Revised 17 May 1997

Report to the Trust For Power Industry Restructuring

PRIORITY PRICING OF ANCILLARY SERVICES

Market Design Inc.
Prepared by Robert Wilson, Vice President

Executive Summary

This report describes the analysis that underlies the new design of the ancillary services markets adopted at the May 8 meeting of the Trust Advisory Committee.[1] The main features are two-part bids, selection of resources based solely on capacity reservation bids, merit ordering of accepted resources in order of energy bids, and energy settlements based on prices in the real-time market. This design conforms to the principles of priority pricing set forth here. I show how the new design promotes overall efficiency and reduces gaming, besides its obvious simplification of settlements. I also comment on the measures required to maintain the competitiveness of these markets.

Introduction

This report describes a unified model for analyzing ancillary services. The exposition is cast in terms of spinning reserves but it applies to other ancillary services sharing the feature that capacity is reserved in advance and then used for generation depending on later contingencies. The formulation is cast in terms of the established theory of priority pricing, which applies to options exercised according to a merit order. The principles of priority pricing developed a decade ago at EPRI were formulated from a demand-side perspective and applied to curtailable loads. Curtailable loads are one kind of ancillary service, of course, and the theme here is that a comparable construction applies to other kinds of ancillary services on the supply side.

I use spinning reserve as the prototype to motivate the formulation and analysis. The presentation focuses first on the case that the markets are competitive; later I comment on the measures required to maintain competitiveness. Due to the availability of a sizable literature on priority pricing, the exposition here is only quasi-rigorous so that mathematical formalities do not detract from the main ideas.[2]

---

[1] A parallel analysis and a comparison of the old and new designs is provided in the report by London Economics, “An Alternative Protocol for the Selection, Pricing, and Settlement of Ancillary Services,” May 1997. This report is a work product of the Inter-Market Efficiency Team. I am indebted to Sam Lovick of London Economics for the basic insight that priority pricing underlies the efficient design of markets for ancillary services.

[2] Hung-po Chao and Robert Wilson, “Priority Service: Pricing, Investment, and Market Organization,” American Economic Review, 1987, 77(5): 899-916; Hung-po Chao, Shmuel Oren, Stephen Smith, and Robert Wilson, “Priority Service: Market Structure and Competition,” Energy

Context of the Pricing Problem for Ancillary Services

I interpret spinning reserve as one option in a cascade of options established to cap the real-time price. For instance, such a cascade might include AGC, balancing-market resources, spinning reserves, non-spinning reserves, replacement reserves, and various cold- or black-start reserves, with load curtailment somewhere in the merit order depending on its cost and response time.[3] These options differ in terms of their costs for capacity reservation and energy generation, as well as technical characteristics such as ramping speeds, sustainable limits on incremental loads and durations, and locational factors affecting transmission. The effect of the technical characteristics, for instance, can be that more expensive spinning units are loaded until less expensive non-spinning or replacement reserves can be mustered for generation over a longer duration.

The characteristic feature of ancillary services is that capacity is reserved in advance, and then incremental (or decremental) generation is ordered in response to later contingencies, such as an unexpected rise in the demand load due to weather, or on the supply side, the failure of a generator or a transmission line. This implies that there are two relevant prices.[4] One is the price paid for reserved capacity, and the other is the price paid for energy. Each of these prices is associated with a merit order, one for reserving capacity, and the other for ordering generation. It is important to realize that these two merit orders should not be the same.[5] One way to see this is to appreciate that a plant that is infra-marginal in the day-ahead energy market typically has a higher opportunity cost for reserving capacity but a lower generation cost, as compared to one that is marginal or extra-marginal in the energy market. This is because reserving capacity of the infra-marginal plant foregoes a greater profit from sales into other markets. For this reason, an efficient design of the combined markets for energy and ancillary services requires that the two merit orders are constructed separately. In particular, capacity reservations should be selected on the basis of offered prices for capacity, and generation should be ordered on the basis of offered prices for energy. The main contribution of the theory of priority pricing is to establish how this can be done to achieve an outcome that is efficient for the combined markets, including both energy and ancillary services.

