

## **Comments on Barriers to Demand Response and the Symmetric Treatment of Supply and Demand Resources**

by

**Frank A. Wolak, Chairman**

**James Bushnell, Member**

**Benjamin F. Hobbs, Member**

**Market Surveillance Committee of the California ISO**

June 30, 2009

### **Executive Summary**

A major cause for many of the problems that have afflicted wholesale electricity markets is the unrealized potential for the demand-side to be a full participant. This fact is widely recognized, for example in the California ISO's recent report on barriers to demand response.<sup>1</sup> However, in an effort to increase demand's role in the wholesale market, "demand response" programs are being promoted and implemented that we believe will actually forego many of the benefits of having final consumers that are truly responsive to system conditions. In these comments, we argue that dynamic pricing that reflects varying system conditions over locations as well as time is the path to realizing the full benefits of active participation of final demand in the wholesale market.

We strongly support the Federal Energy Regulatory Commission's (FERC) continued focus and interest in fully integrating final consumers into wholesale power markets. However, the limits of FERC's regulatory jurisdiction greatly restrict its ability to promote such integration. Being limited to working with the tools at its disposal, FERC has chosen to pursue a model with competitive provision of "demand response" services at the wholesale level. Unfortunately, such services in practice fall well short of meeting FERC's vision of wholesale markets where demand and supply are treated symmetrically. There is a real possibility these efforts would crowd out far superior approaches. Thus, we are concerned that FERC, by pushing too hard on the regulatory levers it does have at its disposal, will end up undermining its own goals of achieving symmetric treatment of supply and demand.

The major remaining barrier to active participation of final demand in California ISO markets is state-level regulatory policy. The necessary metering technology to record the hourly consumption of all final consumers of the three large investor-owned utilities is scheduled to be in place by the end of 2011. The failure to set the default wholesale price paid by all California consumers equal to hourly real-time price, instead setting a fixed default retail price that does not vary with real-time system conditions, significantly dulls the incentive for final consumers to participate actively in California ISO markets. If such a default pricing mechanism were to be adopted, the multi-

---

<sup>1</sup> *California Independent System Operator Demand Response Barriers Study (per FERC Order 719)*, Prepared by Freeman, Sullivan & Co. and Energy and Environmental Economics, Inc., April 28, 2009. This report identifies a large number of barriers to demand response that we also view as significant obstacles to effective price-responsive demand, and are broadly consistent with our comments *infra.*

settlement energy and ancillary services markets that currently exist in the California ISO market would be well-suited to active participation of final demand. We emphasize that, even if the default retail price all consumers face passes through the hourly real-time wholesale price, no consumer is required to pay this retail price. However, in order to avoid paying this price, the consumer must sign a hedging agreement with its retailer that recovers the retailer's cost of managing some or all of this real-time price risk.

In these comments we make an important distinction between traditional demand response (DR) programs and dynamic pricing. Traditional demand response programs typically pay customer to reduce their consumption relative to an administratively set baseline level of consumption. Unfortunately, individual customers will always know more about their true baseline than the administrator of a DR program, and can likely profit from that knowledge. Therefore we fear that efforts to increase traditional demand response programs with an administratively set baseline could very well crowd out more reliable and effective dynamic pricing approaches – simply because customers will prefer them for the wrong reasons.

That said there is a productive role that can be played by the provision of demand response through the traditional approaches. The baseline issue is largely irrelevant for the provision of ancillary services as long as ISOs can accurately measure the actual change in electricity consumption by the load resources providing those services. In addition, baseline issues for demand participating in energy markets can be addressed by requiring that baselines levels be paid for, rather than simply estimated by a statistical model or set through some other administrative mechanism.

