Market Design in a Zero Marginal Cost Intermittent Renewable Future

by

Frank A. Wolak
Director, Program on Energy and Sustainable Development (PESD)
Holbrook Working Professor of Commodity Price Studies
Department of Economics
Stanford University
Stanford, CA 94305-6072
wolak@zia.stanford.edu

Current Draft: June 30, 2020

1. Introduction

The basic features of an efficient short-term wholesale market design do not need to change to accommodate a significantly larger share of zero marginal cost intermittent renewable energy from wind and solar resources. A large share of controllable zero marginal cost generation does not create any additional market design challenge relative to a market with a large share of controllable positive marginal cost generation. In both instances, generation unit owners must recover their fixed costs from sales of energy, ancillary services, and long-term resource adequacy products.

A larger variance in the hourly amount of energy produced by intermittent resources is the primary market design challenge associated with a zero marginal cost renewable future. The past ten years in California has demonstrated that as the amount of wind and solar generation capacity increases, the variance in hourly megawatt-hours (MWhs) produced by these resources increases. This increase in supply uncertainty increases short-term price volatility, which can support needed investments in storage and other technologies that allow consumers to shift their withdrawals of grid-supplied energy away from periods when little wind and solar energy is being produced.

An increased risk of large intermittent energy shortfalls and short-term price volatility resources implies a greater need for risk management activities. The short-term intermittent energy supply risk is likely to require accounting for more transmission and generation operating constraints in the day-ahead and real-time energy markets as well as purchasing more operating reserves and creating additional ancillary services products. Because controllable generation units are likely to have to start and stop more frequently to make up for unexpected renewable energy shortfalls, there will be a greater need to develop short-term pricing approaches that recover the associated start-up and minimum load costs.

The potential for sustained periods of low intermittent energy production creates both a medium and long-term energy supply risk that requires a new long-term resource adequacy mechanism. The traditional capacity-based approach is unlikely to be the least cost mechanism for ensuring the future demand for energy is met. Long-term resource adequacy in a zero marginal cost intermittent energy future must ensure that these resources to re-insure their energy supply risk with controllable generation resources. Cross hedging between these technologies accomplishes two goals. First, it can provide the revenue stream necessary for fixed cost recovery by controllable generation units. Second, it ensures that there is sufficient controllable generation capacity to meet demand under all possible future system with a high degree of confidence.

The remainder of the paper first describes the key features of an efficient short-term wholesale market design—a multi-settlement locational marginal pricing (LMP) market with an automatic local market power mitigation mechanism—the standard for all short-term markets in the United States. This section concludes with a discussion of modifications of this basic design likely to be necessary to accommodate a larger share of intermittent renewables.

The second half of the paper first a new long-term resource adequacy mechanism for an electricity supply industry with a large share of zero marginal cost intermittent renewables. I first explain why a wholesale electricity market requires a long-term resource adequacy mechanism. I then describe a mandated standardized long-term contract approach to long-term resource adequacy that provides strong incentives for intermittent renewable resource owners to hedge their energy supply risk with controllable generation resource owners. I argue that this mechanism ensures long-term resource adequacy at a reasonable cost for final consumers while also allowing the short-term wholesale volatility necessary to finance investments in storage and other load-shifting technologies necessary to manage large share of intermittent renewable resources.

2. Short-Term Market Design

More than twenty-five years of international experience with wholesale electricity markets around the world has identified four crucial features of an efficient short-term market design. First is the extent to which the market mechanism used to set dispatch levels and locational prices is consistent with how the grid and generation units actually operate. Second is a financially binding day-ahead market that prices all transmission and generation unit operating constraints expected to be relevant in real-time. Third is an automatic local market power mitigation mechanism (LMPM) that limits the ability of a supplier to influence the price it is paid when it possesses a substantial ability to exercise unilateral market power. Fourth is retail market policies that foster active participation of final demand in the wholesale market.

The early US wholesale market designs in the PJM Interconnection, ISO New England, California, and Texas employed simplified versions of the transmission network configuration and generation unit operating constraints. Similar market designs currently exist throughout Europe and the rest of the world. They set a single market-clearing price for an hour for an entire control area or for large geographic regions, despite the fact that in real-time there are often generation units with offer prices below this market-clearing price not producing electricity and units with offer prices above this market-clearing price producing electricity. This outcome occurs because of the location of demand and available generation units within the region and the configuration of the transmission network prevents some of these low-offer-price units from producing electricity and requires some of the high-offer-price units to supply electricity. The former units are typically called "constrained-off" units and the latter are called "constrained-on" units.

This approach to short-term market design provides incentives for suppliers to take actions to exploit the fact that in "real-time physics wins," rather than offer their resources into the day-ahead market in an manner that minimizes the cost of meeting demand at all locations in the grid in real-time. Instead, suppliers take actions in the simplified day-ahead market that allow them to profit from knowing they will be needed in real-time or not needed in real-time because of transmission and generation unit operating constraints.

