UNITED STATES OF AMERICA

BEFORE THE

FEDERAL REGULATORY COMMISSION

Wholesale Competition in Regions	)	Docket Nos. RM07-19-000
with Organized Electric Markets	)	and AD07-7-000

COMMENTS ON WHOLESALE COMPETITION IN REGIONS WITH ORGANIZED ELECTRIC MARKETS

WILLIAM W. HOGAN, Harvard University

SUSAN L. POPE, LECG, Cambridge MA[i]

September 17, 2007

Introduction

The Federal Energy Regulatory Commission is considering potential reforms to improve the operation of organized wholesale electric markets.[1] The purpose of this paper is to discuss a larger framework for evaluating issues of regulation and market design in electricity markets. Regulation and competition are essential elements of electricity policy. The special requirements of electricity systems create a dual challenge: First, regulation must address issues of market design; markets cannot solve the problem of market design. Second, regulation must complement competition; inconsistent choices in either can undermine the foundations of reliable electricity supply at market prices and subvert the goals of organized electricity markets.

Organized electricity markets are developing with many advanced features that address the technical requirements of electricity systems. Initial defects in market design are being addressed in a continuous process of learning and improvement. As experience accumulates, inevitably problems arise that present challenges for regulators. In some cases, the problems can be addressed through the use of markets and incentives. In other cases, the problems require regulation and mandates. A critical task for the regulator is to provide a proper balance of regulation and markets. This challenge is complicated through the unintended consequences of decisions in each realm. Market design can have significant effects on the outcome of regulation. In turn, regulation can have significant effects on the operation of markets. Whenever regulators must act, there is a choice in the type of action. Big "R" regulatory solutions often call for mandates and subsidies for

---

[1] Federal Energy Regulatory Commission, "Wholesale Competition in Regions with Organized Electricity Markets," Advanced Notice of Proposed Rulemaking, Dockets RM07-19-000 and AD07-7-000, June 22, 2007, p. 1.

favored programs. Little "r" regulatory solution would emphasize reforms of market design to improve incentives or limits on regulatory mandates to support rather than replace market choices. Regulation may be unavoidable, but there is flexibility in the type of regulation.

A framework for evaluating issues of regulation and market design in electricity markets helps regulators identify regulatory choices that minimize the unintended consequences in markets, and identify market design features that can support the goals of regulation. A concern is that major regulatory decisions are being made without consideration of the interaction with markets and market design. The result is both a failure to resolve the immediate problems and collateral damage to operation of the market. The cycle precipitates more problems and more need for regulatory mandates to counter the effects of poor incentives in market design.

This is an avoidable problem. The discussion here illustrates the type of problems that arise in organized markets and provides examples of innovative approaches addressing the problems that balance regulation and market design.

Coordination for Competition

The Commission has responsibility for regulating wholesale electricity markets. The stated framework emphasizes support for competition in wholesale markets as a clear and continuing national policy:

> "National policy for many years has been, and continues to be, to foster competition in wholesale power markets. As the third major federal law enacted in the last 30 years to embrace wholesale competition, the Energy Policy Act of 2005 (EPAct 2005) strengthened the legal framework for continuing wholesale competition as federal policy for this country.
>
> The Commission's core responsibility is to 'guard the consumer from exploitation by non-competitive electric power companies.' The Commission has always used two general approaches to meet this responsibility—regulation and competition. The first was the primary approach for most of the last century and remains the primary approach for wholesale transmission service, and the second has been the primary approach in recent years for wholesale generation service.
>
> The Commission has never relied exclusively on competition to assure just and reasonable rates and has never withdrawn from regulation of wholesale electric markets. Rather, the Commission has shifted the balance of the two approaches over time as circumstances changed. Advances in technology, exhaustion of economies of scale in most electric generation, and new federal and state laws have changed our views of the right mix of these two approaches. Our goal has always been to find the

> best possible mix of regulation and competition to protect consumers from the exercise of monopoly power."[2]

A task for regulation is to support this policy framework while developing hybrid markets and dealing with both the limits of markets and the failures of market designs. As Chairman Kelliher observed at the Commission's related 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]

In the case of electricity, there are many reasons for employing regulation to constrain and direct market operations. There are monopoly elements in the system and concerns with the same issues that give rise to regulation in other markets, although part of the motivation for electricity restructuring was a view that changes in electricity generation had reduced the scope of monopoly and would allow for greater reliance on markets. There is a great deal of experience in employing regulatory systems to deal with traditional problems of monopoly or other forms of market abuse. Developing good regulations and avoiding unintended consequences is not easy, but the scope and nature of the traditional problems are understood.

In addition to these familiar conditions from other industries, there is a special problem for electricity with its complex and large interactions in electricity flow across the transmission grid. With current technology, unfettered operation of a decentralized electricity market is not possible. There must be a central coordinator that controls operation of the transmission grid to ensure moment-by-moment system balance and reliability. This is the system operator, and every system has a system operator. This is a monopoly function, and there must be rules imposed on the operation of this monopoly to make the system work. *Hence, the issue is not whether there must be regulation coupled with explicit consideration of market design, but rather what should be the nature of the regulations on the electricity grid and the wholesale market.*

In the nature of electricity, there must be coordination for competition. While it might seem like a contradiction viewed from the perspective of other industries, central coordination of the market is a necessity in the case of the wholesale electricity system. This presents a task for the regulator because the need for this central coordination means that the market can't solve the problem of market design. A balance is required, and regulatory stability has its own value. But when problems arise and regulators must act, the first focus should be on the market design.

The challenge for the Commission, the regulator of the wholesale market, is further complicated by the unavoidable technical details that cannot be simplified away. The nature of the high voltage grid at the core of the wholesale market imposes substantial constraints on market design. Natural or obvious approaches to market design

---

2 Federal Energy Regulatory Commission, "Wholesale Competition in Regions with Organized Electricity Markets," Advanced Notice of Proposed Rulemaking, Dockets RM07-19-000 and AD07-7-000, June 22, 2007, pp. 4-5.

3 Endorsing comments attributed to Fred Kahn.

developed by analogy from other markets are in sometimes sharp conflict with actual operation of the grid. The long and intellectually painful discussion of the "contract path" model provides a prominent example, and makes the point that a market design element developed without or despite a larger framework that is specific to electricity can be both defective and dangerous.