The standard theory of priority pricing considers only the special case that opportunity costs for reserved capacity are negligible, so part of our task is to extend this analysis to include the case that opportunity costs are significant. Doing so requires analyses of

---

Journal, 1988, 9(4): 77-104; Robert Wilson, "Efficient and Competitive Rationing," Econometrica, 1989, 40(1): 1-40; Robert Wilson, Nonlinear Pricing, Oxford Press 1993, Chapter 10; Daniel Spulber, "Capacity-Contingent Nonlinear Pricing by Regulated Firms," Journal of Regulatory Economics, 1992, 299-320.

[3] The representation of the merit order for ancillary services as a cascade ordered by cost-reflective prices is a rough approximation, since locational and ramping constraints affect the ordering. In addition, the actual merit order for spinning and replacement reserves reflects the fact that when a spinning unit is loaded there may be additional costs associated with activating replacement reserves. Similarly, when AGC reaches its limit and is replaced by loading a spinning unit, it must be backed down to its set point.

[4] I simply assume that bids have two parts; a comparable analysis of one-part bids has not been undertaken.

[5] This is the fundamental insight that motivates the analysis in London Economics' report.

the markets for both capacity reservations and generation. As we shall see, the matter is straightforward when the markets are fully competitive. This is because the two markets can be separated when the design objective is overall efficiency of the combined markets. In fact, in both markets the merit order is constructed directly from the suppliers' offered prices. Some further measures required to ensure competitiveness are addressed later.

The Basic Model

I take as a datum the ISO's predictions about the likelihoods of various contingencies. In the case of spin these predictions can be represented by the probability $F(Q)$ that the demand for incremental generation ("AS demand") exceeds the quantity Q. This quantity is interpreted as the residual demand for generation from spinning reserves after cheaper or faster-response options in the cascade have been exhausted. Any excess not served by spinning reserves may be served by more expensive or slower-response options, such as non-spinning or replacement reserves or load curtailments.[6] Note that F depends on the position of spin in the cascade of available options; we take this position to be fixed and assume that its effects are already represented in the function F.[7]

The second datum is the economic structure of supply. In the case of spin this can be represented by a supply function $S(p)$ that indicates the offered supply of incremental generation at the price p ("AS supply"). This function indicates both the magnitude and the composition of the aggregate AS supply available from the reserved capacity. Due to the cascade structure, the AS supply function S should be interpreted as including all higher-cost options in the cascade that are available at prices higher than those for spin. For example, if the supply of spin is exhausted at a particular price then at each higher price some other supply or curtailable demand is included along with spin.

Assumptions of the Basic Model

The first task is to determine a supplier's optimal strategy in selecting its offered price for generation, assuming its capacity reservation is accepted. For this analysis I make two assumptions.

- The first assumption is that the ancillary services markets are fully competitive, and in particular each plant's reserved capacity is a small portion of the total. This is carried further by assuming specifically that:
  - (a) the distribution function F has a positive density function over the range of positive AS demands, so there is no lumpiness to AS demands;
  - (b) the AS supply function S is smooth and strictly increasing; and

[6] Even if non-spinning and replacement reserves have lower incremental costs, their start-up costs and slower ramping times may require their assignment to a later position in the cascade. As mentioned previously, the merit order takes account of both direct costs and the costs of replacements necessary to maintain the required reserve margin, as well as the imputed costs of technical constraints. Also, the WSCC requirements do not mandate replacement until the next hour. These are some of the reasons why the merit orders in terms of costs and response times need not be perfectly correlated.

[7] I assume that F applies to a particular hour. I do not develop here the extension to the case that incremental generation can be ordered repeatedly from the same plant during the hour.

(c) a plant's marginal cost is constant over the (small) range of its reserved capacity, so that it is sufficient to consider only how a supplier chooses its offered price for generating each incremental MWh.

