Electricity Market Design and Efficient Pricing: Applications for New England and Beyond

William W. Hogan[i]

June 24, 2014

Electricity markets support open access and non-discrimination to allow competition, entry, and innovation. Investment and operation in the competitive sectors follow the incentives induced by prices. To achieve the intended outcomes of reliability and economic efficiency, it is important to have efficient prices that are consistent with the objectives and operation of the underlying system. The basic design of successful organized electricity markets, built on the principles of bid-based, security-constrained economic dispatch, goes a long way towards meeting this objective. However, the real electricity system involves features that are difficult or impossible to fully reconcile within this core model. This calls for an application of the principles of dispatched-based pricing to move as far as possible to achieving the ideal of efficient pricing and minimizing the need for additional payments through uplift and other interventions to maintain reliability. The challenge is constantly present to match the prices to reflect the actual changing conditions of the dispatch. Motivated by issues under review in New England, a summary of the basic principles and illustrative applications provides examples of seeking the first-best efficient prices to mitigate the unintended consequences of second-best out-of-market payments.

Introduction

An important objective of electricity market design is to provide efficient prices with the associated incentives for operation and investment. In the idealized theory, energy and related reserve scarcity prices would provide all that would be needed to support a market and capture the benefits of competition. In the less than perfect reality, there are many complications in achieving the theoretical ideal. The result is a resort to out-of-market (OOM) interventions and payments that create incentive problems and compromise many of the benefits of efficient markets. Although it is not possible to fully achieve the theoretical ideal, it is possible to achieve a good approximation through application of the basic principles of efficient prices.

The essential ideas apply across all the organized electricity markets operating under the rubric of bid-based, security-constrained economic dispatch—the successful market design. (IEA, 2007) Improvement of the implementation of this model is a subject of constant review in all the organized markets. For example, in response to a number of problems seen in 2012-2013, the New England Independent System Operator (ISONE) undertook a substantial effort to address real-time pricing and related issues as part of an ongoing review of market design. (Coutu & White, 2014a) A persistent problem, in New England and elsewhere, is the apparent assumption that important interventions in the market by system operators are necessary for successful operation of the system but are too hard to incorporate in the associated pricing model. The "too hard" part might be technical or political. It always looks easier to handle the interventions outside the market, and spread the average costs as an uplift allocation. But in markets where participants have discretion to follow the incentives, this superficially easy fix creates perverse incentives that cause larger unintended consequences. While some pricing problems may have no technical solution, the technical problems can be less of a barrier when viewed from the perspective of dispatch-based pricing. The political problems can be less serious once we see the consequences of suppressing reality. Raising prices or increasing volatility always creates pressure on market regulators. But when prices are too low or not volatile enough, there is too little incentive for load management, storage technologies, fast response generation, dual fuel capability, firm fuel supply, and so on. The special interventions to treat the symptoms create new problems and make the market oversight problem more difficult. In the end it is easier and better to emphasize the priority of efficient pricing as the first choice, not the last resort. The principles and practices of dispatch-based pricing illustrate how to adopt and implement this perspective. The goal in the present paper is to contribute to this process and illustrate how getting the prices right can both improve the incentives and market signals, but also simplify some of the more difficult problems that arise in the continuing market review. The ongoing review of market design and pricing in New England motivates much of the discussion and recommendations, but similar issues arise in all organized electricity markets. For example, on June 19, 2014, the Federal Energy Regulatory Commission launched just such an inquiry "to improve the energy and ancillary services price formation process." (Federal Energy Regulatory Commission, 2014)

Electricity Market Challenges

New England operates forward capacity markets plus day-ahead and real-time markets for energy and ancillary services. Success of these markets requires design of components that mesh. In a market with many participants exercising discretion in their energy supply and demand choices, a critical design principle is to *begin* with an efficient and transparent real-time energy market that operates under the principles of security-constrained economic dispatch. The design of forward markets for energy and ancillary services should follow, to be consistent with the principles applied in the real-time market which will be anticipated in forming bids and offers in the forward markets. Similarly, long-term capacity markets should reflect the future operation of forward energy and ancillary service markets. In all cases, a key design principle is to establish prices and pricing rules that are consistent with and reinforce the incentives for efficient operation and investment.

The actual experience in these markets reveals a number of problems that have been a source of concern for many years. For instance, in its review of the accumulated experience, the ISONE Market Monitor summarized the concerns.

> "The goal of the real-time market is to efficiently procure the resources required to meet the reliability needs of the system. To the extent that reliability needs are not fully satisfied by the market, the ISO must procure needed resources outside of the market process. This tends to distort the real-time prices and may indicate that there are reliability needs that are not fully priced. Both of these issues are significant because they undermine the efficiency of the real-time price signals. Efficient real-time price signals are essential because they encourage competitive conduct by suppliers, efficient participation by demand response, and investment in new resources or transmission where it is needed most. Hence, it is beneficial

> to regularly evaluate whether the market produces efficient real-time price signals. … [W]e evaluate several aspects of the market operations related to pricing and dispatch in the real-time market in 2012:
>
> - Prices during the deployment of fast-start generators;
> - Prices during shortages of operating reserves;
> - Prices during the activation of real-time demand response;
> - Efficiency of real-time ex post prices; and
> - Frequency of price corrections." (p.64) (Potomac Economics, 2013)

The collective results of a number of these market aspects tend to suppress prices. Fast start generators are partially accounted for, but "real-time LMPs frequently do not reflect the full cost of deploying the fast-start generator, even if the generator is still economic to be online." "Furthermore, since the minimum output level of most fast-start generators is within 90 percent of their maximum output level, fast-start generators are frequently dispatched at their minimum output levels where they do not set price during the second phase of commitment. In such cases, the resulting LMP may be lower than the incremental offer of the fast-start generator. (p. 66) (Potomac Economics, 2013)

For the treatment of reserves, "[i]n the real-time market, the Reserve Constraint Penalty Factors ("RCPFs") limit the costs that the model may incur to meet the reserve requirements (i.e., marginal dispatch actions that would exceed the relevant RCPF are foregone). Consequently, if the cost of maintaining the required level of a particular reserve exceeds the applicable RCPF, the real-time market model will allow a reserve shortage and set the reserve clearing price based on the level of the RCPF." (p. 73) (Potomac Economics, 2013)

The use of demand response also can have an unintended impact on real-time prices. "The activation of demand response in real time can inefficiently depress real-time prices substantially below the marginal cost of the foregone consumption by the demand response resources, particularly during shortages or near-shortage conditions. Although there is little information available on the marginal cost of foregone consumption for demand response resources, the marginal costs of most demand response resources are likely to be much higher than the marginal costs of most generators. Hence, real-time prices should be very high when demand response resources are activated." (p. 79) (Potomac Economics, 2013)

The incentive problems for real-time pricing in ISONE have been compounded by scheduling and operating problems during severe cold weather which result in part from depressed real-time prices and the inability to schedule and pay for gas supplies at the short-term constrained prices. This has produced a wide array of programs that treat many of the costs in uplift rather than fixing the underlying real-time prices that should rise to incent all market resources to be available, including those that need to schedule gas. (ISONE, 2014c) "…real-time prices often do not fully reflect the cost of satisfying demand and maintaining reliability during tight market conditions, particularly when fast-start resources or demand response resources are deployed in the real-time market." (p. iii) (Potomac Economics, 2013)

The fact that many of these problems have been recognized for many years points to the difficulty of finding simple treatments for the symptoms without addressing cures for the underlying problem of not getting the prices right. Failure to address the fundamentals has the almost inevitable effect of compromising the benefits of efficient markets and creating new challenges. For example, efforts to increase investment by developing capacity markets, rather than fixing the pricing in real-time, raise costs for loads but the costs are socialized and do not provide effective incentives for either demand participation or reliable operation. This leads to the need for supplemental real-time performance incentives, that may help with generation operations but still do not address the opportunities for demand participation. The net effect is move more and more towards administrative prescription and higher average costs, recreating the problems that were intended to be addressed by electricity restructuring. These indirect attempts to create the effects of efficient pricing, without the efficient prices, confront the reality that we do not know how to design regulations for efficient outcomes when the pricing incentives motivate inefficient behavior. If we did know how to accomplish this administrative feat, there would be no need for electricity markets.

Energy Pricing and Revenues

Organized markets operating under the only successful market design have not been completely successful. The markets share a number of problems. Perhaps the best known and most important is the so-called "missing money" problem when energy and ancillary service payments are not sufficient to cover the going forward cost of new generation investment. For a variety of reasons, pricing rules and operating practices of different types combine on average to suppress energy prices to the point that the energy prices are well below the level needed to support new generation investment. (Joskow, 2008)

One approach to addressing this problem is through the creation of forward capacity markets. In the case of New England, the market result for 2013 found energy revenues of $58.14/MWh. In addition, capacity payments were $8.20/MWh, or 12% of the total. (p. 2) (ISONE, 2014a) There is some dispute about whether this capacity value is enough to support adequate investment and the ISONE continues to review investment choices relating to generation, fuel diversity and so on.

The use of the capacity market creates important incentive problems and challenges. For instance, ISONE has recognized that a capacity obligation from the capacity market is not sufficient to ensure the capacity is available in real time under shortage conditions. This created an initiative for changing and improving the real-time incentives for capacity resources. The resulting controversial pay-for-performance program has as its essence that some generators deficient in meeting capacity obligations pay other generators who meet or exceed their commitments. (ISONE, 2014b) (Potomac Economics, 2014) However, the net payments are zero in aggregate and do not provide the same real-time incentives for other participants in the market, such as price responsive load, as would arise from efficient scarcity pricing.

Similar problems arise in reliability related OOM transactions. The excess costs of these transactions are added to uplift payments which are allocated across market participants. In the case of ISONE, in 2013 uplift payments amounted to $158 million. "Approximately 70% of all reliability payments in 2013 were made in January, February, July, and December—months that had unusual operating conditions resulting in tight or uncertain system conditions and causing the commitment of additional resources out of merit order." (p. 3) (ISONE, 2014a) Although much smaller than capacity payments, the uplift charges are important because of the impacts on cost allocation and real-time energy prices. The 2013 uplift charges were lower than they were before 2009 when transmission investments reduced transmission congestion and the associated OOM actions that had been more common in the presence of greater transmission congestion. (p. iii) (Potomac Economics, 2013) In effect, this moved the OOM payments out of the uplift and into transmission cost allocations. But in both cases, the result was to raise costs overall while adding to the pressures that depress market-clearing energy prices.