The mandates for open access and no undue discrimination in the use of the transmission grid, developed by the Commission as fundamental to its policy, reinforce the requirement to recognize the technical details of electricity and the consequential restrictions on the market design. If the Commission gives more than lip service to these principles, then it is critical that its regulations and market design choices accommodate the physical characteristics of the operation of the grid. Simply put, if the market incentives and choices that apply to everyone are not consistent with the physical reality, then open access and non-discrimination would be a threat to reliability and impossible to maintain.

Hence, to remain successful there must be a hybrid for the wholesale electricity market[4] that integrates regulation and competition. The dictates of open access and no undue discrimination combined with the nature of the grid substantially determine the fundamental framework for the market design.

The Market Design Framework

The larger framework integrating the many components and decisions that arise in designing wholesale electricity markets follows from both analysis of economic first-principles and the by now extensive experience in organized wholesale markets: bid-based-security-constrained-economic-dispatch-with-locational-prices-and-financial-transmission-rights.[5] As the Commission knows well, there has been an enormous amount of study and experimentation with electricity markets, 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, the rationale and practical support for it require repetition. 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. For the benefit of new market participants, and new regulators, it is necessary to refresh the general awareness that there is a real problem here—electricity is different—and the proved market design contains

---

4 The term "hybrid market" is used generally here and is not a reference to the hybrid market implemented in Ontario.

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.

many complementary pieces that fit together to provide a solution in light of these differences. Furthermore, for those who have not been taught by the economists, it is easy to forget that the prices and associated incentives flowing from the market design are also complex, and are just as important to get right in order to provide incentives consistent with the engineering details.

The fundamental problem centers on getting market pricing in place to provide the proper decentralized operating and investment incentives while creating the associated property rights to allow market-based investments to go forward. Ironically, despite the importance of long-run investment, for reasons that are peculiar to electricity, the critical pricing rules and conditions arise in the wholesale spot market. In order to provide better incentives for long-term market-based investments that support reliability, it is critical to provide the prices in the spot-market that reflect actual operation of the grid. The regulator concerned with long-run investment should focus first on the spot market.

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

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. Many seemingly innocuous decisions appear isolated and unique, but on closer inspection are seen as fundamentally incompatible with and hence undermine the larger framework. This creates a slippery slope problem, where one ad hoc solution creates the need for another, and regulators are driven more and more to intervene in ever more ad hoc ways. The result is big "R" regulatory micro-management. For example, socialized costs for preferred infrastructure investment can easily reduce the incentives for other market-based investments, thereby increasing the need for regulators expand the scope of their intervention to additional 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. 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 by 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 unavoidable regulatory mandates have some hope of avoiding the slippery slope.

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 apply the test of compatibility with the basic market framework. The discussion that follows of the need to support better demand response and infrastructure investment provides illustrations of the general proposition.

Spot Markets and Investment Incentives

Electricity systems are capital intensive. The most important decisions involve long-term investments. There are substantial uncertainties that create significant financial risks for these investments. One of the most important motivations for electricity restructuring was dissatisfaction with the nature of the investment decisions and associated allocation of risks under the traditional regulatory model.

The most important role of markets is in allowing for new mechanisms for making those investment decisions and allocating the risks. A key presumption of greater reliance on markets is not that it would eliminate the uncertainties and risks, but that there would be a better allocation of the risks that would lead to better investment decisions. In particular, investment decisions would be primarily market-based, with the associated risks and returns accruing to the market participants who make the decisions. The responsibility for investment decisions would shift primarily from the regulatory arena to the market setting.

If this were not the case, and investment decisions could not be left primarily to market choices, there would be little justification for electricity restructuring. Risks do not disappear under a regulated paradigm. However, if regulators and central planners knew better when, where and how to invest, then they should simply make the decision and allocate the costs as before. There would be little need for the wrenching changes that accompany electricity restructuring. A key fundamental is that the assumption that central planners and regulators are at a disadvantage compared to the marketplace when it comes to managing commodity pricing risks.

This argument has two implications. First, to the extent that market forces prove inadequate and there is a move to central planning of investment, this must be seen as a failure of electricity restructuring and the development of the associated hybrid market structure. Second, the primary focus of electricity market design should be on supporting the necessary ingredients to support market-based investment decisions.

It might be surprising that support of market-based decisions for long-term investments depends so critically on the hybrid spot market design. A good spot market design is important for reliable short-term operations. But in part because of the special nature of electricity grids, good spot market design also is essential in providing the incentives and support for long-term investment.

The argument for the importance of good spot market design in supporting reliable operations is straightforward. Under the principles of open access and no undue discrimination, market participants have substantial flexibility in how and where they choose to generate and consume electricity. Yet the complex interactions through the transmission grid dictate that the choices must conform closely to the coordinated solution identified by the system operator. A good market design provides incentives to produce and consume that reflect the actual operational restrictions of the grid. Any other market design would provide incentives that deviate from the dictates of efficient and reliable operations. If market participants respond to those incentives, the system operator will necessarily have to act to undo the effect of those perverse incentives. This means the system operator must deviate from open access and discriminate among the market participants. It follows, therefore, that a key feature of spot market operations is that the incentives in the market should reflect the essential details of efficient and reliable dispatch. Working out the main features of those incentives amounts to analysis of the properties of reliable economic dispatch, and leads immediately to the basic framework of bid-based-security-constrained-economic-dispatch-with-locational-prices.

What is less straightforward is the explanation of why this same framework is necessary for support of market incentives for long-term investments. A full exposition of the details would be lengthy, but the basic argument can be summarized without going into all the details. The argument emphasizes two features of the operation of an electricity market that would be essential for successful market-based investments: the balancing market, and long-term transmission rights.

The role of the balancing market is the easiest to appreciate. The need for instantaneous balancing of aggregate load and generation makes a balancing market necessary. Even under the best of intentions, suppliers will not be able to follow the load profile of their customers exactly and will always be over or under on any bilateral transaction. The difference must be made up in the balancing market. If the balancing market prices and rules reflect the true costs at the margin, the balancing market can provide efficient outcomes without overly restrictive rules.

In making investment decisions, and arranging long term contracts, market participants will know that one alternative to full performance under the contract would be to turn to the balancing market at any given moment over the course of the contract. If the expectation is that the balancing market will be priced efficiently, then this reinforces efficient investments and contracts. But if the balancing market does not reflect real costs and real operations, expectations by market participants will take this into account and bias the choice of investments and contracts.

