**Managing Demand-Side Economic and Political Constraints
on Electricity Industry Re-structuring Processes**

by

Frank A. Wolak
Department of Economics
Stanford University
Stanford, CA 94305-6072
e-mail: wolak@zia.stanford.edu
Web-site: http://www.stanford.edu/~wolak
June 2, 2009

Abstract

This paper identifies the major political and economic constraints that impact the demand-side of electricity industry re-structuring processes. It then describes how these constraints have been addressed and how this has harmed market efficiency and system reliability. Finally, the paper proposes demand-side regulatory interventions to manage these constraints in a manner that limits the harm to wholesale market efficiency.

1. Introduction

Economic and political factors constraint electricity industry re-structuring processes. Powerful entities that existed before re-structuring continue to exercise this political clout in the new regime. Former state-owned or privately-owned vertically-integrated geographic monopolies maintain their dominant position in the new regime. Existing regulatory agencies continue to exercise control over market participant behavior even if these actions adversely impact wholesale market efficiency. New market participants find their demands ignored in favor of those by the more politically power incumbent firms. Conflicts between regulatory agencies arise because of the uncertain boundaries of the authority between these agencies brought about by the restructuring process.

The primary economic factor constraining most re-structuring processes is a physical infrastructure poorly suited to the wholesale market environment. A transmission network with insufficient transfer capacity between generation unit owner locations makes it extremely difficult for competition between suppliers to discipline wholesale electricity prices at all locations in the transmission network. The lack of hourly meters on the premises of final electricity consumers prevents retailers from setting default retail prices that vary with hourly wholesale prices. These political and economic constraints on the re-structuring process exert the greatest influence on the demand-side of the market. For example, the motivation often offered for bid caps and other market power mitigation mechanisms is to protect final consumers from an inadequate physical infrastructure to support competitive market outcomes.

This paper identifies the major political and economic constraints that impact the demand-side of electricity industry re-structuring processes. It then describes how these constraints have been addressed in previous re-structuring processes and how this has harmed market efficiency and system reliability. Finally, the paper proposes demand-side regulatory interventions to manage

these constraints in a manner that limits the harm to wholesale market efficiency. Each of the next three sections of the paper is devoted to one of these tasks.

2. Major Political and Economic Demand-Side Constraints

Because electricity is a necessary input to so many economic activities, there are significant political obstacles to charging business and residential users retail prices that reflect the hourly wholesale price of electricity. A long history of retail electricity prices that do not vary with real-time system conditions makes this task even more difficult. Finally, the lack of hourly meters on the customer's premises makes it impossible to determine precisely how much energy each customer withdraws in a given hour. These factors combine to make it virtually impossible to allow retail prices to allocate the available supply to final consumers willing to pay the market-clearing price as is the case for other energy sources such as oil and natural gas.

2.1. The Political Economy of Electricity Prices

Under the vertically-integrated geographic monopoly regime in the United States (US), retail electricity prices are set by state public utilities commissions (PUCs). Although these geographic monopolies are usually privately-owned firms, they are also among the largest employers in the state, so the PUC must balance the interests of ratepayers and employees of the company in the price-setting process. The usual regulatory bargain in the US is that the vertically-integrated monopoly utility must serve all demand at the prices set by the PUC, and the PUC must set retail prices that allow the utility an opportunity to recover all prudently incurred costs incurred to serve that demand.

This regulatory history has established a public precedent for retail electricity prices that only recover total production costs, or prices equal to the long-run average cost of supplying electricity. However, prices set through market mechanisms can often be vastly in excess of or substantially

below the average total cost of supplying the product. This is particularly true for wholesale electricity because of a number of features of the technology of electricity supply discussed in Wolak (2004) that make these markets extremely susceptible to the exercise of unilateral market power by generation unit owners.

Setting retail prices that pass through hourly wholesale prices is even more difficult in the US because there are explicit regulatory prohibitions against consumers paying wholesale prices that reflect the exercise of unilateral market power. As discussed in Wolak (2003b), the Federal Power Act of 1930 requires that the Federal Energy Regulatory Commission (FERC), the United States wholesale market regulator, to ensure that consumers do not pay unjust and unreasonable wholesale prices. FERC has determined that market prices that reflect the exercise of unilateral market power by suppliers are one example of unjust and unreasonable prices.

This problem is further compounded by the fact that most state PUCs have prohibitions against passing on unjust and unreasonable wholesale prices in the retail prices they set. For example, if the FERC determines that certain wholesale prices are unjust and unreasonable because they reflect the exercise of unilateral market power, then it is illegal for the state PUC to set retail prices that recover these costs. Unjust and unreasonable wholesale prices are imprudently incurred costs and therefore the PUC has no obligation to set a retail price that recovers these costs.

This regulatory structure creates an almost impossible situation for introducing an active demand-side into the wholesale market. Requiring consumers to manage hourly wholesale price risk will create the necessary price-responsive final demand that limits the opportunities for suppliers to exercise unilateral market power in the short-term market. This retail pricing policy will also very likely lead to wholesale prices that are vastly in excess of the average cost of supplying electricity during a number of hours of the year, which may mean that consumers are being charged illegal

prices for wholesale electricity during these hours. However, without the active demand-side participation enabled by charging final consumers prices that reflect hourly wholesale prices, electricity suppliers will face a final demand that is virtually inelastic with respect to the hourly wholesale price and implies significant opportunities for suppliers to exercise unilateral market power.

Although this explicit regulatory conflict between retail prices that pass through hourly wholesale prices and the legality charging these prices to consumers does not exist in other countries, the same precedent exists for setting retail electricity prices equal to the average total cost of supply. In the former state-owned monopoly regime a government agency or regulatory body was charged with setting retail electricity prices to allow the firm to recover its production costs. Outside of the US, there was also a strong aversion to retail electricity price volatility. Substantial input cost increases were slowly phased into retail electricity prices.

In developing countries, there is even greater pressure to keep nominal electricity prices as low as possible because of the crucial role electricity is thought to play in the development process. These concerns have often led to retail prices that only recover the variable costs of supplying electricity. In some of these countries, electricity prices are used to pursue political goals. For example, since 1977 politicians in various regions of India have run on a platform of subsidized or even free electricity for farmers (Mukherjee, 2007).

These political constraints emphasize why it is so difficult for the political process to require final consumers to purchase wholesale electricity at prices that reflect hourly wholesale prices. The fact that few consumers have meters on their premises that measure their consumption on an hourly basis ensures that this situation will not change without significant regulatory intervention.

2.2. The Economics of Hourly Metering

Virtually all electricity meters that exist in the United States and other industrialized countries record the total amount of electricity consumed on a continuous basis. A customer's electricity consumption over any time interval is the difference between the value on the meter at the end of the time period and value at the beginning of the time period. In the US, meters are typically read manually on a monthly or bi-monthly basis. A meter reader must show up at the customer's premises and record the value on the meter. If a meter reader is unable to make it to the customer's premises there are rules for determining the customer's consumption during that billing period.

Another feature of electricity retailing in the US is that customers receive their bill for last month's consumption during the current month. With bi-monthly metering and billing, the delay between consumption and invoicing can be more than one month. If the only information a customer receives about the cost of their consumption during the previous billing cycle is provided at the end of this billing cycle plus a processing delay, it is unclear how hourly retail prices that vary with hourly wholesale prices can be used to cause final consumers to alter their demand in real-time. Some signal about the value of the hourly wholesale price must be provided to final consumers to cause them to alter their real-time demand.

Hourly metering technology can both record consumption each hour of the month and provide information to the customer on the value of hourly prices. There are a variety of technologies available to accomplish this, but all of them share similar cost structures. There are significant up-front costs in terms of infrastructure to install the meters and the technology necessary to read and record the output from the meters. In addition, the average cost of installing meters is much less if

they are installed in volume in over a small geographic area. Once installed in volume, the monthly average cost of operating the system is very low, less than $0.50 per meter-month.

Consequently, the tradeoff for an investment in interval metering is whether the cost saving in terms of the reduced labor costs associated with monthly manual meter reading and wholesale energy purchase costs to serve final consumers are sufficient to recover the up-front costs of installing the meters plus the monthly cost of operating the automated meter reading network.

Figure 1 provides a graphical illustration of an automated meter reading network. The meter must automatically communicate either by wire or by wireless technology to the data center each hour of the day to send consumption information back to the data center. From this data center the information is sent to the retailer, who can also share it with the final consumer. For example, virtually all automated meter reading networks have the capability for consumers to download information on their consumption of electricity as soon as it is recorded at the data center.

The major drivers of the economics of installing of an automated meter reading networks are labor costs and the level and volatility in wholesale electricity prices. In regions where labor costs are higher, the cost savings from eliminating manual meter reading are larger. In areas with higher and more volatile electricity prices, the cost savings in wholesale electricity purchase costs from being able to use price signals to shift demand throughout the day, week, or month are much greater than in a region with low and/or stable wholesale electricity prices. Consequently, a fossil fuel-based system which has substantial price fluctuations within the day, has a much greater potential to realize significant cost savings from an automated meter reading network than a hydro-based system which typically has fairly constant wholesale prices throughout the day.

Wolak (1999) compares the time series behavior of prices in re-structured electricity markets in Australia, New Zealand, England and Wales and Norway and Sweden. Australia and England and

Wales are fossil fuel-based systems with substantial amounts of price variation within the day, whereas New Zealand and Norway and Sweden are hydro-based industries with small price fluctuations within the day-ahead, although there can be substantial price differences across seasons of the year, depending on hydrological conditions. By the above logic, the wholesale electricity purchase savings from an automated meter reading network in Australia and England and Wales should be significantly higher than those in New Zealand or Norway and Sweden.

It is important to emphasize that this statement does not imply that New Zealand and Norway and Sweden would not benefit from retail prices that pass through wholesale prices. Because the major source of price variation in these markets is across seasons of the year or across years, virtually all of the savings from passing through wholesale prices in retail prices could be accomplished with monthly meter reading.

Customers with larger annual electricity bills can expect to realize greater total benefits from hourly meters than smaller customers. Any given percentage saving in wholesale electricity purchase costs from having an hourly meter will translate into a larger total annual dollar savings, which increases the likelihood that the annual total benefits of an hourly meter for that customer will exceed the annual cost. For example, a five percent savings in wholesale electricity purchase costs applied to an annual electricity bill of $1000 only yields $50 in saving. Applying this same percentage to an annual electricity bill of $10,000, yields $500 in saving, which more than covers the annual cost (including fixed costs) of installing and operating an hourly meter.

Another factor determining the benefits a customer might realize from hourly metering is the amount that customer can reduce its demand in response to price signals. The magnitude of a customer's demand responsiveness depends on the mechanism used by the retailer to deliver the price

signal. There has been a substantial amount of recent research on ways to deliver wholesale price signals to final consumers to maximize the cost savings realized from providing these price signals.

Simply passing through the hourly wholesale price in an hourly retail price may not provide the greatest aggregate reduction in wholesale purchase cost savings by customers with hourly meters. By coordinating the demand reduction efforts of consumers with hourly meters it is possible to realize an additional source of benefits from price-responsive final demand besides those that result from customers shifting consumption to lower-priced hours of the day from the high-priced hours of the day. Coordinated actions in the same time interval to reduce demand by all consumers with hourly meters can reduce total system demand which can then lead to lower wholesale prices. These coordinated actions increase the total benefits realized from hourly meters because customers without hourly meters benefit from lower wholesale electricity purchase costs enabled by a lower wholesale electricity price. Section 3 discusses alternate hourly pricing mechanisms that attempt to capture both the load shifting benefits and the wholesale price-reducing benefits of hourly metering of final consumers.