Related proposals would work through capacity market performance payments to alter real-time incentives. (Potomac Economics, 2014) However, everything channeled through the capacity market is indirect and convoluted. The process almost seems driven by a commitment not to fix the actual energy markets prices but rather to find ever new and ever more indirect pathways to reproduce the results of an efficient real-time market without actually implementing an efficient real-time market.

The goal "of the real-time market … to efficiently procure the resources required to meet the reliability needs of the system," is important. Although it may not be possible to eliminate capacity payments, OOM costs and uplift, a companion goal should be to reduce the importance of these interventions and improve the performance of the basic energy market. (Harvey, 2014) And the first-best way to address the deficiencies of the energy market is by addressing real-time pricing. Only after exhausting the consistent improvements in real-time pricing should we resort to the second-best OOM interventions. Achieving the first-best improvements relies on application of the principles of dispatch-based pricing.

Dispatch-Based Pricing

The purpose of ex post or dispatch-based pricing is to determine "prices consistent with the actual usage by applying the marginal tests of economic dispatch." (Hogan, 1992) The essential idea is to let the dispatch determine the prices rather than to change the dispatch to meet some model that might differ from the operational practices of the system operator. By making prices consistent with the dispatch, we create incentives that support rather than oppose the actions of the system operators. (An outline of formulations for different pricing problems appears in the Appendix.)

A starting point relies on a characterization of an underlying security-constrained economic dispatch problem. (Hogan, 2013b). Given the optimality conditions, we can apply the usual analysis to derive the general form of the locational marginal prices (LMP) as in:

$$p = \text{System Energy Cost} + \text{Locational Marginal Loss} + \text{Locational Congestion}\ .$$

The LMP has the interpretation as the effect of the system reference energy price, marginal losses, and congestion.

In this representation, for an interior solution, the LMP is also equal to the marginal benefit of load or the marginal cost of generation at the location.

p = Marginal Benefit of Load = Marginal Cost of Generation.

The details of implementation may affect the characterization of the benefit and cost functions, such as to include capacity constraints, but the basic insight of the LMP formulation will continue to apply. Efficient prices are defined and determined consistent with the economic dispatch formulation. Several types of applications, from ex post LMP to approximating the price impacts of voltage support reliability commitments, illustrate the approach of dispatch-based pricing.

Ex Post LMP

Solving the full economic dispatch problem is complicated, and system operators have long experience in this matter. Trying to improve the quality of the dispatch may be a worthwhile effort. But the intent of dispatch-based pricing is to solve the simpler problem of finding the prices that would apply if we assume we come as close as possible with the actual dispatch as a solution to this optimal dispatch problem. Ex post LMP illustrates dispatch-based pricing because it is a technique to approximate prices that are close as possible to being those that would arise from efficient dispatch, and thus consistent with efficient dispatch.

The earliest implementation of the approach was to extract the implied LMP values from the solution to the linear approximation based on the actual dispatch with binding transmission constraints. The set of binding constraints is not known before the dispatch is determined. But once the dispatch is known, the much smaller set of known limiting constraints produces a simpler and smaller linear approximation of the security-constrained economic dispatch. This reduced model would be inappropriate for choosing the dispatch, but it is ideally suited for calculating the LMP values that would be consistent with the actual dispatch.

Given the dispatch, the information needed to formulate this problem is both relatively simple and readily available. The critical elements would be the "shift factors" that define the marginal flow over transmission lines from an injection and withdrawal of energy at different locations. These are the derivatives of the binding constraints, a small subset of the full list of possible constraints. The actual dispatch is a solution to this model, and the associated prices are the dispatch-based LMPs.

The critical feature of an application of ex post, dispatch-based pricing is to focus on the accuracy of the shift factors, which directly affect the determination of the prices. The ex post approach allows flexibility in the accepting the estimated transmission flows as the best estimate of the limits on the binding transmission constraints. That the actual power flows may differ by some small amount from the announced transmission limits is accepted as part of the approximation. The system operator need only establish the set of limiting constraints without specifying separately the constraints limits, which do not affect the LMPs, and this simplifies finding a solution of the pricing model.

ISONE uses a variant of ex post LMP determination. However, the description of the constraints apparently is not consistent with the actual dispatch. "Large differences between the two prices typically occur when a reserve constraint is binding, creating inaccurate LMP calculator prices during difficult operating conditions when accurate prices are essential to maintaining reliability. [Unit Dispatch System] prices have become reliable enough that the LMP calculator is no longer needed." (p. 15) (ISONE, 2014a) The principle is not new. It is important, however, to make the approximation reflect the real dispatch.

The treatment of reserves and other reliability requirements is an important part of dispatch and pricing. Dispatch-based pricing can integrate these requirements to provide a better approximation of prices that reflect the dispatch.

Operating Reserve Demand Curve

The use of operating reserve demand curves for improving dispatch-based energy prices is an example of an approach to improving the efficiency of prices within the framework of economic dispatch. (Hogan, 2005) This is important for at least three reasons. First, it provides another example of using first principles to derive better approximations of efficient dispatch-based prices. Second, addressing scarcity in efficient prices is widely recognized as a critical challenge in electricity markets. Third, the resulting scarcity pricing mechanism provides a tool for addressing related efficient pricing problems such as for contracted demand response or reliability unit commitments.

The discussion here extracts from the more extensive development in (Hogan, 2013a), which provides further details. Operating reserves for spinning and quick start capacity are a regular feature of all electricity markets. These reserves are distinct from the installed capacity that is the focus of forward capacity markets. Operating reserves are a subset of the installed capacity that is both available and standing by to produce energy on short notice. In any given real-time dispatch interval, reserves are maintained to deal with unpredictable events such as a sudden surge in demand, loss of a generator, or loss of a transmission line. This balancing generation needs to ramp up very rapidly to meet the immediate emergency and to give the system operator time to reconfigure the energy dispatch.

Although it is difficult to forecast requirements for installed capacity many years ahead, it is a comparatively easier and more familiar task to forecast operating reserve requirements and availability for the next instant or parts of an hour. Supply, demand and transmission conditions are known. Weather forecasts are on hand. System operators have experience and procedures for defining and evaluating standby generation capabilities.

The most immediate requirement is for the operating reserves needed to meet security contingency conditions. The flows of electricity respond much faster than operators to sudden events such as the loss of a generator. In order to avoid cascading failures that could blackout most or all of the system, operators must maintain a minimum level of contingency reserves. From an economic perspective, a way to interpret and define the minimum magnitude of the required contingency reserves would be as the level at which the system operator would impose involuntary controlled curtailment on selected loads in anticipation of the possible contingency, in order to maintain minimum adequate capacity that could provide additional energy but must be kept in reserve.

System reliability would be improved if more operating reserves than the minimum were available in terms of response to increase generation or quickly decrease load. Over the next few minutes or parts of an hour, events may arise that deplete operating reserves and bring the system below the minimum contingency requirement, in which case the operator will have to impose involuntary load curtailments to restore the minimum contingency protection.

The importance of operating reserves has always been known, but the requirements for operating reserves were given only a simplified consideration in early wholesale electricity market design. The assumption was that the operating reserve requirement at any moment and location could be represented by a fixed requirement, and that economic dispatch would produce simultaneous optimization that would incorporate the dispatch of energy and reserves. Pricing, especially during shortage conditions, would be provided by demand bidding to voluntarily reduce load at high prices, and the value of operating reserves would be determined by the implied scarcity prices. While this was a workable approximation in theory, it failed in practice when the associated demand bidding did not materialize.

One solution to this problem is to revisit the pricing of operating reserves through a better representation of an operating reserve demand curve (ORDC). To be sure, an operating reserve demand curve is an administrative intervention in the market. But this is already true of the administrative requirement for operating reserves. In the presence of a necessary and inevitable operating reserve requirement, it is clear that the superior administrative rule would be a better model of the demand for operating reserves that goes beyond the fixed quantity requirement. (Hogan, 2005)

Several ISOs have implemented variants of an ORDC, but without the connection to the underlying scarcity principles. The RCPF approach in ISONE already provides a price for operating reserves. This is a scarcity pricing mechanism. However, the level of the payments and conditions under which the payments are made are not derived from the underlying principles of economic dispatch or an explicit model for the reliability requirement. As a result, the prices are too low in at least two ways. The RCPF is not high enough to account for real scarcity costs just when the system is most constrained and the efficient price would be most important. The highest reserve penalty factor cap is only $850/MWh. In addition, in its early implementation the RCPF was not applied during more than 95% of the hours when the value of additional reserves should be lower but not zero. (p. 67) (ISONE, 2014a) The subsequent reforms to increase the penalty payments doubled the number of hours and the value paid for operating reserves. (ISONE Internal Market Monitor, 2014) But the focus remains on a limited number shortage hours that leaves most hours with no scarcity prices. By contrast, small scarcity payments applied across many hours could make a material contribution to reducing the missing money.