To see the importance of the incentives in the balancing market, consider an example that reflects a common argument. With either efficient pricing or regulatory mandates, the volumes handled through the balancing market could be small. It is often proposed that these small volumes in balancing market transactions need not be priced individually, on the grounds that on average the “overs and unders” would cancel out over time. However, with this rule in a market setting without efficient pricing or regulatory mandates, there would be an incentive not to sign long term contracts and simply lean on the balancing market at average prices, especially for those facing high actual balancing market prices. Assuming that the volumes would be small and would not need efficient pricing would be a fallacy leading to a “self-unfulfilling” prophecy, in that the volumes would end up being high and the overs and unders would not cancel out. With open access and no undue discrimination, efficient balancing is required and stands at the center of the market framework.

In addition to providing proper incentives for contracts and long-term investments, an efficient balancing market with the associated locational prices provides the best available mechanism for defining long-term transmission rights. These are the so-called financial transmission rights (FTRs).[6] The nature of grid interactions as power flows along every parallel path precludes any meaningful definition of physical property rights that would determine the flow of power on the grid. The feasibility of any particular transaction depends on the nature of all the other transactions, which are constantly changing. Hence, there is no means consistent with open access and no undue discrimination that could provide workable long-term physical rights to use the grid to deliver power from source to destination.

Without transmission property rights that connect generation to load, long term contracts cannot be supported without leaving the parties exposed to substantial risk. There is a need to define property rights that would capture the benefits of long-term physical transmission rights but without the need to match the individual energy flows to follow the individual rights. The resolution of this dilemma exploits the design of the efficient spot market. If actual usage of the transmission grid is priced according to the differences in the locational prices, then transmission pricing conforms to the physical operation of the grid and transmission decisions can be consistent with reliable operations. Pricing at the difference in locational prices is the efficient solution. This produces a sometimes substantial contribution of congestion revenues. Distributing those revenues to the holders of FTRs produces a result which is economically equivalent to holding moment by moment auctions of physical transmission rights. In effect, the FTRs become the missing property right that hedges the locational cost differential and supports long-term contracts. The physical auction would be impossible to administer, but the FTR payments amount to a relatively simple after the fact accounting exercise.

The basic market framework of an efficient spot market provides both balancing and transmission services. The locational prices for balancing and differences in locational prices for transmission are consistent with each other and support efficient, reliable operations and investment. In addition, the basic market framework provides the ingredients for FTRs necessary to support long-term contracts and long-term investments.

---

[6] In regions with organized day-ahead markets, FTRs settle against day-ahead prices, which also reflect balancing of generation and load. The day-ahead prices are related to the prices anticipated for the real-time balancing market. The distinctions are not important for the present discussion.

Market Defects, Market Failures and Regulatory Mandates

The general framework of bid-based-security-constrained-economic-dispatch-with-locational-prices-and-financial-transmission-rights provides an effective design that conforms to the requirements of open access and no undue discrimination. The framework has complementary components that serve to meet essential short-run and long-run requirements. The pieces fit together, and any analysis of market or regulatory rules should be made with consideration as to how the rules interact with that general framework.

However, the general framework is neither perfect nor complete. Given the state of our knowledge, there is no perfect or complete framework that provides a market-based solution to all problems. Regulation is needed to implement the framework, and regulation has a role in addressing the defects and failures of the market design. The distinction between defects and failures is intended to capture the difference between errors of implementation versus fundamental limitations of a market. Regulations may arise to meet both types of problem, but they are different. However, a key to successful regulation and a working hybrid market is in testing rules and mandates for their compatibility with the basic market design.

Actual operation of the market design will expose implementation defects that may not have been anticipated, or that were not thought serious enough to deserve attention. The defects may undermine operations or investment incentives. Typically, defects that compromise operations and spot markets cry out immediately for modifications in the market design to remove the problem and support the general framework. The repeated unhappy experience with zonal pricing models for the spot market, with the now virtually universal move to locational pricing under the general framework, is a case in point.

When market design defects undermine investment incentives, there is less immediate pressure to fix the problem and much greater interest in adopting regulatory mandates that promise to undo the bad effects of the market incentives. But this tendency ignores the arguments above about the importance of short-term markets in supporting long-term investments. If the market design defect can be identified and a viable alternative developed, then the priority should be on fixing the defect rather than maintaining the poor incentives and creating another structure to undo the effects.

The general framework inevitably confronts market failures in the sense that the markets do not by themselves provide a natural solution to all problems. For example, the general framework assumes a workably competitive market without material monopoly power in ownership and operation of generating facilities. Where this is not true, the basic design does not prevent exercise of market power. Some form of regulatory intervention is required. Little “r” interventions that support the market would be better than big “R” designs that might have appeal in isolation but which would undermine operation of the market.

The examples here illustrate both types of problems with an eye on the impact on investments decisions. For market defects, the little “r” approach is to improve the market design. For market failures, regulatory mandates may be needed and the little “r” approach is to design the regulatory mandates to mitigate conflicts with the general market design. We consider first a market defect with material impacts on spot-market prices that have a detrimental effect on long-term investment incentives: the case of pricing to reflect supply and demand in tight markets. Then we turn to a case of market failure and long-term investment in transmission with substantial free-rider effects.

