use. This in turn provides the opportunity to address some of the obvious questions that arise in considering broader implementation of the concept.

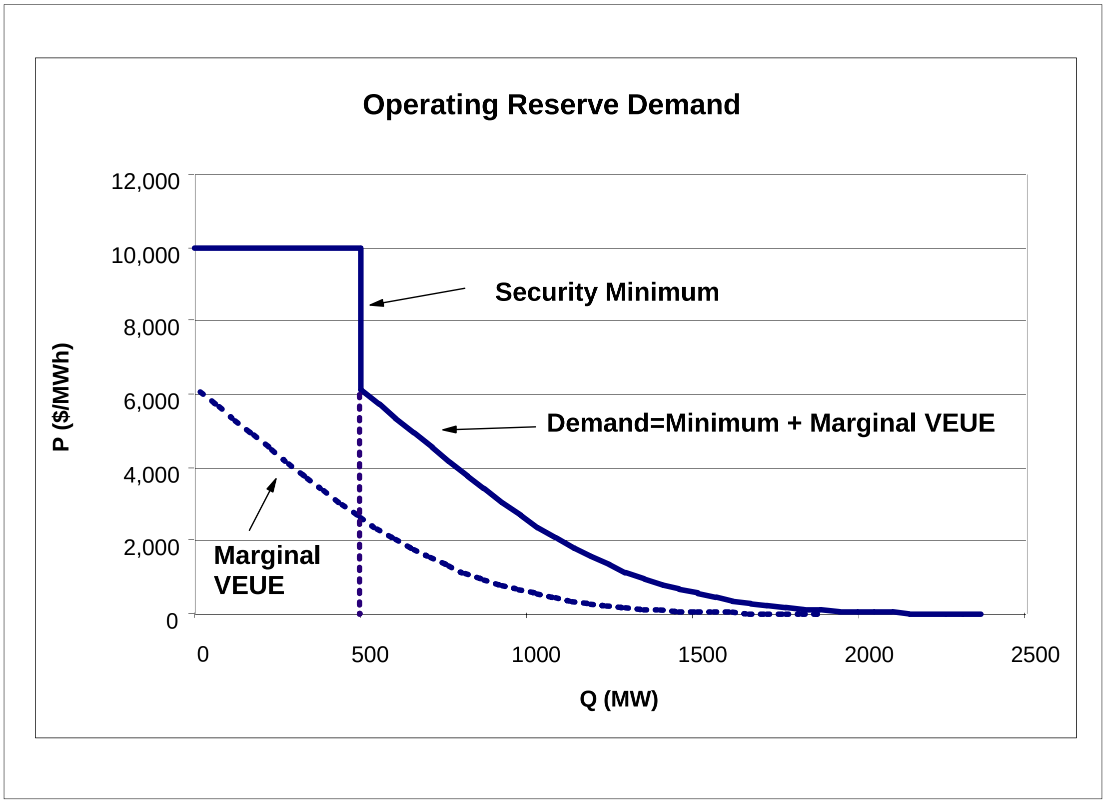
There is no known model that addresses all the complications of formalizing the operating reserve demand curve. For example, there are regional restrictions on reserves that do not lend themselves to the same simplifications of energy pricing that give rise to locational marginal prices (LMP) for energy. There would still be engineering judgment, but the judgment would move from setting the prices to defining regional groupings and translating multiple operating practices into a common metric.

Ignoring the regional grouping, a representative model applied to NYISO data yields an illustrative operating reserve demand curve.[14] This illustrative case is for an expected load of 34,000 MW and representative probabilities of changes in load and generation availability of the next half hour.

With the security minimum set at 500 MW, the remaining demand curve reflects the probabilities and the assumed VOLL of $10,000/MWh for involuntary curtailments based on rolling blackouts. If operating reserve falls below the security minimum, the operator would curtail load, and the price for incremental operating reserves would be $10,000. Above the security minimum, the demand curve reflects the calculated marginal value of expected unserved energy (VEUE).

---

[14] William W. Hogan, "Reliability and Scarcity Pricing: Operating Reserve Demand Curves," Harvard University, March 2, 2006 (available at www.whogan.com).



This example operating reserve demand curve based on representative data illustrates several important points regarding the shape, magnitude and costs. The shape has a simple explanation. As discussed above, there are two underlying demand curves. One is the vertical demand curve from the security minimum defined by the contingency constraint. Second is the more conventional demand curve defined by probabilistic analysis and the value of expected unserved energy. The usual rules apply to yield horizontal addition. Another way of thinking about this is that at the minimum security level of 500 MW, the probability that net demand will exceed expected net demand in the next half hour is less than one. Hence, the curved portion of the demand curve connects at a price below the VOLL.

The magnitude of the illustrativereserve scarcity prices is either very large or very small, depending on the standard of comparison. When considered against the existing maximum offer caps of $1,000 per MWh in most organized markets, the $10,000 figure seems quite high. For example, even in the NYISO case operating reserve demands and prices are determined simultaneously, and have not approached the $10,000 level.