A Structure of Scarcity Pricing Through Operating Reserves

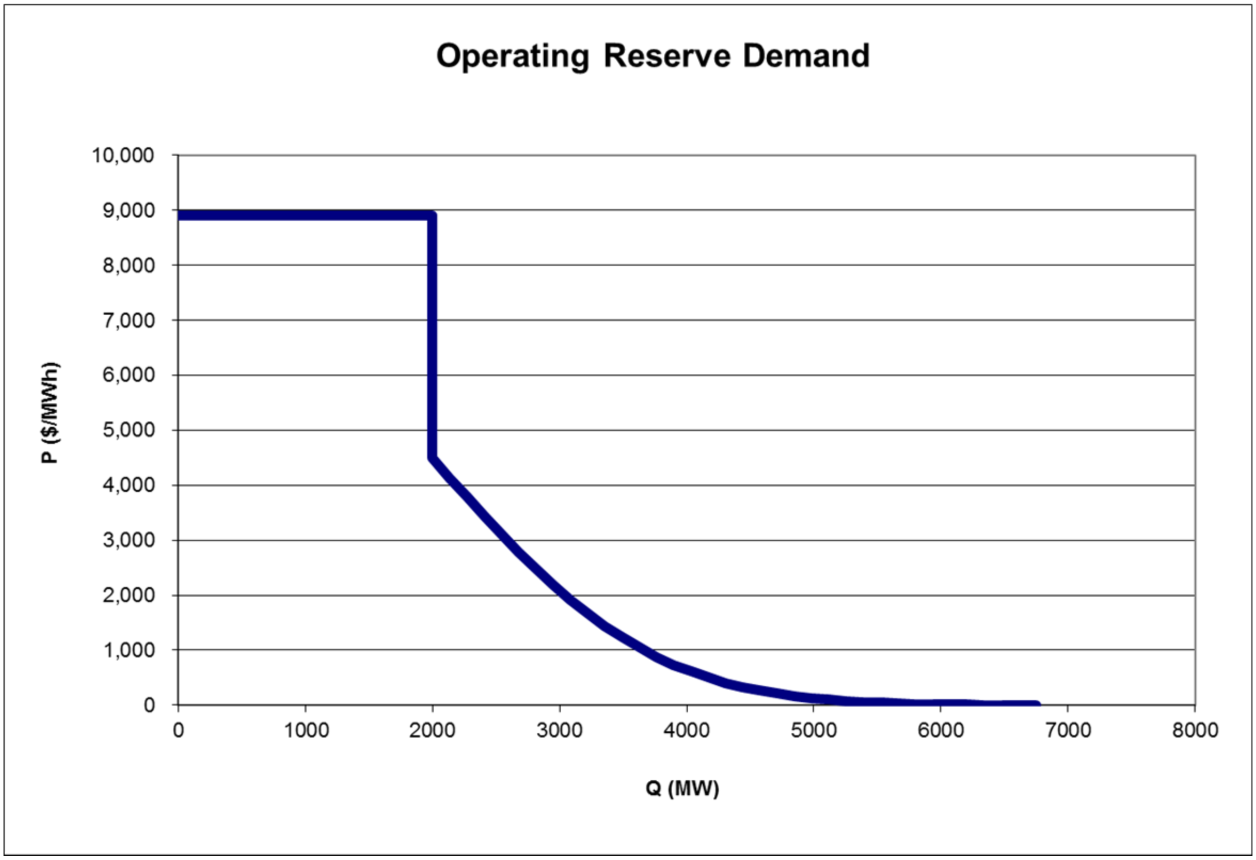
The basic outline of an operating reserve demand curve with efficient prices follows from the description above. Although the ORDC would be integrated with the dispatch model, rather than applied solely in a pricing model, it is in the spirit of dispatch-based pricing in using a simplified scarcity pricing model to approximate the more complicated set of security decisions found in the real dispatch. The key connection is with the value of lost load (VOLL) and the probability that the load will be curtailed. (Potomac Economics, 2014) Whenever there is involuntary load curtailment and the system has just the minimum of contingency operating reserves, then any increment of reserves would correspondingly reduce the load curtailment. Hence the price of operating reserves should be set at the value of lost load during these periods.

At any other level of operating reserves, set to protect the system from events in the immediate future, the value of an increment of operating reserves would be the same VOLL multiplied by the probability that net load would increase enough in the coming interval to reduce reserves to the minimum level where load would be curtailed to restore contingency reserves. Hence the incremental value of operating reserves would be analogous to the product of the loss of load probability (LOLP) and VOLL, or LOLP*VOLL. The clearest early example of the application of this logic is from the implementation by the Midcontinent Independent System Operator (MISO).

> "For each cleared Operating Reserve level less than the Market-Wide Operating Reserve Requirement, the Market-Wide Operating Reserve Demand Curve price shall be equal to the product of (i) the Value of Lost Load ("VOLL") and (ii) the estimated conditional probability of a loss of load given that a single forced Resource outage of 100 MW or greater will occur at the cleared Market-Wide Operating Reserve level for which the price is being determined. … The VOLL shall be equal to $3,500 per MWh." (MISO, 2009)

The ORDC is similar to the operating reserve penalty factors utilized in ISONE. However, the penalty factors in ISONE apply only when there is an operating reserve "shortage," which occurs in relatively few operating hours. By contrast, the ORDC and associated LOLP calculation should apply in all hours. Although many of the scarcity payments may be relatively small, the cumulative effect can have a material effect on the missing money problem. (ERCOT Staff & Hogan, 2013)

Combining this with the treatment of minimum contingency reserves, the resulting operating reserve demand curves would look like the hypothetical illustration in the accompanying figure. Here the assumed VOLL is $9,000/MWh and the minimum contingency reserve requirement is assumed to be 2000 MW, as illustrated in the initial ERCOT implementation. (ERCOT, 2014) The discontinuity at 2000 MW occurs because of the probability that load will reduce over the interval more than the expected generation losses, in which case there is no need for load curtailment. Above the minimum reserve level, the shape of the demand curve follows the LOLP distribution. Importantly, a general property of an operating reserve demand curve derived from first principles is that the demand is not vertical and price does not drop to zero. Scarcity pricing would arise to some degree for all hours.



The same principles could be generalized to include zonal requirements for operating reserves that interact with energy dispatch, incorporate local interface constraints, and provide compatible short-term prices for operating reserves and interface capacity. (Hogan, 2010) With simultaneous optimization in the economic dispatch, the scarcity prices attributed to operating reserves apply as well to energy whenever there is a tradeoff between energy dispatch and operating reserve capacity. Hence, the scarcity prices would contribute to resolving the missing money problem for all generators actually providing energy or reserves. (Hogan, 2013a)

The formulation of the dispatch model with an ORDC, and the dispatch-based pricing, utilize the co-optimization of reserves and energy. This would be true in real-time and would then create a consistent representation of scarcity conditions and prices. Similar comments would apply to a consistent application of the principles in the day-ahead market, which would use an ORDC and co-optimization to reflect consistent scarcity prices. (Hogan, 2013a)

Market Power Mitigation

The many features of an ORDC include an important and sometimes neglected connection with market power mitigation. One of the principal challenges in monitoring electricity markets has been to distinguish between (good) high prices caused by scarcity and (bad) high prices caused by the exercise of market power.

One of the principal tools for mitigating real-time market power is to combine offer requirements with offer caps. The cap sets a limit on generator offers to be close to the variable operating cost. A persistent difficulty is that the offer caps, while effective in mitigating market power, also have the unfortunate effect of preventing scarcity pricing from entering the market in the form of higher generator offers. Given the pricing rules without adequate accounting for scarcity, competitive generation offers could be above estimated variable costs. However, despite the defects in the pricing rules, there is a presumption that any generation offers above a reasonable estimate of variable cost must be an attempt to withhold supply and exercise market power. There is no easy means to separate the good from the bad high price outcomes.

The ORDC formulation provides a simple escape from this important dilemma. Scarcity of operating reserves, and hence a co-optimized energy price increase, enters the market through the ORDC and clearly distinguishes between scarcity and market power. With an ORDC, it is no longer necessary to have high generation offers that look like economic withholding, in order to achieve high market clearing prices. Generators will have an incentive to make offers that approximate their true opportunity cost, and market mitigation rules based on offer caps would be less problematic. A generator with an offer of $75/MWh could be dispatched at capacity and not face the likelihood that its offer would set the market prices during a condition of scarcity. In the presence of even a modest amount of scarcity, the generator would see and be paid the probably much higher ORDC-induced price at the VOLL*LOLP. During times of system stress, generators could be paid thousands of dollars per MWh even though the generation offer was still $75/MWh. And the cause of scarcity and high prices would be easy to identify as being market conditions and not generator offers. Everyone would benefit from this analytical clarity and from the contribution towards making up the missing money.

Augmented ORDC

An ORDC addresses many problems in providing better incentives in the real-time energy market. In addition, the ORDC contributes to reducing the missing money problem and increases the attractions of investing in generation. However, there is no reason in principle why the basic ORDC by itself would provide enough of the missing money to be sufficient to induce investment sufficient to attain a planning reserve margin that would be consistent with a typical one-event-in-ten-year reliability standard. (Newell, Spees, Pfeifenberger, & Karkatsouli, 2014)

A reasonable implication of this disconnect would be to ask whether the cost of the reliability standard is worth the benefit. This could result in policy decisions related to the level of the reliability standard and the associated reserve margins.

Whatever the choice on the reliability standard, a separate question arises as to the best way to provide the required missing money. A common assumption is that the only practical alternative is to establish a forward capacity market. As we know well from experience, designing and operating a capacity market presents its own set of challenges and unintended outcomes. A basic ORDC helps mitigate many of the related problems of a capacity market, and makes the capacity market less important. An augmented ORDC could be a complement or an alternative to a capacity market that could operate through the energy market to provide a further contribution towards the missing money.

The idea of an augmented ORDC is simply to impose conservative assumptions on the basic ORDC model. The intent would be to provide both a greater level of reliability and an associated increase in total operating reserve and energy payments to address the missing money problem.