Demand Response, Price Incentives and Infrastructure Investment

There have been impressive accomplishments through the organization and operation of Regional Transmission Organizations (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] Additional systematic quantitative evidence is accumulating, as reported by others in the first technical conference in the series leading to this proceeding.[8]

It is apparent, however, that most regulators would be hard-pressed to justify the costs and turmoil of the transition of electricity restructuring based only on the results to date. There have been significant benefits, but there should be even greater benefits to come. 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. 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 spot 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.[9]

---

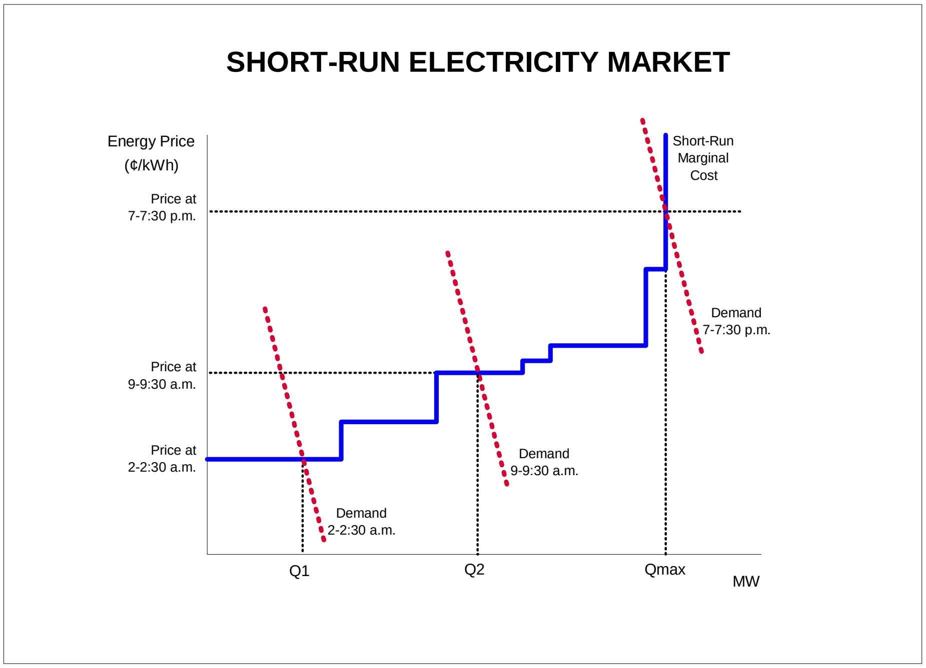
[7] William W. Hogan, "Getting the Prices Right in PJM: Analysis and Summary: April 1998 through March 1999, The First Anniversary of Full Locational Pricing," Harvard University (available at www.whogan.com), April 2, 1999.

[8] Harvey, Scott, Bruce McConihe and Susan Pope. "Analysis of the Impact of Coordinated Electricity Markets on Consumer Electricity Charges." LECG, 18 June 2007. (available at http://www.ksg.harvard.edu/hepg/Papers/LECG_Analysis_061807.pdf.)

[9] 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 set by the intersection of supply and demand and 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.[10] It is important to recognize that this is not a problem of secondary importance that we can defer in deference to other pressing issues. 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 from the energy transactions in the spot market. From this perspective, it is not surprising that there is both a theoretical and actual concern about whether there is inadequate investment in generation and related infrastructure.

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 short-term pricing in tight markets. 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

An obstacle to 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.[11] 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.[12] 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, improved pricing would have a major impact on the nature of such contracts and could greatly simplify matters such as providing a price signal for the deliverability of generation capacity.

---

[10] 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.

[11] The Australian market employs a $10,000 MWh cap and allows very high generator offers that set prices in short-duration periods of scarcity.

[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. 27-33, (available at www.whogan.com).

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 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.[13]

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 over-simplification in the early discussion and implementation of market design for wholesale electricity markets. The stylized spot market figure illustrating equilibrium pricing simplifies 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 market defect 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 inadequate 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.[14]

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. 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

---

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

14 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.

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 serves as a dispatch 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.

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 in the contingency-only model. 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 various practices 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.[15]

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.

This is not a new idea, and it is not simply a conceptual proposal. For example, the New York Independent System Operator (NYISO) and the Independent System Operator New England (ISONE) adopted the concept and implemented an operating reserve demand curve integrated with the energy market design in the spot market.[16] 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 or ISONE operating demand curves 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.

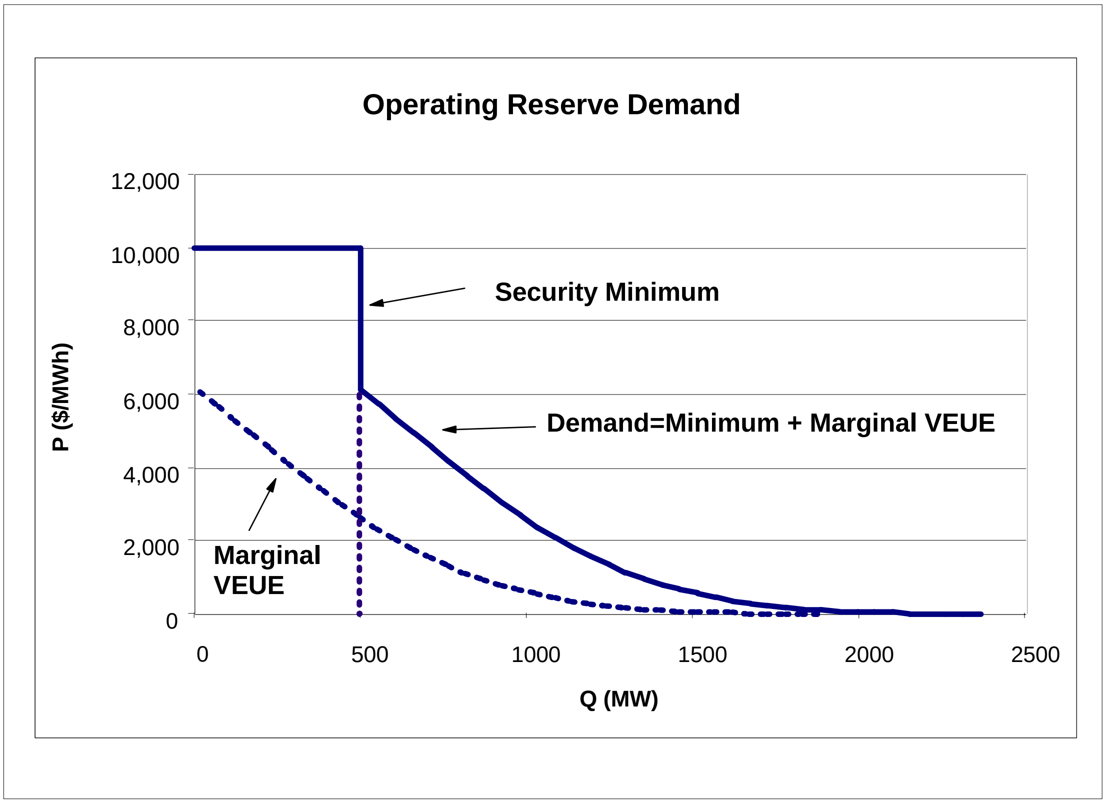
Using a formal albeit simplified model and representative data from the NYISO, it is possible to outline how to obtain a reasonable operating reserve demand curve and to determine the quantitative implications of its use. This in turn provides the opportunity to address some of the obvious questions 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. Operating reserves are intended for short-term use when called and are subject to different transmission limits. These limits are largely probabilistic and based on engineering analysis. 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.