It is useful to realize that part (a) does not assume that there is no chance of a large AS demand due to failure of a generator or a transmission line; rather, it only assumes that there is enough stochastic variability in loads that each specific level of AS demand has negligible probability. This ensures that no supplier can design its energy bid to exploit some particular contingency. This depends on the time frame, of course, but in the actual conduct of the ancillary services market this assumption is quite reasonable since the main markets are conducted a day before AS demand is determined.

Parts (b) and (c) convey the more specific requirement that no individual supplier has any significant role in the aggregate supply. This is a strong assumption that incorporates the major ingredient of a competitive market, namely that each supplier is small relative to the aggregate. The combined effect of (a) and (b) is that the probability distribution $G(p) = 1-F(S(p))$ has a positive density, indicating that each particular price has negligible probability. Note that $G(p)$ is the probability that the price required to serve the AS demand is no more than p. Both (a) and (b) are essential; e.g., if the supply function S is a step function due to discrete supply units then so too is the distribution of the required price.

A problem with (b) is that there could be a gap between the supply prices for energy from spinning reserves and the next option, say non-spinning reserves, when one takes account of the latter's start-up costs in constructing the overall merit order. For now I explicitly exclude this problem, but it is addressed in the final section where I examine the ISO's constructive role in ensuring that such gaps are closed.

A portion of part (c) is innocuous, since a basic feature of ancillary services is that each reserved capacity is limited to the range that can be achieved at its ramping rate within a specific time, and maintained for a specified duration; thus, it is sufficient for practical purposes to consider only the average cost over this limited range. The actual design allows each supplier to offer a supply function for incremental energy that is a step function, but we ignore this complication here.

- The second assumption is that settlements are based on the real-time price, and that a supplier is ordered to generate from its reserved capacity if and only if its offered energy price is no more than the current settlement price. [8]

This assumption states a key feature of the new design, namely that the merit order is based on the energy prices offered by those suppliers whose capacity is reserved, and settlements are based on the cost of the most expensive resource used. We shall see that this assumption is sufficient for efficiency and incentive compatibility. It is also largely necessary: to whatever extent settlements are linked to the offered energy price there is some prospect that suppliers' adaptation of their bidding strategies to exploit this connection will confound the determination of the efficient merit order.

[8] This assumption can be relaxed by assuming that the total supply when the balancing price is p is the sum of $S(p)$ and a random variable that represents the portion of the spin demand that is absorbed by unexpected supplies in the balancing market, or the effect on the merit order of ramping constraints.

The settlement price will typically be the price offered by the most expensive resource used; it may be the price of the last source used from the balancing market or the offered energy price of the most expensive reserved capacity ordered to generate. This reflects the perspective that the cascade of ancillary services is invoked to cap the real-time price.

Henceforth I call the energy price offered by a supplier its reserve price, in the sense that it is the lowest settlement price at which the supplier wants to be ordered to generate. To maintain incentives to fulfill reserve commitments, it might seem necessary to ensure that a supplier's reserve price is no less than its marginal cost, but we shall see that this provision is satisfied automatically, so there is no need to enforce it as a constraint on the design. For ordered generation, the merit order of spinning reserves with comparable response times is obtained by arranging the reserve prices of the accepted capacities in increasing order. In comparing different kinds of resources, such as spinning and replacement reserves, there are additional factors to consider, such as the requirement that when spin is used there may be a cost of replacing it to ensure maintenance of the reserve margin (see fn. 6).

Suppliers' Optimal Reserve Prices for Incremental Generation

The second assumption implies that the settlement price p will be at least as high as the reserve prices of those suppliers ordered to generate. In combination with the first assumption, this implies that each supplier's strategy can be phrased as a decision about where in the merit order it wants its reserve price to be.