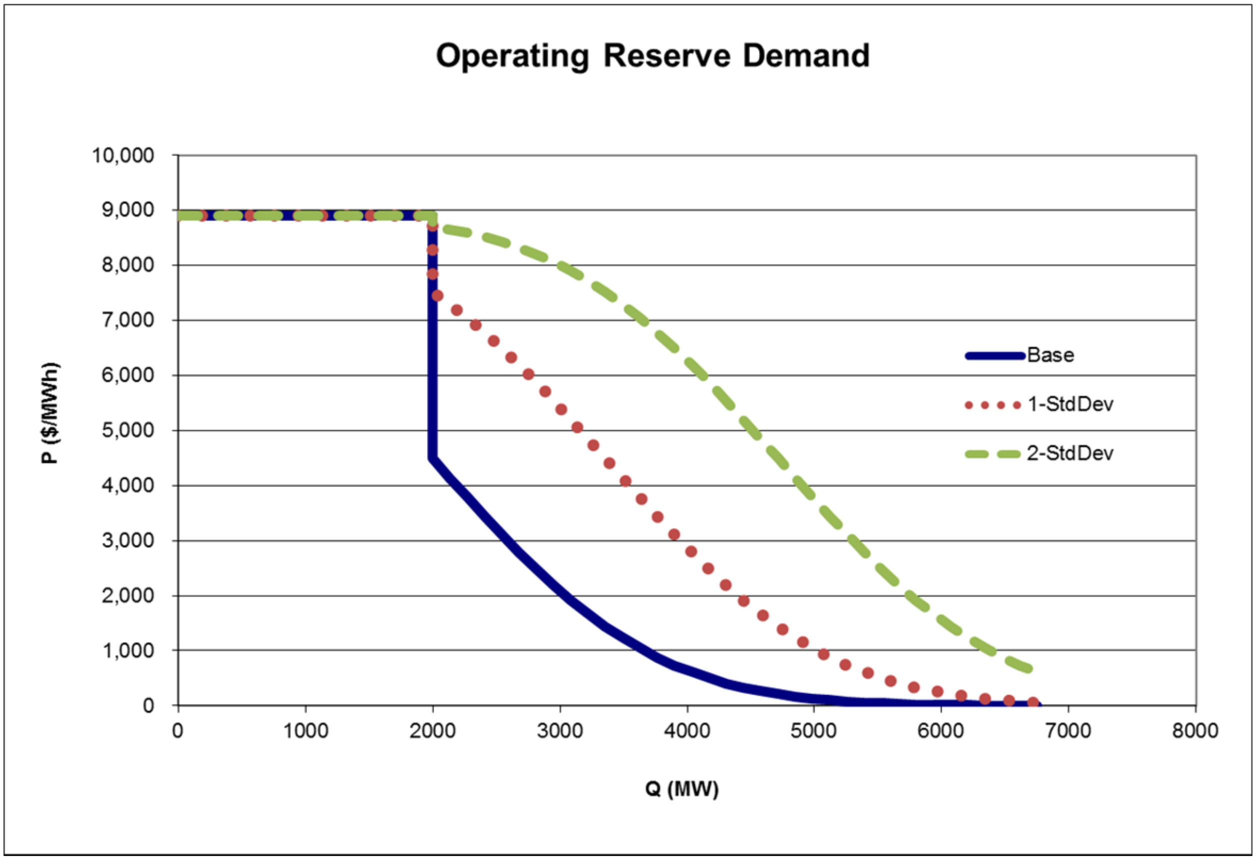
For purposes of this discussion consider the three parameters of the basic ORDC: VOLL, Minimum Contingency Reserves (X), and loss of load probability (LOLP). The VOLL should approximate the average cost of involuntary load curtailments or other non-market actions that would actually be undertaken by the system operator when reserves become scare. The "X" factor should approximate the level of reserves at which the system operator would begin to take these non-market actions in order to restore reserves to an acceptable minimum contingency level. And the LOLP should be the short-run probability (e.g., over the next hour) that unanticipated increases in load or decreases in available generation would reduce operating reserves below the X level. Increases in any of these three parameters would have the effect of increasing operating reserve prices on average and contributing more of the missing money in the reserve and energy markets.

The VOLL price applies when conditions require involuntary load curtailment. It is important that this price be paid to generation and charged to remaining load. Hence, an upper bound on a conservative VOLL would be the maximum price we were willing to charge in the face of load curtailment. It may be better to err in the direction of a higher VOLL, but this may not be enough to address the reliability goal and provide the missing money. Too low a VOLL would create a situation where its application would have the average curtailed load willing to pay more to avoid curtailment. Too high a VOLL would be less problematic because it would eventually stimulate demand response that would avoid the need for the involuntary load curtailment. However, the VOLL price should not be set so high that it becomes not credible that regulators would allow the price actually to be charged or paid, in which case a VOLL at too high a level would undermine the intended incentives and there could be no real contribution to better generation investment.

Similar arguments apply to the minimum contingency level X where the system operator would initiate non-market interventions to raise the level of available operating reserves. The minimum contingency level is more directly connected to reliability. If the level is set too low, then actual interventions would start before reaching the high prices and the resulting impact would be to lower rather than raise market prices as we approach shortage conditions. If the minimum contingency threshold is set too high, then we would produce periods when high prices were being imposed but no non-market interventions were needed and the regulators would have to defend applying the VOLL when it was not required.

The short-term load and generation changes that give rise to the LOLP summarize a complex process. The models of LOLP employ certain assumptions about the accuracy of the system approximations and the ability to avoid problems like human error typically found in events that threaten the stability of the system. A conservative approach to reliability is already part of the motivation for the use of contingency constraints to define secure operations. However, it would be consistent to extend this reliability motivation to a conservative estimation of the LOLP. This would avoid the conflicts that arise with too high a VOLL or too high an X. In particular, if the conditions that define the missing money are motivated by reliability concerns, then the most obvious connection with the ORDC would be through the LOLP estimation.

There are many ways to augment the LOLP calculation, but the simplest approach would be to shift the probability curve by some fixed amount that represents the further margin of safety that motivates a high reliability standard.



The accompanying figure illustrates such an augmented LOLP and resulting augmented ORDC in an example from ERCOT. (ERCOT Staff & Hogan, 2013) Here the standard deviation used in the normal approximation for the LOLP is 1357 MW. The two sensitivity cases illustrate the impact of shifting the mean expected change in net load over the operating reserve period by (i) one and by (ii) two standard deviations. Clearly the resulting augmented ORDC could make a material difference in the implied scarcity prices for many hours, well before we get to the minimum contingency level.

The choice of how much of a margin of safety in the ORDC should be applied is a judgment call. However, expressing the margin of safety in terms of the size of the shift in the LOLP curve seems more intuitive and transparent than setting the planning reserve margin according to the one-event-in-ten-years rule that is not connected to the actual conditions in real-time operations. Simulations would be required to determine the final estimate of the incremental contribution to the missing money under an augmented ORDC. But if the goal is to provide greater reliability, rather than to meet a planning reserve margin per se, the approach of augmenting the ORDC has advantages over a capacity market and may provide a better answer to define how much is reliability is enough.

Demand Response

The use of Demand Response (DR) to provide operating reserves raises the question of how to interact with the ORDC and energy pricing in the dispatch. The focus here is on how to incorporate DR to determine the impact on prices in the electricity market. This is separate from the discussion of the appropriate payments for DR that has been the subject of controversy.

(United States Court of Appeals for the District of Columbia Circuit, 2014) In the case of DR that is offered on a continuing unconstrained basis as price responsive demand, or under the buy-sell model, the response can be treated in a straightforward way with bid prices. However, in many cases the DR program is implemented through a contract that includes features that constrain the cumulative use and therefore affect the opportunity cost from the perspective of the rest of the market.

The details of DR matter somewhat, but a common feature is that the use of DR over some period such as a season may be limited by a total megawatt hour constraint. (Yoshimura & Parent, 2014) The essential basic structure is close to the problems found in the use of an energy limited facility. The analysis of the familiar problem of an energy limited facility provides guidance for the treatment of DR with a cumulative megawatt hour constraint. For the present discussion, assume the direct variable cost of the energy is zero, much like a hydro facility with a limited reservoir of water.

Suppose that we have an energy limited generation facility that is capable of providing operating reserves and energy. It can provide operating reserves as long as there is still available energy. However, the cumulative amount of energy over the current and future periods is limited. Using the facility to provide operating reserve now does not affect the energy available in the future. But using the energy now does affect the energy and reserves that can be provided in the future. Hence, in the current period there is no opportunity cost for using the plant for reserves, but there is an opportunity cost for using the plant to provide energy.

Estimating the opportunity cost and solving for the optimal use of such a facility is possible and familiar from the study of constrained hydro systems. A full optimization would require a dynamic model that goes beyond the limits of the static or near-horizon representation in the usual economic dispatch model used in operations. The reduced summary of the full optimization appears in the "value of water" stored in a pond, which can then be applied in determining the economic dispatch.

However, in the mode of dispatch-based pricing, the problem allows for a simpler approximation. The purpose of dispatch-based pricing is to assign prices to actions the system operator takes as part of the dispatch. To the extent possible, the idea is to assume that the operator is making good decisions consistent with economic dispatch, and to determine the implied prices accordingly. The current operating protocols include operator decisions for deploying DR to provide energy. As long as these rules are in place, the task is to determine prices that are consistent with the operator's decisions.

Ignoring the lumpiness of the decision, for a continuously dispatchable energy limited facility we can compute an upper bound on the implied energy opportunity cost whenever the plant is dispatched to provide energy. If the operator dispatch decision is optimal, it must be that the opportunity cost of the energy is no more than the market clearing price of energy including the cost of scarcity. This market price must be less than the energy offer of the next unit in the dispatch stack plus the scarcity price determined from the ORDC, with the energy limited facility removed from reserves and providing energy. This implies that an upper bound on the implied energy offer for the DR must be the energy offer of the next unit in the stack.

In a dispatch-based framework, applying this energy offer for the energy limited unit ensures that the resulting price is consistent with the resulting energy dispatch and scarcity price. The resulting energy offer plus scarcity price will not produce a price reversal, when deploying the valuable reserve energy reduces the price. And the generator at the margin will be indifferent to providing reserves or energy.

The assumption is that the continuous approximation should be a workable solution for the lumpy DR dispatch decision. When DR is providing reserves, it should be included in the ORDC calculation as reserves. When DR is dispatched to provide "negawatt" energy, it should be removed from reserves and included in the generation stack at the implied energy offer of the first undispatched unit in the stack.

This rule should produce consistent market-clearing prices. Since DR is contracted for in advance, the associated reserve or energy payment would not go to the DR provider. Presumably the reserve payment would be rebated to entities paying for the DR contract.