The costs of hourly metering and the magnitude of the typical annual household electricity bill in most developed countries make it difficult for the expected benefits from universal hourly metering not to exceed the costs, particularly given the significant economies to scale and geographic economies to scope in the installation and operation of hourly meters. This economic logic is consistent with recent regulatory decisions made in a number of jurisdictions. The state of Victoria in Australia, the province of Ontario in Canada and the state of California in the US have all decided to implement universal hourly metering for all consumers. In Victoria, the plan is to install approximately 2.5 million hourly meters by 2013. In Ontario, 5 million meters are to be installed by 2010. In California, no firm deadlines have been set, but the plan is to install hourly meters for all

customers of the three large investor-owned utilities. In all these jurisdictions, the cost of the hourly metering technology will be included in the regulated cost of local distribution services.

For developing countries, the economic case for universal hourly metering is much less favorable because the labor costs associated with manual meter reading are much lower and annual residential and business electricity bills tend to be much lower. However, in the areas where affluent households live there are residential customers with sufficiently large annual electricity bills to pass the cost/benefit test for hourly metering. There are also likely to be many industrial and commercial customers that are viable candidates for hourly meters. Consequently, determining precisely where to draw the line between customers that pass the net benefit test for hourly meters and those that do not is much more difficult in developing countries. This logic also suggests an alternative approach to organizing the retailing segment of the industry in developing versus industrialized countries discussed in Section 4.

2.3. The Political Economy Case Against Real-Time Pricing

The need for explicit regulatory intervention to install the metering infrastructure necessary for the widespread implementation retail pricing plans that reflect real-time wholesale prices has significantly slowed the pace of adoption of these technologies. Particularly in the US, the regulatory framework governing the electricity industry re-structuring process has further conspired against the adoption of interval meters and real-time pricing plans. Because the cost of conventional manual metering reading is sufficiently high that in places like California, the case for the adoption of automated meter reading technology can be largely made using the costs savings associated with the elimination of manual meter reading.

Despite the adoption of automated meter reading technology, most state PUCs have been extremely reluctant to implement retail pricing plans that reflect hourly wholesale market conditions

because of the apparent contradiction with their regulatory goals. Mandating that all customers face an hourly retail price that passes through the hourly wholesale price would create strong incentives for final consumers to respond to hourly wholesale prices. A real-time demand for electricity that responds to hourly wholesale electricity prices is crucial to limiting the opportunities for suppliers to exercise unilateral market power in the wholesale market and ensuring competitive wholesale market outcomes. However, such a requirement would also expose final customers to prices that reflect the exercise of significant unilateral market power during a number of hours of the year. For this reason, a state PUC might argue that setting a pass through of the hourly wholesale price as the default retail price is inconsistent with its regulatory mandate to protect consumers from unjust and unreasonable retail prices.

The fact that this very straightforward solution to the lack of a price elastic wholesale electricity demand has been rejected by all US state PUCs, suggests these entities view a default retail price that passes through hourly wholesale prices in hourly retail prices as explicitly or implicitly inconsistent with their regulatory mandate. As we discuss in the next section, the response of state and federal regulators in the US to the risk of very high hourly wholesale prices has been to implement regulatory interventions that limit price volatility but very likely increase average wholesale prices and reduce system reliability.

The situation in wholesale markets in other industrialized countries is not much better. There is very little penetration of hourly metering technology in most of these markets because of the reluctance of the regulators to mandate its adoption. As noted above, recently this trend has begun to reverse, but it remains to be seen if once these meters are in place default prices that pass through hourly wholesale prices will be adopted.

3. Mechanisms Adopted to Deal With Constraints on Demand-Side Participation

The desire of policymakers to shield final consumers from wholesale price risk has led to a number of regulatory interventions that significantly degrade the efficiency of the re-structuring industry in the short-term and long-term. The first is the implementation of the bid caps and other market power mitigation mechanisms in the short-term market. This has led retailers and final consumers to engage in an inadequate amount of hedging of short-term price risk and claims by generation unit owners that the existence of bid caps and other market power mitigation mechanisms prevent them from full revenue recovery. These claims have led to a number of regulatory interventions that provide additional revenue to generation unit owners. However, these revenue flows also raise total wholesale energy costs to consumers and decrease the likelihood they will receive any benefits from industry re-structuring.

The one bright spot in these regulatory adaptions to political and economic constraints on demand-side participant is experimental evidence on real-time pricing programs that final electricity consumers find politically palatable. As discussed in Section 3.4, these programs share the real-time price risk between electricity retailers and final consumers in a manner that can share the benefits between the two parties.

3.1. Offer Caps and Market Power Mitigation Mechanisms

Virtually all short-term wholesale markets operating in the world have offer caps that limit highest price offer a supplier can submit or a price cap that limits the value the market clearing price can take on. In the US, FERC has set the maximum offer cap in the eastern US markets at $1,000/MWh. In California, the current offer cap is $400/MWh, but it is scheduled to increase to $1,000/MWh to be consistent with the offer caps in the eastern US markets. All US markets also

have local market power mitigation mechanisms that limit the maximum bid a generation unit owner can submit when it is determined to possess local market power.

Wholesale markets in other parts of the world also have offer caps. For example, the Australian market currently has an offer cap of 10,000 Australian dollars per MWh. The Alberta electricity market currently has a price cap equal to 1,000 Canadian dollars per MWh. In the Nord Pool, the market operator sets a maximum and minimum offer price each day as part of the operation of the day-ahead energy market.

These bid caps and price caps and local market power mitigation mechanisms are proposed to address the fact that the wholesale demand for electricity is completely price inelastic because of the lack of hourly meters and default retail prices that pass through hourly wholesale prices. One of the standard tests to determine whether a supplier possesses local market power worthy of mitigation in US wholesale markets is whether that supplier or a small number of suppliers is pivotal to meet demand in a congested portion of the transmission network. A supplier or group of suppliers is pivotal if removing their supply implies that demand could not be met by the remaining suppliers alone. For example, with five firms each owning 100 MW of capacity, if the demand for electricity is above 400 MW, then each of the five suppliers is pivotal. If the demand is not completely price inelastic then no supplier could be pivotal, because there would always be a price at which the demand would equal the available supply.

There would be significantly less need for bid caps and local market power mitigation mechanisms if final consumers were required to manage short-term wholesale price risk. If all final consumers had hourly meters and were required to pay the hourly wholesale price as part of their default hourly retail price, consumers would sign fixed-price forward contracts for their essential demand so that they could consume this quantity of electricity each hour regardless of the hourly

price. These consumers could then to alter their hourly demand around this contracted essential demand in response to hourly price signals. In this way, the need for bid caps and other market power mitigation mechanisms would be significantly reduced.

3.2. Inadequate Hedging of Short-Term Prices and the Reliability Externality

These offer caps and market power mitigation mechanisms create incentives for market participant behavior that can significantly degrade market efficiency and system reliability. Offer caps limit the potential downside to electricity retailers and large consumers (able to purchase from the short-term market) delaying their purchases of electricity until real time operation. These offer caps also create the possibility that real-time system conditions can occur where the amount of demanded at or below the offer cap is less than the amount suppliers are willing to offer at or below the offer cap. This outcome implies that the system operator must be forced to either abandon the market mechanism or curtail load until the available supply offered at or below the offer cap equals the reduced level of demand.   Because random curtailments are used to make demand equal to the available supply at or below the bid cap, this mechanism creates a reliability externality that further increases the incentive of retailers to rely short-term market purchases.

Particularly for markets with very low offer caps, retailers have little incentive to engage in sufficient fixed-price forward contracts with generation unit owners to ensure a reliable supply of electricity for all possible realizations of demand. For example, a 200 MW generation unit owner that expects to run 100 hours during the year with a variable cost of $80/MWh would be willing to sign a fixed-price forward contract to provide up to 200 MWh of energy for up to 100 hours of the year to a retailer. Because this generation unit owner is essentially selling its expected annual output to the retailer, it would want a $/MWh price that at least exceeds its average total cost of supplying energy during that year. This price can be significantly above the average price in the short-term

wholesale market during the hours that this generation unit operates because of the offer cap on the short-term market and other market power mitigation mechanisms. This fact implies that the retailer would find it expected profit-maximizing not to sign the forward contract that allows the generation unit owner full cost recovery but instead wait until the short-term market to purchase the necessary energy at prices that are the result of offer caps and the market power mitigation mechanism.

Although this incentive for retailers to rely on a mitigated the short-term market is most likely to impact generation units that run infrequently, if the level of demand relative to the amount of available supply is sufficiently large, it can even impact intermediate and baseload units. Because of the expectation of very low prices in the short-term market and the limited prospect of very high prices because of offer caps or market power mitigation mechanisms, retailers may decide not to sign fixed-price forward contracts with these generation unit owners and purchase their energy in the short-term market. By this logic, a mitigated short-term energy market always creates an incentive for retailers to delay purchasing some of their energy needs until real-time, when the market power mitigation mechanisms on the short-term market can be used to obtain this energy at a lower price than the supplier would willingly sell it in the forward market.

The lower the offer cap and the more stringent the market power mitigation mechanisms are, the greater is the likelihood that the retailer will delay their electricity purchases to the short-term market. Delaying more purchases to the short-term market increases the likelihood of the event that insufficient supply will bid into the short-term market at or below the offer cap to meet demand. Because of the lack of hourly metering, there is no way to determine precisely how much electricity each customer is consuming during these time periods. For this reason, system operators manage these shortfalls by curtailing sufficient load to allow the available supply meet the remaining demand.

If retailers know this is how supply shortfalls in the short-term market will be met, this creates an additional incentive for them to rely on the short-term market.

If a retailer knows that part of the cost of its failure to purchase sufficient fixed-price forward contracts will be borne by other retailers and large consumers, then it has an incentive to engage in less fixed-price forward contracts than it would in a world where all customers had hourly meters and all customers could be charged hourly prices high enough to cause them to reduce their demand to equal the amount of supply available at that price. As discussed in Wolak (2003), all of the wholesale markets in Latin American recognize this incentive to purchase too much energy in the short-term wholesale market when it is subject to offer caps or other market power mitigation mechanisms. These countries address this incentive to under contract by mandating forward contract coverage ratios for retailers and large consumers that have the option to purchase from the short-term market. For example, in the Brazilian market all retailers and large consumers are required to have 100% of their final demand covered in a fixed-price forward contract.

Without these forward contracting requirements on retailers and large consumers, a wholesale market with offer caps and stringent market power mitigation mechanisms and final consumers without hourly meters face significant reliability challenges in both the short-term and long-term. In the short-term market the lower the bid caps and more stringent the market power mitigation mechanism the greater the likelihood of that there will be insufficient supply offered into the short-term market at or below the offer cap to meet demand. Because of the mitigated short-term market and inadequate fixed-price forward contracting by retailers, there is a likelihood that new generation entrants will be unable to earn sufficient revenues from the selling in the short-term market and therefore unwilling to construct new generation units to serve load growth, which increases the likelihood of future supply shortfalls.