Last, we believe that truly price responsive demand should depend on prices that accurately reflect the underlying real-time market conditions. This means that locational as well as time-varying aspects of prices would be reflected in the prices faced by consumers. It also means that customers should be responding to an energy price, rather than a periodic capacity payment. Remuneration of demand response resources through capacity payments simply reinforces existing flaws in traditional approaches to DR. Existing capacity payments mechanisms have continually struggled to provide suppliers with performance incentives that properly reflect the underlying hour-by-hour fluctuations of market conditions. Forcing demand response resources as well as the supply side of the market into this paradigm only perpetuates this problem and will make it make it more difficult to correct.

## **1. Demand Response versus Dynamic Pricing**

In order to assess the barriers to demand response, first it is helpful to consider the question of what defines demand response. A second and more important question is whether the term is becoming obsolete. One rarely hears the term used outside of the electricity industry because the notion that consumers must pay and make decisions based on a real-time price is a fact of life in all industries without explicit price regulation. For example, in the airline industry, consumers always have the option to show up at the airport the day and time they would like to fly and pay the current price for a ticket,

which could be extremely high if the flight is sold out. As we discuss below, the failure to treat electricity like other products is the major barrier to active participation of final demand in the wholesale market.

For the purposes of this discussion, we will make the distinction between the *demand response* as the term has traditionally been used in the electricity industry and *dynamic pricing*. Traditionally, demand response has represented a specific paradigm for integrating the consumption decisions of certain types of customers into wholesale electricity markets. This paradigm involves identifying a potential *reduction* in consumption and treating that reduction as the service provided. Historically, this service has typically been called upon by system operators only for reliability reasons, rather than for economic reasons. Specifically, many demand response resources can only be called upon to provide a reduction in consumption in response to the system operator declaring a system emergency, rather than because the final consumer faces a real-time price above its willingness to pay for electricity. For the reasons we discuss below, this paradigm has always been problematic for the efficient integration of final consumers into wholesale electricity markets.

The term *dynamic pricing* refers to a form of consumer interaction with conditions in the market that is commonly found in other industries. The idea behind dynamic pricing is that customers pay to consume the product at a price that varies with real-time supply and demand conditions. Within the electricity industry various models of dynamic pricing have been applied, including real-time pricing (RTP) and less dynamic and more restrictive forms such as critical peak pricing (CPP).

In past decades, the argument could be made that technological constraints, such as the lack of availability of interval meters in the distribution network, limited the potential for integrating consumers in through dynamic pricing, and that the traditional DR paradigm was therefore a necessary second-best alternative. Today, although some infrastructural barriers to locational pricing for load remain, as mentioned in the ISO's barriers report, these technological barriers have been greatly reduced. Specifically, California's three investor-owned utilities all have plans to install interval meters for all of their customers by the end of 2011. In addition, many technologies allow customers to respond automatically to pricing signals currently exist and many more are being developed.

## **2. Problems with Current Demand Response Paradigm**

The fundamental problem with what we will refer to as the *demand response paradigm* is that electricity consumption is *not* treated symmetrically with production. Symmetric treatment would set a default price for electricity consumption that is equal to the default price that suppliers are paid for producing electricity—the real-time hourly price for electricity. In addition, under true symmetric treatment, the amount supplied and the amount consumed during each period would be actually metered, not measured by comparison to some estimated baseline level.

Put another way, the baseline for suppliers in wholesale electricity markets is zero, and they earn the hourly price for each MWh of energy or operating reserves they provide above zero. If suppliers choose, they can contract ahead of time to sell a pre-specified amount of energy. This amount defines a baseline in later markets which will pay them an hourly price if they produce more than this amount, or require them to buy power from others if they fall short. As a result, suppliers can hedge risks but still have efficient incentives for short-run operation.

### ***2.1. Symmetric Treatment of Load and Generation***

The symmetric treatment of demand, by definition, involves setting a zero baseline for final consumers, as for suppliers, and charging for each unit of consumption at the hourly price. If a demand resource would like to provide operating reserves, such as spinning or non-spinning reserves, then it would need to schedule a quantity of consumption in the day-ahead market and be willing to curtail consumption relative to this day-ahead schedule if its offer to consume less energy is accepted in the real-time market.

