Financing the Energy Transition in a Low-Cost Intermittent Renewables Environment

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: November 7, 2021

Abstract

Declines in the up-front costs of both wind and solar generation units over the past decide has significantly closed the gap between the levelized cost of energy (LCOE) for these resources and the LCOE of natural gas and coal-fired generation. This outcome has the potential to reduce the cost of increasing the share of intermittent renewable resources in a region significantly. The experience of regions with significant shares of intermittent renewables is used to provide recommendations for short-term wholesale market design, a long-term resource adequacy mechanism and a renewables support mechanism to achieve a substantial intermittent renewable energy share at least cost to electricity consumers. A multi-settlement locational marginal pricing short-term market design, a standardardized fixed-price forward contract approach to long-term resource adequacy and a renewables energy certificates market are the major market design elements proposed to achieve this goal.

1. Introduction

Until recently, transitioning from a fossil fuel-dominated electricity supply industry to an intermittent-renewables-dominated low carbon electricity supply industry has required significant above-market financial support for investments in wind and solar generation resources because the levelized cost of energy (LCOE) from these resources was greater than the average market price at which the electricity produced could be sold, largely because the LCOE of natural gas and coal-fired generation units was less than that for wind and solar units.

Declines in the up-front costs of both wind and solar over the past decide has significantly closed the gap between the LCOE for these resources and the LCOE of natural gas and coal-fired generation. Figure 1 plots the annual global quantity weighted-average LCOE for grid scale wind and solar photovoltaic (PV) generation units that began operation during the for each year from 2010 to 2020 (IRENA, 2021). This graph also plots the annual quantity-weighted-average LCOE for residential and commercial solar PV generation units that began operation during the year (IRENA, 2021). As this graph demonstrates, in 2020 the LCOE of grid scale wind and solar is roughly one-third to one-quarter LCOE distributed solar energy.

The United States Energy Information Administration (EIA) estimates that the LCOE for a new combined cycle natural gas generation unit entering service in 2023 is $33.21/MWh versus $30.44/MWh and 30.63/MWh for grid scale wind and solar resources (Table A1a, EIA 2021). These LCOE differences signal a new regime for investments in wind and solar resources. However, because of the intermittency of wind and solar resources, there is still the need for a significant amount of dispatchable generation capacity to supply energy when the wind is not blowing or the sun is not shining. This low-cost intermittent renewables regime and the desire of policymakers to significantly increase the share of their jurisdiction's energy consumption coming from intermittent renewables argues for a paradigm shift in how these investments are financed.

The purpose of this paper is to explain why this paradigm shift is necessary and outline a short-term market electricity market design, long-term resource adequacy mechanism, and renewables support mechanism for this low-cost intermittent renewables regime. A multi-settlement locational marginal pricing (LMP) market design with the co-optimized procurement of ancillary services rewards quick response dispatchable resources and appropriately prices the intermittency of wind and solar resources. This market design also allows additional reliability constraints on system operation required by a larger share of intermittent renewables to be

incorporated into the energy and ancillary services market in a straightforward manner. Finally, this short-term market design prices electricity across both the space and time to provide financial incentives to locate storage and load flexibility investments where they can provide the greatest benefits to system reliability.

A fixed-price forward contract for energy approach to long-term resource adequacy is the most important change necessary to support large renewable energy shares under a low-cost intermittent renewables regime. A significant amount dispatchable generation capacity will still be required to produce energy when the underlying renewable resources are unavailable. However, these generation resources will start-up and shutdown more frequently and operate at increasingly smaller annual capacity factors as the share of intermittent renewables increases. Consequently, the long-term resource adequacy mechanism must encourage cross-hedging between intermittent renewables and dispatchable generation units. The intermittent renewable owners implicitly or explicitly purchase price spike insurance from dispatchable thermal resources for time periods when these renewables are unlikely to produce energy. This price spike insurance provides an annual revenue stream to dispatchable resources that contributes to their financial viability that they are available to supply energy despite operating with significantly lower annual capacity factors.

The ultimate success of this approach to long-term resource adequacy requires phasing out common approach to financing investments in intermittent renewables—paid as-delivered power purchase agreements. These long-term contracts pay the intermittent renewable generation unit owner a fixed-price or according to a fixed price or price schedule for all energy produced by the generation unit, regardless of when this energy is produced. These contracts provide an implicit subsidy to intermittent renewable resources because similar contract terms are not offered to dispatchable generation units. Moreover, paid-as-delivered contracts dull the financial incentive for intermittent renewable resource owners to pair their investments with storage capacity to manage the uncertainty in energy production from their units. As explained below, requiring all resources to sell standardized fixed-price and fixed-quantity forward contracts provides strong incentives for market mechanisms to find the least cost solution to meeting a given renewable energy target.

To provide the least cost amount of above-market revenues to intermittent renewables resources to meet a given renewable energy share goal, a renewables portfolio standard mechanism

is necessary. This mechanism separately prices the renewable attribute separate from the energy the intermittent renewable resource produces. This renewables support mechanism and a multi-settlement LMP market design with five-minute settlement in the real-time market provides strong incentives for investments in storage facilities necessary to achieve renewable energy shares greater than 50%.

The remainder of the paper proceeds as follows. Section 2 discusses the essential features of the short-term market design to support the least cost deployment of a large share intermittent renewables. Section 3 discusses the necessity of a long-term resource adequacy mechanism for all wholesale electricity market with finite offer caps on the short-term market and why the traditional capacity-based approach to long-term resource adequacy is poorly suited to regions with significant intermittent renewable energy goals. Section 4 introduces our proposed standardized fixed-price forward contract approach to long-term resource adequacy and explains why it is superior to solution to long-term resource adequacy for intermittent renewables dominated electricity supply industry. Section 5 explains why a renewable energy certificates market is necessary to achieve renewable energy shares above 50 percent. This section also explains why paid-as-delivered forward contracts for intermittent renewable energy for contrary to achieving large shares of renewable energy at least cost to electricity consumers.

2. Short-Term Market Design

An important lesson from electricity market design processes around the world is the extent to which the market mechanism used to dispatch and operate generation units is consistent with how the grid is operated. In the early stages of wholesale market design in the US, all the regions attempted to operate wholesale markets that used simplified versions of the transmission network. The single zone or zonal markets assumed infinite transmission capacity between locations in the transmission grid or only recognized transmission constraints across large geographic regions. These simplifications of the transmission network configuration and other relevant operating constraints creates opportunities for market participants to increase their profits by taking advantage of the fact that in real time the actual configuration of the transmission network and other operating constraints must be respected.

These markets set a single market-clearing price for a half-hour or hour for an entire country or large geographic region even though there were generation units with offer prices below the market-clearing price not producing electricity and units with offer prices above the 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' or 'must-run' units.

A market design challenge arises because how generation units are compensated for being constrained-on or constrained-off impacts the offer prices they submit into wholesale energy market. For example, if generation units are paid their offer price for electricity when they are constrained-on and the unit's owner knows that it will be constrained-on, a profit-maximizing unit owner will submit an offer price significantly higher than the variable cost of the unit and be paid that price for the incremental energy it supplies, which raises the total cost of electricity supplied to final consumers.

A similar set of circumstances arises for a constrained-off generation unit, which is usually paid the difference between the market-clearing price and the unit's offer price for not supplying electricity that the unit would have produced if not for the configuration of the transmission network. This market rule creates an incentive for a profit-maximizing supplier that knows its unit will be constrained-off to submit the lowest possible offer price in order to receive the highest possible payment for being constrained-off and raise the total cost of electricity supplied to final consumers.