---

15 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.

16 The NYISO and ISONE implementations are discussed and compared in the white paper by the California Independent System Operator, "California ISO Straw Proposal: Reserve Scarcity Pricing Design," September 5, 2007, http://www.caiso.com/1c51/1c51b3ab4fea0.pdf.



Ignoring the regional grouping, a representative model applied to NYISO data yields an illustrative operating reserve demand curve.[17] 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 for operating reserves, the remaining demand curve reflects the probabilities and the assumed value of lost load (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).

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

---

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

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 illustrative reserve 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 measure 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. The current market designs have not ignored generation costs as this 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 the cost of this market defect 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-term 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 cases of the NYISO and ISONE dispose of any argument that it would be impractical to implement an operating reserve demand curve. The price assumptions and parameters could be revisited, but the basic existence test has been completed. The experience 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 locational pricing 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 locational price 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 and ISONE, 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 the locational price 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, reveals 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.[18] But there is a simpler argument that the problem of market power would be substantially reduced because its 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 generation offer caps in order to better approximate scarcity prices, since the high prices could result from the operating reserve demand curve given the existing price caps. 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.[19] However, as previously stated, there are ready models available that would be highly compatible with improved scarcity pricing. The Basic

---

[18] With the same demand slope, higher prices reduce the marginal incentive to exerciser market power. Hence, the net effect on the incentive to exercise market power would depend on the facts.

[19] 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.

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 and aggregators 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. The providers of the hedges would solve the deliverability problem through the market, without this otherwise complex task falling on the regulator. These commercial firms would address the substantial volatility in the spot market which would better reflect the real operating conditions, without requiring the hedged customers to face that volatility.

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 appropriately 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 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 alternative means to get the "missing money" to suppliers. It also would provide real time (i.e. in increments of five minutes) 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 had been selected as part of the long-term resource adequacy program. The operating reserve demand curve could help reduce the real total costs.

Argentine Approach to Transmission Investment

Within the U.S., there has been extended debate concerning the most efficient and effective way to identify and pay for investments in the electricity system; this debate applies broadly to transmission, generation and demand-side resources. Conventional regulatory approaches to mandating investment are colliding with the successful creation of day-ahead and real-time electricity markets in many regions of the U.S., pushing these regions toward a slippery slope of increasing reliance on regulation accompanied by a degradation of the price incentives that drive the competitive sector. Recognizing the need to make the approach to investment, transmission in particular, compatible with the design of competitive electricity markets, some have emphasized market-based investment, while others have recommended the use of competitive solicitation processes.

In the case of economic transmission investment, there is an argument that market-based approaches cannot be relied upon to produce much or any of the needed infrastructure.[20] The reasons for this market failure include a combination of regulatory uncertainty and the inability of property rights defined by incremental FTRs to capture the benefit of transmission investment. The typical example revolves around the case of a large, lumpy transmission project with many potential beneficiaries. Because the project is lumpy, the investment, if made, would eliminate congestion and eliminate the value of the FTRs. Because there are many beneficiaries that will benefit from the investment, and those that do not pay cannot be easily excluded, the free-rider problem arises and everyone waits for someone else to incur the cost of the investment. In this stylized circumstance, some form of regulatory mandate appears as a natural and perhaps inevitable solution.

The appeal of the argument moves its proponents to further apply the regulatory mandate to all forms of transmission investment, even to those projects that are of a scale and nature that should not give rise to the free-rider problem. Once this extension occurs, a natural question arises as to how to treat alternatives to transmission such as generation and demand-side investments. If these compete with transmission, should they also be eligible for support under the regulatory mandate? Absent a principled answer that can distinguish between the different types of investment, the pressure pushes the regulator down the slippery slope of greater use of big "R" regulatory mandates and more centralization of investment decisions of all kinds, undoing the market incentives for long-term investment without explicitly reverting to the traditional regulatory model.

The little "r" alternative is not to deny the potential need for regulatory mandates for the cases that give rise to a serious free-rider problem. Rather, the need is for a practical mechanism to distinguish the cases that give rise to serious free-rider problems without along the way moving the bulk of the investment decisions and risk allocations back into the regulated model with central planning.

Such an approach is available from the experience in Argentina. The Argentine approach to transmission expansion is a variant that relies on users rather than regulators to identify investment projects, vote them in, and fund them.[21] This is a beneficiary-pays model that works.

The Argentine experience illustrates a way forward to more effectively integrate market-driven and regulated transmission investment within the competitive electricity markets now established within many regions of the U.S. This experience has not been examined closely within the public debate concerning electricity market design. It provides a framework to avoid the slippery slope that appears when the pressure to provide regulatory approaches to transmission investment creates the unintended consequence of undermining market decisions. The Argentine experience provides an example of the application of an approach that recognizes and responds to the pressures,

---

[20] Transmission investment necessary to meet a reliability standard presents a separate issue. Although all transmission investments have both reliability and economic impacts, the focus here is on making investment decisions for economic reasons while assuming that reliability constraints are met.

[21] Stephen C. Littlechild and Carlos J. Skerk, "Regulation of Transmission Expansion in Argentina Part I: State Ownership, Reform and the Fourth Line," CMI Working Paper 61, 2004, and "Regulation of Transmission Expansion in Argentina Part II: Developments Since the Fourth Line," CMI Working Paper 62, 2004. The account in this paper is largely drawn from these sources.

but is in the form of a little "r" framework of regulatory design in support of market decisions.