2.1. Locational Marginal Pricing

Starting with PJM in 1998 and ending with Texas in late 2010, all US wholesale markets adopted a multi-settlement locational marginal pricing (LMP) market design that co-optimizes the procurement of energy and ancillary services and includes an automatic local market power mitigation mechanism built into the market software. This design has a day-ahead financial market which satisfies the locational demands for energy and each ancillary service simultaneously for all 24 hours of the following day. A real-time market then operates using the same network model as the day-ahead market adjusted to real-time system conditions. Deviations from purchases and sales in the day-ahead market are cleared using these real-time prices. Both of these markets price all relevant transmission network and other relevant operating constraints on the transmission network and generation units. Combined with the appropriate retail market technical and regulatory infrastructure discussed below, this market design fosters active participation of final demand in the wholesale market.

Only generation unit output levels that are physically feasible will be accepted in both the day-ahead and real-time markets. Prices for the same hour vary depending on whether the location is in a generation-deficient or generation-rich region of the transmission network. The locational marginal price or nodal price at each location is the increase in the minimized value of the "as-offered costs" of serving the locational demands for energy and all ancillary services as a result of a one unit increase in the amount of energy withdrawn at that location in the transmission network. The price of each ancillary service is defined at the increase in the optimized value of the objective function as a result of a one unit increase in the demand for that ancillary service.

The recent experience of many European countries with significant wind and solar resources indicates that the cost of making the final schedules that emerge from their zonal markets physically feasible is likely to get even larger as the amount of intermittent renewable generation capacity increases. According to the European Network of Transmission System Operators for Electricity, these costs were over 1 billion Euros Germany, more than 400 million Euros in Great Britain, over 80 million Euros in Spain, and approximately 50 million Euros in Italy in 2017.

2.2. Multi-Settlement LMP Market

A multi-settlement LMP market has at least a day-ahead forward market and a real-time market each of which employs same market-clearing mechanism. The day-ahead market typically allows generation unit owners to submit three-part offers to supply energy--start-up costs, minimum load costs, and energy offer curves to compute hourly generation schedules and ancillary service quantities and LMPs for energy and ancillary services for all 24 hours of the following day. A generation unit will not be accepted to supply energy in the day-ahead market unless the combination of its offered start-up costs, minimum load costs and energy production costs are part of the least as-offered-cost solution to serving hourly locational demands for all 24 hours of the following day.

The energy schedules that arise from the day-ahead market do not require a generation unit to produce the amount sold or a load to consume the amount purchased in the day-ahead market. Any production shortfall relative to a day-ahead generation schedule must be purchased from the real-time market at that location. Any production greater than a day-ahead schedule is sold at the real-time price at that location. For loads any additional consumption beyond the load's day-ahead purchase is paid for at the real-time price at that location and the surplus of a day-ahead purchase relative to actual consumption is sold at the real-time price at that location.

2.3. Mitigating Local Market Power

The configuration of the transmission network, the level and location of demand, as well as the level of output of other generation units can create system conditions when almost any generation unit or group of generation units has a significant ability to exercise unilateral market power. The constrained-on generation problem is an example of this phenomenon. The unit's owner knows that it must be accepted to supply energy regardless of its offer price. Without a local market power mitigation mechanism, there may be no limit to what offer price the unit's owner could submit and have it accepted to supply energy.

This logic is why *ex ante* market structure-based market power mitigation mechanisms typically used in Europe and other industrialized countries and initially employed in the US, which designate in advance the offers of certain generation units for mitigation for an entire year, miss many instances of the exercise of substantial unilateral market power.

An automated LMPM mechanism built into the market software that relies on actual system conditions to determine whether any supplier has a substantial ability and incentive to exercise unilateral market power is likely to be significantly more effective. This regulator-approved administrative procedure determines: (1) when a supplier has an ability to exercise local market power worthy of mitigation, (2) the value of the supplier's mitigated offer price, and (3) the price mitigated supplier is paid. It is increasingly clear to regulators around the world, particularly those that operate markets with a finite amount of transmission capacity and significant intermittent renewable generation capacity, that an automatic LMPM is necessary for any short-term market design.

2.4. Benefits of a Multi-Settlement LMP Market

A multi-settlement LMP market design can facilitate the active participation of final consumers in the wholesale market and reduce both the input fuel and total variable cost of producing the same amount of thermal energy relative to a multi-settlement zonal market design. The presence of an automatic LMPM mechanism and make-whole payments that guarantee start-up, minimum load, and energy cost recovery for the day for all generation units committed to operate in the day-ahead market reduces the incentive of suppliers to exercise unilateral market power.

Because day-ahead purchases are firm financial commitments, a retailer can sell energy purchased in the day-ahead market at the real-time price by consuming less than its day-ahead energy schedule. This eliminates the need for the regulator to set an administrative baseline relative to which a retailer sells demand response reductions. The day-ahead market also allows retailers and large consumers to submit price-sensitive bid curves respectively, into the day-ahead market to reduce the price and the quantity of energy purchased in the day-ahead market.