3.3. Capacity Markets and Other "Cures"

A number of "remedies" have been proposed for bid caps and market power mitigation mechanisms necessitated by the lack of hourly metering and the pass-through of hourly wholesale prices in the default retail prices. Capacity payment mechanisms are the most common. The major rationale for capacity markets in the US appears to be a holdover from the vertically-integrated regulated regime when capacity payments compensated generation units for their capital costs, because the regulatory process typically reimbursed unit owners for their variable operating costs.

It is important to emphasize that in a wholesale market regime all generation unit owners have the opportunity to earn the market-clearing price which is typically above a generation unit's average variable cost when the unit is operating. In this way, the generation unit earns a return to capital during each hour it produces electricity. This paradigm for earning a return on capital from the difference between the market price and the firm's average variable cost of production has managed to provide the appropriate incentives for investment in new productive capacity all workably competitive industries. There is little reason to expect that it could not work in the wholesale electricity industry with an active demand side.

Capacity payments typically involve a dollar per kilowatt year ($/kW-year) payment to individual generation units based on some measure of the average amount of their capacity available to produce electricity within the year. For example, a baseload coal-fired unit would have a capacity value very close to its nameplate capacity, whereas wind generation facility would have a capacity value significantly below its nameplate capacity.

Capacity payment mechanisms differ along a number of dimensions. In some regions, the payment is made to all generation unit owners regardless of how much total generation capacity is needed to operate the system. In other regions, the independent system operator (ISO) specifies a

system-wide demand for capacity equal to peak system demand plus some planning reserve, typically between 15 to 20 percent, and only makes capacity payments to enough generation units to meet this demand.

There have been attempts to use market mechanisms to set the value of the $/kW-year payment to the generation units needed to meet the total demand for capacity. However, these market mechanisms have been largely unsuccessful because they are extremely susceptible to the exercise of unilateral market power, because of the pivotal supplier problem created by the inelastic demand for capacity. In the eastern US markets, there have been numerous instances of the exercise of the enormous market power in these capacity markets. During the off-peak months of the year when no single supplier is pivotal in the capacity market, the price of paid for capacity is very close to zero, which is the marginal cost of a supplier providing an additional megawatt (MW) of available capacity from existing generation capacity. During the peak and shoulder months when one or more suppliers are pivotal in the capacity market, there is no limit on the price a supplier can charge. For example, suppose a market has 10 suppliers each of which owns 1200 MW and the peak demand for the system during the peak month is 10,000 MW. Under these circumstances all suppliers know that the aggregate available capacity requirement of say 11,500 MW (=1.15 x 10,000 MW) cannot be met without some of their capacity. As consequence in all of the Eastern US markets, very stringent market power mitigation measures have had to be put in place. Consequently, capacity prices typically fluctuate from very close to zero to the regulatory price cap. It is difficult to see how these very volatile prices provide very useful signals about the need for new investment in generation capacity.

This market power problem leaves open the question of how to determine the value of the $/kW-year capacity payment. In most regions, the value of the capacity payment is based on the

regulator's estimate of annual $/kW fixed cost of a peaking generation unit. This is backed by the logic that because of the offer cap on the short-term market and other market power mitigation mechanisms this peaking unit could only set a price slightly higher than its variable operating costs. Because this generation unit and all other generation units are missing the hours when the market price would rise above its variable operating costs because a price-responsive final demand would set the market price, the annual $/kW cost of the peaking unit is needed to compensate all generation units for the revenues they do not receive because of the offer cap and market power mitigation mechanisms. This logic for the value of $/kW-yr capacity payment explicitly assumes that the real-time demand for electricity is completely price inelastic and that suppliers are unable to exercise significant amounts of unilateral market power in the short-term market. Both of these assumptions are clearly false.

In addition, it is unclear why electricity is so fundamentally different from other products that it requires paying suppliers for their generation units to exist. Consumers want cars, not automobile assembly plants; point-to-point air travel, not airplanes; and a loaf of bread, not a bakery. In these markets producers do not receive capacity payments for owning the facilities needed to provide these products. All of these industries are also high fixed cost, relatively low marginal cost production processes, yet all of these firms earn their return on capital invested by selling the good that consumers want at a price above the variable cost of producing it. Cars, air travel, and bread are in many way essential commodities, yet capacity payments are not needed to ensure that there is sufficient productive capacity for these products to meet our needs.

Capacity payment mechanisms virtually guarantee that consumers will pay more for electricity their annual electricity consumption than they would in a world with active demand-side participation in the wholesale market. Recall that the capacity payment is made to either all

generation units in the system or all generation units needed to meet the ISO's demand for capacity. On top of this, all suppliers receive a market-clearing price set by the highest generation offer needed to meet system demand. Thus to the extent that suppliers are able to exercise unilateral market power in the short-term market, they can raise energy prices significantly above the variable cost of the highest cost unit operating within the hour for all hours of the year.

For a number of reasons, a wholesale market with a capacity payment mechanism makes it more likely that suppliers will be able to exercise unilateral market power in the short-term wholesale market relative to a market with active demand-side participation and no capacity payment mechanism. Capacity payment mechanisms are typically accompanied by offer caps and market power mitigation mechanisms that significantly limit the incentive for final consumers to become active participants in the short-term wholesale market. For example, if the maximum wholesale price in an hour in $400/MWh because of an offer cap at this level, then a 1 KWh reduction in demand for a residential customer (a very large demand reduction) during an hour only saves the customer 40 cents, which seems unlikely to be sufficiently attractive to cause that consumer to reduce its demand. This lack of an active demand-side of the wholesale market impacts how generation unit owners offer their generation units into the wholesale market. Active participation by final demand substantially increases the competitiveness of the short-term wholesale market because all suppliers know that higher offer prices will result in less of their generation capacity being called upon to produce because the offers of final consumers to reduce their demand are accepted instead. Without an active demand-side of the wholesale market suppliers know that they can submit offers that are farther above their variable cost of supplying electricity and not have these offers rejected. Therefore, a market with a capacity payment mechanism charges consumers for the $/kW-year fixed cost of a

peaker unit for their entire capacity needs and then give suppliers greater opportunities to exercise unilateral market power in the short-term market.

Another argument given for capacity payments is that they reduce the likelihood of long-term capacity inadequacy problems because of the promise of a capacity payment provides incentives for new generation units to enter the market. However, until very recently capacity payments in most markets around the world were only promised for at most a single year and only paid to existing generation units. Both these features substantially dulled the incentive for new generation units to enter the market, because the unit that entered often had no guarantee of receiving the capacity payment for one year and no guarantee that if it received it the first year it would continue to receive it. This has led the eastern US ISOs to focus on the development a long-term capacity product that is sold two to three years in advance of delivery to provide the lead time for new generation units to participate. As we discuss in Section 4, this solution unlikely to lead to a lower cost solution for consumers than the long-term contract adequacy approach described in that section.

3.4. Politically Palatable Real-Time Pricing

One benefit of the political and economic constraints associated with implementing an active demand-side in wholesale markets in the US is that there have been a number of experiments to determine the real-time price-responsiveness of retail electricity consumers. These experiments typically install hourly metering on a sample of customers and require a fraction of these customers to pay retail prices that vary with hourly system conditions and the remainder to pay according the standard retail price schedule.

These experiments have been run in a number of jurisdictions and found statistically significant evidence that retail customers are able to substantially alter their consumption of

electricity in response to hourly retail prices. Although these results are not surprising, the more surprising conclusion from this research is that how real-time wholesale price signals are provided to final consumers can impact the magnitude of the price response.

Regulators and many final consumers often argue that responding to hourly price signals would be too complex and time-consuming for most retail customers. Customers would have to continually monitor the price of electricity each hour of the day to determine whether it makes economic sense to alter their consumption. In addition, hourly real-time electricity prices can be extremely volatile and customers are likely to find it difficult to determine how long price spikes are likely to last and whether it is worth taking actions to reduce their consumption in response to a very high price during a single hour. For example, an electricity intensive industrial customer, may only be able to reduce their consumption significantly by shutting down an entire 8-hour production shift. This customer would need to have an intimate understanding of the time series behavior of real-time electricity prices to be able to make an informed decision about whether it makes economic sense to shut down production for an entire 8-hour shift in response to high prices in a single hour. Patrick and Wolak (1997) study responsiveness of large industrial and commercial customers in England and Wales to retail prices that pass-through half-hourly wholesale prices and find significant diversity in the magnitude and pattern of the demand responses with the day. All of these customers have extremely large monthly wholesale electricity bills, in the thousands of dollars, so they have a strong financial incentive to invest in the expertise needed to respond to half-hourly wholesale prices.

For smaller customers with the flexibility to reduce their consumption during a few hours of the day, it may not make sense to do so in response to every hour with a high real-time price. There are 744 hours (= 24 hour/day x 31 days) in a month. If there is a fixed-cost to taking action to reduce

demand in an hour or number of hours of the day, the price increase with a single hour of the month must be very high to cause the customer to take action. Attaching some numbers to this calculation is helpful. Suppose for simplicity that the household's average hourly consumption is 2.5 KWh (roughly equal to the average hourly consumption of a California household) in all 744 hours of the month and the average wholesale price is $0.05/KWh. Suppose the household is able to reduce its consumption by 0.5 KWh by taking the actions that cost $5 in psychic or actual costs. The wholesale price in this hour would have to rise to at least $100/KWh = ($5 cost of taking action)/(0.5 KWh saving) for the customer to find the cost of taking action for that single hour to exceed the benefit. This $100/KWh price translates into $100,000/MWh, which is vastly in excess of the bid cap on any market currently operating in the world and is 100 times higher than the offer cap on the wholesale market in the eastern US.

This calculation illustrates a very important point about nature of hourly price spikes required for small electricity consumers with relatively small fixed-costs of taking actions to reduce their consumption in response to a price spike in a single hour of the day. If this fixed cost of taking action is $1, then the required wholesale price to take action falls to $5,000/MWh, which is still five times the offer cap on all of the eastern US markets.

Another factor that can reduce the magnitude of the wholesale price spike required to cause customers to take action is the duration of the price spike. In our simple example, the longer the customer expects the price spike to last the greater the likelihood the customer will take action, because once the customer pays the cost to act it can reduce its consumption by 0.5 KWh for as many hours as it would like. Therefore, at the $5 cost to take action, a two-hour price spike would only need the price to average $50,000/MWh or more.

This cost of taking action to reduce hourly electricity demand expressed as a fraction of the customer's typical monthly electricity bill is likely to be smaller the larger is the customer's monthly bill. For example, a large industrial user with an hourly consumption of 1 MWh is likely to have many ways to reduce its hourly consumption by 5 percent that cost significantly less than $50 to implement per event. This 0.05 MWh reduction in consumption only requires a wholesale price of $1000/MWh or more to yield energy savings sufficient to justify the $50 cost of taking actions to reduce demand within the hour.   This logic implies that larger customers need a smaller price spike to find it profitable to take actions to reduce their hourly demand for electricity.

If there are offer caps and other market power mitigation mechanisms that limit the level of wholesale electricity prices, other mechanisms for passing through real-time price signals must be devised to reduce the cost of customers responding to real-time prices or increase the benefits they receive from responding. This logic has led to the design of critical peak pricing (CPP) programs that share the risk of responding to real-time prices between retailers and final consumers in order to both reduce the customer's cost of responding and the benefits it expects to receive from responding.