Transmission Investment Framework in Argentina

From 1992 to 2002, Argentine regulators implemented a novel approach to major expansions of the transmission system that depended on users to propose, adopt, and fund the expansions. Perception of the relative success or failure of this approach is largely entangled with judgments about whether the ensuing delay in the construction of a major transmission line to Buenos Aires, the "Fourth Line," was or was not in the public interest. A review of the particular debate and of the associated cost-benefit analyses for that particular investment is beyond the scope, and not the point of this paper. Those who are interested may read the careful analysis of Littlechild and Skerk, which argues that the delay of the Fourth Line was probably efficient and may be taken as evidence of the success of the Argentine's novel approach to voting-in major transmission investment.[22]

The point here is that the remainder of the Argentine experience, with less prominent transmission investments, is the more important story for consideration in the United States. The Argentine model was a great success at supporting both voluntary market-based investments and mandatory regulated investments in a way that minimized the unintended consequences for markets. The Argentine example illustrates a way to avoid the slippery slope without foregoing the principal protections provided by the regulatory option.

One of the primary reasons for the Argentine's approach to market-driven transmission investment resonates strongly in the U.S.: a desire for increased efficiency in the location of transmission and generation facilities. In Argentina, as in the U.S., there was a realization that decisions concerning transmission investment were inextricably related to decisions concerning the location of new generation plants, and thus were central to the long-run efficiency and security of electricity supply. However, other drivers are less strongly felt here. The Argentine's approach to transmission investment also was chosen because they did not trust their conventional regulatory framework to provide efficient transmission investment decisions. The Argentines had experienced a decade of uneconomic and extremely costly investments in high voltage transmission that undermined the credibility of the traditional regulatory approach. While this suspicion of regulatory approaches to transmission investment is not as strongly felt in the U.S., a parallel experience does exist in the realm of generation planning, and experience which lies at the core of the interests in greater reliance on market-based decisions.

Because of their interest in supporting market-based approaches in the early 1990's, the Argentine's response was to create a competitive process to drive transmission investment that was as far as possible independent of regulatory or

---

22 The Fourth Line ultimately was voted-in and built, but this result reflected the contribution of extra funds that reduced the cost of the line and altered the incentives of some of the voting beneficiaries. Littlechild and Skerk conclude, "Subject to these limitations, the Public Contest method worked well in that it increased the efficient use of transmission, and forced a much-needed reappraisal of the most economic way to supply electricity to Buenos Aires…" (Part I, p. 4)

government involvement, but maintained the technical competence of the transmission system as a whole. Fundamental elements of the Argentine approach are accepted in many regions of the U.S., such as separation of transmission ownership and control and the operation of a coordinated spot market for electricity organized under an ISO and settled using Location Marginal Pricing (LMP).[23] The Argentine approach embraced the idea that existing and new transmission facilities could be regulated separately and differently, but needed to include necessary rules for third-party access. The Argentines also accepted the idea that there could be competition for the ownership, operation and maintenance of new transmission lines, for example among construction companies and existing transmission companies. Transmission and sub-transmission companies were not allowed to buy or sell energy.

Sitting on top of this basic framework, the Argentine regulators designed an approach to new transmission investment with four separate paths[24]:

- Expansion of Transmission Capacity by Contract Between Parties. This path allowed new merchant transmission to be added to the Argentine system through voluntary participant funding.
- Minor Expansions of Transmission Capacity (<$2M). The Argentine model included special rules to reduce the transactions costs for minor transmission expansions for which there was little threat of uneconomic behavior. These investments continued to be regulated and the costs were assigned to the beneficiaries, either through negotiation or as determined by the regulator, with mandatory participant funding.
- Article 31 of the 1992 Act enabled the Secretary of Energy to authorize a generator, distributor or large user to build a transmission line at its own cost for its own use.
- Major Expansions of Transmission by "Public Contest" Method.

The Public Contest method, the focus here, is relatively simple in concept. Under the Public Contest method, beneficiaries of major transmission expansions had to propose, approve and pay for the expansions. Once a transmission expansion was proposed, the regulator was required to apply the "Golden Rule" to check that the total costs of generation, transmission and outages would be expected to be lower as a result of the proposed transmission expansion. Next, the system operator identified the set of beneficiaries in the "Area of Influence" for the proposed expansion, and each beneficiary's estimated usage of the new line. The value estimated for each beneficiary determined its weight in the vote held to determine whether or not the proposed expansion would be constructed. The value for each beneficiary also defined its mandatory percentage obligation to pay for the expansion. Only the beneficiaries

[23] The details of this description are from Part I, pp. 18-24.

[24] In this paper we focus on changes to transmission expansion policy implemented in Argentina in 1992. Subsequent refinements and a second round of reforms considered in 1998 are not central to the discussion.

identified by the system operator were eligible to vote on whether or not the proposed transmission expansion would move forward. The outcome of the vote was determined based on the 30%-30% Rule. Under this rule, at least 30% of the beneficiaries had to vote in favor of the proposed expansion, and no more than 30% of the beneficiaries could vote in opposition to the expansion. If the proposed major expansion passed these hurdles, construction, operation and maintenance of the new facilities was put out to bid. Once built, all the beneficiaries would be required to pay.

There were, in addition, two context-specific elements of the Public Contest method; these are not integral to the Public Contest method and would not need to travel with the main ideas, if these were to be considered within the U.S. First, the Argentines did not award FTRs to parties that paid for the new transmission expansion. The award of FTRs is a logical extension of the application of competitive market principles to transmission investment and, in fact, the tariffs of virtually all of the U.S. where ISOs operating locational pricing markets call for the award of FTRs to parties that fund so-called merchant transmission expansions. The FTRs would provide the property rights that would accompany transmission investment. It is remarkable that the Argentine system was able to support a beneficiary funded system even without the allocation of these property rights, but use of the FTRs would only help make the system work better.

Second, because there were no FTRs assigned, there was an accumulation of congestion payments under the LMP market design with the existing grid. Under the U.S. approach, these congestion payments are disbursed to the holders of FTRs. Instead, the Argentines allocated accumulated congestion rents, called the "Salex" funds, to reduce the cost of new transmission expansions proposed under the Public Contest method. This allocation was not consistent with the market solution and clearly had the potential to tilt the economic incentives explicitly designed into the method. Fortunately, recreation of the Salex fund is unnecessary in the presence of FTRs for the existing grid.