When compared to the standard set by existing resource adequacy programs, however, there is a different story. There have been regular calculations to show that the long-term installed capacity reserve requirements imply a VOLL of $200,000 (or much more) per MWh, yielding a corresponding maximum price for operating reserves. Hence the illustrative operating reserve demand curve prices would be modest by comparison, and this suggests that there is either something wrong with the VOLL assumption or something wrong with the installed capacity reserve requirement.

The demand curve defines the price and this is related to marginal cost. We might be interested in the total cost of expected curtailment implied by our operating reserves rules. The area under the demand curve to the right defines this total cost. If we integrate the area under the demand curve to the right of the security minimum, we find that for the NYSIO case the estimated total cost of operating at the security minimum is of the same order of magnitude as the total cost of the energy generation. We can't ignore the generation costs as that is an explicit part of the dispatch. But it is possible to ignore, or at least not account for, the value of expected unserved energy. The example shows that this cost is not trivial.

Apparently the operating reserve demand curve is important in its own right, and it would be crucial for improving scarcity pricing while we work to expand demand side bidding. There is real money here. If we take this seriously, then there is a series of related issues.

Demand Response

Better scarcity pricing implemented through the operating reserve demand curve would provide an important signal and incentive for flexible demand participation in spot markets. Those market participants that already have access to wholesale market prices and the necessary metering to reflect hourly or shorter price changes would now have an opportunity to react to prices that better reflect the real value of demand response. And those market participants without access or meters, but who have a significant capability to react to prices, would see an incentive to overcome the barriers and react to spot prices. The greater the demand response, the less important will be the operating reserve demand curve. But we should not wait for the demand response before implementing the operating reserve demand curve. Without demand response, the operating reserve demand curve is more important, and it could catalyze an accelerated expansion of demand response. This would be especially true when prices would be highest, and demand response would be most valuable.

Price Spikes

Introduction of an operating reserve demand curve with a maximum price at the VOLL raises the specter of regular price spikes. While this may happen on occasion, to focus on this is a mistake. When price spikes do occur under this model, there is a real shortage of capacity and the price signal must be needed. The higher price would be part of the solution.

A more often overlooked feature is that the operating reserve demand curve implies that there should be some scarcity price adder in virtually every hour of operation, not just when reserves get dangerously close to the security minimum. This would be consistent with the experience in NYISO, and would be a reasonable conjecture for other systems. The contribution to the "missing money" from better scarcity pricing would involve many more hours and smaller price increases.

Practical Implementation

The case of the NYISO disposes of any argument that it would be impractical to implement an operating reserve demand curve. The NYISO price assumptions and parameters could be revisited, but the basic existence test has been completed. The experience from NYISO shows that this operating reserve demand curve, fully integrated with economic dispatch and energy pricing, is feasible and important.

Operating Procedures

Implementing an operating reserve demand curve does not require changing the practices of system operators. The assumption is that the same principles that were followed in developing LMP would be followed here. In other words, reserve and energy prices would be determined as though the decisions by the operators were consistent with the adopted operating reserve demand curve. This would require some translation from the practices into equivalent operating reserve quantities. For example, a small voltage reduction would be mapped into the pricing rule as though it were a reduction of operating reserves. Hence, prices would go up during voltage reductions, not down as they do under current rules. Similar comments would apply to appeals to reduce load, exercise interruptible contracts, and so on.

Multiple Locations

Transmission limitations mean that there are locational differences in the need for and efficacy of operating reserves. There is not as yet a simple way to delineate these requirements to the degree we model LMP differences. However, this is not an insurmountable problem and a workable zonal system appears in the NYISO case. Reserve requirements are different in New York City, but in-city reserves can also contribute to meet total NYISO needs. The pricing model implemented includes these interactions and prices cascade to reflect the combined value of locational reserves.

Multiple Reserves

There are different kinds of operating reserves, from spinning reserves to standby reserves. These are familiar to system operators. With a few simple rules, similar to the mechanism for cascading prices across locations in NYSIO, there could be consistent pricing of multiple categories of reserves.

Reliability

There are and will be operating reserve requirements to meet reliability standards. The same minimum security requirements for contingencies would remain to meet the same reliability requirements. The pricing mechanism provides a stream of revenue even when operating reserves exceed the strict minimum. In addition, the pricing rules make the generator indifferent between generating energy and providing operating reserves. Both features should enhance reliability in the same way that pricing energy at LMP improved reliability and system operations. Market operating incentives would be better aligned with reliability requirements.

Market Power

Introducing an operating reserve demand curve would increase scarcity prices. A natural assumption is that this would increase the problem of market power in electricity markets. Looking a little further, however, would reveal that the reverse may be true.

The analysis of the change in incentives induced to exercise market power would be complicated because the change would affect both the level and the slope of the aggregate demand curve. But there is a simpler argument that the problem of market power would be substantially reduced because mitigation would be easier.