Under this sort of real-time pricing program customers pay according to a single fixed-price or an increasing block tariff during the month with a fixed price for each block of the household's monthly consumption.   The retailer is then allowed to call a certain number of critical peak days within a given time interval. Typically, this is done the day before by a telephone or e-mail, but the program could be modified to notify the customer closer to the time of CPP event. During an agreed-upon peak period of a CPP day the customer must pay a substantially higher price. For example, if the customer normally pays 8 cents/KWh for energy, during the peak period of a CPP day it would

pay 35 cents/KWh. This mechanism does not require the final consumer to follow the hourly wholesale price or know anything about wholesale market conditions. The retailer declares CPP events on the days that it would like customers to reduce their consumption. Another benefit of the CPP program is that the peak period of the day during which a CPP customer pays the higher retail price is typically between four to six hours long. This implies a longer period over which a CPP customer has to accrue benefits by reducing its consumption.

If the retailer has enough customers on the CPP pricing program, then the structure of the program causes all CPP customers to focus their demand-reduction efforts during the same time period, which increases the likelihood that declaring a CPP event will result in lower wholesale prices during the CPP period because of the reduced system-wide demand for electricity. This further increases the benefits realized from implementing real-time pricing because it reduces the cost to the retailer of serving its remaining customers.

One variation on the standard critical peak-pricing program that is very popular with customers involves a rebate for consumption reductions relative to a reference level on critical peak days. Under this scheme the customer is paid a $/KWh rebate for every KWh of consumption less than some reference level during critical peak periods instead paying for all consumption at the higher price. For example, if a customer's peak period reference level is 8 KWh and the customer consumes 6 KWh, then it is paid the $/KWh rebate for 2 KWh. If the customer does not reduce its consumption below this reference level then it does not receive any rebate and its does not have to make any payment. Mathematically, the payment received by the customer during CPP days is $p_{rebate}*\max(0, q_{ref} - q_{actual})$, where p_{rebate} is the $/KWh rebate, q_{ref} is the reference level for rebates, and q_{actual} is the customer's actual consumption during the peak period.

This mechanism implies greater risk for the retailer because it could pay out more in rebates than it saves in wholesale energy purchase costs. This real-time pricing program is more attractive to customers than the conventional CPP program because the customer cannot lose from participating the program. At worst, the customer does not receive any rebate payments.

Wolak (2006a) analyzes household-level price responsiveness under a CPP program with a rebate for the City of Anaheim in southern California. This program paid customers a $0.35/KWh rebate for reductions in consumption relative to their reference level during peak period of CPP days. During all other hours, the customer pays a price of 6.75 cents/KWh for monthly consumption less than 240 KWh and 11.02 cents/KWh for monthly consumption above 240 KWh. The peak period of the day for the purposes of the Anaheim CPP mechanism with a rebate is noon to 6 pm. Wolak (2006a) found that during CPP days the mean difference in the difference in consumption between the CPP customers and the control group of customers is a reduction of approximately 13 percent. If this mean consumption reduction associated with a CPP event could be scaled to all residential consumers in California, approximately one-third of the consumption in California, this would imply slightly more than a 4 percent reduction in system demand as a result of a CPP event. Applying this to a peak demand in California of 50,000 MWh implies a 2,000 MWh reduction in demand, which means that California can avoid building and paying for almost 2,000 MW of new generation capacity as a result of this demand response capability.

The magnitude of the response to a critical peak day estimated in Wolak (2006a) is likely to underestimate the potential demand reduction possible, because of a number of new technologies to monitor and control electricity consumption automatically. There are a number of standards for allowing advanced meters to communicate with appliances throughout a geographic area using both

wireless and wireline technologies. For example, a household could program a personal computer to alter electricity use based on wholesale prices or other signals provided by the retailer. The ZigBee Alliance (www.zigbee.org) is perhaps the most popular of these standards. It is a wireless network designed to monitor and control appliances and was organized as a nonprofit corporation in 2002. A number of companies are offering appliance control networks that are compliant with the ZigBee standard. Homeplug Powerline Alliance is powerline-based open standard for communications (www.homeplug.org) aimed at providing, among other services, monitoring and control of appliances. These technologies are likely to reduce overall electricity consumption as well as reduce the cost of responding to real-time price signals and the magnitude of the demand response.

The development of politically attractive real-time pricing plans and technologies that reduce the cost and increase the magnitude of demand response strongly argues in favor of introducing mechanisms that require final consumers to manage real-time price risk. The nontrivial cost of hourly meters and the technologies to reduce the cost of demand response favor a phased-in approach that focuses on customers realizing the greatest net benefits from these technologies and respects the political constraints facing regulators and policymakers in allowing active demand-side participation in wholesale electricity markets.

4. Managing Demand-Side Economic and Political Constraints

This section proposes a retail market regulatory structure that addresses the economic and political constraints described in Section 2 with minimal harm to wholesale market efficiency and system reliability. This retail market structure emphasizes the necessity of hedging short-term wholesale price risk either through fixed-price forward contracts or active demand-side participation to ensure a reliable supply of electricity and the long-term financial viability of the industry. Another

guiding principle is symmetric treatment of generation unit owners and final consumers in the sense that both sets of market participants face a default price that reflects all real-time price risk. Finally, this regulatory structure recognizes that hourly meters may not make economic sense for all retail customers at the present time, but these circumstances may change in the future as the price of electricity rises and the cost of hourly meters falls.

4.1. Hedging Short-Term Wholesale Price Risk

There are two types of wholesale price risk that can harm electricity consumers. The first is prices persistently above competitive levels. This pattern of wholesale prices is typically the result of suppliers exercising unilateral market power in the short-term market by withholding output. The second is a short duration of very high prices usually accompanied by stressed system conditions because of a generation unit or transmission line outage or an extreme unexpected weather event. Each form of wholesale price risk is best dealt with using a different set of actions by final consumers.

The risk of short-term prices persistently above competitive levels is best managed with fixed-price forward contracts between generation unit owners and retailers or large consumers able to purchase directly from the wholesale short-term market. As discussed in detail in Wolak (2000), fixed-price forward contract commitments by generation unit owners reduce their incentive to exercise unilateral market power in the short-term energy market because the supplier only earns the short-term price on any energy it sells in excess of its forward contract commitment.

To understand this logic, let pc equal the forward contract price at which the supplier agrees to sell energy to an electricity retailer and qc equal to the quantity of energy sold. This contract is negotiated in advance of the date that the generation unit owner will supply the energy, so that the

value of pc and qc are pre-determined from the perspective of the supplier's behavior in a short-term wholesale market. The quantity of fixed-price forward contract obligations held by the supplier impact what price the firm finds profit-maximizing given its marginal cost of producing energy, the supply offers of its competitors, and the level of aggregate demand. Incorporating the payment stream a generation unit owner receives from its forward contract obligations, its variable profit function for a given hour of the day is:

$$\pi(p) = (pc - c)qc + (qs - qc)(ps - c), \tag{4.1}$$

where qs is the quantity of energy produced by the generation unit owner, ps is the price of energy sold in the short-term market and c is the supplier's marginal cost of producing electricity. The first term in (4.1) is the variable profit from the forward contract sales and the second term is the additional profit or loss from selling more or less energy in the short-term market than the supplier's forward contract quantity. Because the forward contract price and quantity are negotiated in advance of the delivery date, the first term is a fixed profit stream to the supplier from the perspective of its participation in the day-ahead market. The second term depends on the price in the short-term market, but in a way that can significantly limit the incentive for the supplier to raise prices in the short-term market.

For example, if the supplier is too aggressive in its attempts to raise prices by withholding output, it could end up selling less in the short-term market and than its forward contract quantity, and if the resulting market-clearing price is greater than the firm's marginal cost, c, the second term in the firm's variable profit function will be negative. Consequently, only in the case that the supplier is confident it will produce more than its forward contract quantity in the short-term market does it have an incentive to withhold output in order to raise short-term prices.

The quantity of forward contract obligations held by a firm's competitors also limits incentive of that supplier to exercise unilateral market power in the short-term market. If a supplier knows that all of its competitors have substantial fixed-price forward contract obligations, then this supplier knows these firms will be bidding very aggressively to sell their output in the short-term wholesale market. Therefore, attempts by this supplier to raise prices in the short-term market by withholding output are likely to be unsuccessful because of the aggressiveness of the offers into the short-term market by its competitors with substantial fixed-price forward contract obligations.

Short periods of extremely high prices are best managed through active demand-side participation in the wholesale market, because many of these price spikes are driven by unexpected events that occur too quickly for the supply side of the market to respond to. The outage of a large generation unit can often be managed by the generation units providing operating reserves increasing their output. However, the outages are sometimes severe enough that the only way to manage them is to reduce the demand electricity.

Although it is possible to manage the risk of the exercise of unilateral market power in the short-term market with demand response alone, this could impose significant hardship on consumers. For example, in a hydro-dominated system where water comes primarily in the form of winter snow-pack, if the amount of water available to produce electricity is much less than normal, then the fossil-fuel suppliers will have a greater opportunity to exercise unilateral market power until the following year. As discussed in Wolak (2003b), this describes the initial conditions in the western US immediately before the start of the summer of 2000. To limit the ability of suppliers to exercise unilateral market power under these system conditions, consumers would likely have to reduce their demands for long periods of time period until the next year's snowfall melted, which could impose

significant hardship on electricity consumers.   Consequently, a strategy that involves a lower downside to consumers would be to hedge their expected demand for electricity each period in fixed-price long-term contracts. That way if low hydro conditions arise the fossil fuel suppliers will have less of an incentive to exercise unilateral market power in the short-term wholesale market because of their substantial fixed-price forward contract obligations.

Fixed-price long-term contracts can be used to protect consumers against short-term price spikes, but this is likely to be more expensive for consumers than managing this risk with active demand-side participation in the wholesale market.   To hedge against the risk of price spikes, consumers or their retailers would have to purchase fixed-price forward contract coverage for 100% of their demand requirements.   Because the realized demand for electricity is unknown at the time a retailer signs the fixed-price forward contracts, it would have to purchase more forward contracts for more than 100% of its expected demand. This implies that during many hours, the retailer would be selling back energy purchased in the forward contract at a low spot market price because its actual demand is less than the amount it purchased in the forward contract. This further increases the effective price consumers pay for the electricity. A numerical example helps to illustrate this point. Suppose the distribution of the retailer's demand has a mean of100 MWh and a standard deviation of 20 MWh. For this reason, the customer purchases 130 MWh in a fixed price forward contract at a price of $50/MWh, to guard against the risk of paying very high spot prices if its demand is unexpectedly high. If a retailer's realized demand is 100 MWh and the real-time price is $20/MWh, then the retailer makes a loss of $900 by selling the 30 MWh it bought for $50/MWh at a price of $20/MWh.   This implies an effective price for the 100 MWh consumed of $59/Mwh = ($50/MWh*100 MWh + $900)/100 MWh, almost a 20% price increase. A lower cost strategy for

the retailer is simply to purchase the expected demand of 100 MWh in the forward market and manage the remaining short-term price risk by altering the demand of its customers in response to real-time prices.