Transmission Investment Lessons from Argentina

The Argentine approach to transmission investment has the desirable property that the responsibility for identifying and voting-in new transmission rests exactly with the beneficiaries, and in the same proportions, as the obligation to pay for the new transmission if it is approved. It allocates the responsibility, risks, rewards and the payment obligation to the same parties, which goes a long way to incenting economic behavior. In the story that is told, the Argentines appear to have backed into this result. They decided early-on that the construction and operation of new lines could be put out to bid. They then faced the question of who should determine which new transmission facilities should be built and who should pay for them. They turned to economic theory that suggests that setting charges, for example for new transmission, in proportion to the benefits received by users would minimize welfare loss due to the overall charges being above the level indicated by the locational prices.[25] Once they decided to charge the costs to the beneficiaries, they then saw that it would be consistent to choose network expansions or reinforcements to maximize the net benefits to these beneficiaries. This benefit test became the "Golden Rule," which reassures regulators and policy-makers that

[25] Part I, p. 24.

a proposed expansion is in the public interest without relying on centralized planning to advocate specific transmission investments in the first place.

The next step was the realization that if beneficiaries were going to be required to pay for new transmission investments, then they would have an incentive to propose and adopt investments that they would expect to bring them net benefits, and reject those that do not. The Argentine model does not rely solely on incumbent transmission companies to propose new transmission investments, although they must supply information and support for the process.

In the early 1990's, the decision to place responsibility for instigating transmission expansion on beneficiaries was novel. Today this initiative is encouraged or accommodated within the tariffs of the U.S. with ISOs operating organized markets, but there is always a regulatory backstop to provide for transmission expansion in the event that fears of market failure materialize. In Argentina, on the other hand, placing responsibility for action on the users of the network was a daring and assertive step away from relying on regulatory approaches, and towards designing an approach to transmission investment that was consistent with the rest of their competitive market design. The Argentines sought improved performance through support of markets and competition, and giving users the opportunity to identify and determine the direction of transmission expansion given the allocation of the costs involved.

The identification of beneficiaries in Argentina addressed the issue of dispersed benefits by attempting to include all users potentially affected by a proposed expansion. In practice, the identification of beneficiaries – generators located in export areas that are not dispatched because of transmission constraints or loads located in import areas that do not have access to remote generation – is determined from a simulation model corresponding to that used to set nodal prices. Roughly speaking, the model is used to estimate the change in flow over a line due to a change in consumption or generation by a beneficiary. The voting share of a beneficiary is the weighted average of its estimated usage over the first two years that the new transmission facilities are in service.

The particulars of calculating the benefits are details that could differ in different regions and for different types of investments. The important point is that the Argentine model demonstrates that support of market-based approaches, and constraints on mandated regulatory investments, can be made to work in a framework where the beneficiaries pay for transmission expansion. This framework allows for mandated investments in some circumstances without the unintended consequences of undermining market incentives and decisions.

Putting aside the debate over the Fourth Line, the transmission investment policy reforms appear to have resulted in substantial investment in transmission in Argentina. By the end of 1994, just two years after the implementation of the Public Contest method, "three major high voltage lines totaling 853 km had been successfully put out to competitive tender."[26] "[O]ver the period 1993 to 2003 the length of transmission lines increased by 20 percent, main transformers by 21 per cent, compensators by 27 per cent and substations by 37 per cent, whereas series capacitors increased by 176 per cent. As a

[26] Part I, p. 31.

result, transmission capacity limits increased by 105 per cent, more than sufficient to meet the increase in system demand of over 50 per cent and several connections totaling 60 km had been made to the high-voltage system."[27]

Translating to Organized Markets in the United States

In terms of the current policy debate in the U.S., the Argentine experience can be taken as evidence that participant-funded transmission expansion, along with clearly defined rules about who pays, can lead to substantial investment in transmission. The central step of identifying beneficiaries and aligning decision-making responsibility and the obligation to pay helps unlock the puzzle of how to integrate market-based and regulated investments. This alignment allows competitive market incentives to function in determining efficient levels of and locations for new transmission and makes this process compatible with short-run competitive markets for energy. In the U.S., the award of FTRs would also be included in the process, and would not be a controversial step given the general level of acceptance and understanding of financial transmission rights.

The Argentine model provides a principled answer to the question of extending regulatory mandates beyond large transmission projects to generation and demand-side investments. The beneficiary-pays principle and the voting rules differ from conventional market-based investments only in the case of large lumpy projects with substantial free-rider effects. For most generation and virtually all demand-side investments the projects are inherently small and the beneficiaries limited. These projects do not require regulatory mandates and under the Argentine model would not be affected. For the few generation or demand-side investments that were large and lumpy, the argument would include these investments under the Public Contest rubric without requiring the extension of regulatory mandates to all investments.

It does not appear that there any major structural dissimilarities between U.S. electricity markets and the Argentine market in the early 1990's that would impede application of some or all of the Argentine innovations in the U.S. For instance, competitive electricity markets in the U.S. already are organized under ISOs with locational pricing and financial transmission rights. In terms of the three major paths described under the Argentine model:

- Expansion of Transmission Capacity by Contract between Parties. The basic market model of the ISOs can accommodate voluntary participant-funded merchant transmission expansion and allocate long-term FTRs to the parties funding the expansion.
- Minor Expansions of Transmission Capacity (<$2M). Small, regulated investments could be left to the initiative of the existing wires companies. The incremental revenue for these expansions should implicitly or explicitly flow to the rate payers paying for the expansion, if possible, and the incremental FTRs could be auctioned along with the FTRs for the existing system.

---

27 Part II, p. 56.

- Major Expansions of Transmission by “Public Contest” Method. This would be the primary innovation in the U.S. Proposed transmission investments would need to pass both the Golden Rule, cost-benefit test, and the 30%- 30% voting rule. The rules would award either long-term FTRs to beneficiaries along with costs.

Identifying the beneficiaries and calculating the expected benefits would not be easy or uncontroversial. However, any cost benefit analysis of a transmission investment inherently involves simulations of hypothetical conditions with and without the investment. The simulations must address the impacts of transmission, and by the very nature of transmission the benefits will be different by location. Thus the cost-benefit analysis provides the basic elements needed to identify the beneficiaries and the distribution of the benefits. The change in benefits would include production costs, congestion and value of demand. The distribution of benefits could accrue to both suppliers and loads.

The particular 30%-30% voting rule is a compromise to capture the wisdom of super majority decisions. It would be hard to maintain that a major investment was both economic and necessary if not even 30% of the beneficiaries would support it, or more than 30% of the purported beneficiaries would oppose it. Without such a super-majority rule, the method reverts to the traditional regulatory model where central planners decide and regulators mandate.