On April 1, 2009, California market transitioned to a multi-settlement nodal-pricing market design from a multi-settlement zonal-pricing market. The hourly total quantity of input fossil fuel energy used to produce thermal generation electricity fell by 2.5 percent and the total hourly variable cost fell by 2.1 percent after the implementation of nodal pricing. This hourly variable cost reduction implies a roughly $100 million reduction in the total annual variable cost of producing electricity in California associated with the transition to a LMP market design.

The existence on an automatic local market power mitigation mechanism and make-whole payments that compensate generation unit owners for under-recovery on their start-up and minimum load costs on a daily basis implies that a supplier with no ability to exercise unilateral market will submit an offer price equal to its marginal cost of production. The supplier knows that if its units are committed in day-ahead market they are guaranteed recovery of start-up, minimum load and energy production costs, so it expected profit maximizing for this supplier to submit an offer price for energy equal to the unit's marginal cost.

2.5. Modifications for Large Scale Intermittent Renewables Deployment

A multi-settlement LMP market design is capable of managing a generation mix with a significant share of intermittent renewables. However, modifications are likely to be required as the share of intermittent renewable resources increases. Additional operating constraints will need to be incorporated into the day-ahead and real-time market models for reliable system operation with an increased amount of intermittent renewables.

There is also likely to be a need to introduce additional ancillary services to accommodate a larger share of intermittent renewable energy. For example, the California market introduced a fast-ramping ancillary service product that compensates controllable generation units for not supplying energy during certain hours of the day in order to have sufficient unloaded capacity to meet the rapid increase in net demand (the difference between system demand and renewable generation) in the early evening when the state's solar resources stop producing.

Because controllable resources are likely to have to start and stop more frequently as the share of intermittent resources increases, implementations of convex hull pricing and other mechanisms to limit the magnitude of make-whole payments will need to be developed.

One complaint often leveled against LMP markets is that they increase the likelihood of political backlash from consumers because wholesale prices can differ significantly across locations. Most regions with LMP markets have addressed this issue by charging all customers in a state, region, or utility service territory a weighted average of the LMPs at all load withdrawal points in that geographic region. Under this scheme, generation units are paid the price at their location, but all loads pay a geographically aggregated hourly price. In Singapore all generation units are paid the LMP at their location, but all loads are charged the Uniform Singapore Electricity

Price (USEP), which is the quantity-weighted average of the prices at all generation nodes in Singapore. This approach to pricing load captures the reliability and operating efficiency benefits of an LMP market while addressing the equity concerns regulators often face with charging customers different locational prices.

3. Resource Adequacy with Significant Intermittent Renewables

Why do wholesale electricity markets require a regulatory mandate to ensure long-term resource adequacy? Electricity is essential to modern life, but so are many other goods and services. Consumers want cars, but there is no regulatory mandate that ensures enough automobile assembly plants to produce these cars. They want point-to-point air travel, but there is no regulatory mandate to ensure enough airplanes to accomplish this. Consumers want bread, but there is no regulatory mandate to ensure sufficient bakeries to meet this demand. All of these goods are produced using high fixed cost, relatively low marginal cost technologies, similar to electricity supply. Nevertheless, all of these firms recover their cost of production including a return on the capital invested by selling the good at a market-determined price.

So what is different about electricity that requires a long-term resource adequacy mechanism? The regulatory history of the electricity supply industry and the legacy technology for metering electricity consumption results in what I call a *reliability externality*.

3.1. The Reliability Externality

Different from the case of wholesale electricity, in the market for automobiles, air travel, and even bread, there is no regulatory prohibition on the short-term price rising to the level necessary to clear the market. Airlines adjust the prices for seats on a flight over time in an attempt to ensure that the number of customers traveling on that flight equals the number of seats flying. This ability to use price to allocate the available seats is also what allows the airline to recover its total production costs.

Using the short-term price to manage the real-time supply and demand balance in a wholesale electricity market is limited by a finite upper bound on a supplier's offer price and/or a price cap that limits the maximum market-clearing price. Although offer caps and price caps can limit the ability of suppliers to exercise unilateral market power in the short-term energy market, they also reduce the revenues suppliers can receive during scarcity conditions. This is often referred to as the *missing money* problem for generation unit owners. However, this missing money problem is only a symptom of the existence of the "reliability externality."

This externality exists because offer caps limit the cost to electricity retailers of failing to hedge the cost purchases of electricity from the short-term market. Specifically, if the retailer or large consumer knows the price cap on the short-term market is $250/MWh, then it is unlikely to be willing to pay more than that for electricity in any earlier forward market. This creates the possibility that real-time system conditions can occur where the amount of electricity 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, as occurred a number of times between January 2001 and April 2001 in California.

Because random curtailments of supply---also known as rolling blackouts---are used to make demand equal to the available supply at or below the bid cap under these system conditions, this mechanism creates a "reliability externality" because no retailer bears the full cost of failing to procure adequate amounts of energy in advance of delivery. A retailer that has purchased sufficient supply in the forward market to meet its actual demand is equally likely to be randomly curtailed as the same size retailer that has not procured adequate energy in the forward market. For this reason, all retailers have an incentive to under-procure their expected energy needs in the forward market.