4.2. Contract Adequacy in Wholesale Electricity Markets

Adequate fixed-price forward contracting by electricity retailers and large customers able to purchase from the short-term wholesale market is a necessary condition for both competitive short-term market outcomes and adequate generation capacity to meet future demand. These fixed-price forward contracts must be negotiated far enough in advance of delivery for all possible sources of supply to compete. Signing a fixed-price forward contract a day, month, or even a year ahead of delivery can limit the number of suppliers and modes of supply that are able to provide this energy. For example, a contract negotiated one day in advance limits the sources of supply to existing generation unit owners able to produce energy the next day. Even a year in advance limits the sources that can compete to existing generation unit owners, because it takes longer than a year to site and build a substantial new generation unit. To obtain the most competitive prices, at a minimum, the vast majority of the fixed-price forward contracts should be negotiated far enough in advance of delivery to allow new entrants to compete with existing suppliers.

Regulators should focus on ensuring contract adequacy, not on generation adequacy. Specifically, retailers and large consumers should have adequate fixed-price forward contract coverage for their expected future demand signed far enough in advance of delivery to obtain the most competitive prices. By purchasing a hedge against the spot price risk at the locations in the network where the retailer or large consumer withdraws energy, the buyer can rely on the financial incentives that the seller faces to provide the contracted for energy at least cost.

A major mistake made by the California Department of Water Resources (CDWR) in negotiating the forward contracts signed by the State of California during the winter and spring of 2001 is that it focused on purchasing power plants instead of hedges against the spot price of energy at the locations where the three large electricity retailers withdrew energy from the transmission network. This procurement strategy created a number of market inefficiencies that significantly increased the cost of these forward contracts and prices in the wholesale market, because they often called for more expensive generation units to operate (than those required for a least-cost dispatch of California's generation resources) in order for the seller's contractual obligations to met.

By focusing on contract adequacy rather than building generation facilities, California would have had a portfolio of forward contracts that provided incentives for least cost production of electricity in the short and long term. Firms that sold these forward financial contracts would have strong incentives to ensure that the spot prices at the locations in the California ISO control area where these contracts clear are as low as possible. That is because as equation (4.1) demonstrates, once a supplier has signed a fixed-price a forward contract that clears against the spot price at a given location in the network, the supplier's revenue stream is fixed for this quantity of energy, so it has the strongest possible incentive to ensure that the cost of meeting this forward contract obligation in real-time is as low as possible. Most of the contracts signed by the State of California had durations of eight years and longer. If these contracts were hedges against short-term wholesale prices at locations where the major California retailers withdraw electricity, the sellers of these forward contracts would want to construct any new generation units needed to meet these obligations to limit the magnitude of transmission congestion the new generation units face.

An active forward market has other hedging instruments besides swap contracts where a supplier and a retailer agree to a fixed price at a location in the transmission network for a fixed quantity of energy. Cap contracts are also very effective instruments for guarding against price spikes in the short-term market and for funding the appropriate amount of peak generation capacity. For example, a supplier might sell a retailer a cap contract that says that if the price at a specific location exceeds the cap's exercise price the seller of the contract pays the buyer of the contract the difference between the spot price and the cap exercise price times the number of MWh of the cap contract sold. For example, suppose the cap exercise price is \$300/MWh and market price is \$400/MWh, then the payoff from the cap contract is \$100/MWh = \$400/MWh - \$300/MWh times the number of MWh sold. If the spot price is less than \$300/MWh, then the buyer of the cap contract does not receive a payment.

Because the seller of a cap contract is providing insurance against price spikes, it must make payments when the price exceeds the cap exercise price. This price spike insurance obligation implies that the buyer must make a fixed up-front payment to the seller in order for the seller to be willing to take on this obligation. This payment can then be used by the seller of the cap contract to fund a generation unit that provides a physical hedge against price spikes at this location, such as a peaking generation unit. The Australian electricity market has an active financial forward market where these types of cap contracts are traded. These contracts have been used to fund peaking generation capacity to provide the seller of the cap contract with a physical hedge against this insurance obligation.

One question often asked about the contract adequacy approach is whether sufficient generation resources will be built to meet demand if consumers only buy forward financial hedges

against spot price risks at their location in the network. In this regard it is important to bear in mind that the incentives faced by the seller of the forward financial contract once this contract has been sold. The supplier has an obligation to insure that the forward contract quantity of energy can be purchased at the agreed upon location in the spot market (or whatever market the forward contract clears against) at the agreed upon forward price or less. The seller bears all of the risk associated with higher spot prices at that location. In order to prudently hedge this risk, the seller has a very strong incentive to construct sufficient generation capacity to ensure that the risk associated with guaranteeing the price in the short-term market at that location in the network is minimal for the quantity of energy sold in the fixed-price forward contract.

This logic implies that if a supplier signs a forward contract guaranteeing the price for 500 MWh of energy for 24 hours a day and 7 days per week at a specific location in the network, it will construct or contract for more than 500 MWh of generation capacity to hedge this spot price risk. Building only a 500 MW facility to hedge this risk would be extremely imprudent and expose the supplier to significant risk, because if this 500 MW facility is unavailable to provide electricity, the supplier must effectively purchase the energy from the spot market at the price that prevails at the time. If this generation unit is unavailable, it is very likely that the spot price will be extremely high.

Different from the case of a capacity market, the contract adequacy approach does not require the regulator to specify the amount of total generation capacity needed to meet demand. Instead the regulator ensures that retailers and large customers have adequate fixed-price forward contract coverage of final demand and then relies on the incentives that the suppliers of these contracts face to provide sufficient generation capacity to meet these forward contract obligations.

Implementing the contract adequacy approach in a world with offer caps and market power

mitigation mechanisms is complicated by the fact that retailers and large consumers have an incentive to rely on the short-term market as discussed in Section 3. To address the incentives caused by these regulatory distortions, the regulator must mandate certain levels of fixed price forward contract coverage at various horizons to delivery.

For example, the regulator could require that a large fraction of the retailer's year ahead and two-year ahead demand forecasts be covered by fixed-price forward contract obligations. How large this fraction needs to be depends on a number of factors. First, the larger the fraction of final demand paying a retail price that passes through the hourly wholesale price, the smaller this fraction needs to be. Second, the greater the share of electricity coming from hydroelectric sources, the greater this fraction needs to be because hydroelectric energy has an additional supply shortfall risk not relevant for fossil fuel-based sources: insufficient water behind the turbine to meet the unit's forward contract obligations. Higher electricity prices will not cause more water to show up behind the turbine, but it is very likely to increase the amount of fuel that can be profitably sold to a fossil fuel-fired generation unit owner. As Wolak (2003a) emphasizes, the vast majority of Latin American markets mandate minimum fractions of fixed-price forward contract coverage of the retailer's or large consumer's demand at various horizons to delivery as way to deal with the incentive of retailers to rely on the short-term wholesale market.

It is important to emphasize that mandating these contracting levels should not impose a financial hardship on retailers that lose customers in a competitive wholesale market regime. If a retailer purchased more fixed-price forward contract coverage than it ultimately needs because it lost customers to a competitor, it can trade this obligation in the secondary market. Unless the market demand in the future is unexpectedly low, this retailer is just as likely to make a profit on this sale

as it is to make a loss, because one of the retailers that gained customers is going to need a forward

contract to meet its regulatory requirements for coverage of its final demand. Only in the very

unlikely case that the aggregate amount of forward contracts purchased is greater than the realized

demand, will there be a potential for stranded forward contracts held by retailers than lose load.

4.3. Symmetric Treatment of Load and Generation

As noted in Section 2, the economic and political constraints on demand-side participation

in wholesale electricity markets in the US have led state PUCs to set fixed default retail prices that

have a significant risk of failing to cover the retailer's wholesale energy purchase costs. In addition,

many states allow customers taking service from a competing retailer to switch back to the regulated

retail price whenever they would like. This further increases the regulated supplier's wholesale

energy price risk, because customers are most likely to switch back to the regulated retail price when

it benefits them to do so and these benefits are greatest when the wholesale price of electricity is

extremely high. This ability to switch back at will leaves the regulated retailer with an enormous

unhedged risk against movements in the short-term price of wholesale electricity.

The best way to solve this problem is to make the default retail price pass-though the hourly

real-time price of electricity. Any attempt to set a fixed retail price that consumers can switch to at

their own discretion is an invitation to create a "California Problem," in the sense that there is a risk

that the implicit fixed wholesale price in the regulated retail price is less that the wholesale price of

electricity. Treating all final consumers like generation unit owners in the sense that their default

price is equal to the hourly real-time price of electricity solves this problem. This is the same default

rate faced by all electricity generation unit owners. Unless owners of generation units enter into

forward market agreements, they will receive the hourly spot price for all electricity they deliver in

real-time. Similarly, all final customers, including residential and small business customers should have to purchase all of their consumption a retail price that reflects the hourly real-time wholesale price plus the relevant transmission and distribution charges. However, all customers should also be able to enter into forward contracts and other forward market hedging agreements with competitive retailers, if they desire, just as generators are permitted to do. No final consumer must purchase any of its energy at the real-time price if it is willing to pay for spot price risk management services.

It is important to emphasize that this mechanism would not require any customer to purchase even a fraction of their consumption at the hourly real-time price, only that this is the default price that the customer pays for wholesale electricity if he does not enter into a hedging arrangement. This requirement is no different from what occurs in other markets, such as air travel where the customer always has the option to purchase the ticket at the airport at the time they would like to fly. Customers rarely do this because of a desire to hedge the spot price associated with this real-time purchasing strategy.

An important necessary condition for providing valid economic signals for customers to manage real-time price risk is to set a default rate that requires customers to manage this risk and sets the price of insurance against short-term wholesale price volatility appropriately. Figure 2 assumes that final customers have a expected utility functions, $U(E(P_r),\sigma(P_r))$, that are decreasing functions of the expected hourly retail price, $E(P_r)$, and standard deviation of the hourly retail price, $\sigma(P_r)$ for the retail pricing plans offered. Indifference curves for consumer 0 and consumer 1 are plotted in the figure. Consumer 0 is less risk-averse than consumer 1. This figure also plots the set of feasible pairs $(E(P_r),\sigma(P_r))$ that the retailer can offer in their retail pricing plans without facing a significant risk of going bankrupt. The "Feasible Expected Price and Price Risk Frontier" implies that the retailer

must increase the value $\sigma(P_r)$ in order to offer a pricing plan with a lower value of $E(P_r)$. Finding the point of tangency between each customer's indifference curve and this frontier yields that customer's optimal pricing plan choice. For customer 0 this process yields the point $((E(P_r)^0, \sigma(P_r)^0)$ and for customer 1 the point $((E(P_r)^1, \sigma(P_r)^1)$. It is important to emphasize that the reason each customer chose a plan that required it to take on some hourly price risk is because it faces the default retail rate that is a pass through of the hourly wholesale price.