This problem occurred so frequently in the early US zonal markets that it has acquired the name "the DEC game," because it involves a supplier selling energy in the day-ahead market that it knows is very likely to be infeasible to inject to the transmission grid in real-time. The supplier then agrees to buy decremental (DEC) energy at a price below the day-ahead market price and earns the difference between these two prices times the amount of decremental energy purchased for producing little or no energy in real-time. Bushnell, Hobbs and Wolak (2008) discuss this problem and the market efficiency consequences in the context of the initial zonal-pricing market in California. Graf, Quaglia, and Wolak (2020) document the incentives for generation unit owner offer behavior created by the divergence between the day-ahead zonal market model and full network model used to operate the Italian market in real-time. However, this outcome is not unique to markets in industrialized countries. Wolak (2009) discusses these same issues in the context of

the Colombian single-price market with its negative and positive reconciliation payments mechanism.

2.1. Locational Marginal Pricing

As described in the previous section, almost any difference between the market model used to set dispatch levels and market prices and the actual operation of the generation units needed to serve demand creates an opportunity for market participants to take actions that raise their profits at the expense of overall market efficiency. Wholesale electricity markets that use locational marginal pricing (LMP), also referred to as nodal pricing, largely avoid these constrained-on and constrained-off problems, because all transmission constraints and other relevant operating constraints are respected in the process of determining dispatch levels and locational marginal prices. Consequently, different from single-zone or zonal market designs, locational marginal pricing markets can allow multiple settlements without creating the opportunities for suppliers to degrade the efficiency of the short-term market by taking advantage of the constrained-on and constrained-off process discussed in the previous section.

All LMP markets in the US co-optimize the procurement of energy and operating reserves. This means that all suppliers submit to the wholesale market operator their generation unit-specific willingness-to-supply schedules for energy and any operating reserve the generation unit can provide. Likewise, large loads and load-serving entities submit their willingness-to-purchase energy schedules. Locational prices for energy and ancillary services and dispatch levels and ancillary services commitments for generation units at each location in the transmission network are determined by minimizing the as-offered costs of meeting the demand for energy and operating reserves at all locations in the transmission network, subject to all transmission network and other relevant operating constraints. No generation unit will be accepted to supply energy or an operating reserve if doing so would violate a transmission or other operating constraint.

An important distinction between an LMP market design and the standard zonal market design is the centralized commitment of generation units to provide energy and ancillary services. The zonal markets throughout Europe do not typically require generation units to submit energy offer curves into the day-ahead market and instead allow individual producers to make the commitment decisions for their generation units using simplified single-zone or multiple-zone models of the transmission network. A self-commitment market can result in higher cost generation units operating because of the differences among producers in their assessment of the

likely market price. Self-commitment markets also do not allow the simultaneous procurement of energy and operating reserves and instead rely on sequential procurement of operating reserves before or after energy schedules have been determined. As Oren (2001) demonstrates, sequential clearing of energy and operating reserves markets increases the opportunities for generation unit owners to exercise unilateral market power in the energy or operating reserves markets, because suppliers know that capacity sold in an earlier market cannot compete with suppliers in a subsequent market.

Another advantage of a centralized LMP market that co-optimizes the procurement of energy and operating reserves ensure that each generation unit is used in the most cost-effective manner based on the energy and operating reserves offers of all generation units, not just those owned by a single market participant. Specifically, the opportunity cost of supplying any operating reserve a unit can provide will be explicitly considered in deciding whether to use the unit for that ancillary service. For example, if the market-clearing price of energy at that generation unit's location is $40/MWh, the unit's offer price for energy is $30/MWh, and the unit's offer price for the only operating reserve the unit can supply is $5/MW, then the unit will not be accepted to supply that operating reserve. It is accepted to supply the operating reserve only if the price of this operating reserve is greater than or equal to $10/MW because of the $10/MWh (= $40/MWh - $30/MWh) opportunity cost of energy for that unit.

In contrast, self-commitment markets or sequential operating reserves markets such as those that exist in Europe and other industrialized countries must rely on individual market participants to make the efficient choice between supplying energy or ancillary services from each generation unit. This is possible for a supplier to do within its portfolio of generation units, but it unlikely to be the case across suppliers. Consequently, there are likely to many instances when a resource is taken to supply an operating reserve at a $/MW price that turns out to be less than unit's opportunity of providing energy. There are also likely to be instances when a resource is providing energy at price that has smaller opportunity cost of energy than the prevailing price of an operating reserve the unit can provide.

The nodal price at each location is the increase in the minimized value of the 'as-offered costs' objective function because of a one unit increase in the amount of energy withdrawn at that location in the transmission network. In a co-optimized energy and operating reserves locational marginal pricing market, the price of each operating reserve is defined as the increase in the

optimized value of the as-offered costs objective function due to a one unit increase in the demand for that operating reserve. In most LMP markets, operating reserves are procured at a coarser level of spatial granularity than energy. For example, energy is typically priced at the nodal level and operating reserves are priced over larger geographic regions. Bohn, Caramanis, and Schweppe (1984) provide an accessible discussion of the properties of the LMP market mechanism.

Another strength of the LMP market design is the fact that other constraints that the system operator considers in operating the transmission network can also be accounted for in setting dispatch levels and locational prices. For example, suppose that reliability studies have shown that a minimum amount of energy must be produced by a group of generation units located in a small region of the grid. This operating constraint can be built into the LMP market mechanism and reflected in the resulting locational prices. This property of LMP markets is particularly relevant to the cost-effective integration of a significant amount of intermittent renewable generation capacity in the transmission network because additional reliability constraints may need to be formulated and incorporated into LMP market to account for the fact that this energy can quickly disappear and re-appear.

An important lesson from the US experience with LMP markets is that explicitly accounting for the configuration of the transmission network in determining dispatch levels both within and across regions can significantly increase the amount of trade that takes place between the regions. Mansur and White (2012) dramatically demonstrate this point by comparing the volume of trade between two regions of the Eastern US, what the authors call the Midwest and East of PJM, before and after these regions were integrated into a single locational marginal pricing market that accounts for the configuration of the transmission network throughout the entire integrated region. Average daily energy flows from the Midwest to East of PJM almost tripled immediately following the integration of the two regions into an LMP market. There was no change in the physical configuration of the transmission network for the two regions. This increase in energy flows was purely the result of incorporating the two regions into a single LMP market that recognizes the configuration of the transmission network for the two regions in dispatching generation units.

2.2. Multi-Settlement Markets

Multi-settlement nodal-pricing markets have been adopted by all US jurisdictions with a formal short-term wholesale electricity market. A multi-settlement market has a day-ahead

forward market that is run in advance of real-time system operation. Generation unit owners submit unit-level offer curves for each hour of the following day for energy and operating reserves as well as technical characteristics of their generation units, such as ramp rates, minimum and maximum safe operating levels, and other operating characteristics required by the system operator. Large consumers and electricity retailers submit demand curves for energy for each hour of the following day. The system operator set the demands for each operating reserve and then minimizes the as-offered cost to meet the demand for energy and each operating reserve simultaneously for all 24 hours of the following day subject to the anticipated configuration of the transmission network and other relevant operating constraints. This gives rise to LMPs and firm financial commitments to buy and sell energy and each operating reserve each hour of the following day for all generation unit and load locations.