This approach is quite general, and would apply whenever the operator is dispatching DR. The pricing rule is not affected by whether or not we are at, above or below the minimum contingency level. This dynamic pricing rule should be at least feasible, and perhaps easy to implement. The resulting LMPs will incorporate the scarcity value of the DR in the energy price, and use of DR will not produce a price reversal. The pricing model does not tell the operator when to use DR, but given that the dispatch calls on the DR the model determines the ex post prices consistent with the dispatch.

The rule could apply to all kinds of activities that are approximately like energy limited facilities.

Reliability Unit Commitment

In day-ahead markets the bid-in load that clears provides one estimate of the forthcoming demand in the real-time dispatch. As part of their responsibility for reliability monitoring, system operators regularly prepare an independent forecast of the prospective load. When the two forecasts differ, the system operator faces the possibility that there will not be enough generation capacity to meet the requirements for forecast load and the associated reserves. In ISONE this analysis is referred to as the Resource Adequacy Assessment (RAA). This conforms to a common practice, which is to undertake a related Reliability Unit Commitment (RUC) to make sure that there is sufficient capacity to meet the system operator's forecast load.

The effect of the incremental RUC units is to increase the available capacity relative to the bid-in and scheduled load. Absent some compensating adjustment, the effect is to lower the market-clearing prices in the day-ahead dispatch. This has the perverse effect of sending the market signal that there is less scarcity than implied by the actual reliability requirements adopted by the system operator.

One way to apply the approach of dispatch-based pricing for the day-ahead market would be to treat the RUC requirement as a need for increased reserves relative to the bid-in load indicated requirement. A direct approach would be to utilize an augmented ORDC to incorporate the reliability requirement of the RUC. The locational or regional ORDC calculated from the first principles would be augmented by shifting the probabilistic portion of the non-spin demand curve by the RUC amount. The increased capacity committed would be part of the available reserves, and the augmented ORDC would incorporate the value of the increased reserves.

In real-time, the uncertainty would be resolved to be higher or lower load than the system operator forecast. In any event, the RUC commitment would be sunk, and there would be no purpose in a separate treatment of the generation capacity committed by the RUC.

Voltage Support

Full description of all the many transmission constraints that limit an economic dispatch can be a daunting task. Often the dispatch models are functionally equivalent to DC load models that address real power only. This modeling approach assumes that reactive power is readily available, unconstrained, and unimportant in determining the dispatch. However, when voltage constraints are important and these simplifying assumptions do not apply, it may be necessary to restrict load or operate certain generators in order to avoid reactive power losses or ensure reactive power supply. (Hogan, 1993)

The dispatched generation or load constraints often are then treated as OOM transactions. The decisions are fixed in the dispatch model and excluded from the price calculations. This results in increased total capacity and lower prices in the associated dispatch.

Incorporating voltage constraints and reactive power flows in the full dispatch model would require a full AC, security-constrained, optimal power flow representation of the dispatch. This might be a good idea, someday, but it would be a great challenge with existing software for controlling economic dispatch. Real and reactive power flows and the impact on local voltage constraints are highly nonlinear, and thus difficult to incorporate in the formal dispatch model. The practice of reliability evaluations and dispatch of a generator for reactive power depends more on the discretion of the system operator.

However, the software limitation on economic dispatch does not apply to the application of the principles of dispatch-based pricing. A linear representation of the impact of real power flows on the local voltage would not be a good guide to the dispatch, but a linear approximation of the reactive power and voltage impacts on the margin for prices for the actual dispatch would be both appropriate in theory and much easier to estimate in practice. In effect, this linear approximation would be included as one of the binding transmission constraints modified to include the independent effect of local generation.

Extended Locational Marginal Pricing

The examples of dispatch-based pricing illustrate the simplicity and flexibility available for improving both real-time and day-ahead pricing models. Incorporating refinements in dispatch-based pricing should go a long way in supporting the incentives and integrity of electricity markets. However, the underlying economic dispatch pricing formulation has limitations in dealing with the full complexity of electricity markets.

In the core model of economic dispatch for electricity markets, energy prices derived from marginal costs support the equilibrium solution. This is true in the limited sense that within the formal structure included in the model the energy prices provide the appropriate charges to loads and payments to suppliers. However, prices that support the equilibrium solution for energy do not provide the necessary payments for products and services not included in the core model. Ancillary serves such as reactive support, black start capability and so on are not priced in the same way and must be paid for in a matter separate from the formal structure embedded in the core model. The particular rules for determining these payments are often ad hoc and not derived from an inclusive model. Charges applied to customers to cover these costs are similarly based on reasonable but ad hoc rules that often approximate some pro rata allocation across customers. These charges applied in addition to energy payments go under the heading of the "uplift" following a nomenclature established in the UK electricity market restructuring.

Hence, an uplift payment is an inherent part of electricity markets. The total cost of uplift payments is usually relatively small and the effects on market incentives are often ignored in formal analysis as being de minimis. However, this may not be true as more and more charges are included in the uplift. An example of an added source of uplift charges is associated with costs for start-up and minimum load requirements associated with unit commitment.

Unit commitment with economic dispatch introduces discontinuities into the electricity market model; e.g., the plant is committed or not. The usual interpretation of market-clearing prices has the economic dispatch as consistent in the sense that the dispatch is a profit-maximizing solution for the individual generators and load that take the prices as given. However, this market-clearing definition ignores the effects of the discrete decisions that define the discontinuities. In the presence of theseconditions, there may be no set of prices that supports the economic dispatch as the market clearing solution. In other words, given the prices some generators or loads may have an incentive to deviate from the economic dispatch. For example, a generator may not recover its startup costs at the given prices, and would prefer not to operate. Or another generator that should not be committed and dispatched may see a foregone opportunity cost because that generator would earn a profit at the given energy prices. One solution to make the economic dispatch consistent with the incentives in the market is to make up any losses for dispatched generators or to pay the lost opportunity profits for the constrained off generator.

These extra payments constitute an addition to the uplift that balances the difference between optimal profits and actual profits at the given energy prices. A reasonable goal would be to set the energy prices so as to minimize the extent of these uplift payments. This approach goes under the heading of Extended Locational Marginal Prices (ELMP). The ELMP derive from the so-called convex hull of the total cost of the unit commitment and economic dispatch problem. These ELMP prices minimize the uplift associated with the difference between the market clearing solution given the prices, and the optimal commitment and dispatch. This minimum uplift includes transmission congestion and loss revenues that determine the revenue adequacy of financial transmission rights. (Gribik, Hogan, & Pope, 2007)

The ISONE has been examining the implications of ELMP and its dispatch-based pricing approximations as part of its market discussions. (Coutu & White, 2014b) The essential idea is that many activities related to start-up, block loading of generation units, minimum output requirements, minimum run times, and so on, create various kinds of lumpy choices (the plant is off or on) that produce non-convexities in the total cost curve for unit commitment and dispatch. The results can sometimes be counterintuitive, but the theory and examples developed by the ISONE illustrate the use of ELMPs to provide a pricing regime that comes as close as possible to the core model and that reduces to the core model when the best convex approximation leaves no gap relative to the total cost curve. (Coutu & White, 2014c) The balance of payments appears in the uplift charges, and this contribution to uplift is as small as possible.

The ELMP approach deals with an application of the unit commitment and economic dispatch framework that goes beyond some of the simplifying assumption of otherwise convex problems. The intuition for ELMP can be summarized as recognizing the difference between the market-clearing solution at the ELMP and the economic dispatch. The market-clearing solution is the outcome that would prevail if all load and generation took energy prices as given, and individually chose their own profit-maximizing solution. If the problem has no lumpy decisions, essentially assuming convexity, then the economic dispatch is a market-clearing solution. There is no gap. But if the problem does not meet these convexity conditions, the two solutions can differ. Then the ELMP supports the market-clearing solution, and the associated revenue effects, impacts on financial transmission rights, the nature of binding constraints, and so on, apply to the market-clearing solution, but not to the actual economic dispatch. (Cadwalader, Gribik, Hogan, & Pope, 2010) The difference creates the need for the uplift payments.

The ELMP approach creates another type of requirement for dispatch-based pricing. Essentially, we have to know the solution and associated costs of the unit commitment and economic dispatch in order to calculate the ELMP. This does not require a perfect solution to the unit commitment and economic dispatch. The minimum uplift property of ELMP applies relative to any dispatch, including an approximate solution to the underlying problem. However, the ELMP is not produced as a byproduct of the economic dispatch in anything other than the convex case. We need to know the underlying costs and model in order to then calculate the ELMP. So ELMP is inherently a case of dispatch-based pricing.

In addition, actual calculation of ELMP involves finding a solution to an appropriate mathematical dual version of the unit commitment and dispatch problem. This dual formulation has long been recognized as an approach applicable to a wide range of optimization models. (Falk, 1969) There is a great deal of experience with methods for solving for these dual prices. The cumulative experience is that it is easy to get a good approximation of the prices but it is more challenging to get an exact solution. Combined with the complexity of the underlying model, there is a strong interest in providing what the ISONE terms as Approximate Extended Locational Marginal Prices (AELMP). (Coutu & White, 2014b)

One approach that has a great deal of appeal is to utilize a relaxed version of the unit commitment and economic dispatch model. This is referred to as the “dispatchable” model in (Gribik et al., 2007). This is a much easier problem to solve, and with the usual assumption would produce a convex model with easy-to-obtain prices. This convex problem looks like the core model where it is easy to compute prices. This approximation may not be a good way to determine the commitment and dispatch, but this AELMP formulation may provide a workable way to approximate the prices. The resulting dispatch-based prices would not be exactly the uplift minimizing ELMP, but they could be a workable approximation.

Day-Ahead and Real-time Interaction

A day-ahead electricity market creates financial commitments that are settled at the real-time price. There are a few decisions such as unit commitment actions that involve irrevocable physical effects, but the vast majority of the day-ahead transactions create obligations and rights for physical flows in real time that permit deviations from the day-ahead contracts to be discharged at real-time prices.

In addition, the typical day-ahead markets include so-called “virtual” transactions which are explicitly acknowledged to be purely financial. The virtual bids and offers are treated essentially the same as bids and offers from entities that own generation or serve load. However, unlike the entities with these real-time physical transactions, the virtual traders need have no connection to physical transactions and will settle as an imbalance the full amount of their contract at the real-time price.