Facing a default price equal to the hourly wholesale price does not require any final consumer to pay this price or any electricity supplier to receive this price for all their electricity. Similar to the markets for other products, this is the default price that the consumer must pay and producers will receive if they do not make any forward market arrangements. In many markets, consumers reserve some amount of the product in advance at a fixed price. However, if they want to consume more than they reserved, they have to pay at the current price. If they are willing to consume less than what was reserved, they are instead paid that price for the amount they do not consume. In this way, the incremental consumption decision is based upon the real-time price.

Returning to the airline industry, few passengers pay the default real-time price for a flight because they find the fixed-forward contract offered by the airline—an advance-purchase ticket—more attractive. However, the reason that consumers find the advance-purchase ticket more attractive is because the real-time price of a flight can be extremely volatile (in particular, the flight can be sold out) and the advance-purchase ticket provides insurance against this uncertainty. For the same reason, we would expect few customers to purchase all of their electricity consumption at the hourly real-time wholesale price because they are likely to find pricing plans that provide price certainty for a significant fraction of their expected hourly consumption more attractive.

### ***2.2. The Problem of Paying for Demand Reductions***

Most of what are traditionally described as demand response (DR) programs pay consumers to reduce their consumption relative to some administratively set level. Any initiative that pays customers *not* to consume something, be it water, CO<sub>2</sub> emissions, or electricity, faces a serious challenge of measuring what the consumer *would have* done without the payment. It is impossible to observe this counterfactual consumption level,

because we cannot measure something that did not happen. If the customer receives payment for a demand reduction during the hour, the only quantity that can be observed is the customer's actual consumption with the payment. Determining what the customer's consumption would have been without the payment must rely on an economic or statistical model of the customer's behavior to estimate this counterfactual consumption.

This problem is often described as the "baseline" problem. The baseline problem can often be broken down into two classic difficulties that are found in markets where information is not complete: an adverse selection problem and a moral hazard problem. The adverse selection problem arises from the fact that, when paying for reductions, the "buyer" of demand response does not know precisely what the consumer would have consumed in the absence of a DR payment. Even the best economic or statistical models of a customer's hourly electricity consumption behavior as function of hourly prices and all observable customer and weather characteristics are only able to explain a small fraction of the variation in that customer's consumption of electricity across hours of the year. Consequently, even assuming that the best model of the determinants of each customer's electricity consumption is available, it is impossible to determine accurately what that customer would have consumed in the absence of the payment.

For example, California's 20-20 program was designed to encourage conservation during the summer of 2001 by rewarding consumers whose monthly consumption fell more than 20% relative to the same month in the previous year. Although it is very likely that this program stimulated electricity conservation, it also rewarded customers who would have reduced their consumption even without the rebate program. It is widely acknowledged that historically a sizeable fraction of residential customers experienced a fall in their monthly consumption of 20% percent or more *without any incentive to reduce their consumption*. These customers received rebates under the 20-20 program simply because a child graduated or they took a long vacation, not because they actively took steps to reduce their electricity consumption.<sup>2</sup>

The moral hazard problem arises whenever customers are rewarded for having higher baselines. Put simply, firms and customers have a strong incentive to inflate the level of their baseline, because they are paid based upon the comparison of their actual consumption to this baseline. In many cases, customers can be given a perverse incentive to over-consume as a means to inflate their baselines. For example, when customers have the opportunity to consume all they want at a standard fixed retail rate, and then sell back reductions from that level at the real-time price in an ISO market, there is an incentive to raise consumption in order to increase the level of the "reduction," which is rewarded at a much higher price. It is critical to recognize that this is more than just a