The day-ahead market typically allows generation unit owners to submit their start-up and minimum load cost offers as well as energy offer curves, and both costs enter the objective function used to compute hourly generation schedules and locational marginal prices for all 24 hours of the following day. This logic implies that a generation unit will not be dispatched in the day-ahead market unless the combination of its start-up and no-load costs and energy costs are part of the least cost solution to serving hourly demands for all 24 hours of the following day.

To the extent that generation unit owners do not receive sufficient revenues from energy and operating reserves sales to recover their as-offered costs to provide these products throughout the day, they are provided with a make-whole payment to recover these costs. Total make-whole payments are recovered from all loads through a $/MWh charge. For example, if a generation unit owner with a start-up cost of $5,000 and a variable cost of energy offer of $40/MWh sells 100 MWh at price of $42/MWh, the unit's make-whole payment would be $5,000 - $4,200 = $800. If system demand was 4,000 MWh and this was the only make-whole payment made, then the per unit charge to demand would be $0.50/MWh.

The energy schedules that arise from the day-ahead market do not require a generation unit to supply the amount sold or a load to consume the amount purchased in the day-ahead market. The only requirement is that any shortfall in a day-ahead commitment to supply energy must be purchased from the real-time market at that same location or any production greater than the day-ahead commitment is sold at the real-time price at that same location. For loads, the same logic applies. 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. Both buyers and sellers of energy in the day-ahead market bear the full financial consequences of failing to meet the day-ahead sales and purchases. However, it can often be the case that deviating from day-ahead schedules is profit-maximizing for a generation unit owner, distribution utility, or free consumer, particularly as the share of intermittent renewable resources increases.

In all United States wholesale markets, real-time LMPs are determined from the real-time offer curves from all available generation units and dispatchable loads by minimizing the as-offered cost to meet real-time demands (rather than bid-in demand) at all locations considering the current configuration of the transmission network and other relevant operating constraints. This process gives rise to LMPs at all locations in the transmission network and the actual hourly operating levels for all generation units. Real-time imbalances relative to day-ahead schedules are cleared at these real-time prices.

To understand how a two-settlement market works, suppose that a generation unit owner sells 50 MWh in the day-ahead market at 60 $/MWh. It receives a guaranteed $3,000 in revenues from this sale. However, if the generation unit owner fails to inject 50 MWh of energy into the grid during the specified delivery hour of the following day, it must purchase the energy it fails to inject at the real-time price at that location. Suppose that the real-time price at that location is 70 $/MWh and the generator only injects 40 MWh of energy during the hour in question. In this case, the unit owner must purchase the 10 MWh shortfall relative to its day-ahead schedule at 70 $/MWh. Consequently, the net revenues the generation unit owner earns from selling 50 MWh in the day-ahead market and only injecting 40 MWh is $2,300, the $3,000 of revenues earned in the day-ahead market less the $700 paid for the 10 MWh real-time deviation from the unit's day-ahead schedule.

If a generation unit produces more output than its day-ahead schedule, then this incremental output is sold in the real-time market. For example, if the unit produced 55 MWh, then the additional 5 MWh beyond the unit owner's day-ahead schedule is sold at the real-time price. By the same logic, a load-serving entity that buys 100 MWh in the day-ahead market but only withdraws 90 MWh in real-time, sells the 10 MWh not consumed at the real-time price. Alternatively, if the load-serving entity consumes 110 MWh, then the additional 10 MWh not purchased in the day-ahead market must be purchased at the real-time price.

A multi-settlement nodal-pricing market is well-suited to regions that do not have an extensive transmission network because it explicitly accounts for the configuration on the actual transmission network in setting both day-ahead energy schedules and prices and real-time output levels and prices. This market design eliminates much of the need for ad hoc adjustments to generation unit output levels that can increase the total cost of wholesale electricity to final consumers because of differences between the prices and schedules that the market mechanism sets and how the actual electricity network operates.

Wolak (2011) quantifies the magnitude of the economic benefits associated with the transition to a two-settlement nodal pricing market from a two-settlement zonal-pricing market that was very similar to the standard market design currently in Europe and other industrialized countries. Wolak (2011) find total hourly BTUs of fossil fuel energy consumed to produce electricity is 2.5 percent lower, the total hourly variable cost of production for fossil fuels units is 2.1 percent lower, and the total number of hourly starts is 0.17 higher after the implementation of nodal pricing. This 2.1 percent cost reduction implies a roughly $105 million reduction in the total annual variable cost of producing electricity from fossil fuels in California associated with the introduction of nodal pricing. Triolo and Wolak (2020) study the transition from a European-style zonal market design with self-scheduling and self-commitment to a multi-settlement nodal market design in the Electricity Reliability Council of Texas (ERCOT) on December 1, 2010. They find a 3.9% reduction in the total variable cost of fossil-fuel generation for the first year of operation of this market, or an annual cost savings of $323 million.

2.3. Multi-Settlement LMP Market with Significant Intermittent Renewables

A multi-settlement LMP market design is also particularly well suited to managing a generation mix with a significant share of intermittent renewable resources. The additional operating constraints necessary for reliable system operation with an increased number of renewable resources can easily be incorporated into the day-ahead and real-time market models. Therefore, the economic benefits from implementing a multi-settlement LMP market relative to market designs that do not model transmission and other operating constraints are likely to be greater the larger is the share of intermittent renewable resources. Bjorndal et. al (2018) shows that in region with significant wind resources even embedding a nodal market design within a larger zonal market design outperforms a full zonal market design. The authors also demonstrate that a nodal design for the entire region yield even greater savings relative to zonal design. Consequently,

any region with significant renewable energy goals is likely to realize significant economic benefits from implementing a multi-settlement LMP market.

This short-term market design values the dispatchability and flexibility of generation units even though it pays all resources at the same location in the grid the same price in the day-ahead and real-time markets. Suppose that a wind unit sells 50 MWh and a thermal resource sells 40 MWh in the day-ahead market at $30/MWh. If in real time not as much wind energy is produced, the dispatchable thermal unit must make up the difference. Suppose that the wind unit produces only 30 MWh, so that the thermal unit must produce an additional 20 MWh. Because of this wind generation shortfall, the real-time price is now $60/MWh. Under this scenario, the wind unit is paid an average price of $10/MWh = (50 MWh x $30/MWh – 20 MWh x $60/MWh)/30 MWh for the 30 MWh it produces, whereas the dispatchable thermal unit is paid an average price of $40/MWh = (40 MWh x $30/MWh + 20 MWh x $60/MWh)/60 MWh for the 60 MWh it produces.

A similar logic applies to the case that the wind resource produces more than expected and the thermal resource reduces its output because the real-time price is lower than the day-ahead price due to the unexpectedly large amount of wind energy produced. For example, suppose the wind unit sells 30 MWh and the thermal resource sells 60 MWh in the day-ahead market at $30/MWh. However, in real time there is significantly more wind, so that the wind unit produces 50 MWh at a real-time price of $10/MWh. Because of this low real-time price, the thermal resource decides to produce 40 MWh and purchases the additional 20 MWh from its day-ahead energy schedule from the real-time market. The average price received by the wind unit is $22/MWh = (30 MWh x $30/MWh + 20 MWh x $10/MWh)/50 MWh and the average price received by the thermal unit is $40/MWh = (60 MWh x $30/MWh – 20 MWh x $10/MWh)/40 MWh. Despite paying the same price to all energy in the day-ahead and real-time markets, a multi-settlement market pays a higher average price to the dispatchable generation unit for the energy it provides during the same hour as the wind unit.