These ingredients imply a formula for the expected profit (per MW of reserved capacity) for a supplier whose marginal cost is c and who offers the reserve price P. The probability that its plant is ordered to generate is $F(S(P)) = 1- G(P)$ and given the event that it is so ordered, its expected profit is the conditional expectation $E[\, p-c \mid p > P \,]$ of the profit margin given that the settlement price p exceeds the supplier's reserve price P. This conditional expectation is calculated using the probability density of the settlement price p implied by the distribution function $G(p)$ of the settlement price. Thus, the overall expected profit per MW is

$$\mathrm{Pft}(P,c) = F(S(P)) \times E[\, p - c \mid p > P \,] .$$

Using this formula, it is straightforward to show that the supplier's optimal strategy is to set the reserve price equal to its marginal cost, namely $P = c$. It is important to note however that the key step in deriving this conclusion stems from the first assumption, which in effect says that the distribution of the settlement price has a positive density. This property states in effect that the supplier cannot set its reserve price to exploit any specific contingency.

This result is also demonstrated in the analysis by London Economics. I realize that it seems counter-intuitive to some, but most arguments against it amount to the assertion that if the settlement price is set by the most expensive generator then necessarily that generator has market power. This would be true if a large AS supplier could plan for the event that its generator is the marginal one, but the above assumptions say essentially that no plant is large enough to have a significant probability of being marginal; thus this potential opportunity to affect the settlement price is ignored by suppliers when setting their reserve prices in advance.[9] It is true, of course, that in the actual event the marginal supplier could balk and demand a higher price, but my maintained assumption here is that sanctions against non-performance are sufficient to exclude this behavior.

The net conclusion from the above result is that in a fully competitive market, which was the first assumption, the second assumption implies that a supplier's optimal reserve price is its marginal cost. This implies that the merit order based on the suppliers' offered reserve prices is the efficient one.

Suppliers' Offered Prices for Reserved Capacity

Our next task is to go back to the initial stage when suppliers are selected for capacity reservations and show that the market outcome is also efficient at that stage. Thus we need to determine the optimal pair (R,P) that a supplier wants to submit, specifying its reserve price R per MW of reserved capacity, and its reserve price P per MWh for incremental energy.

To do this we must take account of the crucial role of opportunity costs. A supplier whose plant is uneconomical has no opportunities to sell energy into other markets, but one whose plant is infra-marginal in some other market must forego profits from sales there in order to reserve capacity for spin. If we represent these foregone profits per MW of capacity as V(c) depending on its marginal cost c, then net of profits in the ancillary services market, its actual reserve price is

$$Res(P,c) = V(c) - Pft(P,c) ,$$

since at any lower price the supplier prefers to use its capacity in other markets. Recall too our previous conclusion that <u>after</u> knowing its capacity reservation is accepted the supplier's preferred reserve price for energy is P = c . Thus, conditional on acceptance of its bid, the supplier's minimal reserve price is Res(c,c) , and this is as low as Res(P,c) for any other choice of P.

<u>The Role of the Bid Evaluation Rule: The New Rule</u>

The question we now address is how the bid evaluation rule for capacity reservations affects the supplier's strategy. Taking the simplest case first, suppose that the bid evaluation and settlement rule for capacity reservations are versions of the ones used in the May 8 design:

- **Bid Evaluation:** Offers of capacity reservations are accepted in the merit order determined by their capacity reserve prices, up to the total required by the ISO.

- **Settlements:** All accepted offers are paid the capacity reserve price, R*, that is the minimum among those rejected. [10]

---

[9] A large supplier at the margin obtains not only the advantage of setting the settlement price for its marginal MWh but also, due to the settlement rule, this price applies to all the energy it provides.

[10] In the actual implementation the settlement rule for capacity reservations is slightly different in that the settlement price is the maximum accepted bid, rather than the minimum rejected bid. In a

Thus for an individual bidder its offer (R,P) is accepted if $R < R^*$. In this case it is easily seen that its optimal strategy is to offer its actual capacity reserve price, namely $R = Res(P,c)$. Further, because this bid is lowest, and therefore provides the greatest chance of acceptance, when also the optimal energy reserve price is offered too, the supplier's optimal strategy is to offer the pair (R,P) for which

$$R = Res(c,c) \qquad \text{and} \qquad P = c \ .$$

Note that with this strategy the marginal rejected supplier, the one with the marginal cost c^* for which $R^* = Res(c^*,c^*)$, is indifferent whether its offer is accepted. If rejected this supplier gets its opportunity cost $V(c^*)$, and if accepted its expected profit is

$$R^* + Pft(c^*,c^*) = Res(c^*,c^*) + Pft(c^*,c^*) = V(c^*) \ ,$$

which is exactly the same. In general, those accepted profit and those rejected have no regrets.