Figure 3 illustrates the choices of consumer 0 and 1 if a low regulated retail price is set that completely eliminates all retail price risk, as is currently the case in all US wholesale markets. The original indifference curve for consumers 0 and consumer 1 are drawn as U_{01} and U_{11}. Two indifference curves with a higher level of utility for each consumer are drawn as U_{02} and U_{12}. These represent the utility levels that consumers 0 and 1 would achieve if a default fixed retail price, $E(P_r)^d$, was set that eliminated all price risk faced by these two consumers. Because $U_{01} < U_{02}$ and $U_{11} < U_{12}$, both consumers would achieve a higher level of expected utility by choosing $E(P_r)^d$ instead of any point along the Expected Price and Price Risk Frontier. This diagram illustrates the necessity of setting a default retail price that is a pass through of the hourly wholesale price or setting a fixed default price that contains a substantial risk premium so that it does not interfere with the choices the customers make along the Expected Price and Price Risk Frontier. This suggested fixed default price is given by the vertical line on the far right of the graph.

It is important to emphasize that requiring the default retail price to at least pass through the hourly real-time wholesale price is only making explicit something that must be true on a long-term basis: All wholesale electricity costs paid by the retailer must be recovered from retail rates. If this is not the case, then the retailer cannot remain in business over the long-term because it will be

charging a price that is less than the amount it pays for wholesale electricity.

Therefore, a prohibition on hourly meters and real-time pricing in the name of protecting consumers from real-time wholesale price volatility does not mean that consumers do not have to pay these volatile wholesale prices. On an annual basis they must or the retailer supplying them will go bankrupt. The regulatory prohibition on hourly meters and a default retail price that passes through the real-time wholesale price only prevents consumers from obtaining a lower annual electricity bill by altering their consumption in response to hourly wholesale prices. A default fixed retail price requires the consumers to pay the same wholesale price for electricity every hour of the year regardless of the wholesale price.

A final point to emphasize with respect to the question of symmetric treatment of load and generation is that all retail customers must face the real-time hourly price as their default price unless they find an entity willing to provide a hedge against this risk. The same logic applies to electricity generation unit owners. Unless they are able to find an entity willing to provide a hedge against short-term wholesale price risk, they will sell all output they produce at the hourly real-time price.

Symmetric treatment of load and generation creates the following sequence of market efficiency-enhancing incentives. First, final consumers must sign long-term contracts to obtain a fixed-price hedge against their wholesale market spot price risk. Retailers then would attempt to hedge their wholesale market risk associated with selling this fixed-price retail contract to the final consumer. This creates a demand for fixed-price forward contracts sold by generation unit owners. Therefore, by requiring both generation unit owners to receive and final consumer to pay the hourly real-time price by default, each has a strong incentive to do their part to manage this real-time price risk.

4.4. A Core/Non-Core Approach to Retail Market Operation

This section proposes a core/non-core customer approach to organizing the retail segment of the industry that recognizes the economic and political constraints on active demand-side participation in wholesale electricity markets described in Section 2. This approach recognizes the need for adequate fixed-price forward contracting by electricity retailers and large customers and the fact that with offer caps and market power mitigation mechanisms there is less of an incentive for these agents to sign the necessary quantity of fixed-price forward contracts. It also recognizes that there are very few regions with hourly meters in place at the start of re-structuring so it is necessary to determine which customers will receive these meters and what prices these customers will face once they have hourly meters.

The core/non-core distinction refers to the fact that core customers remain with the regulated retailer and are not required to have hourly meters and the non-core customers are required to have hourly meters and purchase directly from the wholesale market or from a competitive retailer. All non-core customers face a default retail price that passes though the hourly wholesale price. The regulated retailer is required to take a non-core customer back at this retail price if the competitive retailer serving that customer goes bankrupt or terminates service with that customer. The regulated retailer does not have an obligation to offer this customer any other retail price that provides some short-term risk management services.

In order to switch from the core segment to the non-core segment, a customer must have an hourly meter installed on their premises. As discussed in Section 2, it seems likely that hourly metering will soon replace conventional meters for most jurisdictions in the industrialized world and that metering services will be provided as a regulated distribution service. However, the process of

installing these meters will take time so it is important to emphasize that a customer cannot switch to the non-core segment without an hourly meter. This is necessary because of the requirement that the default retail rate for all non-core customers is a pass through of the hourly wholesale price and without an hourly meter it is impossible to measure the customer's consumption during each hour of the day.

Customers in the core segment would not be required to have hourly meters, but those with hourly meters could remain in the core segment. A major challenge faced by the regulatory process is to set tariffs that define the Feasible Expected Price and Price Risk Frontier presented in Figures 2 and 3. The regulator must guard against setting a fixed retail price at an unrealistically low level to drive out any incentive by core customers to manage wholesale price risk as described in the previous section. This is the most important factor to consider in setting the default price for core customers, because if this price is set too low, the sequence of events outlined in Figure 3 will occur and the risk of bankruptcy for the regulated retailer will be significantly higher.

The regulator must set a fixed retail price for a year that guarantees that the retailer will have sufficient revenue to meet its core customer wholesale energy costs for the following year. The regulator must be confident that even if it is fixed for a year, this retail price will provide the retailer with sufficient revenue to cover its wholesale energy costs. The expectation is that this retail price will be adjusted only once a year. The regulator should also mandate 100% forward contract coverage of the expected hourly demand of its core customers signed one year in advance of delivery. Following the process of validating adequate forward contract coverage, the regulator can set the fixed retail price for the year taking the total forward contracting costs divided by the retailers annual load forecast as the average wholesale price in the retail rate.

Under this scheme, the regulated retailer then faces only the quantity risk associated with serving an uncertain retail load. It is free to manage the remaining revenue risk through real-time pricing programs offered to its customers. For example, the retailer can offer its core customers a CPP rate or CCP rate with a rebate to ensure that its total demand during certain hours of the year is consistent with its forward contracting purchases made one year in advance.

This core customer retail pricing scheme encourages active demand-side participation in the wholesale market because it sets the fixed retail price sufficiently high to leave room for customers to choose expected price and standard deviation of price combinations that provide higher levels of expected utility for final consumers either from the regulated retailer or its competitors. Consistent with the economic and political constraints on active demand side participation in the wholesale market, all market participants will take on this wholesale price risk voluntarily. The retailer serving core customers must offer programs that core customers find beneficial relative to the fixed-price retail rate to manage the quantity risk associated with this fixed wholesale price.

As discussed in Section 3, offer caps and market power mitigation mechanisms create the possibility that the wholesale market price cannot rise to a level where amount supplied at this price equals the amount demanded. For this reason, it is important to specify what will happen when there are supply shortfalls in the short-term market. As noted earlier, the usual approach to solving this problem involves random curtailment. This outcome is unavoidable because the technology to switch off certain customers is not universally available. However, to limit the risk of this outcome, all customers are required to pay a penalty rate for their consumption during hours of system emergency. This penalty rate is designed to provide both core and non-core customers with the strongest possible incentive to reduce their demand during these periods and to take preventive

actions to ensure that supply shortfalls do not occur. For example, if the offer cap on the ISO's real-time market is $1,000/MWh, the penalty rate for consumption during these periods should be sufficient to ensure that non-core consumers will make the greatest possible efforts to reduce their consumption. For example, a penalty price of $5,000/MWh would provide strong incentives for non-core customers to reduce their demand during system emergency periods so that random curtailment of load is not necessary to manage a temporary supply shortfall.

It is important to emphasize that this penalty rate need never be active. It is only imposed to ensure the credibility of the offer cap in the wholesale market. Specifically, in order to avoid paying the penalty rate, both non-core customers and retailers serving core customers could be expected to bid demand response into the ISO's real-time market at or below the offer cap to ensure that economic curtailment (less demand clears the day-ahead and real-time market) takes places before it is necessary to invoke random curtailment. If insufficient demand is offered into the day-ahead and real-time markets at or below the offer cap to prevent system emergencies, this should be taken as strong evidence that the offer cap is set to low or the penalty price is too low.

Large retailers can even use their customers with hourly meters to reduce the wholesale prices they pay to serve all of their customers. This requires that retailers charge real-time pricing customers a different wholesale price in a given hour than the retailer is actually paying for power in that hour. Both the CPP and CPP with a rebate pricing mechanisms are simple examples of this sort of program. Because all real-time pricing programs offered in this core/non-core scheme are voluntary, the regulator does not need to set these real-time pricing rates. For core customers, the retailer must offer the fixed retail rate set by the regulator and the retailer is free to offer any other retail pricing contracts it would like. For the non-core segment, the retailers are free to offer

whatever plan customers would like, the only requirement is that the core customer's default rate on return to the regulated retailer is an hourly pass through of the wholesale price.

Retailers can reduce their total wholesale purchase costs for a given number of total MWh by reducing their total demand during hours when the aggregate bid supply curve is very steep and increases its demand in hours when the aggregate bid supply curve is flat. Consider the following two-period example of a single retailer exercising its unilateral market power as a buyer. Suppose this is a core retailer is serving customers on a fixed price retail rate and paying a real-time pricing rate.

Let PW_i equal the wholesale price in period I and PR_i the price charged to retail customers on the real-time pricing program in period 1. Let $D_i(p)$ equal the demand of real-time pricing customers at price p in period 1. Suppose that the retailer commits to guaranteeing that demand served on the real-time pricing contract will provide no marginal contribution to retailer's profits. This imposes the following constraints on the expected profit-maximizing values of PR_i for i=1,2:

$$PR_1(D_1(PR_1) + PR_2(D_2(PR_2)) = PW_1(D_1(PR_1) + PW_2(D_2(PR_2)), \qquad (4.2)$$

The total payments by customers facing real-time prices, PR, equals the total payments the retailer makes to the wholesale market to purchase this energy, because PW is wholesale price in that hour that the retailer pays for all its wholesale market purchases.

Suppose the retailer maximizes the profits associated with serving customers on fixed retail rates. Let PF equal the fixed retail rate and QF_i the demand for customers facing price the PF in period i. Let $S_i(p)$ equal the aggregative bid supply curve in period i. The profit function for the firm assuming the constraint (4.2) is:

$$\Pi(PR_1, PR_2) = PF(QF_1 + QF_2) - PW_1 QF_1 - PW_2 QF_2$$

44

The wholesale price for each period, PW is the solution to $S_i(PW_i) = D_i(PR_i) + QF_i$. This equation

implies that PW_i can be expressed as:

$$PW_i = S_i^{-1}(D_i(PR_i) + QF_i),$$

which implies that PW_i is a function of PR_i.

The simple two-period model of choosing PR_i to maximize the retailers expected profits can

be illustrated graphically. Figure 4 makes the simplifying assumption that $D(p)$ and $S(p)$ are the same

for periods 1 and 2. The only difference is the amount of fixed-price load the retailer must serve in

each period. I assume that $Q_1 < Q_2$. I define P_i as the value of the wholesale price in period i if the

retailer passively bids in the real-time demand function $D(p)$ in each period. In this figure, PW_i is

the wholesale price in period i assuming that the retailer chooses PR_i, the price charged to real-time

pricing customers, to maximize daily profits. The large difference in PR_2 and PW_2 shows the

tremendous benefit in high demand periods from the retailer exercising its market power. In order

to satisfy the constraint that the retailer makes less than or equal to a zero profit from serving real-

time pricing load, the retailer must set PR_1 below PW_1. The two lighter shaded areas in the Period

1 and 2 diagrams are equal, illustrating that the constraint (4.2) given above is satisfied. The large

difference between P_2 and PW_2 versus the relatively small difference PW_1 and P_1 illustrates the large

reduction in daily average wholesale prices from the retailer using its real-time pricing customers to

exercise market power versus simply using their demand curves non-strategically. The darker shaded

rectangles in the Period 1 and Period 2 figures, shows the profit increase achieved by the retailer as

a result of exercising its buying power. Some of the difference between the large dark rectangle in

Period 2 and the small dark rectangle in period 1 can be given to the real-time consumers as payment

for their price response efforts.