An additional way to reward flexibility in a multi-settlement LMP market is to clear the real-time market as frequently as possible within the hour. For example, all United States wholesale markets clear--set real-time prices and dispatch levels—every five minutes. This means that real-time prices can increase rapidly across 5-minute intervals when net system demand—the difference between system demand and intermittent renewable generation—rapidly increases. This rewards generation units that can quickly increase their output with substantially higher prices for

the output they supply within that 5-minute interval. Units that can rapidly reduce their output in response to an increase in net demand during a 5-minute interval can sell back energy scheduled in the day-ahead market at substantially lower prices.

Shorter settlement intervals can also reduce the demand for frequency response operating reserves, because more fast-response units are moving up and down according to 5-minute dispatch instructions within the hour, so that less secondary frequency up and less secondary frequency down is needed to maintain system balance within the hour. More frequent settlement of the real-time market rewards dispatchable resources for the quick response and flexibility that they provide, particularly if share of intermittent renewable generation increases significantly.

2.4. A Cost-Based Multi-Settlement LMP Market

The transition to formal market mechanisms in a number of developing and small countries has been slow. Several regions in Africa, Latin America and Asia proposed wholesale markets in the early 2000s, but these regions have yet to begin operating a formal market mechanism. These regions frequently face significant challenges because of limited transmission capacity between and within their member countries. Consequently, any attempt to operate an offer-based market for most developing countries is likely to run into severe local and system-wide market power problems. In addition, the almost complete absence of hourly meters in these regions limits the opportunities for active demand-side participation, which makes implementing an offer-based wholesale market even more challenging.

Building on the experience of Latin American countries discussed in Wolak (2014), a viable market design for these regions is a cost-based short-term market that uses locational marginal pricing. This market design is straightforward to implement because it simply involves solving for the optimal dispatch of generation units in the region based on the market operator's estimate of each unit's variable cost subject to the operating constraints implied by the actual regional transmission network and other reliability constraints.

All generation unit owners would submit the characteristics–for instance the heat rate and the amount of fuel required to start up the unit--of their generation units to the market operator, which then determines the variable cost for each generation unit using a publicly available price index for the unit's input fossil fuel. For example, for a coal-fired generation unit, the market operator could use a globally traded price for coal and a benchmark delivery cost to the generation unit to determine the fuel cost of the unit. This would be multiplied by the unit's heat rate to

compute its variable fuel cost. An estimate of the variable operating and maintenance cost for the unit could be added to this variable fuel cost to arrive at the total variable cost of the unit. In order to provide incentives to minimize their actual total variable cost of producing electricity, the values of the components of the total variable cost could be based on benchmark values for the technology used by the generation unit owner, rather than an estimate of that unit owner's variable cost.

The variable cost computed by the market operator along with the configuration of the transmission network would be used to set day-ahead schedules and prices for each location in a multi-settlement version of this market design. In real time, the dispatch and locational marginal pricing process would be completed using the actual system demand and actual configuration of the transmission network with these same generation unit-level variable cost figures.

It is important to emphasize that this short-term market is only for settling imbalances relative to long-term contracts for energy. Joskow (1997) argues that the majority of the economic benefits from the electricity industry restructuring are likely to come from more efficient investment decisions in new generation capacity. The combination of a cost-based short-term market and fixed-price forward contract mandates on electricity retailers as discussed in Section 4 is a low-cost and low-regulatory burden approach to realizing more efficient investments in new generation capacity.

This market design also has the advantage that it can easily transition to an offer-based market once the transmission network in the region is expanded, hourly meters are deployed and the regulator is able to design an effective local market power mitigation mechanism. The LMP market is in already place and generation unit owners' costs as computed by the market operator can easily be replaced by the offers of these producers. Starting from a cost-based market and transitioning to an offer-based market is a low-risk approach to introducing an offer-based market. The PJM Interconnection in the Eastern US followed this strategy during the early stages of its development. It ran one year as a cost-based market before transitioning to an offer-based market.

3. The Reliability Externality and Long-Term Resource Adequacy

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. Many goods are produced using

high fixed cost, low marginal cost, technologies similar to electricity supply. Nevertheless, these firms recover their production costs, including a return on the capital invested, by selling their output 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 Wolak (2013) calls a *reliability externality.*

3.1. The Reliability Externality

Different from the case of wholesale electricity, the market for automobiles and air travel does not have a regulatory limit on the level of the short-term price. 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 and can result in as many different prices paid for the same flight as there are customers on the flight.

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 set by the regulator 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 their expected purchases 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 firm load until the available supply offered at or below the offer cap equals the reduced level of demand, as occurred several times in California between January 2001 and April 2001, and most recently on August 14 and 15 of 2020. A similar, but far more extreme set

of circumstances arose from February 14 to 18, 2021 in Texas and this required significant demand curtailments from February 15 to 18.[1]

Because random curtailments of supply to different distribution grids served by the transmission network—also known as rolling blackouts—are used to make demand equal to the available supply at or below the offer cap under these system conditions, this mechanism creates a *reliability externality* because no retailer bears the full cost of failing to procure adequate energy to meet their demand 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 another retailer of the same size 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. When short-term prices rise because of the supply shortfalls, retailers that do not hedge their wholesale energy purchases will go bankrupt. If they attempt to pass these short-term prices on to their retail customers, many are likely to be unable to pay their electricity bills. As discussed in Section 4.4.2 of Wolak (2021), both outcomes occurred in Texas following the events of February 14 to 18, 2021.

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, there is a missing market for long-term contracts for energy with long enough delivery horizons into the future to allow new generation units to be financed and constructed to serve demand under all future conditions in the short-term market. Therefore, a regulator-mandated long-term resource adequacy mechanism is necessary to replace this missing market.

Regulatory intervention is necessary to internalize the resulting *reliability externality* unless the regulator is willing to eliminate the offer cap and commit to allowing the short-term price to clear the real-time market under all possible system conditions. There are no short-term wholesale electricity markets in the world that make such a commitment. All of them have either explicit or implicit caps on the offer prices suppliers can submit to the short-term market. The Electricity Reliability Council of Texas (ERCOT) has a $9,000/MWh offer cap, which is highest

[1] See http://www.ercot.com/content/wcm/key_documents_lists/225373/Urgent_Board_of_Directors_Meeting_2-24-2021.pdf.

in the United States. The National Electricity Market (NEM) in Australia, has a 15,000 Australia dollars per MWh offer cap, which is currently the highest in world.

As the experience of February 14-18, 2021 in Texas demonstrated, an extremely high offer cap on the short-term market does not eliminate the *reliability externality*. It just shrinks the set of system conditions when random curtailments are required to balance real-time supply and demand. For the same reason, there also have been a small number of instances when the NEM of Australia has experienced supply shortfalls despite having an extremely high offer cap.