The important conclusion is that efficiency is also obtained. We saw previously that the resulting the merit order for incremental generation is correct, and now we see that the capacity reservations accepted are from precisely those suppliers whose resources are less valuable in other markets.

The Role of the Bid Evaluation Rule: The Old Rule

We now re-examine these conclusions in the context of the old rule for bid evaluation. Recall that under this rule the merit order for capacity reservations was constructed from a linear combination, say $R + XP$, of the reserve prices for capacity and incremental energy. The parameter X was interpreted as the probability that generation is ordered from reserves.

To understand the motivation for this rule, consider the case of a supplier whose plant is infra-marginal in another market, such as the PX. For such a supplier, its marginal cost c is below average and its opportunity cost $V(c)$ is above average. This produces two offsetting effects in determining its capacity reserve price: $V(c)$ tends to be large when also $Pft(c,c)$ tends to be large, because both depend on the marginal cost common to both. This reflects the fact that at any particular settlement price for energy such a supplier has both a higher profit margin and a greater chance of being ordered to generate. Thus, there is no general tendency for those suppliers with the lowest reserve prices for energy to have the lowest reserve prices for capacity. The greater expected profit of a supplier whose marginal cost is low might be offset so much by a high opportunity cost that its offered price to reserve capacity exceeds the cutoff price and therefore is rejected.

This situation can be represented as an inherent negative correlation between suppliers' opportunity costs and marginal costs. Low cost-suppliers tend to have high opportunity

---

competitive market this makes no material difference. There some advantages from using the maximum accepted bid; e.g., it deters collusion, and it avoids the appearance of paying more than seems necessary – although in fact this is false since when one takes account of how the bidders adapt their strategies the expected cost is the same.

costs, and high-cost suppliers tend to have low opportunity costs, so that in a scatter diagram of their pairs (c,V(c)) one expects a regression line with a negative slope. This correlation represents the selection effect manifested in the composition of the supply sources for ancillary services. Most importantly, it implies that low generation costs cannot be obtained unless high prices are paid for capacity reservations.

This composition effect creates a significant problem in markets for ancillary services. It accounts for designs that evaluate bids using some combination of a supplier's reserve prices for capacity reservation and incremental generation. Such rules attempt to strike a balance between the costs of capacity reservations and incremental energy. A bid evaluation rule of roughly this sort is necessary to minimize the expected total cost of ancillary services whenever the correlation is significantly negative.

In sum, minimizing the ISO's total expected cost of ancillary services requires a bid evaluation rule akin to the old rule; in particular, bids cannot be evaluated solely on the basis of the prices offered for capacity reservations.

The Design Objective for Bid Evaluation

The contrast between the old and new bid evaluation rules shows that the correct rule depends on the design objective. If the objective is to minimize the ISO's costs then something like the old rule is required. If the objective is to promote the efficiency of the markets overall then the new rule is sufficient. The adoption of the new rule endorsed the efficiency objective rather than the cost-minimization objective.

To appreciate the consequences of retaining the old rule, the following scenario may be helpful. If the ISO were to evaluate bids based on a combination of the reserve prices for capacity and energy, then some additional suppliers with low marginal costs would be attracted away from the energy markets. This would reduce the ISO's costs but it would also increase prices in the energy markets, since these suppliers would be replaced there with higher-cost suppliers who were previously extra-marginal. Thus, the reduction in the ISO's costs would be obtained at the expense of consumers in the energy markets.

Competitiveness

The foregoing analysis depends on the assumption that the ancillary services markets are fully competitive. In this section I examine the validity of this assumption. First I examine the prospects for vigorous competition in the initial auction of capacity reservations, and then I examine the measures required to ensure competition in the subsequent energy market.