The lower the offer cap, 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 insufficient supply in the short-term market at or below the offer cap. Because retailers do not bear the full cost of failing to procure sufficient energy in the forward market to meet their future demand, there is a missing market for long-term contracts for long enough delivery horizons into the future to allow new generation units to be financed and constructed to serve demand under all possible future conditions in the short-term market. Therefore, a regulator-mandated long-term resource adequacy mechanism necessary to replace this missing market.

Specifically, unless the regulator is willing to eliminate or substantially increase the offer cap so that the short-term price can be used to equate available supply to demand under all possible future system conditions, some form of regulatory intervention is necessary to internalize the resulting reliability externality. If customers do not have interval meters that can record their consumption on an hourly basis, they have a very limited ability to benefit from shifting their consumption away from high-priced hours, so raising or having no offer cap on the short-term market would not be advisable. Even in regions with interval meters there would substantial political backlash from charging hourly wholesale prices that cause real-time demand to equal available supply under all possible future system conditions.

Currently, the most popular approach to addressing this reliability externality is a capacity payment mechanism that assigns a firm capacity value to each generation unit based on the amount of energy it can provide under stressed system conditions. Sufficient firm capacity procurement obligations are then assigned to retailers to ensure that annual system demand peaks can met.

Capacity-based approaches to long-term resource adequacy rely on credibility of the firm capacity measures assigned to generation units. This is a relatively straightforward process for thermal units. The nameplate capacity of the generation unit times its annual availability factor is a reasonable estimate of the amount energy the unit can provide under stressed system conditions. For the case of hydroelectric facilities, this process is less straightforward. The typical approach uses percentiles of the distribution of past hydrological conditions for that generation unit to determine its firm capacity value.

Assigning a firm capacity value to a wind or solar generation unit is extremely challenging for several reasons. First, these units only produce when the underlying resource is available. If stressed system conditions occur when the sun is not shining or the wind is not blowing, these units should be assigned little, if any, firm capacity value. Second, because there is a high degree of contemporaneous correlation between the energy produced by solar and wind facilities within the same region, the usual approach to determining the firm capacity of a wind or solar unit assigns a smaller value to that unit as the total MWs of wind or solar capacity in the region increases. These facts imply that a capacity-based long-term resource adequacy mechanism is poorly suited to zero marginal cost intermittent renewable feature.

3.2. Supplier Incentives with Fixed-Price Forward Contract Obligations for Energy

The standardized fixed-price forward contract (SFPRC) approach to long-term resource adequacy recognizes that having the ability serve demand at a reasonable price, does not imply that it will occur if suppliers have the ability exercise unilateral market power. Because a supplier with a significant ability to exercise unilateral market power with a fixed-price forward contract obligation finds it expected profit maximizing to minimize the cost of supplying that quantity of energy, this long-term resource adequacy mechanism requires retailers to hold hourly fixed-price forward contract obligations for energy that sum to the hourly value of system demand. This implies that all suppliers find it expected profit maximizing to minimize the cost of meeting fixed-price forward contract obligations that sum to hourly system demand for all hours of the year.

To understand the incentives created by this mechanism, let PC equal the fixed price at which the supplier sold energy in the standardized forward contract and QC_h equal to quantity of energy sold in hour h. The values of PC and QC_h are predetermined from the perspective of the supplier's behavior in a short-term wholesale market because this contract has been sold advance of the date that the contract clears. In terms of this notation, the supplier's variable profits from selling energy in the short-term market and clearing its forward contract obligations for hour h is:

$$\pi(PS_h) = (PC - C)QC_h + (QS_h - QC_h)(PS_h - C), \tag{1}$$

where QS_h is the quantity of energy sold in the short-term market by the generation unit owner in hour h, PS_h is the price of energy sold in the short-term market in hour h and C is the supplier's marginal cost of producing electricity, which for simplicity is assumed to be a constant.

The first term in (1) is the variable profit from SFPRC 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 SFRPC quantity. Because the SFRPC price and quantity are known in advance of the delivery date, the first term, $(PC - C)QC_h$, is a fixed profit stream to the supplier submits offers to supply energy into the day-ahead market. Because the second term depends on PS_h, the value of QC_h significantly limits the incentive the supplier has to raise short-term market prices.

If the supplier attempts to raise PS_h by submitting a high offer price, it could end up selling less in the short-term market than its SFRPC quantity ($QC_h > QS_h$), and if the resulting market-clearing price is greater than the firm's marginal cost ($PS_h > C$), the second term in (1) will be negative. Consequently, only if the supplier is confident that its offer price will result in short-term market sales greater than QC_h does it have an incentive to submit an offer price above its marginal cost.

The quantity of forward contract obligations held by a firm's competitors also limits ability a supplier has 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 that each competitor will find it unilaterally profit-maximizing submit offer prices into the short-term market equal their marginal cost for quantities of energy up to QC_h. 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 SFPFC obligations limits the price increase a supplier can expect to achieve from these actions.