3.2. Conventional Solution to Reliability Externality with Intermittent Renewables

Currently, the most popular approach to addressing the *reliability externality* is a capacity procurement mechanism that assigns a firm capacity value to each generation unit based on the amount of energy it can provide under stressed system conditions. Retailers are then required to demonstrate that they have purchased sufficient firm capacity to meet their monthly or annual demand peaks. Having sufficient firm capacity typically means that the retailer has purchased firm capacity equal to between 1.10 and 1.20 times its annual demand peak. The exact multiple of peak demand chosen by a region depends on the mix of generation resources and the reliability requirements of the system operator.

Under the current long-term resource adequacy mechanism in California, firm-level capacity procurement obligations are assigned to retailers by the California Public Utilities Commissions (CPUC) to ensure that monthly and annual system demand peaks can be met. Electricity retailers are free to negotiate bilateral capacity contracts with individual generation unit owners to purchase firm capacity to meet these obligations. The eastern United States wholesale electricity markets in the PJM Interconnection, ISO-New England, New York ISO, and Midcontinental ISO (MISO) markets all have a centralized market for firm capacity. These involve periodic capacity auctions run by the wholesale market operator where all retailers purchase their capacity requirements at a market clearing price. ERCOT does not currently have formal long-term resource adequacy mechanism besides its $9,000/MWh offer cap and an ancillary services scarcity pricing mechanism.

All capacity-based approaches to long-term resource adequacy rely on the credibility of the firm capacity measures assigned to generation units. This is a relatively straightforward process for dispatchable thermal units. As noted earlier, the nameplate capacity of the generation unit times its annual availability factor--the fraction of hours of the year a unit is expected to be

available to produce electricity--is the typical starting point for estimating the amount of energy the unit can provide under stressed system conditions. As discussed below, if all retailers have met their firm capacity requirements in a sizeable market with only dispatchable thermal generation, there is a very high probability that the demand for energy will met during peak demand periods.

A simple example helps to illustrate the logic behind this claim. Suppose that the peak demand for the market is 1,000 MW and the market is composed of equal size generation units and each unit has a 90% annual availability factor, meaning that it is available to produce electricity any hour of the year with a 0.90 probability. Suppose that the event that one generation unit fails to operate is independent of the event that any other generation unit fails to operate. This independence assumption is reasonable for dispatchable thermal generation units because unavailability is typically due to an event specific to that generation unit. If each generation unit has a nameplate capacity of 100 MW, each has a firm capacity of 90 MW (= 0.90 x 100 MW). If there are 13 generation units, then with probability 0.96 demand peak will be covered.[2] In this case, a firm capacity requirement of 1.17 times the demand peak would ensure that system demand is met with 0.96 probability. Assuming that each generation unit is one-tenth of the system demand peak is unrealistic for most electricity supply industries, but it does illustrate the important point that smaller markets require firm capacity equal to a larger multiple of peak demand to achieve the desired level of reliability of supply.

Suppose that each generation unit is now 50 MW and each still has the same availability factor, so the firm capacity of each unit is now 45 MW. In this case, the same firm capacity requirement of 1.17 times the demand peak, or 26 generation units, would ensure system demand is met with 0.988 probability. If each generation unit had a nameplate capacity of 20 MWs with the same availability factor, each unit would have a firm capacity of 18 MW. This 1.17 times peak demand firm capacity requirement, or 65 generation units, would ensure that system demand is met with 0.999 probability. This example illustrates that an electricity supply industry based on dispatchable thermal generation units, where each unit has an independent 10 percent probability of being unavailable, the system demand peak will be met with a very high probability with a firm

[2] The number of generation units available is a binomial random variable with probability p = 0.9 and with number of trials N = the number of generation units. The probability of meeting the demand peak is the probability the available capacity is greater than or equal to the peak demand.

capacity requirement of 1.17 times peak demand if all the generation units are small relative to the system demand peak.

Introducing renewables into a capacity-based long-term resource adequacy mechanism considerably complicates the problem of computing the probability of meeting system demand peaks for two major reasons. First, the ability to produce electricity depends on the availability of the underlying renewable resource. A hydroelectric resource requires water behind the turbine, a wind resource requires wind to spin the turbine, and a solar facility requires sunlight to hit the solar panels. Second, and perhaps most important, the availability of water, wind, or sunshine to renewable generation resources is highly positively correlated across locations for a given technology within a given geographic region. This fact invalidates the assumption of independence of energy availability across locations that allows a firm capacity mechanism to ensure system demand peaks can be met with a very high probability. For example, if the correlation across locations in the availability of generation units is sufficiently high, then a 0.9 availability factor at one location would imply only a slightly higher than a 0.9 availability factor for meeting system demand, almost regardless of the amount intermittent renewable capacity that is installed.

Hydroelectric facilities have been integrated into firm capacity regimes by using percentiles of the distribution of past hydrological conditions for that generation unit to determine its firm capacity value. However, this approach only partially addresses the problem of accounting for the high degree of contemporaneous correlation across locations in water availability in hydroelectric dominated systems. There is typically a significant amount of data available on the marginal distribution of water availability at individual hydroelectric generation units. However, the joint distribution of water availability across all hydro locations is likely to be more difficult to obtain. The weather-dependent intermittency in energy availability for hydroelectric resources is typically on an annual frequency. There are low-water years and high-water years depending on global weather patterns such as the El Nino and La Nina weather events as discussed in McRae and Wolak (2016).

Incorporating wind or solar generation units into firm capacity mechanism is even more challenging for several reasons, and increasingly so as the share of energy produced in a region from these resources increases. The intermittency in energy supply is much more frequent than it is for hydroelectric energy. There can be substantial differences across and within days in the output of wind and solar generation units. Moreover, if stressed system conditions occur when it

is dark, the firm capacity of a solar resource is zero. Similarly, if stressed system conditions occur with the wind is not blowing, a likely outcome on extremely hot days, the firm capacity of a wind resource is zero.

The contemporaneous correlation across locations in the output of solar or wind generation resources for a given geographic area is typically extremely high. There is even a high degree of correlation across locations in the output of wind and solar resources. Wolak (2016) demonstrates the extremely high degree of contemporaneous correlation between the energy produced each hour of the year by solar and wind facilities in California. Again, information on marginal distribution of wind or solar energy availability at a location is much more readily available than the joint distribution of wind and solar energy availability for all wind and solar locations in a region. For these reasons, calculating a defensible estimate of the firm capacity of a wind or solar resource that is equivalent to the firm capacity of a dispatchable thermal generation resource is extremely difficult, if not impossible.

The high degree of contemporaneous correlation across locations in hourly capacity factors requires a methodology for computing firm capacity that accounts for the joint distribution of hourly capacity factors across locations throughout the year. Not only does this methodology need to account for the contemporaneous correlation in capacity factors across locations, but also the high degree of correlation of capacity factors over time for the same locations and other locations. California currently uses an Effective Load Carrying Capacity (ELCC) methodology for computing the firm capacity values of wind and solar generation units. The ELCC methodology was introduced by Galvin (1966), and it measures the additional load that the system can supply from a specified increase in the MWs of that generation technology with no net change in reliability. The loss of load probability, which is the probability that system demand will exceed the available supply, is the measure of reliability used in the ELCC calculation. Consistent with the results of Wolak (2016), the ELCC values for solar generation resources in California have declined as the amount of solar generation capacity in the state has increased.

For example, a recent study prepared for California's three investor-owned utilities (Carden et al. (2020)), Southern California Edison, Pacific Gas and Electric, and San Diego Gas and Electric, recommended ELCC values for a MW of fixed-mount solar photovoltaic capacity for 2022 of approximately 5 percent of the nameplate capacity. Their estimates for 2026 are less than

half that amount and those for 2030 are less than one-fourth that amount. These declines in ELCC values are due to the forecast increase in the amount of solar generation capacity in California.