These characteristics of electricity markets make clear the importance of efficient real-time pricing. All traders who will have to settle in the real-time market will try to anticipate the real-time price. Even if the full pricing rules in the day-ahead market are not perfect, with no transactions costs the effect of virtual transactions will tend to arbitrage any anticipated differences between the day-ahead and real-time price. This price convergence function provides a number of benefits to the market. The existence of virtual transactions also guarantees that defects in real-time pricing design will not be corrected by the day-ahead market. The day-ahead market comes chronologically before the real-time market, but in terms of the incentive effects the day-ahead market will follow the incentives established by the real-time pricing rules.

In the electricity system, a day is a long time. Hence, at the time the day-ahead market clears there is a good deal of uncertainty about the actual supply and demand conditions that will appear in real-time. This means that the arbitrage and equilibrium between day-ahead and real-time should operate in terms of the expected prices in real-time. On any given day, the prices in real-time could be quite different than the day-ahead prices. One of the reasons the day-ahead market is valuable for market participants is the ability to hedge the uncertain real-time price. But on average there should be convergence between day-ahead and real-time price, other than for what should be relatively small aggregate effects of risk aversion and transaction costs.

The connection between day-ahead and the expected real-time prices applies in the case of ELMP. The values of prices will depend on the choices of which commitment decisions are treated as variable. In real-time, many of these commitment decisions will be fixed in advance. If treated as fixed in the real-time ELMP pricing model, the prices will differ to some degree from the day-ahead prices. Sometimes they could be higher or lower in real-time, but on average the day-ahead prices and real-time prices should converge due to the effects of arbitrage and virtual bidding. In addition, virtual bids by design add convex elements to the day-ahead problem. This should reduce the lumpiness inherent in the model and tend to reduce the need for uplift.

Although virtual bidding offers a powerful tool to overcome defects in the day-ahead market, there is no good reason to impose defects that can be avoided. The implication is that the design of the day-ahead market should be established to reflect as much as possible the anticipated conditions in the real-time market. Hence, the description of the transmission grid should reflect the real grid expected to be in operation. Energy and ancillary services commitments should be included in the day-ahead clearing. Pricing for energy, operating reserves, congestion and so on, should all be the best possible approximation of the real-time conditions. The solution of the day-ahead model should be co-optimized just as in the real-time model. The same principles of dispatch-based pricing can be applied in the day-ahead and the real-time.

Forward Markets

More efficient pricing in the real-time and day-ahead spot markets would serve many purposes that will provide better incentives for forward markets for electricity and related markets for scheduling fuel supply.

The real-time market is the most important market. The rules for dispatch and pricing in real-time affect everything that comes before. Forward markets provide hedging and scheduling activities, but the forward markets look ahead to what will happen in real-time. If efficient prices do not appear in association with the actual dispatch, this inevitably creates incentives to either "lean-on" the rest of the system, when prices are too low, or self-schedule to avoid providing flexibility for economic dispatch, when penalties are too high. This applies to short-term scheduling of generation and fuel supply. It applies to long-term investment in generation capacity, energy efficiency and load management. The regulatory and administrative efforts to compensate for bad incentives in real-time add costs and complexity that are unnecessary and self-reinforcing.

Better spot prices that capture more of the scarcity impacts would help reduce the missing money problem and, therefore, make capacity markets less important. Better scarcity pricing would address directly the real-time performance problem for capacity resources that has been the subject of controversy and attention in ISONE.

Better real-time scarcity pricing would also provide incentives for generators to arrange for natural gas supplies and dual fuel capabilities. The ISONE plan is to allow revisions of offer prices to reflect real-time gas supply conditions and this would interact with dispatched-based pricing to provide the right incentives at the right time.

There are important related scheduling issues for arranging natural gas supplies, and these are being addressed. But better scheduling will not be as effective as it could be if it also accompanies improved implementations of real-time dispatch-based pricing.

Uplift Cost Allocation

Improved pricing would reduce some of the sources of uplift. The remaining uplift must be collected in some way to cover the costs incurred by the market operator. This allocation decision problem is a separate question with its own complications. There is an important dimension, however, in which the uplift allocation rules interact with the objective of achieving efficient prices.

A common appeal related to efficiency is to assign uplift according to some principles of cost causation. This logic is likely misplaced in the case of many uplift components. The cost causation arguments are motivated by the incentives at the margin, to the extent the effects can be identified. A basic objective of dispatch-based pricing is to move as much of these costs as possible into the efficient locational prices for energy and ancillary services. From this perspective, if there are marginal impacts of a transaction that contribute to uplift, this is a signal to identify and fix the ex post pricing rules, and not a signal to guide the allocation of uplift.

By this construction, uplift costs are those costs which cannot be attributed at the margin and therefore cannot be allocated according to a marginal principle of cost causation. The attempts to move away from the marginal to lumpy decisions such as simulations of the market with and without various transactions run into similar problems of lumpy decisions that are so difficult or impossible to solve as discussed in the case of ELMP.

For costs that cannot be identified at the margin, the principles of efficiency imply that cost allocation should be done in a way that minimizes the change in decisions, such as a decision not to schedule an efficient import due to the transmission access charge, while covering the total costs. The cost uplift allocation challenge is similar to or the same as allocating transmission access charges, connections fees and so on. The basic guideline is similar to more restrictive case of Ramsey pricing, where the effect is to allocate such costs to the transactions that will be changed the least. (Transpower, 2002)

The implication of this rule would be to allocate uplift charges more to final physical load and less or not at all to day-ahead virtual transactions. A problem arises through the allocation of uplift charges to virtual transactions. In ISONE, these uplift charges average around $2/MWh and provided a significant impediment for virtual trading to equilibrate the market. The independent market monitor recommended eliminating the charges for all virtual load and others that could not reasonable be assigned cost causation responsibility. (pp. viii-ix) (Potomac Economics, 2013) Removing uplift charges for virtual transactions would increase day-ahead market participation, improve price convergence and promote efficient day-ahead prices that properly reflect real-time expectations.

Conclusion

Dispatch-based pricing principles and associated practices work to create efficient price signals consistent with economic dispatch. A fundamental assumption of markets in electricity is that market participants will make operating and investment decisions based on market prices. The simplified theory implies that efficient prices are all that would be needed to provide the right incentives. That real systems do not conform to the simple theory does not imply that efficient prices are less important. To the contrary, the practical implication is that efficient prices become even more important in minimizing the out-of-market interventions and costs that are so difficult to design and which produce perverse incentives. The most important step for ISONE and other organized markets would be to change the priority from compensating for deficiencies in efficient real-time pricing to removing the impediments to efficient prices. Only when everything available had been done to move to efficient real-time pricing, would we turn to other market interventions. The examples of dispatch-based pricing applied to transmissions constraints, operating reserves, reliability commitments, voltage support, and unit commitment problems illustrate the ideas, the flexibility of the approach, and some of the limitations. Efficient real-time and day-ahead pricing models should be the first choice, not the last resort.

Appendix
Dispatch-Based Pricing Formulations

The purpose of ex post or dispatch-based pricing is to determine "prices consistent with the actual usage by applying the marginal tests of economic dispatch." (Hogan, 1992) The essential idea is to let the dispatch determine the prices rather than to change the dispatch to meet some model that might differ from the operational practices of the system operator. By making prices consistent with the dispatch, we create incentives that support rather than oppose the actions of the system operators.

The basic formulation utilizes a security-constrained economic dispatch problem. (Hogan, 2013b). Let $B(d)$ define the benefits of bid-in load (d) and $C(g)$ the cost of generation (g) offers. Incorporate other relevant variables such as unit commitment decisions in the control variables in u . The net load at each location is defined as the vector $y = d - g$. Aggregate losses are $L(y,u)$. Finally the transmission constraints appear in the vector function $K(y,u)$. With these definitions, we treat the underlying security-constrained economic dispatch problem as

$$
\begin{aligned}
&\underset{d\in D, g\in G, u\in U}{Max}\; B(d) - C(g)\\
&s.t.\\
&\quad d - g = y,\\
&\quad L(y,u) + \iota^t y = 0,\\
&\quad K(y,u) \le 0.
\end{aligned}
$$

The notation sets up the corresponding optimality conditions as

$$
\begin{aligned}
&\textit{There exists}\left(d^*, \mathrm{g}^*, y^*, u^*, \lambda, \eta, p\right), \textit{such that}\\
&d^* - g^* = y^*,\\
&L\left(y^*, u^*\right) + \iota^t y^* = 0,\\
&K\left(y^*, u^*\right) \le 0, \quad \eta^t K\left(y^*, u^*\right) = 0,\\
&\eta \ge 0, \quad u^* \in U,\\
&\left(d^*, g^*, y^*, u^*\right) \in \arg \max_{d\in D, g\in G, y, u\in U} \left[B(d) - C(g) - \lambda\left(L(y,u) + \iota^t y\right) - \eta^t K(y,u) - p^t (d - g - y) \right].
\end{aligned}
$$

Given these optimality conditions, we can apply the usual analysis to derive the general form of the locational marginal prices (LMP) as in:

$$
p = \lambda\left(\nabla L_y\left(y^*, u^*\right) + \iota^t\right) + \eta^t \nabla K_y\left(y^*, u^*\right).
$$

Hence, the LMP has the interpretation as the effect of the system reference energy price, marginal losses, and congestion.

In this representation, for an interior solution, the LMP at a location is also equal to the marginal benefit of load or the marginal cost of generation at the location.

$$p = \nabla B\left(d^*\right) = \nabla C\left(g^*\right).$$

The details of implementation may affect the characterization of the benefit and cost functions, such as to include capacity constraints, but the basic insight of the LMP formulation will continue to apply. Efficient prices are defined and determined consistent with the solution of the economic dispatch formulation. For further details on cases with multiple sets of efficient prices, see (Hogan, 2012).