The operating reserve demand curve is not likely to eliminate concerns about market power. Hence, the preferred little "r" methods of mitigation through the use of offer caps would continue to apply. But with the operating reserve demand curve there would be no need to raise offer caps in order to better approximate scarcity prices. Unlike the plan in Texas and the practice in Australia, more realistic scarcity pricing would not require higher or no limits on the offers by generators. Scarcity pricing would be driven by the operating reserve demand curve and not solely by the generators' offers. This would remove ambiguity from the analysis of high prices and distinguish (inefficient) economic withholding through high offers from (efficient) scarcity pricing derived from the operating reserve demand curve.

Hedging

The operating reserve demand curve would likely raise both the average level of energy prices and the volatility of these prices in the spot market. It is difficult to imagine that this change would be politically feasible absent some mechanism to provide average hedges for small consumers.[15] However, there are ready models available that would be highly compatible with improved scarcity pricing. The Basic Generation Service auction in New Jersey provides a prominent example that would yield an easy means for hedging small customers with better scarcity pricing in PJM. Large customers with access to the wholesale market could arrange their own contracts to provide energy hedges. Importantly, this would avoid some of the vexing "deliverability" requirements that complicate other resource adequacy proposals.

Increased Costs

The higher average energy costs from use of an operating reserve demand curve do not automatically translate into higher costs for customers. In the aggregate, there is an argument that costs would be lower.

The "missing money" problem has given rise to various resource adequacy mandates that often involve contracting forward for "capacity." Assuming these are perfectly designed and work well, introduction of the operating reserve demand curve would not increase costs. The well-designed capacity programs are intended to net out the net energy market revenues to reduce the required capacity payments and just provide

---

[15] William W. Hogan, "On An 'Energy Only' Electricity Market Design For Resource Adequacy," Center for Business and Government, John F. Kennedy School of Government, Harvard University, September 23, 2005, (available at www.whogan.com), pp. 26-33.

the "missing money" needed. Hence, on average the use of the operating reserve demand curve would simply reallocate revenues from the capacity payment to the energy payment. The reallocation would result in a better match of prices and incentives to reflect operating conditions.

However, if the resource adequacy models do not work as well as hoped, the operating reserve demand curve would provide an important tool to compensate for the mistakes and provide better incentives to meet the real reliability and operating requirements. Furthermore, the operating reserves demand curve and associated scarcity pricing would apply to all supply and demand, not just to those who were part of the resource adequacy program. The operating reserve demand curve could help reduce the real costs.

Conclusion

Inadequate demand response participation and infrastructure investment are both compelling problems in organized wholesale electricity markets. A common feature is the lack of adequate scarcity pricing and the resulting problem of missing money. The big "R" regulatory solutions often call for mandates and subsidies for favored programs. The little "r" regulatory solution would emphasize reforms of market design to improve scarcity pricing and provide other initiatives to support rather than replace market choices. Analysis of market designs points to the operating reserve demand curve as a missing piece of the picture that could precipitate a virtuous circle of complementary improvements in markets and the associated incentives. Furthermore, the operating reserve demand curve example illustrates the principle of designing and evaluating regulatory interventions to be compatible with the general wholesale electricity market framework. Absent the little "r" action to implement better scarcity pricing, the pressure will continue for more and bigger big "R" interventions. Leadership must come from the Commission. There is a well-established framework for crafting and evaluating market design initiatives. There is a clear problem, and the analysis outlined here points to a workable solution that uses regulation through design of the operating reserve demand curve to support a market approach with flexible incentives rather than mandates.

---

[i] William W. Hogan is the Raymond Plank Professor of Global Energy Policy, John F. Kennedy School of Government, Harvard University and a Director of LECG, LLC. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. The author is or has been a consultant on electric market reform and transmission issues for Allegheny Electric Global Market, American Electric Power, American National Power, Australian Gas Light Company, Avista Energy, Barclays, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator, Calpine Corporation, Canadian Imperial Bank of Commerce, Centerpoint Energy, Central Maine Power Company, Chubu Electric Power Company, Citigroup, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, Conectiv, Constellation Power Source, Coral Power, Credit First Suisse Boston, Detroit Edison Company, Deutsche Bank, Duquesne Light Company, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, GPU Inc. (and the Supporting Companies of PJM), Exelon, GPU PowerNet Pty Ltd., GWF Energy, Independent Energy Producers Assn, ISO New England, Luz del Sur, Maine Public

Advocate, Maine Public Utilities Commission, Merrill Lynch, Midwest ISO, Mirant Corporation, JP Morgan, Morgan Stanley Capital Group, National Independent Energy Producers, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, NRG Energy, Inc., Ontario IMO, Pepco, Pinpoint Power, PJM Office of Interconnection, PPL Corporation, Public Service Electric & Gas Company, Reliant Energy, Rhode Island Public Utilities Commission, San Diego Gas & Electric Corporation, Sempra Energy, SPP, Texas Genco, Texas Utilities Co, Tokyo Electric Power Company, Toronto Dominion Bank, TransÉnergie, Transpower of New Zealand, Westbrook Power, Western Power Trading Forum, Williams Energy Group, and Wisconsin Electric Power Company. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (Related papers can be found on the web at www.whogan.com).