Opinion on Reserve Scarcity Pricing Design

by
Frank A. Wolak, Chairman
James Bushnell, Member
Benjamin F. Hobbs, Member
Market Surveillance Committee of the California ISO

November 26, 2009

Executive Summary

In this opinion, we comment on the California ISO's Reserve Scarcity Pricing proposal. Scarcity pricing is a term that technically applies to any circumstance where prices are set at levels above the offer price of the most expensive resource available. The California market has had de-facto scarcity pricing policy since start of the new market design in April of 2009. The current ISO proposal in fact deals with only a narrow aspect of scarcity pricing, that pertaining to the pricing of ancillary services under periods of insufficient market supply. We support the continued development of scarcity pricing in general, and this proposal in particular. However, we also lament the lack of a cohesive and consistent framework for valuing the scarcity of resources in a market context. Scarcity pricing is fundamentally about sending price signals about the value of additional resources, whether those resources are needed to supply energy, provide ancillary services, or relieve transmission congestion. We believe the CAISO–as well as other ISOs–should work towards a framework that can assess the system benefits of additional resources under consistent set of principles regardless of the specific services provided by those resources. These principles would apply both the transparent demand-curve values for reserves in Reserve Scarcity Pricing proposal, as well as to the more opaque parameter values used to value the relaxation of operating constraints when calculating locational energy and ancillary services prices.

1. Background

This opinion comments on the ISO's Reserve Scarcity Pricing proposal. Scarcity pricing is an administrative procedure that allows market prices to rise beyond the bid cap when there is insufficient supply offered into the market to meet the inelastic demand for that product. Reserve scarcity pricing will occur when the ISO is unable to procure sufficient Regulation or Operating Reserves through market mechanisms to meet its requirements for these products. The Federal Energy Regulatory Commission (FERC) has directed the ISO to implement Reserve Scarcity Pricing within 12 months after the start of the new market. The ISO's Reserve Scarcity Pricing proposal responds to this FERC order.

The Market Surveillance Committee (MSC) has discussed versions of ISO's Reserve Scarcity Pricing proposal at previous MSC meetings over the past few years. Most recently, the MSC was briefed on the ISO's latest Reserve Scarcity Pricing proposal during the MSC's October 15, 2009 meeting. This was followed by a discussion with stakeholders of the ISO's

proposal. The MSC also received written comments on the ISO's proposal from the Ellen Wolfe, the California ISO consultant for the Western Power Trading Forum (WPTF).[1] On November 4, 2009, the ISO released its final Reserve Scarcity Pricing Design, which this opinion comments on.[2] We are grateful to the ISO staff and stakeholders that participated in these meetings for their comments and insights.

Although the MSC supports the use of scarcity pricing in the ISO's markets, we believe that it is important to emphasize that the ISO has had a scarcity pricing process in place since the start of new market in April of 2009. The price-setting process in the day-ahead, hour-ahead, and real-time markets uses penalty parameters that allow a number of operating constraints, including the requirement that the ISO's operating reserve demands are met, to be relaxed if the optimized cost of honoring one of these constraints exceeds its penalty parameter. Consequently, the ISO's current Reserve Scarcity Pricing proposal can be thought of as simply a refinement of the current penalty pricing process for ancillary services.

We have one major comment on the ISO's reserve scarcity pricing proposal. This relates to the reliability and economic justification for how the operating reserves demand curve values are defined and how the values of all penalty prices are set. In addition, we raise two other issues. The first issue, similar to concerns raised by WPTF, is with the justification for setting the scarcity reserve demand curve value for sub-region reserve scarcity at a lower level than it is set for region-wide reserve scarcity. The second issue concerns the consistency between energy and reserve scarcity pricing. This opinion explains those concerns. We then conclude by pointing out that active participation of final demand in the wholesale market is the most effective and efficient of scarcity pricing.

We hope that the ISO will initiate a process for addressing the issues raised in this opinion as soon as is practical. This is because the choice of penalty parameters for system-wide and locational constraint violations can exert a significant impact on hour-ahead and real-time prices, even though the current constraints on operating reserves have yet to be violated locationally or on system-wide basis. We believe that a process for setting the values of penalty parameters based on reliability-based criteria can be accomplished without delaying the implementation of the ISO's reserve scarcity pricing proposal.

2. Scarcity Pricing Under Current Market Design

In theory, a locational marginal pricing market should reflect all relevant operating constraints in determining the feasibility of day-ahead energy and ancillary services schedules, real-time pre-dispatch at the interties, and real-time dispatch of internal resources and, furthermore, the settlement prices should be consistent with those constraints. However, particularly in the real-time pre-dispatch process and real-time market, it can be the case that there is no way to meet demand at all locations in the transmission network and still respect all of these operating constraints. For this reason, markets that use locational marginal pricing have mechanisms for relaxing certain operating constraints in order to obtain a feasible dispatch. This is accomplished by assigning a penalty parameter times the size of constraint violation in the

[1] Wolfe, Ellen, Sub-regional Scarcity Pricing White Paper, November 4, 2009.
[2] Final Proposal: Reserve Scarcity Pricing Design, November 4, 2009.

objective function used to determine the amount of energy and ancillary services each generation unit provides.