Ex Post LMP

One of the earliest implementations of the dispatched-based approach was to extract the implied LMP values from the solution to the linear approximation of the full dispatch problem, where the linear approximation is based on the actual dispatch with binding transmission constraints in K^*. The set of binding constraints is not known before the dispatch is determined. Given the linear approximation of the binding constraints, the ex post LMP model would be:

$$\begin{aligned}
&\underset{d\in D, g\in G, u\in U}{Max} B\left(d\right) - C\left(g\right) \\
&\quad s.t. \\
&\qquad d - g = y, \\
&\qquad L\left(y,u\right) + \iota^t y = 0, \\
&\qquad \nabla K^*\left(y^*,u^*\right)\left(y - y^*, u - u^*\right) \leq 0.
\end{aligned}$$

Given the dispatch, the information needed to formulate this problem is both relatively simple and readily available. The critical elements would be the "shift factors" that define the derivatives of the binding constraints K^*, which is a small subset of the full list of possible constraints in K. The actual dispatch is a solution to this model, and the associated prices are the dispatch-based LMPs. The derivatives in ∇K^* directly affect the determination of the prices. However, the actual transmission flows may differ by some small amount from the announced limits is accepted as part of the approximation. The system operator need only establish the set of limiting constraints without specifying separately the constraints limits, which do not affect the LMPs.

Operating Reserve Demand Curves

The dispatch-based pricing formulation to include the ORDC for reserves r is a simple modification of the generic economic dispatch problem to separate energy generation from reserves subject to the generation total capacity Cap. The operating reserve benefit is just the corresponding area under the ORDC. Hence, the economic dispatch problem could be seen as:

$$
\begin{aligned}
&\underset{d\in D, g\in G, r, u\in U}{Max} B(d) - C(g) + ORBenefit(r) \\
&s.t. \\
&\quad d - g = y, \\
&\quad g + r \leq Cap, \\
&\quad L(y,u) + \iota^t y = 0, \\
&\quad K(y,u) \leq 0.
\end{aligned}
$$

Application of dispatch-based pricing with this model would account for simultaneous co-optimization of reserves and energy, incorporating scarcity prices defined by the ORDC in both the reserve prices and energy prices in real time. Similar comments would apply to a consistent application of the principles in the day-ahead market, which would use an ORDC and co-optimization to reflect consistent scarcity prices. (Hogan, 2013a)

Augmented ORDC

In the case of the ERCOT implementation, the ORDC incorporates both spinning and non-spinning reserves. Introducing an augmented ORDC by shifting the full ORDC raises the criticism that we would create an incentive to incur the costs of "too much" spinning reserve (R_S). As shown in (ERCOT Staff & Hogan, 2013), given the value of scarcity $(v = VOLL - mc)$ and the loss of load probability $(LOLP)$, the shift of the full ORDC produces a difference between the price of spin (P_S) and non-spin (P_{NS}) of:

$$P_S - P_{NS} = v * 0.5 * LOLP_S(R_S).$$

If we shift the full ORDC then this increases the LOLP and the difference between the price of spin and non-spin increases accordingly. This would create an artificial incentive to convert from non-spin to spin. Since non-spin (R_{NS}) is less costly, an adaptation would be to maintain the base difference between spin and non-spin, but apply the conservative reliability adjustment to the price of non-spin.

For example, if we made the change by simply adding a capacity availability payment to the price of all reserves, then the differential $P_S - P_{NS}$ would not change, and there would be no added incentive to convert from non-spin to spin.

A better approach would be to shift the LOLP for non-spin only. This would preserve the same differential between spin and non-spin, but would raise the price of both. The amount of the price increase would be a function of the total level of reserves in real time, preserving the property that when capacity is scarce the prices should be higher. In terms of the notation in the ERCOT backcast report, this would imply:

$$P_S = v * 0.5 * LOLP_S(R_S) + P_{NS}$$

$$P_{NS} = v * (1 - 0.5) * {LOLP^{\#}}_{NS}(R_{\mathrm{SNS}})$$

Here $LOLP^{\#}_{NS}$ would be calculated by shifting the distribution by a specified amount and calculated with the total spinning and non-spinning reserves. Hence, the total ORDC shifts up and the difference between spin and non-spin remains as

$$P_S - P_{NS} = v * 0.5 * LOLP_S(R_S).$$

There would be no increased incentive to incur the costs of spinning above the economic benefit. The conservative scarcity pricing would affect the total value of spin and non-spin, but the increase in availability would be for non-spin capacity.

Demand Response

The use of Demand Response (DR) to provide operating reserves lends itself to dispatch-based pricing. The implementation for DR reserves called to provide energy would be to incorporate the marginal cost of the next unit of undispatched generation, say mc , and apply this to the DR used and include the remaining demand response held reserves, DRr, as part of the benefit function.

$$\begin{aligned}
&\underset{d\in D, g\in G, r, DR, DRr, u\in U}{Max} B(d) - C(g) + ORBenefit(r + DRr) - mc * DR \\
&\quad s.t. \\
&\qquad d - g = y, \\
&\quad DRr + DR = TotalDR, \\
&\qquad g + r \leq Cap, \\
&\qquad L(y,u) + \iota^t y = 0, \\
&\qquad K(y,u) \leq 0.
\end{aligned}$$

The resulting LMPs will incorporate the scarcity value of the DR in the energy price, and use of DR will not produce a price reversal. The pricing model does not tell the operator when to use DR, but given that the dispatch calls on the DR the model determines the ex post prices consistent with the dispatch.

Reliability Unit Commitment

The effect of the incremental RUC units is to increase the available capacity relative to the bid-in and scheduled load. One way to apply the approach of dispatched based pricing for the day-ahead market would be to treat the RUC requirement as a need for increase reserves relative to the bid-in load indicated requirement. A direct approach would be to utilize an augmented ORDC to incorporate the reliability requirement of the RUC. The locational or regional ORDC calculated from the first principles would be augmented by shifting the probabilistic portion of the non-spin demand curve by the RUC amount. The increased capacity committed would be part of the available reserves, and the augmented ORDC would incorporate the value of the increased reserves. If we take the area under the augmented ORDC as $ORBenefit_{RUC}$. Then the revised dispatch-based adjustment would be as in:

$$
\begin{aligned}
&\underset{d\in D,g\in G,r,RUC,u\in U}{Max} B(d)-C(g)+ORBenefit_{RUC}(r+RUC)\\
&\quad s.t.\\
&\qquad d-g=y,\\
&\qquad g+r+RUC\le Cap,\\
&\qquad L(y,u)+\iota^t y=0,\\
&\qquad K(y,u)\le 0.
\end{aligned}
$$

This formulation for the day-ahead should avoid the problem of price reversals in the day-ahead market.

Voltage Support

When voltage constraints are important, it may be necessary to restrict load or operate certain generators in order to avoid reactive power losses or ensure reactive power supply. (Hogan, 1993) A linear representation of the impact of real power flows on the local voltage would not be a good guide to the dispatch, but a linear approximation of the reactive power and voltage impacts on the margin for prices for the actual dispatch would be both appropriate in theory and much easier to estimate in practice. In effect, this linear approximation would be included as one of the binding transmission constraints in K^* modified to include the independent effect of local generation. (Caramanis, 1982) (Baughman & Siddiqi, 1991) Let this constraint be K^+ .

The resulting addition to the implied pricing model would be to incorporate the effect of the voltage constraint.

$$
\begin{aligned}
&\underset{d\in D,g\in G,u\in U}{Max} B(d)-C(g)\\
&\quad s.t.\\
&\qquad d-g=y,\\
&\qquad L(y,u)+\iota^t y=0,\\
&\qquad K(y,u)\le 0,\\
&\qquad \nabla K^+(y^*,\mathrm{g}^*)(y-y^*,\mathrm{g}-g^*)\le 0.
\end{aligned}
$$

This version of the dispatch pricing model would then yield prices that incorporate an effect of voltage constraint on real power LMP results.

Extended Locational Marginal Pricing

Calculation of the exact ELMP involves finding a solution to an appropriate dual version of the unit commitment and dispatch problem. Although this is feasible, there is a strong interest in providing what the ISONE terms as Approximate Extended Locational Marginal Prices (AELMP). (Coutu & White, 2014b) One approach that has a great deal of appeal is to utilize a relaxed version of the unit commitment and economic dispatch model.

Discrete variables can capture a great deal of the complexity in real unit commitment and economic dispatch models. The discrete decisions, such as to start a plant, can be modeled as controlled by a variable that can be zero or one. Suppose the basic model with discrete decisions where the control variables in u must be zero or one.

$$
\begin{aligned}
&\underset{d\in D, g\in G, u\in U}{Max}\ B(d)-C(g)\\
&s.t.\\
&\quad d-g=y,\\
&\quad L(y,u)+\iota^t y=0,\\
&\quad K(y,u)\le 0\\
&\quad u=0,1.
\end{aligned}
$$

The discrete variables present a significant challenge in even solving these problems. But the solution is a separate topic. The importance for pricing is that the presence of these discrete variables creates the need for ELMP.