The Role of Contestability

A central feature of the overall system is that the ancillary services markets lie between the day-ahead and hour-ahead energy markets. This has the important consequence that a supplier can withhold capacity from the day-ahead energy markets to offer it in the ancillary services markets and then, if its offer is rejected, sell into the later hour-ahead and balancing markets. That is, a "failed" strategy in the ancillary services markets can be redeemed or "unwound" in later markets. Thus, there is no substantial sunk cost or irreversibly foregone opportunities from offering to reserve capacity for ancillary services.

Markets with this feature are said to be contestable. Even if a small fraction succeed in selling capacity reservations, there are many who stand ready to offer capacity reservations at lower prices if there is any systematic tendency for prices (i.e., R*) to exceed competitive levels. The most prominent of those ready to swoop in are the infra-marginal suppliers in the day-ahead energy markets: their lower marginal costs assure higher profit margins and priority in the merit order, and if the capacity reservation price R* is high then they can recover their opportunity costs.

This scenario indicates that the market for capacity reservations is inherently competitive. Those suppliers who are marginal in the energy markets have the lowest opportunity costs but also the highest marginal costs. Compared to the infra-marginal suppliers, they are easily undercut in the merit order, and their capacity reservation offers can be undercut if they don't offer competitive prices.

Contestability plays an equally important role ensuring a competitive market for incremental energy. This occurs in two ways. In one scenario, suppose there is a persistent gap in the distribution of accepted reserve prices, implying that their merit order creates a supply function with a vertical segment, and correspondingly the distribution of settlement prices has such a segment. If suppliers with slightly lower marginal costs anticipate this then they have incentives to raise their reserve prices; on the other hand, the mitigating competitive force is an influx of new suppliers whose offers can fill the gap and undercut those who did raise prices. In a second scenario a supplier persistently offers a price low enough to ensure acceptance of its capacity reservation but sets a very high reserve price for its energy, counting on a small chance that eventually it will be ordered to generate at this highly profitable price. This too attracts other suppliers who can profit from a lower energy price for the same capacity price. But there is more to this story that we address next.

The ISO's Role

The magnitudes of the ISO's requirements for ancillary services play an important role in sustaining the competitiveness of the energy part of ancillary services. For example, too small a requirement for spinning reserves makes it probable that spinning reserves will be exhausted. Anticipating this, and recognizing the ISO's extra cost of invoking non-spinning or replacement reserves, some suppliers of spin could raise their energy reserve prices to exploit this gap. This is like the scenarios above, but it need not be mitigated by new entrants since they too might seek to exploit this gap. This indicates that the ISO's requirement for one option, such as spin, should be sufficient to eliminate any gap between the prices for that option and the next option in the cascade. Closing gaps between successive options is essential to avoid the second scenario above. Ideally, the supplier who is last in the merit order for spin should obtain the capacity reservation price R* only, with no revenues from incremental energy because in the actual dispatch it is just marginally displaced by the least-cost supplier in the next option, say non-spin. This supplier's capacity reservation is accepted mainly to discipline the reserve prices offered by other suppliers of spin.

Alternatively, given a fixed requirement for spinning reserves, the ISO's intended policy of using auxiliary purchases of non-spinning and replacement reserves to discipline the market for energy from spinning reserves will close this gap. For instance, it is desirable to purchase enough resources from economical substitutes for spin to dispel any illusion of spin providers that higher energy reserve prices could capture extra profits. In addition, the premier option on a regular basis will be to use resources from the balancing market, which do not incur costs of replacement or start-up.

Conclusion

The main features of the new design for the ancillary services markets include two-part bids, selection of resources based on capacity reservation bids, merit ordering of accepted resources with comparable response times in order of energy bids, and energy settlements based on prices in the real-time market. This design conforms to the principles of priority pricing. It promotes overall efficiency, reduces gaming, and simplifies settlements. The competitiveness of these markets stems from contestability as well as the ISO's options to obtain sufficient resources to close gaps in the overall merit order for incremental generation.