For example, the capacity of a transmission link can be relaxed in the scheduling and pricing runs in California if the minimized cost of honoring that constraint exceeds the administratively set penalty price for relaxing transmission constraints. Similar logic holds under the current ISO market for the requirement that the amount of operating reserves supplied equals the amount of operating reserves demanded by the ISO.

Although the ISO engaged in an extensive stakeholder process to set the values of these penalty parameters, no well-articulated economic justification was used. To provide the most efficient locational price signals under these system conditions, the values of the penalty parameters should be related to the consequences of constraint violation in terms of increases either in costs (for example, increased maintenance as a result of temporary overloads) or risk of load interruption. These consequences can depend upon many factors, including the particular load and generation configuration (which determine the likelihood of particular contingencies), the sequence of operator corrective actions that would take place in the case of such contingencies, and the effects of temporary overloads upon material properties and future maintenance costs, among many others. Although the complexity and uncertainty of these factors means that determining penalties cannot be an exact science, we believe that it should be possible to describe the relative consequences of broad classes of constraint violations and use that understanding as a guide to set the relative sizes of parameters.

We should stress that our concern is not unique to the California ISO market, but is a common issue for all formal markets. The lack of an economic basis for setting penalty parameters in formal markets is a symptom of the broader environment in which these constraints were developed. The operating constraints implemented by ISO's are driven by NERC and WECC reliability standards. These standards are in turn based upon engineering analyses that did not anticipate the kinds of trade-offs between reliability and prices that an LMP framework forces one to confront. A common problem in LMP markets is that designers must make explicit the trade-offs between different constraints, and these trade-offs need to be articulated in terms of values that significantly impact prices. The problem also presents an important opportunity; the opportunity to reconcile the economic and reliability standards that are applied to the regulation and operation of power markets.

We believe that a superior mechanism for setting these penalty parameters is to determine the extent to which a violation of a constraint significantly increases the likelihood of a system failure or equipment maintenance costs. For example, if a 1 MW violation of a transmission link significantly increases the loss-of-load probability and a 1 MW violation of another transmission link has little impact on the loss of load probability, the penalty price for the first constraint should be much greater than the penalty price for the second constraint. Specifically, if the change in expected unserved energy from relaxing the constraint on the first link is twice the change in expected unserved energy from relaxing the constraint on the second link, then the penalty parameter on the first transmission link should, in theory, be twice the penalty parameter on the second transmission link.

Setting penalty parameters based on the change both the change in the loss-of-load probability and the value of lost load would more accurately reflect the reliability implications of relaxing operating constraints in the locational prices for both energy and ancillary services. Any extreme hour-ahead and real-time prices that result from relaxing operating constraints could then be explained in terms of the increased likelihood of a system failure due to the relaxation of the constraint that led to the extreme price. We believe that the reserve scarcity demand curve values in the ISO's proposal should be set using this same logic. These demand curve values should be set based on changes in expected unserved energy associated relaxing the relevant operating reserve demand times the value of lost load. Although these numbers will always be imprecise, it should still be possible to assess the relative consequences of different violations in terms of possible magnitudes and extent of outages under various contingencies or changes in maintenance costs.

We recognize that making these calculations is extremely difficult and very time-consuming. Loss of load probabilities are dynamic and location-specific and depend on many aspects of real-time system conditions. Nevertheless, we believe that even are crude analysis of these issues that is continuously updated and refined should produce demand curve values and penalty prices that better reflect the cost of unserved load and therefore provide superior locational price signals for energy and ancillary services.

3. Locational Reserve Scarcity Pricing

A value-of-reliability-change justification for setting the penalty price for relaxing operating reserve requirements also provides guidance on setting reserve scarcity demand curve value for sub-regional versus regional requirements. Under the new market design, the ISO sets regional operating reserve requirements and minimum and maximum sub-regional requirements for regulation reserve and operating reserves. Each of these requirements can be relaxed at penalty price, which can lead to region-wide scarcity pricing or sub-region scarcity pricing for regulation and operating reserves, even without implementing the ISO's current reserve scarcity pricing proposal.

The ISO's reserve scarcity pricing proposal sets lower reserve scarcity demand curve values for relaxing sub-regional requirements relative to system-wide operating reserve requirements. Consistent with the logic of the previous section, this ordering of penalty parameters should be justified on a value-of-reliability basis. Relaxing a sub-regional operating reserve constraint could result in a larger or smaller increase in unserved energy than relaxing a system-wide operating reserve constraint, depending on feasible operator responses to such violations and the spatial extent and magnitude of load loss if particular contingencies occur. Assuming that the value of lost load is the same for all consumers in the California ISO control area, the changes in the loss-of-load probability would then imply a larger or smaller demand curve value, respectively, for violating sub-regional versus regional regulation and operating reserves. On one hand, a larger value could result if, for instance, there are many fewer options available within a particular sub-region to respond to a contingency. On the other hand, a smaller value would be appropriate if, for instance, emergency energy could be imported to a

sub-region from elsewhere by temporarily overloading transmission, whereas that option might not be available in the case of a CAISO-wide shortage.[3]