This strategy for retailers to exercise market power extends in a straightforward manner to multiple time periods within the day, week or month. It represents a major source of potential benefits from a price responsive final demand in the retail segment.

A final aspect of this core/non-core model for electricity retailing is a change in the mission of the industry regulator. Although the regulator's primary role in the former vertically integrated regime was setting retail prices, there is less need for this role in the core/non-core model, particularly if there is universal interval metering. In fact, if the regulator sets the fixed-retail price too high this will only encourage more customers to manage real-time wholesale price risk with a competing retailer or the core retailer. For this reason, the regulator should focus its attention on providing information to retail customer to help them better manage their real-time price risk. For example, the regulator might manage a web-site that has all of the plans offered and illustrates the price risk and standard deviation of price tradeoff inherent in each plan.

If there are a significant number of core customers without hourly meters the regulator's job becomes more difficult because a moral hazard problem in electricity retailing arises that is similar to the one that exists in retail banking. The fear in retail banking is that the bank will take customer deposits and invest them in extremely risky assets in an effort to deliver a very favorable return to the investor and the bank's shareholders. However, engaging in this risk-taking behavior may lead to outcomes that render the bank unable to meet certain future obligations to its depositors. An analogous chain of events can happen in the electricity retailing industry. The retailer has a strong incentive to under-invest in forward contracts to cover their future load obligations when it sells a fixed-price commitment to a customer for one or two-year period. It may be able to earn a higher expected return by taking risks that increase the probability of bankruptcy but also have the prospect

of very high positive profit levels due to low wholesale prices.

Consequently, similar to the retail banking sector regulation, state PUCs must change its focus from retail rate setting to monitoring the forward contract procurement process and ensuring forward contract coverage requirements of all retailers relative to their forecasted retail market commitments. Clearly, if firms are always required to hold the 100% of their forecast demand in fixed-price forward contracts one year in advance, then these firms will find it profit-maximizing to honor their retail market commitments.

This market monitoring process should require all retailers to submit to their state PUC on a monthly basis a list their retail market commitments by duration and price and their wholesale market coverage by quantity and price. The role of the PUC would be to verify that the retailer met these risk management prudency standards and assess penalties and sanctions for violations.

Consider the following example of how this might work. The second and third column of Table 1 contains a list of the quantity-weighted average wholesale price implicit in the fixed retail price retail and quantity obligations that the retailer has agreed to supply for various delivery months in the future. The fourth and fifth columns of Table 1 contain the quantity-weighted average fixed wholesale price and quantity commitments the retailer has signed with wholesale energy suppliers. The sixth columns contain the desired percentage of the total monthly quantity of fixed-price wholesale quantity commitments that the state PUC deems that it is prudent for the retailer to hold as a hedge against its fixed price retail commitments for each future delivery date. The last column contains the product of the percentage in the sixth column and the fixed price retail obligation quantity given in the second column.

In this example there are several delivery horizons where the desired hedge quantity is greater

than the amount given in the fourth column. In these instances there are several actions that the state could take. First, it could assess a substantial penalty per MWh on the positive part of difference between desired quantity in the seventh column and the actual quantity in the fourth column. The PUC could also prohibit this retailer from selling more fixed-price retail obligations at this time horizon or shorter until the retailer submits a monthly report that is not out of violation for all months longer than this delivery horizon.

For the case given in Table 1, the first month the retailer is out of compliance is month 4. This means that retailer is prohibited from signing fixed price commitments for deliveries longer than 3 months in the future during the next month unless it submits proof of compliance in the next month for all delivery horizons up to 3 months. There are other prudency standards that state PUCs could impose on hedging behavior of retailers that uses risk measures based on the prices of retail obligation versus the price of wholesale commitments that cover them. Fortunately, these hedging standards do not need to be set using very sophisticated methods in order provide a reasonable level of assurance that all retailers will be able to meet their fixed price retail obligations with a high degree of certainty.

The other role of the state PUC in a competitive retail market is to ensure that all retailers have equal access to the billing and metering services provided by the regulated monopoly local distribution company. The PUC must establish rules that prevent the local distribution company from favoring its competitive retailing affiliate.

4.5. Developing Country Issues

Developing countries complicate several features of this core/non-core model. First, in many developing countries a significant fraction of customers lack of any sort of meter on their premises.

Second, substantial fractions of customers in a number of countries do not pay their bills. Third, a significant fraction of the population do not have access to electricity. Although crafting a satisfactory solution to all of these problems is beyond the scope of this paper, a few promising directions to consider are suggested.

Electricity networks are well-suited to implementing group payment programs for electricity bills because all customers in a given geographic area typically take their energy from the same location in the high voltage transmission network. The lower voltage distribution network that serves a given geographic area typically interconnect at this location and the system operator is able to meter total withdrawals from these locations in real-time. This fact suggests allocating the liability for the cost of all wholesale energy withdrawn at the lowest voltage location in the transmission network that the system operator is able to meter withdrawals to all customers taking service from this location in the transmission network.

The wholesale market operator could be made responsible for terminating service for all customers at this location after a certain period of nonpayment. Because it is impossible to determine how much electricity was consumed by each customer in a given time period because of the lack of meters or the lack of hourly meters, assigning payment liability to each customer in the geographic region and collecting payment from them is an extremely complex task. This problem should be easier to solve by asking other customers in the same area to ensure that all other customers in the area pay their bills and do not steal electricity. Allowing the wholesale market operator to curtail power at lowest level in the network at which it has this capability provides credibility to threat that nonpayment will result in curtailment. Credible demonstration of this threat by the system operator

will make it easier for electricity retailers to address the problem of nonpayment, because a substantial fraction of nonpayments in many countries is due to theft.

The use of social pressure to ensure prompt payment has been successfully used most notably in the area of providing microfinance. Johnson and Rogaly (1997) describe the successful use of group liability in the provision of microfinance. Borrowers are formed into groups by the microfinance banks and these groups assume joint liability for repayment of each member's loan. By the same logic, the set of electricity consumers connected to the transmission network at a given location must assume joint liability for payment for the total amount of electricity withdrawn at that location, or jointly face the risk of no electricity for all consumers in the geographic area until the liability is paid.

Although this may seem like a drastic measure to ensure payment, as Wolak (2006b) emphasizes for the case of India, without a change in the attitude of consumers toward paying for electricity, it is unlikely that India will ever be able to attract private investment in the electricity sector. Even spending government money on this sector seems misguided if final consumers do not pay for the electricity that is produced. Determining the magnitude of the total amount of KWh consumed and assigning it to all customers in that geographic area and alerting these customers to the joint liability nature of their electricity supply costs should help to improve payment rates.

The second issue concerns the need to built out the transmission and distribution network to serve more customers in many developing countries. These customers should be treated as core customers and their retail prices determined as describe above for non-core customers. As these regions grow, it may make sense to install hourly meters and convert some of these customers to the non-core segment.

5. Concluding Comments

All existing electricity markets in the US and virtually all markets that exist in other jurisdictions have failed to introduce the necessary demand-side incentives for setting the lowest possible prices for wholesale electricity consistent with the long-term financial viability of the industry. In the name of protecting financial consumers, state PUCs in the US have denied consumers the ability to benefit from being active participants in the spot market. We have argued in this paper, by handicapping the demand side of the market the PUCs are only increasing the likelihood that wholesale suppliers will be able to raise prices through their own unilateral bidding and scheduling behavior.

Final consumers must bear the full cost of high wholesale prices and have the ability to realize the full benefits from taking actions in the forward and spot markets to respond to these high prices. Investments in hedging instruments and demand-responsiveness technology will then lead to a more competitive wholesale market that will, in turn, lead to lower average prices than the former vertically integrated monopoly regime when final demand was a passive participant in the wholesale market.

The well-known dictum of "there's no such thing as a free lunch" applies to the case of introducing competition into a formerly regulated industry. Unless competition changes the behavior of some market participants, it cannot benefit consumers relative to the former monopoly regime. For example, if generation unit owners continue to produce the same amount of electricity in the same manner as they did under the former monopoly regime and all input costs for all companies remain the same, then total production costs will not change. Similarly if consumers continue to

demand the same amount of electricity in each hour of the year their annual electricity bills cannot decrease.

Only by providing incentives for more efficient operation of generating facilities and more efficient consumption signals can a market result in lower annual average prices than under the former monopoly regime. The retail market infrastructure presented in this paper provides the strongest possible incentives for consumers to alter their behavior to reduce the cost of producing wholesale electricity and making most efficient use of the generating capacity that currently exists.

References

Johnson, Susan, and Rogaly, Ben (1997) *Microfinance and Poverty Reduction*, Oxfam.

Mukherjee, Andy (2007) "Why you need to run a power plant at home," Livemint, May 25, 2007 at http://www.livemint.com/2007/05/25001738/Why-you-need-to-run-a-power-pl.html

Patrick, Robert H. and Wolak, Frank A. (1997) "Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices," August 1997, (available from http://www.stanford.edu/~wolak/).

Wolak, Frank A., Nordhaus, Robert, and Shapiro, Carl, "Preliminary Report on the Operation of the Ancillary Services Markets of the California Independent System Operator (ISO)," August 1998 (available from http://www.caiso.com/docs/2000/09/14/200009141610025714.html).

Wolak, Frank A., Nordhaus, Robert, and Shapiro, Carl, "Report on the Redesign of the Markets for Ancillary Services and Real-Time Energy," March 1999 (available from http://www.caiso.com/docs/2000/09/14/200009141610025714.html).

Wolak, Frank A., (1999) "Market Design and Price Behavior in Restructured Electricity Markets: An International Comparison," in *Competition Policy in the Asia Pacific Region, EASE Volume 8*, Takatoshi Ito and Anne Krueger (editors) University of Chicago Press, 79-134.

Wolak, Frank A., (2000) "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market," *International Economic Journal*, Summer 2000, 1-40.

Wolak, Frank A. (2003a) "Designing Competitive Wholesale Markets for Latin American Countries," available from http://www.stanford.edu/~wolak.

Wolak, Frank A. (2003b) "Diagnosing the California Electricity Crisis," *The Electricity Journal*, August/September, 11-37.

Wolak, Frank A. (2003c) "Regulating Wholesale Electricity Markets in the Aftermath of the California Crisis," *The Electricity Journal*, August/September, 50-55.

Wolak, Frank A., (2004) "Managing Unilateral Market Power in Wholesale Electricity," in *The Pros and Cons of Antitrust in Deregulated Markets*, edited by Mats Bergman, Swedish Competition Authority.

Wolak, Frank A., (2006a) "Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment, available from http://www.stanford.edu/~wolak.

Wolak, Frank A. (2006b) "Reforming the Indian Electricity Supply Industry," available from http:/www.stanford.edu/~wolak.