---

<sup>2</sup> A discussion of the 20/20 program can be found in Faruqui, A. and George, S. (2006) "Pushing the Envelope on Rate Design." *The Electricity Journal*, March, Vol. 19, Issue 2. It is important to appreciate, however, that this payment to people who would have reduced consumption anyway is not *per se* inefficient. Rather, inefficiencies can arise if consumers have an ability and incentive to alter their decisions at other times in order to affect that baseline (leading to the moral hazard problem), or can choose whether or not to participate in that program (leading to the adverse selection problem). See our discussion in Footnote 5, *infra.*, and elsewhere in this opinion.

measurement problem. Even *perfect* measurement of consumption does not eliminate the moral hazard problem with regards to baselines. The problem is created by an underlying rate structure that provides asymmetric rewards for consumption and reductions.

We are concerned that many of the proposed demand response products are vulnerable to both adverse selection and moral hazard. This would be a serious problem for several reasons. It can continue to undermine faith in the reliability of DR among system operators, who would come to view price-responsive demand as less reliable than an equivalent amount of generation capacity. This, in turn, can result in customers having to purchase more capacity resources and operating reserves to replace the demand response resources that do not in fact provide the promised amount demand reductions. This will inflate costs to consumers and increase uncertainty as to the actual reliability of the system. Such DR programs can also be artificially attractive to consumers relative to dynamic pricing, and thus discourage its adoption and resulting, in the long run, in the loss of the benefits that dynamic pricing could provide.

### **3. The Role of Price-Responsive Demand**

An examination of the behavior of day-ahead and real-time prices under the Market Redesign and Technology Upgrade (MRTU) shows that the energy market has the greatest value for quickly responsive adjustments during periods with large demand changes, rather than simply high levels of demand. The experience of the ISO to date with MRTU also reveals that this value is highly localized. For this reason, we believe that the current focus of using DR as a reliability tool is counter-productive. Trimming a few demand peaks during the year, as is envisioned for demand response resources under the current resource adequacy (RA) paradigm, greatly undervalues the role of price-responsive demand.

Therefore, we believe that an emphasis on remunerating DR resources primarily with capacity payments would not be economically efficient or yield the greatest benefits to system reliability. The capacity paradigm exacerbates all the weaknesses of current DR approaches and, as with generation, tends to reward the *potential* for provision of a service rather than the actual provision of that service. By far the most desirable form of remuneration for demand in both limiting real-time price volatility and enhancing system reliability is through energy payments rather than capacity payments.

As we argue above, it is far more effective to implement these payments through an approach that treats demand and supply resources symmetrically. The multi-settlement markets that exist in all of the ISOs in the United States are ideally suited to allowing symmetric treatment of demand and supply. Final consumers can schedule a given level of consumption in the day-ahead market and then sell day-ahead ancillary service capacity or real-time energy reductions relative to this day-ahead schedule in the real-time market. Similarly, suppliers can schedule from their generation units in the day market and then sell day-ahead ancillary services or additional energy in the real-time market beyond that day-ahead energy schedule. This “buy your baseline” approach to selling demand reductions in a subsequent market ensures that retailers and curtailment

service providers (CSPs) face the full financial consequences of their baseline choice in the same way that suppliers face the full financial consequences of their final energy schedules in the real-time market.

We readily admit the quantity of demand response resources that market participants are willing to provide is likely to be less if demand response resources are required participate in the ISO markets under the same terms and conditions as generation unit owners. We do not believe that the success of DR programs should be judged by the amount of MWs or MWhs sold, if these magnitudes are in fact not financially binding or directly verifiable. For example, the ISO could purchase a large quantity of MWhs of demand reductions that are purely the result of an artificially high baseline. These demand reductions provide no economic or reliability benefits, but consumers must still pay for them. To judge a program as a success because it has a large number of participants and a large number of MWhs sold fails to recognize the primary goal of symmetric treatment of demand and supply resources—to improve market efficiency and system reliability.