3.3. SFPFC Approach to Resource Adequacy

This long-term resource adequacy mechanism requires all electricity retailers to hold SFPFCs for energy for fractions of realized system demand at various horizons to delivery. For example, retailers in total must hold SFPFCs that cover 100 percent of realized system demand, 95 percent of system demand one year in advance of delivery, 90 percent two-years in advance of delivery, 87 percent three years in advance of delivery, and 85 percent four years in advance of delivery. The fractions of system demand and number of years in advance that the SFPFCs must be purchased are parameters set by the regulator to ensure long-term resource adequacy. In the case of a multi-settlement LMP market, the SFPFCs would clear against the quantity-weighted average of the hourly locational prices at all load withdrawal nodes.

SFPFCs are shaped to the hourly system demand within the delivery period of the contract. If QD_h is equal to the system demand in hour h of the delivery period and QT_j is the total MWhs of the standardized contract sold by supplier j during the delivery period for j=1,2,..J, then the forward contract obligation of supplier j during hour h is $QC_{jh} = \left(\frac{QD_h}{\sum_{h=1}^{H} QD_h}\right) QT_j$, the system demand shaped hourly profile during the delivery period of QT_j, the total MWhs of SFPFCs sold by supplier j for the delivery horizon. Because the mechanism requires retailers to hold total MWhs of SFPFCs equal to their total demand during the delivery period, the total amount SFPFC obligations sold by suppliers for the delivery horizon is equal to total system during that delivery horizon. To the extent that the total amount of QT purchased in the forward market is more or less than realized system demand during the delivery horizon, a true-up auction run after annual demand has been realized would buy or sell the amount of QT needed to make $\sum_{j=1}^{J} QT_j = \sum_{h=1}^{H} QD_h$. By allocating each supplier's total SFPRC obligations according to the actual hourly load shape during the delivery horizon, total hourly SFPFCs obligations across all suppliers is equal system demand for all hours of the delivery horizon.

These standardized fixed-price forward contracts are allocated to retailers based on their share of system demand during the month. Let QR_{kh} equal the realized demand of retailer k during hour h for h=1 to M, where M is the number of hours in the month and k=1 to K, where K is the number of retailers. Because all retailers are assigned standardized fixed-price forward contract obligations, $QD_h = \sum_{k=1}^{K} QR_{kh}$, hourly system demand is equal to the hourly demand of all retailers for each hour of the month. Shares of the aggregate quantity of SFPFCs obligations for each hour of the month, QTC_h, are assigned to each retailer each hour of the month according to the following rule: $QRC_{kh} = \left(\frac{\sum_{h=1}^{M} QR_{kh}}{\sum_{h=1}^{M} QD_h}\right) QTC_h$. Under this scheme, retailer k's variable profits in hour h from selling energy to final consumers at PR is:

$$\Pi_R(PS_h) = (PR - PS_h)QR_{kh} + (PS_h - PC)QRC_{kh}. \quad (2)$$

The hourly variable profits of generator j in hour h in equation (1) can be re-written as:

$$\Pi_G(PS_h) = (PS_h - C)QS_{jh} - (PS_h - PC)QC_{jh}, \quad (3)$$

Because the sum of hourly SFPFCs allocated to retailers is equal to the sum of the hourly SFPFC obligations of suppliers, the sum of the second term in (3) over all suppliers equals the sum of the second term in (2) over all retailers on an hourly basis.

The SFPRCs only influence the magnitude of payments between suppliers and retailers each hour. Because the sign and magnitude of these payments depends on both the short-term market price and quantity sold, they affect the supplier's incentive to offer that quantity of energy into the short-term market. A generation owner with a SFPRC obligation of QC maximizes expected profits by submitting an offer price equal to marginal cost for this quantity of energy.

This offer price ensures that supplier j makes the economically efficient "make versus buy" decision to supply QC_{jh} units of output. If the PS_h is above C, then it is cheaper to produce QC_{jh} from its generation units, but if PS_h is below C then it is cheaper to QC_{jh} from the short-term market. Submitting an offer price equal to the supplier j's marginal cost for all output levels up to QC_{jh} ensures that the market-clearing mechanism yields the efficient outcome. If all suppliers submit offer prices that yield the efficient "make versus buy decision" for their QC_{jh}, then all generation unit owners will find it unilaterally profit maximizing to produce system demand in a cost-effective manner possible because the sum of the QC_{jh} over the J suppliers is equal to system demand for the hour. Because this equality holds for all hours of the year, this incentive to produce system demand in a cost-effective manner also applies on a yearly basis.

3.4. Mechanics of Standardized Forward Contract Procurement Process

The SFPFCs are purchased through auctions several years in advance of delivery in order to allow new entrants to compete to supply this energy. Because the aggregate hourly values of these contract obligations are allocated to retailers based on their actual share of system demand during the month, this mechanism can easily accommodate retail competition. If one retailer loses load and another gains it during the month, the share of the aggregate hourly value of SFPFCs allocated to first retailer falls and the share allocated to the second retailer rises.

The wholesale market operator would run the auctions with oversight by the regulator. One advantage of the design of the SFPFC products is that a simple auction mechanism can be used to purchase each annual product. A multi-round auction could be run where suppliers submit the amount of QT they would like to sell for a given delivery period at the price for the current round. Each round the price would decrease until the amount suppliers are willing to sell at that price is less than or equal to aggregate amount of QT demanded.