To be clear, CAISO staff *has* offered explanations for the scarcity value differences in the context of stakeholder meetings. The staff believe that the relaxation of local ancillary service requirements would indeed imply less severe consequences on local and system reliability than would the relaxation of a systemwide AS requirement. This provides a *qualitative* justification for a difference in scarcity values. However, the present scarcity pricing proposal would be enhanced by a clearly articulated rationale for setting relative penalty values for sub-regional versus regional constraints. In the absence of the consistent framework for evaluating all the various constraints, it is difficult to know whether the exact ratio of scarcity values offered by the ISO is *quantitatively* appropriate. We therefore recommend that the ISO consider undertaking studies to determine the reliability implications of relaxing all operating constraints, including those for operating reserves at the regional and sub-regional levels. We expect the results of these studies to yield penalty parameters that produce prices that better reflect the reliability costs of relaxing operating constraints than the current penalty prices and demand curve values for reserve scarcity pricing.

4. Inconsistency Between Reserve and Energy Scarcity Pricing

One potential flaw with the ISO's reserve scarcity pricing proposal is pointed out in Section 2.3 of the "Quarterly Report on Market Issues and Performance" by the Department of Market Monitoring (DMM). The fact that real-time ancillary services prices are set before real-time energy prices are determined is a significant source of potential market inefficiencies in the ISO's current reserve scarcity pricing proposal.[4] One hundred percent of the ISO's ancillary service requirements are purchased in the day-ahead market, but a small amount is purchased in real-time, if needed, for 15-minute intervals in the real-time pre-dispatch run (RTPD). The RTPD market co-optimizes energy and ancillary services, but only the ancillary services quantities are financially binding. The real-time dispatch (RTD) market is run 7.5 minutes prior to real-time, resulting in binding energy schedules for a 5-minute period. In addition, the awarded ancillary schedules are pre-determined in the RTD market. The DMM notes that, "(t)he disconnect in real time between procurement of ancillary services in RTPD and procurement of imbalance energy in RTD weakens the link between these two products and impedes price signals reflecting higher cost or scarcity in one product from being reflected in the other product."[5]

The analysis of market outcomes in the DMM's quarterly report demonstrates that this temporal mismatch between real-time ancillary services and energy prices is a potentially serious

[3] We note, for example, that in Section 5.2.1.2 of the MISO Business Practices that the Zonal Operating Reserves scarcity prices ($1100/MWh up to $2500/MWh) are appreciably above the proposed incremental value for CAISO regional reserves (Midwest ISO, *Energy and Operating Reserve Markets Business Practices Manual*, MO-BPM-001-r4, October 2, 2009, MKO-12 Public Manual No. 002, "Business Practices Manual: Energy and Operating Reserve Markets"). It appears that MISO zonal scarcity prices can reach such levels even if there is no market-wide shortage, unlike the CAISO proposal. There may be differences in operating procedures and consequences of contingencies that would justify the lower CAISO values, and if so, they should be described in the proposal.

[4] "Quarterly Report on Market Issues and Performance," Department of Market Monitoring, October 30, 2009.
[5] Ibid., p. 13.

market efficiency problem, because there is often little correlation between the real-time ancillary services prices and real-time energy prices. This implies that "scarcity conditions" could arise in the RTPD market which could result in extremely high ancillary prices despite the fact that energy prices in RTD market do not reflect scarcity conditions. The DMM report suggests that the ISO consider determining both ancillary services prices and energy prices in the RTD market. We recommend that the ISO consider starting stakeholder process for addressing this potential shortcoming in the current scarcity pricing proposal. We understand that the ISO plans to revisit this aspect of the current market design in a future stakeholder proceeding.

5. Active Demand-Side Participation in the Wholesale Market

In conclusion, we believe that it is important to emphasize that the least cost approach to scarcity pricing for electricity consumers is active participation of final demand in the wholesale market. Specifically, if consumers submit the price at which they are willing purchase additional energy from each ISO market—day-ahead, hour-ahead and real-time—this will appreciably diminish the likelihood that operating constraints will have to be relaxed in order to serve demand at all locations in the transmission network. Moreover, the frequency and magnitude of extreme hour-ahead and real-time prices will be significantly reduced.

As emphasized in our previous opinion on the ISO's Proxy Demand Resource (PDR) proposal, we believe that the ISO's current approach to involving final demand in the wholesale market is likely to require more frequent use of administrative scarcity pricing relative to the mechanism recommended in our opinion that requires consumers to purchase any quantity of energy sold in a subsequent market in an earlier market.[6] For example, in order to sell demand reductions in the real-time market, a load would have to purchase at least that amount of energy in the day-ahead market. We hope that the ISO will address these shortcomings of the PDR proposal before these market inefficiencies create a set of beneficiaries that make these market rules difficult to change.

[6] "The California ISO's Proxy Demand Resource (PDR) Proposal," Wolak, Frank A.; Bushnell, James; and Hobbs, Benjamin F., May 1, 2009, available at http://www.caiso.com/239f/239fc54917610.pdf.