#### **4. Using DR to Provide Ancillary Services**

We also believe that DR resources should be allowed to provide ancillary services under the same terms and conditions as generation unit owners. In this regard, it is important to distinguish between the provision of Ancillary Services (AS) and the provision of demand response (non-consumption) for energy. In the former, the customer is providing an actual service, the ability to alter consumption if the system needs it due to a contingency. As long as measurement is accurate, the baseline issue is not a major concern. The ISO needs to be confident that the *change* in consumption it expects upon calling for it will occur.<sup>3</sup>

In the case of AS, the ISO does not necessarily need to worry about the *level* of consumption from which this change is occurring, as long as it can measure the level of consumption when change is requested and the change in consumption actually does occur. We therefore advise that AS performance simply be measured based upon the actual metered consumption of a customer before and after the demand is called upon to perform. While reliability concerns may necessitate the ability to measure telemetry, we see no incentive problem if it is lacking. It should be understood that the provider should not sell a level of AS in excess of its measured demand (or a fraction of it), and that performance will be measured by the change in consumption from the appropriate interval before to after the resource is called upon. If this change is in fact less than what was offered as reliable AS, then penalties comparable to non-performance by a generator could be applied.

---

<sup>3</sup> We note, for example, that the recent Lawrence Berkeley Laboratory evaluation of Southern California Edison's use of air conditioner load controllers as a spinning reserve resource has provided convincing evidence that provision of AS by demand resources can be accurately monitored and verified (J.H. Eto, J. Nelson-Hoffman, C. Torres, S. Hirth, B. Yinger, J. Kueck, and B. Kirby, Demand Response Spinning Reserve Demonstration, Prepared for Energy Systems Integration Public Interest Energy Research Program, California Energy Commission, LBNL-62761, May 2007).

We can imagine testing the ancillary services sales by demand resources in the same manner that ancillary services sales by generation units are tested. If the ISO accepts 10 MW from the demand resource in the non-spinning reserve market, then the ISO could test the provision of this service by asking the demand resource to reduce its load by this amount with the appropriate advance notice. If it turns out that the customer is unable to do so, then the ISO would be assess the standard penalty for failing to provide non-spinning reserve when called upon that is assessed to generation units that fail to provide this ancillary service.

## **5. Regulatory Barriers to Symmetric Treatment**

California's ambiguous policies toward retail and wholesale market organization further complicate the issues and implications of demand response policies. Political resistance to dynamically-varying and locationally-varying prices for load is highlighted in the California ISO's barriers report. Among the policies that create difficulties are several provisions from California's 2001 Assembly Bill 1X, and certain aspects of the current implementation of the CAISO's wholesale market rules.

Perhaps the most relevant and distinct issue for the California market, relative to other ISO markets, is the suspended status of third-party retail access in AB 1X. While we will not discuss here the relative merits of this policy, it does greatly limit the mechanisms that non-utility providers can use to implement dynamic pricing for final customers. The only model left for non-utility providers is the more traditional, highly flawed, demand response model through what is called curtailment service providers (CSP). Unfortunately, the suspension of third-party retail access creates a stark choice for final consumers served by the California ISO that does not exist in other ISO markets: final consumers can either (1) pursue a more effective model of consumer participation through dynamic rates that would likely have to be implemented primarily through regulated LSEs, or (2) implement the competitive provision of a much less effective and reliable form of DR through the CSP model. We believe that implementation of dynamic pricing, done right, is the most important goal, whether it be accomplished through competitive provision or through traditional utility channels.

Another aspect of AB 1X that is often identified as a barrier to dynamic pricing is the provision that the retail price not increase for residential customers consuming at or below 130% of "baseline" consumption levels. This limitation has been interpreted as precluding the option of charging higher prices during critical hours unless those increases are completely offset by comparable reductions during other hours. However, it is important to remember that this restriction applies only to residential customers. Whether this restriction is equitable has been a continuing source of debate, but we note that most of the "low hanging fruit" of dynamic pricing response is likely found in the commercial and industrial sectors.