The wholesale market operator would also run a clearinghouse to manage the counterparty risk associated with these contracts. All wholesale market operators currently do this for all participants in their short-term markets. In a number of US markets, the market operator also provides counterparty risk management services for long-term financial transmission rights, which is significantly different from performing this function for SFPFCs.

SFPFCs auctions would be run on an annual basis for annual producing making deliveries starting two, three and four years in the future. In steady state, auctions for incremental amounts of each annual contract would also be needed so that the aggregate share of demand covered by each annual SFPFC could increase over time. The eventual 100 percent coverage of demand occurs through a final true-up auction that takes place after the realized values for hourly demand for the delivery period are known.

Each purchase of the same annual SFPFC product is allocated to retailers according to their load shares during the delivery month. If three different size purchases are made for a given annual product at different prices then each retailer is allocated their load share for the month of these three purchases. This ensures a level playing field for retailers with respect to their long-term resource adequacy obligation. All retailers face the same average price for the long-term resource adequacy obligation associated with their realized demand for the month.

The advance purchase fractions of the final demand are the regulator's security blanket to ensure that system demands can met for all hours of the year for all possible future system conditions. If the regulator is worried that not enough resources will be available in time to satisfy this requirement, it can increase the share of final demand that it purchases in each annual SFPFC auctions. If too much QT is purchased in an annual auction, it can be sold back to generation unit owners in a later auction or the final true-up auction. The costs to these decisions will then be allocated to all retailers as part of their long-term resource adequacy obligation.

Cross hedging between controllable generation units and intermittent renewable resources under this mechanism is enforced by tying the amount of QT a generation unit owner can sell on an annual basis to the value of their firm energy. The system operator would assign firm energy values for each generation unit using a mechanism similar what is currently used to compute firm capacity values. Multiplying a unit's MWs of firm capacity by the number of hours in the year would be the unit's firm energy value, which is the upper bound on the amount of QT the unit owner could sell in all auctions for an annual compliance period. Because the firm capacity of a generation unit is defined as the amount of energy it can produce under stressed system conditions, this limitation on annual sales of QT, implies that intermittent wind and solar resources would sell much less QT than the total MWhs they expect to produce in an average year and controllable generation unit owners would sell significantly more QT than the total MWhs they expect to produce in an average year.

In most years, controllable resource owner j would be producing energy in a small number of hours of the year, but earning $(PC - PS_h)QC_{jh}$ in all the hours that it produces no output. Intermittent renewables owners would typically produce more than their SFPFC obligation in energy and sell the additional energy at the short-term price. In years with low renewable output near their SFPRC obligations, controllable resource owners would produce close to their QC_{jh} values and average short-term prices would be significantly higher because of that. Because of their SFPFC holdings, aggregate retail demand would be shielded from these high short-term prices.

3.4. Advantages of SFPRC Approach to Long-Term Resource Adequacy

This mechanism has a number of advantages relative to a capacity-based approach. There is no regulator-mandated aggregate capacity requirement. Generation unit owners are allowed to decide both the total MWs and of mix of technologies to meet their SFPFC energy obligations. There is also no prohibition on generation unit owners or retailers engaging in other hedging arrangements outside of this mechanism. Specifically, retailers could enter into a bilateral contract for energy with a generation unit owner or other retailer to manage the short-term price and quantity risk associated with the difference between their actual hourly load shape and their hourly values of QR_{kh}. This mechanism provides a nudge to market participants to develop a liquid market for these bilateral contract arrangements at horizons to delivery similar to the SFPRC products. Instead of starting from the baseline of no fixed-price forward contract coverage of system demand by retailers, this mechanism starts with 100 percent coverage of system demand, which retailers can unwind at their own risk.

For the regulated retail customers, the purchase prices of SFPRCs can be used to set the wholesale price implicit in the regulated retail price over the time horizon that the forward contract clears. This would provide retailers with a strong incentive to reduce their average wholesale energy procurement costs below this price through bilateral hedging arrangements, storage investments or demand response efforts.

There are a number of reasons why this mechanism should be a more cost-effective approach to long-term resource adequacy in a zero marginal cost intermittent future than a capacity-based mechanism. First, the sale of SFPFC energy starting delivery two or more years in the future provides a revenue stream that will significantly increase investor confidence in recovering the cost of any investment in new generation capacity.

Because retailers are protected from high short-term prices by total hourly SFPFC holdings equal to system demand, the offer cap on the short-term market can be raised in order to increase the incentive for all suppliers to produce as much energy as possible during stressed system conditions. The possibility of higher short-term price spikes will support investments in storage and load-shifting technologies and encourage active participation of final demand in the wholesale market, further enhancing system reliability in a market with significant intermittent renewable resources.