Use of the public contest method in the U.S. would be a significant shift in perspective and would require some undoing of rules and tariffs requiring ISOs and regulators to mandate transmission investment based on considerations of “reliability” or “economics”. However, it would have the critical advantage of providing a principled definition for regulators to apply in drawing the line between market-based investments and regulatory mandates.

Other Examples of Little “r” Regulation

The case of demand response and pricing operating reserves illustrates a market defect that has a market design solution compatible with the general framework. The case of transmission investment illustrates a market failure where a regulatory mandate may be required, forcing free-riders to pay, but where the mandate can be limited to avoid undermining market-based investments. These are important examples in their own right, and they provide concrete illustrations of the distinction between little “r” and big “R” regulation.

There are other examples where the same principles have played out or could be applied. For instance, the potential exercise of market power is a market failure that calls for some type of mitigating regulatory rule. While there have been many proposals, including preventing generators who have such market power from participating in the market by keeping them under cost-of-service regulation, it has never been clear how such big “R” proposals would work in the context of a market system. A little “r” innovation that has addressed the problem was the institution of so-called “bid-caps” for generators that are in a position of exercising market power. Widely used, bid-caps are quite distinct from price caps or cost of service regulation. The essential idea is that generator bids in the market are constrained to be at or close to an estimate of their marginal cost. The generators participate in the market and under the general framework, are paid the market-clearing price, which may be greater than their bid cap. Whenever they are dispatched, this price would be at or above their bid and the payment would not be constrained by the bid. The bid-cap mitigates market power exercised by economic withholding, but does not remove the generator from the market. The bid-cap regulatory solution for mitigating market power is fully compatible with the general framework.

Another example of a little “r” intervention comes from a retail market mechanism for making wholesale market purchases for default customers. Even where retail choice is allowed and encouraged by state regulation, there will be a fairly sizeable category of customers, typically smaller consumers (e.g., residential and small commercial) who, given the option of “not choosing” an alternative supplier, would elect to remain with the incumbent utility or simply not express any preference for change. There may be various reasons for their choice to remain with the incumbent utility, including the absence of sufficiently attractive alternatives or prices, the fact that their usage (and bills) may be too small to make it worthwhile to shop for alternatives (or for competitive retailers to compete for their business), or simply satisfaction with current service. Yet these “default supply” customers may need special consideration, especially where retail competition is not sufficiently developed to give these customers meaningful alternatives.

The concern is not whether the default customer would be reliably served; they will be served in any event, just as they are now. The issue is: at what price? In principle, it would always be possible to ensure that all customers were reliably served simply by maintaining their connection to the grid and charging them the spot price of energy. However, the volatility of spot prices would be problematic for many customers. A market answer would call for the emergence of hedging contracts between final customers and suppliers. A hedging contract could be as simple as an agreement for the supplier to accept all the risk of the price volatility, with the customer paying a mutually agreed fixed price for the term of the contract. Given a well-designed wholesale market such as in PJM, electricity would be available and the hedging contract would be a simple financial arrangement. The concern is that hedging contracts may not be available for these smaller customers, the customers may not be equipped to handle the process, or the transaction costs of arranging the contracts may be too large. Hence, the problem in presenting customers with the opportunity to choose is not that those who do not choose will not receive power, but that they may face price risks that they can’t easily hedge.

A workable solution adopted in many places that provides such a hedge is modeled after the Basic Generation Service in New Jersey. The basic idea is to have a rolling forward contract to meet the full wholesale market requirements of customers in the default category. The auction divides the customers into tranches of load, and therefore allows many suppliers to participate. The focus on full requirements means that the auction (and in turn the regulator) do not need to handle the details of the wholesale market and all that would be involved in portfolio decisions and cost allocation. All those details are handled by the participating suppliers operating in the wholesale market. The details of the New Jersey auction include many important features to improve the efficiency of the auction, and the experience is that the auction mechanism has worked well at providing the intended forward contracts to hedge volatile spot market prices on behalf of retail customers. The contract price is determined competitively in the auction.

The design meets the needs in New Jersey and provides a compelling example of how to craft an effective little "r" regulatory mechanism that is compatible with the general market framework.

A further example would apply to the analysis of unit commitment decisions and pricing such as needed in day-ahead markets. For both reasons of efficiency and reliability, day-ahead markets may involve unit commitment decisions that cover multipart bids with minimum operating levels. An inherent characteristic of these models is that there is no set of energy prices that is capable of supporting the optimal unit commitment and dispatch. This can lead to revenue deficiencies and sometimes perverse incentives that could undermine the spot market. A common approach to such unit commitment designs in day-ahead markets is to pay generators for revenue shortfalls and collect the costs from loads in the form of an uplift payment. The little "r" challenge is to find a pricing mechanism that works best in this environment and provides workable incentives for market participants. One approach would be to determine energy prices compatible with the general market design that minimize the required uplift.[28] Various approximations of this idea have been implemented in the organized markets. The common theme of these implementations has been to find the best among the workable approximations that supports and is compatible with the overall market design.

Conclusion

Support for competition in wholesale markets is a clear and continuing national policy. Market defects and market failures both lead to a need for regulatory mandates to complement market operations. In a hybrid system Big "R" regulatory solutions often call for mandates and subsidies for favored programs. Little "r" regulatory solutions would emphasize reforms of market design to improve incentives and provide other initiatives to support rather than replace market choices. A general framework is available both for designing organized markets and testing the impact of regulatory mandates. In pursuit of the responsibility to integrate regulation and markets, regulators should apply the general framework to balance the needs of the hybrid market. Absent little "r" actions to correct market defects or address market failures, the pressure will continue for more and bigger big "R" interventions that undermine the larger enterprise.

---

[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. Susan L. Pope is a Principal of LECG, LLC. Part of the support for this paper was provided by PSEG Companies. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. The authors are or have been a consultants 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

---

[28] William W. Hogan and Brendan J. Ring, "On Minimum-Uplift Pricing For Electricity Markets," Center for Business and Government, Harvard University, March 19, 2003, (www.whogan.com).

Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, Entergy, 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 and IESO, Pepco, Pinpoint Power, PJM Office of Interconnection, PPL Corporation, Public Service Electric & Gas Company, PSEG Companies, 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 authors. (Related papers can be found on the web at www.whogan.com).