Even for residential customers this limitation is not insurmountable. In fact one could interpret the restriction as fitting well with the concept of "buying your baseline."

The key to this concept is that customers have the right to buy only a set amount of electricity at the fixed rate, rather than the right to buy all they want at that rate.<sup>4</sup> A dynamic tariff could be structured so that customers essentially reserve a baseline level of consumption at the AB 1X price and any consumption in excess of that level is priced at the dynamic wholesale price, while consumption below that level is refunded at that same wholesale price.<sup>5</sup> Thus customers who consume less than the baseline essentially earn rebates based upon the hourly wholesale price. The key difference in this approach from other proposed DR programs is that the baseline is set exogenously and is not costless.

One last current policy that we will highlight as problematic for the implementation of both dynamic pricing and traditional demand response is the current aggregation of the locational marginal prices paid by final consumers to load aggregation points (LAPs). The disconnection between underlying wholesale prices at specific locations and the average prices paid by customers within LAPs is analogous to the disconnection between hourly fluctuations in wholesale prices and the annual averages of those prices paid by consumers. In other words, true demand response would respond to both the temporal and locational components of wholesale prices.

More importantly, the aggregation of consumer prices to LAPs creates difficulties for providing the proper incentive to a specific single customer in a given location. In a DR paradigm, LAP pricing can create the opportunity to inflate baselines simply to take advantage of the difference between a locational price and the LAP price. Essentially customers can be paid at a nodal price for reductions from levels they purchased at the LAP price.<sup>6</sup> Proposals to address this problem really only limit the potential damage rather than fix the underlying incentive problem which can only be addressed by charging consumer at a given node the price at location for its day-ahead schedule and paying that same customer the real-time price at that location for its reduction in consumption in real-time relative to this day-ahead schedule.

## **6. Conclusion**

Any paradigm that sells “reductions” from an exogenous baseline will crowd-out the adoption of direct pricing options such as critical peak and real-time pricing. Thus, we fear that the adoption of this weak form of demand response will ultimately work against the adoption of a truly symmetric treatment of load and generation that is an essential component of an efficient wholesale electricity market. We believe that there is a significant risk of creating conditions that will crowd out true price response by

---

<sup>4</sup> A detailed discussion of the impacts of dynamic pricing on customer bills and the concept of buying baselines can be found in Severin Borenstein, "Customer Risk from Real-Time Retail Electricity Pricing: Bill Volatility and Hedgability," *The Energy Journal*, 28(2) 2007.

<sup>5</sup> The baseline quantity, which is defined as a monthly total KWh quantity would have to be translated to an hour-by-hour allocation. There are several ways to do this and as long as the translation is based upon an exogenous formula and not the ongoing consumption of specific customers, the proper incentives for responding to prices are preserved.

<sup>6</sup> This issue is discussed in detail in our opinion, "The California ISO's Proxy Demand Resource (PDR) Proposal," April 29, 2009, available at <http://www.aiso.com>

focusing too much on DR programs with unverifiable baselines and reliability-based rather than price-based mechanisms for obtaining consumption reductions.

This crowding out can also occur by inflating the attractiveness to consumers of such DR programs relative to responsive pricing by overpaying for reductions that don't actually occur. Even customers who are fully capable and willing to participate in dynamic pricing programs might prefer to instead participate as DR customers, simply because the baseline problems could work to their advantage. Thus, the current paradigm of demand response if it comes to dominate industry practice could become the single largest barrier to truly price-responsive demand.<sup>7</sup>

---

<sup>7</sup> The recent California ISO report on barriers to demand response (*op. cit.*, Footnote 1) identified “existing programs (being) poorly aligned with CAISO markets” as an important obstacle; we believe that proposed demand response programs perpetuate rather than overcome this barrier. By insulating demand from time- and spatially-varying prices, proposed programs entrench an asymmetric and relatively ineffective treatment of demand and supply.