If SFPFC energy is sold for delivery in four years based on a proposed generation unit, the regulator should require that construction of the new unit to begin within a pre-specified number of months after the signing date of the contract or require posting of a substantially larger amount of money in the clearinghouse with market operator. Otherwise, the amount of SFPFC energy that this proposed unit sold would be automatically liquidated in a subsequent SFPFC auction and a financial penalty would be imposed on the developer. Other completion milestones would have to be met at future dates to ensure the unit is able to provide amount firm energy that it committed to provide in the SFPRC contract sold. If any of these milestones were not met, the contract would be liquidated.

4. Final Comments

There is no perfect wholesale market design. There are only better wholesale market designs and what constitutes a better design depends many factors specific to the region. Although there is general agreement on the key features of a best-practice short-term market design, many details must be adjusted to reflect local conditions. For this reason, wholesale market design is a process of continuous learning, adaption, and hopefully, improvement. The standardized energy contracting approach to long-term resource adequacy described in this paper is an example of this process. It has a number of features that are likely make it significantly better suited to a zero marginal cost intermittent renewables electricity supply industry, there are many details of this basic mechanism that should be adapted to reflect local conditions.

Further Reading

Bushnell, J., Hobbs, B. F., & Wolak, F. A. (2009). When it comes to demand response, is FERC its own worst enemy?. *The Electricity Journal*, 22(8), 9-18.

Gribik, P. R., Hogan, W. W., & Pope, S. L. (2007). Market-clearing electricity prices and energy uplift. *Cambridge, MA*.

Hogan, W. W. (2002). Electricity market restructuring: reforms of reforms. *Journal of Regulatory Economics*, *21*(1), 103-132.

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., & Bohn, R. E. (2013). *Spot pricing of electricity*. Springer Science & Business Media.

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, 1-20.

Wolak, Frank A. (2011) “Measuring the Benefits of Greater Spatial Granularity in Short-Term Pricing in Wholesale Electricity Markets,” *American Economic Review*, May, 247-252.

Wolak, F. A. (2016). Level versus Variability Trade-offs in Wind and Solar Generation Investments: The Case of California. *The Energy Journal*, *37*(Bollino-Madlener Special Issue).

Wolak, F. A. (2020) Wholesale Electricity Market Design, in *Handbook on the Economics of Electricity Markets*, J.M. Glanchant, P.L. Joskow, and M. Polllitt (editors), Cambridge University Press.