Table 1: Sample Monthly Forward Contract Filing						
	Retail Obligations		Forecast Wholesale Purchases		Compliance Levels	
Future Delivery Date for Energy (months)	Total Quantity (MWH)	Average Implicit Wholesale Price ($/MWH)	Total Quantity (MWH)	Average Purchase Price ($/MWH)	Hedge Factor (%)	Desired Hedge Quantity (MWH)
1	10000	44.56	10000	40.12	100	10000
2	10000	45.60	10000	45.00	100	10000
3	10000	42.00	11000	40.21	100	10000
4	12000	50.00	11000	49.00	100	12000
5	13000	54.00	12000	52.00	100	13000
6	11000	51.00	9000	50.12	100	11000
12	10000	48.00	10000	45.29	100	10000
18	10000	44.23	9000	39.56	85	8500
24	12000	44.00	10000	42.03	80	9600

Figure 1

Advanced Metering Communication Networks

Local Area Networks

Wide Area Networks

Consumer

Local power lines

Gateway

Utility User

Telephone

Internet

Wireless Network

Wireless

Gateway

Data Center

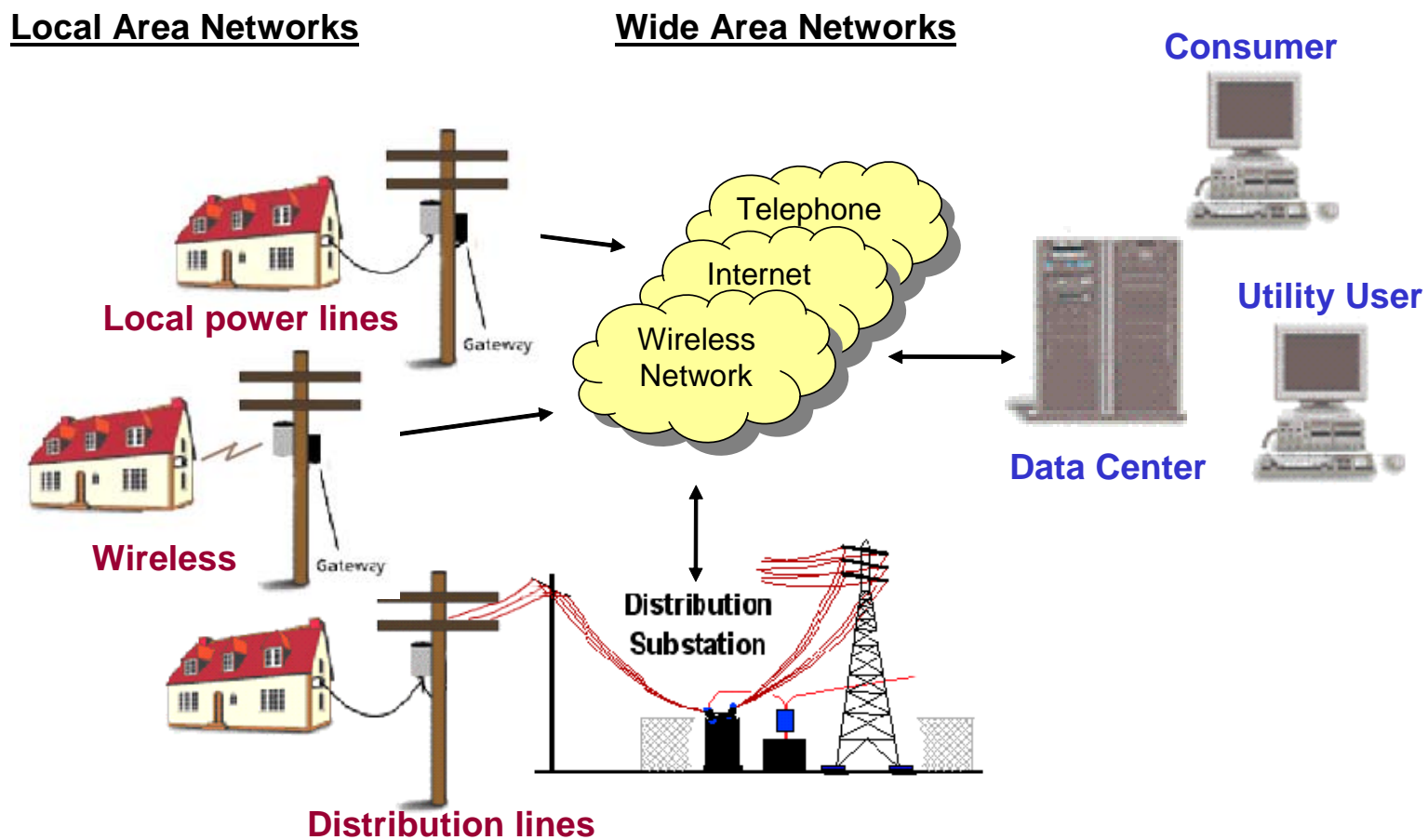Distribution Substation

Distribution lines

Figure 2

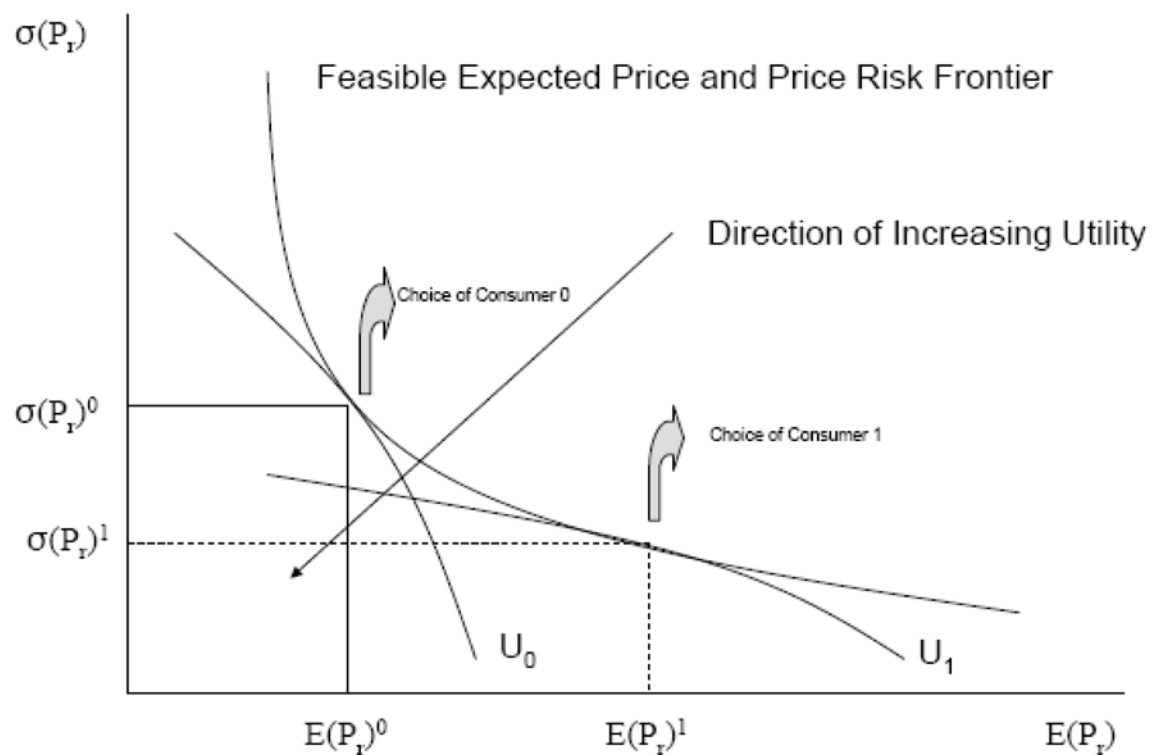Expected Retail Price (E(P_r)) and Standard Deviation of Retail Price (σ(P_r)) Frontier

Figure 3

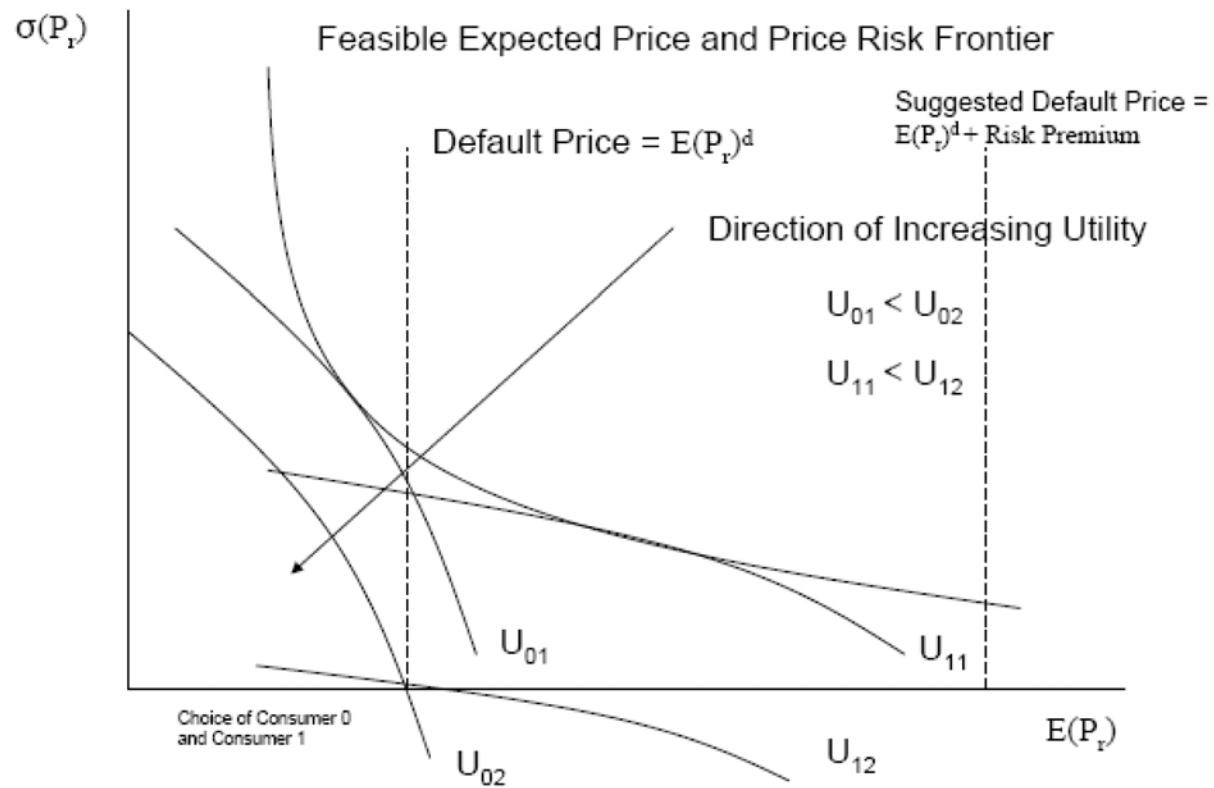## Consumer Choices with Default Rate Set at Average Wholesale Price

$\sigma(P_r)$

Feasible Expected Price and Price Risk Frontier

Default Price = $E(P_r)^d$

Suggested Default Price = $E(P_r)^d$ + Risk Premium

Direction of Increasing Utility

$U_{01} < U_{02}$

$U_{11} < U_{12}$

U_{01}

U_{11}

Choice of Consumer 0 and Consumer 1

U_{02}

U_{12}

$E(P_r)$

Figure 4

Period 1

Period 2