An additional problem with computing the firm capacity of a solar or wind generation resources using the ELCC methodology is that the same investment in megawatts of wind or solar capacity is likely to be able to serve different increments to system demand depending on the location of the investment, the location of the increment to demand, and the size and location of other renewable resources in the region. This leaves the system operator with two difficult choices for setting the value of firm capacity for solar and wind resources. The first would be to set different values of firm capacity for resources based on their location in the transmission network. This would likely be a very politically contentious process because of the many assumptions that go into computing the ELCC of a resource. The second approach would set the same firm capacity value for all resources employing the same generation technology. This means that two resources with very different ELCC values could sell the same product to the potential detriment of overall system reliability.

Wolak (2021) evaluates the performance of California's capacity-based long-term resource adequacy (RA) mechanism based on the experience of August 14 to 18, 2020. Except for May for wind and July for solar, the monthly values of firm capacity computed using the ELCC methodology are slightly below the average capacity factors for the month. However, it is important to bear in mind that the firm capacity of a generation unit is supposed to measure what the facility can reliably produce under extreme system conditions, not what it produces on average. Consequently, a monthly average capacity factor less than the firm capacity value assigned to wind or solar generation resources provides further evidence against the viability of a capacity-based long-term resource adequacy mechanism with a large share of intermittent renewables. This outcome implies there are many hours in the month when the intermittent wind or solar resource is producing less than its firm capacity. Given the unpredictable intermittent nature of these resources, there is a non-zero probability this outcome will occur during a time with stressed system conditions, similar to those that occurred in August of 2020.

These facts, and the fact that what is predicted to be the major source of electricity in the future in California has been estimated to have a little firm capacity value in the future, imply that it would be prudent for California to consider alternatives to its capacity-based long-term resource mechanism if it intends to meet its goals of obtaining 60 percent of the state's energy from

renewable sources by 2030 and increase the use of electricity in space heating and personal transportation.

4. Standardized Fixed-Price Forward Contract (SFPFC) Approach to Long-Term RA

As the previous sections have demonstrated, a capacity-based approach to long-term resource adequacy is poorly suited to a region with significant intermittent renewables. The primary reliability challenge is not adequate generation capacity to serve demand peaks, but adequate energy available to serve realized demand during all hours of the year. As the example of California on August 14 and 15 of 2020 demonstrates, supply shortfalls do not necessarily occur during system demand peaks, but during net demand peaks.

Because of the substantial contemporaneous correlation in hourly output across locations and across renewable energy technologies ensuring sufficient supply to meet demand throughout the year will require taking full advantage of the mix of available generation resources. Intermittent renewable resources must reinsure the energy they sell in the forward market with dispatchable generation resources and storage devices. The long-term resource adequacy mechanism must also recognize the increasing weather dependence of electricity demand with more customers heating and cooling their homes with electricity.

The Standardized Fixed Price Forward Contact (SFPFC) mechanism results in the realized system demand each hour of the compliance period being covered by a fixed-price forward contract. The SFPFC approach to long-term resource adequacy recognizes that a supplier with the ability to serve demand at a reasonable price may not do so if it has the ability to exercise unilateral market power in the short-term energy market. As Wolak (2000) demonstrates, an expected profit-maximizing supplier with the ability to exercise unilateral market power with a fixed-price forward contract obligation would like to minimize the cost of supplying the quantity of energy sold in forward contract. The SFPFC long-term resource adequacy mechanism takes advantage of this incentive by requiring retailers to hold hourly fixed-price forward contract obligations for energy that sum to the hourly value of system demand. The SFPFC mechanism implies that all expected profit-maximizing suppliers would like to minimize the cost of meeting their hourly fixed-price forward contract obligations, the sum of which equals the hourly system demand for all hours of the year.

To understand the logic behind the SFPFC mechanism, consider the example of a supplier that owns 150 MWs generation capacity that has sold 100 MWh in a fixed-forward contract at a

price of $25/MWh for a certain hour of the day. This supplier has two options for fulfilling this forward contract: (1) produce the 100 MWh energy from its own units at their marginal cost of $20/MWh or (2) buy this energy from the short-term market at the prevailing market-clearing price. The supplier will receive $2,500 from the buyer of the contract for the 100 MWh sold, regardless of how it is supplied. This means that the supplier maximizes the profits it earns from this fixed-price forward contract sale by minimizing the cost of supplying the 100 MWh of energy.

To ensure that the least-cost "make versus buy" decision for this 100 MWh is made, the supplier should offer 100 MWh in the short-term market at its marginal cost of $20/MWh. This offer price for 100 MWh ensures that if it is cheaper to produce the energy from its generation units—the market price is at or above $20/MWh—the supplier's offer to produce the energy will be accepted in the short-term market. If it is cheaper to purchase the energy from the short-term market—the market price is below $20/MW—the supplier's offer will not be accepted and the supplier will purchase the 100 MWh from the short-term market at a price below $20/MWh.

This example demonstrates that the SFPFC approach to long-term resource adequacy makes it expected profit maximizing for each seller to minimize the cost of supplying the quantity of energy sold in this forward contract each hour of the delivery period. By the logic of the above example, each supplier will find it in its unilateral interest to submit an offer price into the short-term market equal to its marginal cost for its hourly SFPFC quantity of energy, in order to make the efficient "make versus buy" decision for fulfilling this obligation.

The incentives for supplier offer behavior in a short-term wholesale electricity market created by a fixed-price forward contract obligation are analyzed in Wolak (2000). Consider the case of a single hour in the short-term market. Let QS equal the amount of energy produced and sold in the short-term market by the supplier, PS is the short-term wholesale price, PC is the price of SFPFC energy, and QC is the quantity of SFPFCs sold by the supplier for this hour. The supplier's variable profit for the hour is:

$$\text{Profit} = PS \times QS - C(QS) - (PS - PC) \times QC = PS \times (QS - QC) + PC \times QC - C(QS) \quad (1)$$

where C(QS) is variable cost of producing QS. The first term in the first expression in (1) shows the supplier's variable profits from selling QS MWhs at PS in the short-term market. The second term is the net payment to the seller of QC SFPFC contracts at price PC. The second expression in the above equation demonstrates that a supplier only has an incentive to raise the short-term price if it sells more energy in the short-term market, QS, than its fixed-price forward contract

obligation, QC. This expression also demonstrates that the supplier wants the lowest possible price when it sells less energy in the short-term market than its fixed price forward contract obligation.

Under the SFPFC mechanism, each supplier knows that the sum of the values of the hourly SFPFC obligations across all suppliers is equal the system demand. This means that each supplier of SFPFCs knows that its competitors have substantial fixed-price forward contract obligations for that hour. This implies that all suppliers know that they have limited opportunities to raise the price they receive for short-term market sales beyond their hourly SFPFC quantity.

As discussed below, a supplier's fixed price forward quantity for an hour under the SFPFC mechanism increases with the value of hourly system demand. Therefore, the supplier that owns 150 MWs of capacity in the above example has a strong incentive to submit an offer price close to its marginal cost for the capacity of its generation unit to ensure that its hourly production is higher than the realized value of its SFPFC energy for that hour. Therefore, the SFPFC mechanism not only ensures that system demand is met every hour of the year, but it also provides strong incentives for this to occur at the lowest possible short-term price.