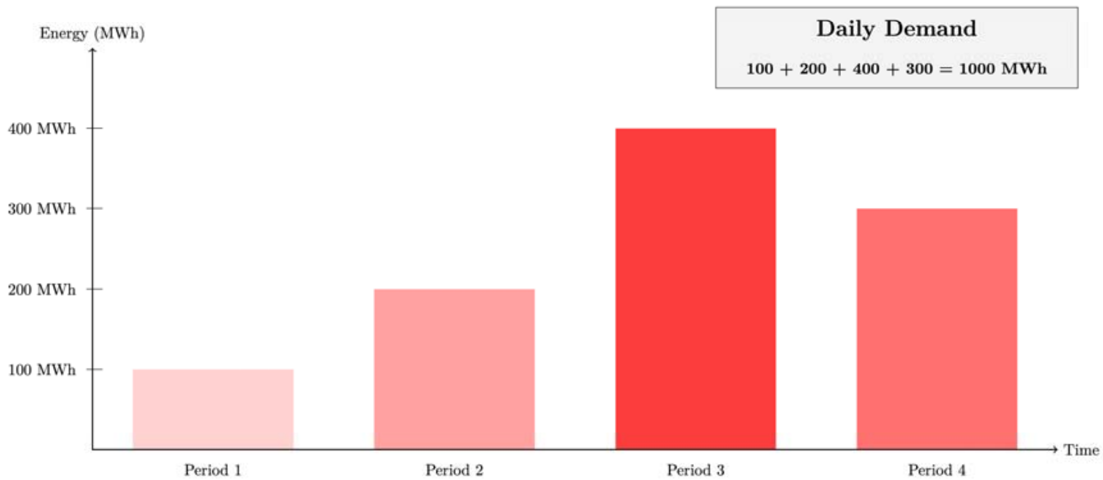

Figure 1: Hourly System Demands for Single Day

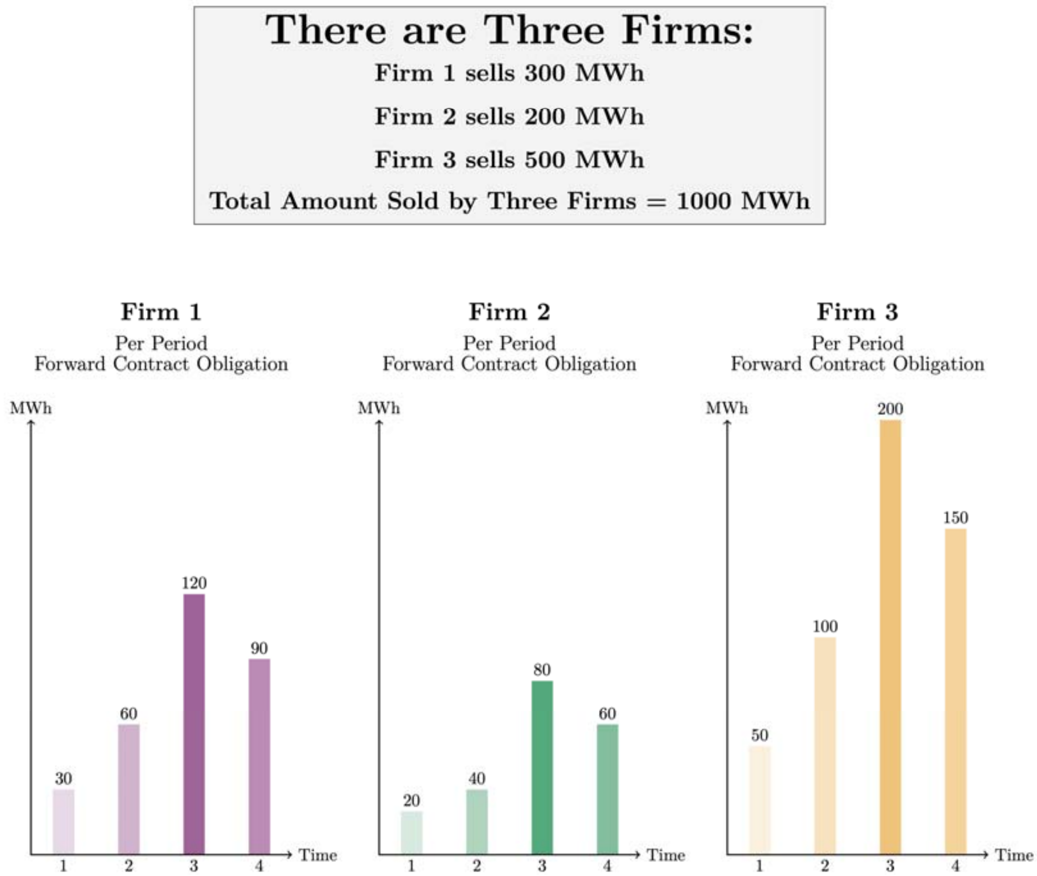

Figure 2: Hourly Forward Contract Quantities for Three Suppliers

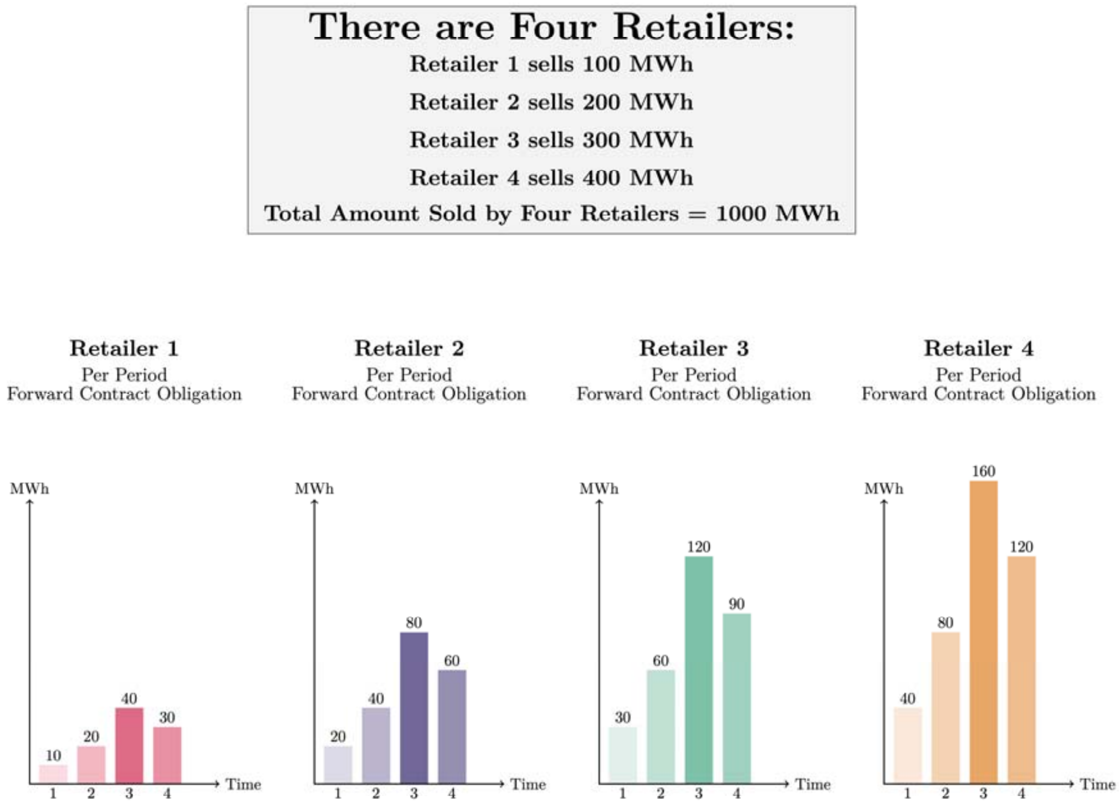

Figure 3: Hourly Forward Contract Quantities for Four Retailers