An approximate description of the same unit commitment and dispatch problem would treat these discrete variables as continuous, constrained only to be between zero and one.

$$
\begin{aligned}
&\underset{d\in D, g\in G, u\in U}{Max}\ B(d)-C(g)\\
&s.t.\\
&\quad d-g=y,\\
&\quad L(y,u)+\iota^t y=0,\\
&\quad K(y,u)\le 0\\
&\quad 0\le u\le 1.
\end{aligned}
$$

This “dispatchable” model is a much easier problem to solve, and with the usual assumptions would produce a convex model with easy-to-obtain prices. This approximation may not be a good way to determine the commitment and dispatch, but this AELMP formulation may provide a workable way to approximate the prices. The resulting dispatch-based prices would not be exactly the uplift minimizing ELMP, but they could be a workable approximation. (Gribik et al., 2007)

References

Baughman, M., & Siddiqi, S. (1991). Real-time pricing of reactive power: theory and case study results. *Power Systems, IEEE Transactions …*, (1), 23–29. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=131043

Cadwalader, M., Gribik, P., Hogan, W. W., & Pope, S. (2010). Extended LMP and Financial Transmission Rights. Retrieved from http://www.hks.harvard.edu/fs/whogan/CHP_ELMP_FTR_060910.pdf

Caramanis, M. (1982). Investment decisions and long-term planning under electricity spot pricing. … *Apparatus and Systems, IEEE Transactions on*, *75*(12), 4640–4648. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4111303

Coutu, R., & White, M. (2014a). *Real-Time Price Formation: Technical Session #1*. ISONE. Retrieved from http://www.iso-ne.com/support/training/courses/energy_mkt_ancil_serv_top/price_information_technical_session1.pdf

Coutu, R., & White, M. (2014b). *Real-Time Price Formation: Technical Session #4*. ISONE. Retrieved from http://www.iso-ne.com/support/training/courses/energy_mkt_ancil_serv_top/price_information_technical_session_session4.pdf

Coutu, R., & White, M. (2014c). *Real-Time Price Formation: Technical Session #5*. Retrieved from http://www.iso-ne.com/support/training/courses/energy_mkt_ancil_serv_top/price_information_technical_session5.pdf

ERCOT. (2014). *About the Operating Reserve Demand Curve and Wholesale Electric Prices* (p. 2014). ERCOT. Retrieved from http://www.ercot.com/content/news/presentations/2014/ORDCUpdate-FINAL.pdf

ERCOT Staff, & Hogan, W. W. (2013). *Back Cast of Interim Solution B + to Improve Real-Time Scarcity Pricing White Paper*. Retrieved from http://www.ercot.com/content/news/presentations/2013/White Paper - Back Cast of Interim Solution B+ to Improve Re.pdf

Falk, J. E. (1969). Lagrange Multipliers and Nonconvex Programs. *SIAM Journal of Control*, *7*(4), 534–545. Retrieved from http://siamdl.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=SJCODC000007000004000534000001&idtype=cvips

Federal Energy Regulatory Commission. (2014). *Price Formation in Energy and Ancillary Services Markets Operated by Regional Transmission Organizations and Independent System Operators, Docket No. ADI4-14-000*. Retrieved from http://www.ferc.gov/industries/electric/indus-act/rto/AD14-14-000.pdf

Gribik, P. R., Hogan, W. W., & Pope, S. L. (2007). Market-Clearing Electricity Prices and Energy Uplift. Retrieved from http://www.hks.harvard.edu/fs/whogan/Gribik_Hogan_Pope_Price_Uplift_123107.pdf

Harvey, S. (2014). *Is the California ISO Becoming an Uplift Market? Pricing , Uplift and Commitment*. FTI Consulting. Retrieved from http://www.caiso.com/Documents/BriefingISO_MarketPricing-MSCPresentation-May19_2014.pdf

Hogan, W. W. (1992). Contract networks for electric power transmission. *Journal of Regulatory Economics*, *4*(3), 211–242. Retrieved from http://www.springerlink.com/index/wh70283126518105.pdf

Hogan, W. W. (1993). Markets in real electric networks require reactive prices. *The Energy Journal*, *14*(3), 171–200. Retrieved from http://econpapers.repec.org/article/aenjournl/1993v14-03-a08.htm

Hogan, W. W. (2005). On an “Energy Only” Electricity Market Design for Resource Adequacy. Retrieved from http://www.hks.harvard.edu/fs/whogan/Hogan_Energy_Only_092305.pdf

Hogan, W. W. (2010). Scarcity Pricing and Locational Operating Reserve Demand Curves. Retrieved from http://www.hks.harvard.edu/fs/whogan/Hogan_FERC_060210.pdf

Hogan, W. W. (2012). Multiple Market-Clearing Prices, Electricity Market Design and Price Manipulation. *The Electricity Journal*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1040619012000917

Hogan, W. W. (2013a). Electricity Scarcity Pricing Through Operating Reserves. *Economics of Energy & Environmental Policy*, *2*(2), 65–86. doi:10.5547/2160-5890.2.2.4

Hogan, W. W. (2013b). Financial Transmission Rights: Point-to Point Formulations. In J. Rosellón & T. Kristiansen (Eds.), *Financial Transmission Rights* (Vol. 7, pp. 1–48). Springer London. doi:10.1007/978-1-4471-4787-9_1

IEA. (2007). *Tackling Investment Challenges in Power Generation in IEA Countries: Energy Market Experience*. *Energy*. Paris. Retrieved from http://www.iea.org/textbase/nppdf/free/2007/tackling_investment.pdf

ISONE. (2014a). 2013 Annual Markets Report. Retrieved from http://www.isone.org/markets/mkt_anlys_rpts/annl_mkt_rpts/2013/2013_amr_final_050614.pdf

ISONE. (2014b). 2014 Wholesale Markets Project Plan. Retrieved from http://www.iso-ne.com/pubs/whlsle_mkt_pln/2014_may_wmpp.pdf

ISONE. (2014c). *Winter Reliability Program*. ISONE. Retrieved from http://www.iso-ne.com/committees/comm_wkgrps/mrkts_comm/mrkts/mtrls/20

ISONE Internal Market Monitor. (2014). Review of Real-Time Replacement Reserve. Independent System Operator New England. Retrieved from http://www.iso-ne.com/committees/comm_wkgrps/mrkts_comm/mrkts/mtrls/2014/jun10112014/a02_imm_memo_05_19_14.docx

Joskow, P. L. (2008). Capacity payments in imperfect electricity markets: Need and design. *Utilities Policy*, *16*(3), 159–170. doi:10.1016/j.jup.2007.10.003

MISO. (2009). FERC Electric Tariff: Schedule 28. Retrieved from https://www.misoenergy.org/_layouts/MISO/ECM/Download.aspx?ID=19240

Newell, S., Spees, K., Pfeifenberger, J. P., & Karkatsouli, I. (2014). Estimating the Economically Optimal Reserve Margin in ERCOT. Retrieved from http://assets.fiercemarkets.com/public/sites/energy/reports/ercotreport.pdf

Potomac Economics. (2013). *2012 Assessment of the Electricity Markets in New England*. Retrieved from http://www.isone.org/markets/mktmonmit/rpts/ind_mkt_advsr/isone_2012_emm_rprt_final.pdf

Potomac Economics. (2014). *Comment on ISONE Pay-forPermormance Proposal, Docket ER14-1050-000*. Retrieved from http://elibrary.ferc.gov/idmws/common/opennat.asp?fileID=13462958

Transpower. (2002). *Confirmed Pricing Methodology: Final Design Principles*. Transpower, NZ. Retrieved from http://www.hks.harvard.edu/hepg/Standard_Mkt_dsgn/Transpower_design_prin_3-29-02.pdf

United States Court of Appeals for the District of Columbia Circuit. (2014). *Electric Power Supply Association v. Federal Energy Regulatory Commission*. Retrieved from http://www.hks.harvard.edu/hepg/Papers/2014/EPSA v FERC (Order No 745) No 11-1486 Decided may 23 2014.pdf

Yoshimura, H., & Parent, C. (2014). Full Integration of Demand Response Resources into the Energy and Reserves Markets. ISONE. Retrieved from http://www.iso-ne.com/committees/comm_wkgrps/mrkts_comm/mrkts/mtrls/2014/may672014/a03_iso_presentation_05_06_14.pptx

Endnote

[i] William W. Hogan is the Raymond Plank Professor of Global Energy Policy, John F. Kennedy School of Government, Harvard University. This paper was supported by Vitol Inc. This paper draws on research for the Harvard Electricity Policy Group and for the Harvard-Japan Project on Energy and the Environment. The author is or has been a consultant on electric market reform and transmission issues for Allegheny Electric Global Market, American Electric Power, American National Power, Aquila, Atlantic Wind Connection, Australian Gas Light Company, Avista Corporation, Avista Utilities, Avista Energy, Barclays Bank PLC, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator, California Suppliers Group, Calpine Corporation, CAM Energy, Canadian Imperial Bank of Commerce, Centerpoint Energy, Central Maine Power Company, Chubu Electric Power Company, Citigroup, City Power Marketing LLC, Cobalt Capital Management LLC, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, COMPETE Coalition, Conectiv, Constellation Energy, Constellation Energy Commodities Group, Constellation Power Source, Coral Power, Credit First Suisse Boston, DC Energy, Detroit Edison Company, Deutsche Bank, Deutsche Bank Energy Trading LLC, Duquesne Light Company, Dyon

LLC, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, Energy Endeavors LP, Exelon, Financial Marketers Coalition, FTI Consulting, GenOn Energy, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., GDF SUEZ Energy Resources NA, Great Bay Energy LLC, GWF Energy, Independent Energy Producers Assn, ISO New England, Koch Energy Trading, Inc., JP Morgan, LECG LLC, Luz del Sur, Maine Public Advocate, Maine Public Utilities Commission, Merrill Lynch, Midwest ISO, Mirant Corporation, MIT Grid Study, Monterey Enterprises LLC, MPS Merchant Services, Inc. (f/k/a Aquila Power Corporation), JP Morgan Ventures Energy Corp., 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 Attorney General, Ontario IMO, Ontario Ministries of Energy and Infrastructure, Pepco, Pinpoint Power, PJM Office of Interconnection, PJM Power Provider (P3) Group, Powerex Corp., Powhatan Energy Fund LLC, PPL Corporation, PPL Montana LLC, PPL EnergyPlus LLC, Public Service Company of Colorado, Public Service Electric & Gas Company, Public Service New Mexico, PSEG Companies, Red Wolf Energy Trading, Reliant Energy, Rhode Island Public Utilities Commission, Round Rock Energy LP, San Diego Gas & Electric Company, Secretaría de Energía (SENER, Mexico), Sempra Energy, SESCO LLC, Shell Energy North America (U.S.) L.P., SPP, Texas Genco, Texas Utilities Co, Tokyo Electric Power Company, Toronto Dominion Bank, Transalta, TransAlta Energy Marketing (California), TransAlta Energy Marketing (U.S.) Inc., Transcanada, TransCanada Energy LTD., TransÉnergie, Transpower of New Zealand, Tucson Electric Power, Twin Cities Power LLC, Vitol Inc., Westbrook Power, Western Power Trading Forum, Williams Energy Group, Wisconsin Electric Power Company, and XO Energy. Thanks to Susan Pope who provided valuable contributions. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (Related papers can be found on the web at www.whogan.com).