4.1. 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 in the current year, 95 percent of realized 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. The SFPFCs would clear against the quantity-weighted average of the hourly locational prices at all load withdrawal locations in the short-term wholesale market.

SFPFCs are shaped to the hourly system demand within the delivery period of the contract. Figure 2 contains a sample pattern of system demand for a four-hour delivery horizon. The total demand for the four hours is 1000 MWh, and the four hourly demands are 100 MWh, 200 MWh, 400 MWh and 300 MWh. Therefore, Firm 1 that sells 300 MWh of SFPFC energy has the hourly system demand-shaped forward contract obligations of 30 MWh in hour 1, 60 MWh in hour 2, 120 MWh in hour 3 and 90 MWh in hour 4. The hourly forward contract obligations for Firm 2 that sold 200 MWh SFPFC energy and Firm 3 that sold 500 MWh of SFPFC energy are also shown in

Figure 3. These SFPFC obligations are also allocated across the four hours according to the same four hourly shares of total system demand shown in Figure 2. This ensures that the sum of the hourly values of the forward contract obligations for the three suppliers is equal to the hourly value of system demand. Taking the example of hour 3, Firm 1's obligation is 120 MWh, Firm 2's is 80 MWh and Firm 3's is 200 MWh. These three values sum to 400 MWh, which is equal to the value of system demand in hour 3 shown in Figure 2.

These standardized fixed-price forward contracts are allocated to retailers based on their share of system demand during the month. Suppose that the four retailers in Figure 4 consume 1/10, 2/10, 3/10, and 4/10, respectively, of the total energy consumed during the compliance month for SFPFCs. This means that Retailer 1 is allocated 100 MWh of the 1000 MWh SFPFC obligations for the four hours, Retailer 2 is allocated 200 MWh, Retailer 3 is allocated 300 MWh, and Retailer 4 is allocated 400 MWh. The obligations of each retailer are then allocated to the individual hours using the same hourly system demand shares used to allocate the SFPFC energy sales of suppliers to the four hours. This allocation process implies Retailer 1 holds 10 MWh in hour 1, 20 MWh in hours 2, 40 MWh in hour 3 and 30 MWh in hour 4. Repeating this same allocation process for the other three retailers yields the remaining three hourly allocations shown in Figure 4. Similar to the case of the suppliers, the sum of allocations across the four retailers for each hour equals the total hourly system demand. For period 3, Retailer 1's holding is 40 MWh, Retailer 2's is 80 MWh, Retailer 3's is 120 MWh, and Retailer 4's is 160 MWh. The sum of these four magnitudes is equal to 400 MWh, which is the system demand in hour 3.

4.2. Mechanics of Standardized Forward Contract Procurement Process

The SFPFCs would be 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 SFPFC 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 the 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 total amount of annual SFPFC energy they would like to sell for a given delivery period at the price for

the current round. Each round of the auction the price would decrease until the amount suppliers are willing to sell at that price is less than or equal to the aggregate amount of SFPFC energy demanded.

The wholesale market operator would also run a clearinghouse to manage the counterparty risk associated with these contracts. All US wholesale market operators currently do this for all participants in their energy and ancillary services markets. In several US markets, the market operator also provides counterparty risk management services for long-term financial transmission rights, which is not significantly different from performing this function for SFPFCs. Both buyers and sellers would be required to post collateral with the wholesale market operator to ensure that each market participant finds it unilaterally profit-maximizing to meet its financial commitments for the SFPFC energy that it has purchased or sold.

SFPFCs auctions would be run on an annual basis for 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. The mechanics of true-up auction is described in Wolak (2021).

4.3. Incentives for Behavior by Intermittent Renewable and Controllable Resources

Because all suppliers know that all energy consumed every hour of the year is covered by a SFPFC in the current year and into the future, there is a strong incentive for suppliers to find the least cost mix of intermittent and controllable resources to serve these hourly demands. To the extent that there is concern that the generation resources available or likely to be available in the future to meet demand are insufficient, features of the existing capacity-based resource adequacy mechanism can be retained until system operators have sufficient confidence in this mechanism leading to a reliable supply of energy. The firm capacity values from the existing capacity-based long-term resource adequacy approach can be used to limit the amount of SFPFC energy a supplier can sell.

The firm capacity value multiplied by number of hours in the year would be the maximum amount of SFPFC energy that the unit owner could sell in any given year. Therefore, a controllable thermal generation unit owner could sell significantly more SFPFC energy than it expects to produce annually, and an intermittent renewable resource owner could sell significantly less

SFPFC energy than it expects to produce annually. This upper bound on the amount of SFPFC energy any generation unit could sell enforces cross-hedging between controllable in-state generation units and intermittent renewable resources. This mechanism uses the firm capacity construct to limit forward market sales of energy by individual resource owners to ensure that it is physically feasible to serve demand during all hours of the year.

Cross-hedging between a controllable resource and an intermittent resource implies that in most years, the controllable resource owner would be producing energy in a small number of hours of the year but earning the difference between the price at which they sold the energy in the SFPFC auction and the hourly short-term market price times the hourly value of its SFPFC energy obligation for all the hours that it does not produce energy. Intermittent renewables owners would typically produce more than their SFPFC obligation in energy and sell an energy produced beyond this quantity at the short-term price. In years with low renewable output near their SFPRC obligations, controllable resource owners would produce close to the hourly value of their SFPFC energy obligation, thus making average short-term prices significantly higher. However, aggregate retail demand would be shielded from these high short-term prices because of their SFPFC holdings.

4.4. Empirical Evidence on the Performance of the SFPFC Mechanism

Although the SFPFC mechanism in the form described above does not exist in any currently operating electricity supply industry, the long-term resource adequacy mechanisms in Chile and Peru create the same set of incentives for supplier behavior as the SFPFC mechanism by assigning system-wide short-term price and quantity risk during all hours of the year to suppliers. Both Chile and Peru operate a supplier-only, cost-based short-term wholesale electricity market. The system operator employs regulated variable cost estimates for each generation unit and an opportunity cost of water for hydroelectric generation units to dispatch generation units to meet demands throughout each country. All consumers or their retailers are required to purchase full requirements contracts from suppliers to meet their retail load obligations. Suppliers financially settle imbalances between the amount of energy they produce and the amount of energy their customers consume under these full-requirements contracts. Suppliers that produce more

energy than their customers consume receive payments from the suppliers that produce less energy than their customers consume.[3]

To see the equivalence of the incentives created for supplier behavior under the market designs in Chile and Peru and the SFPFC mechanism, let QR_i equal the consumption of customers served by the supplier i and PR_i the quantity-weighted average price paid for full-requirements contracts by customers served by supplier i. Let system demand equal QD, which is also equal to $\sum_{i=1}^{N} QR_i$, the sum of the consumption of all customers served by the N suppliers. The variable profit of supplier i is equal to

$$Profit_i = PS \times QS - C(QS) - (PS - PR_i) \times QR_i = PS \times (QS - QR_i) + PR_i \times QR_i - C(QS), \quad (2)$$

which is identical to equation (1) present earlier by setting QR_i equal to QC and PR_i equal to PC. Moreover, because $QD = \sum_{i=1}^{N} QR_i$, all short-term price and quantity risk is borne jointly by the N suppliers that have sold full requirements contracts.

The long-term resource adequacy mechanisms in Chile and Peru have delivered a reliable supply of electricity for at least the past 15 years in each country in the face of significant hydroelectric energy supply uncertainty and an increasing share of the energy consumed coming from intermittent wind and solar generation units. This outcome has been achieved through a cost-based short-term market in two countries with average annual load growth rates that are three to four times that in regions in the United States with formal wholesale electricity markets. Consequently, the experience of Chile and Peru provides a strong argument in favor of the SFPFC mechanism for regions of the United States with significant intermittent renewable energy goals.

5. Mechanisms that Support Large Renewable Energy Shares

This section describes two mechanisms that support large renewable energy shares at least cost to electricity consumers. The first mechanism is a renewable energy certificates market for a region to meet its renewable energy goals. This is followed by the discussion of the need to integrate intermittent renewable resources into the standardized long-term contract approach to long-term resource adequacy as the share of intermittent renewables increases. Finally, this section discusses how a cost-based market can foster the development of renewable resources.

[3] See Section 3.2 of Wolak (2021) for more details on this settlement mechanism.

5.1. Renewable Energy Certificates Market

A renewable energy certificate (REC) market is a significantly lower cost approach to achieving a given renewable energy goal than other available mechanisms because it creates a competitive market for the renewable energy attribute. Under this mechanism the relevant regulatory authority would set up a registry of qualified renewable resources for the region. The set of generation resources that are qualified to sell RECs that would be established and overseen by this regulatory authority. Once a resource is qualified to sell RECs, the energy production by these resources would be compiled by the registry established by the regulatory authority and each of these resources would be issued RECs equal to the MWhs of energy the resource produced during the compliance period.

Assuming an annual compliance period for a renewables mandate retailers would be required to purchase the mandated percentage of their annual consumption of energy in RECs. For example, if the renewables mandate was 30 percent for 2024, free consumers and distributors would have to surrender RECs produced during 2024 equal to 30 percent of their annual consumption in 2024. For example, a retailer with an annual consumption of 20,000 MWh, would be required to surrender 6,000 RECs or pay a per $/MWh penalty set by the regulatory authority or any shortfall relative to this magnitude. For instance, if the retailer only held 5,900 RECs for the 2024 compliance period, it would be liable for a penalty of 100 RECs times this penalty price. The penalty price should be set sufficiently high so that all free consumers and distributors find it expected profit-maximizing to meet their renewable energy requirement.

Renewable resource owners would be allowed to sell RECs that their units have not yet produced, but they would be subject to the financial penalty for any shortfall between the quantity of RECs they have sold for the compliance period and the amount of RECs their units produced during the compliance period. For example, if renewable resource owner sold 1,000 RECs and only produced 900 MWh of energy during the compliance year, the resource owner to be assessed a penalty for the 100 REC shortfall times the per REC penalty.

Unused REC from the previous compliance year could be used in the following compliance year, but not in any subsequent year. For example, a RER unit that produced 100 RECs in 2024 and only sold 90 of these RECs for compliance in 2024 could sell the remaining 10 RECs for the 2025 compliance period. Similarly, if a free consumer or distributor only needed 95 RECs for compliance in 2024, but it held 105 RECs for the 2024 compliance period, the unused 10 RECs

could be used for compliance in 2025. This ability to carryover RECs would only be possible for consecutive compliance years, so a REC produced in 2024 could not be used in the 2026 compliance year or subsequent years.

Unless a jurisdiction establishes a legal commitment to renewable energy targets into the distant future, there is no reason to establish a REC market. Moreover, this regulatory commitment would increase the likelihood that a forward market for RECs would develop to support investments in renewable resources to meet this goal. A centralized forward market procurement mechanism similar to the SFPFC mechanism for long-term resource adequacy could be implemented to ensure retailers purchase sufficient RECs into the distant future to provide the revenue stream necessary to meet the region's renewable energy goals. For example, centralized auctions for RECs could be run at similar time horizons to delivery to the SFPFC auctions. A guaranteed four-year future revenue stream from future REC sales would provide the above market revenues to the quantity of RERs necessary to achieve given long-term RER goal.

It is important to emphasize that without legally mandated commitment by the relevant jurisdiction to meet a specific renewable energy target, such as 20 percent of electricity consumption from these resources by 2030, establishing a RPS is unnecessary. Intermittent renewable resources can compete with conventional generation resources in the long-term resource adequacy mechanism selling SFPFCs. Special procurement processes for intermittent renewable resources or specific technologies should be avoided. They simply reduce the extent of competition suppliers of these products face, which increases costs to consumers, with no accompanying economic or environmental benefit that could not be achieved at lower cost through an RPS.

5.2. Transitioning Renewables to SFPFCs

As the share in intermittent renewable energy resources (RERs) increases, it is increasing costly to place the burden of managing their intermittency on buyers of the renewable power purchase agreement (PPA). A contract that pays a renewable resource owner a fixed price for all MWhs produced whenever this energy is produced, provides an implicit subsidy to the RER owner in a multi-settlement LMP market. In terms of notation of equation (1) of Section 4, the period-level variable profit of the RER unit owner is (PC – C)QC, because QS = QC for all periods under the terms of a contract that pays the RER owner PC for every MWh produced whenever it is produced. This PPA completely insulates the RER unit from the short-term market price, which means it has no financial incentive to manage it intermittency. This contract form is not offered

to conventional dispatchable resources for precisely this reason. Clearly, a thermal or hydroelectric resource owner would prefer a contract that transfers all of its outage or energy shortfall risk to the buyer of the contract. For this reason, all fixed-price and actual production PPA contracts should be eliminated for all generation resources.

Under the proposed multi-settlement LMP market without these PPAs contracts, RER resources that schedule energy in the day-ahead market be responsible for the cost or revenues associated with any deviation between their day-ahead schedule and real-time output level. If the RER unit does not schedule any energy in the day-ahead market, then the energy the unit produces would be paid the real-time price.

Facing intermittent renewable resources with the full cost of their intermittency will foster the development of cross-hedging arrangements between intermittent renewable resources and dispatchable resources. For example, a solar resource owner might purchase price spike insurance against high short-term prices during hours of the day when the resource cannot or is unlikely to produce energy. In this case, the solar resource owner would make an up-front payment to the dispatchable resource owner in exchange for the following hourly payment stream of max(0,(P(spot,h)-P(strike))) times the number of MWhs sold during the term of the "cap contract," where P(spot,h) is spot price during hour h and P(strike) is the negotiated strike price of the financial contract and max(x,y) is a function that chooses the maximum of x and y. The solar resource owner would earn (P(spot,h) – P(strike)) per MWh purchased from this cap contract when P(spot,h) > P(strike) and zero otherwise. The dispatchable resource that sold this contract is liable for this payment stream. For this reason, the dispatchable resource has a strong incentive to produce as much output as possible during periods when P(spot,h) is likely to exceed P(strike) to avoid making this payment.

This across-technology hedging accomplishes two goals. First, it provides up-front revenues to dispatchable generation resources to cover their annual fixed costs in a world in which they operate during fewer hours of the year because of the increasing amount intermittent RERs. Second, it ensures that intermittent RERs account for the full cost of their intermittency in the prices they offer for SFPFC energy and RECs. If intermittent renewable resource owners are unable to recover these costs from selling SFPFC energy or energy in the short-term market, these above-market costs must then be recovered from sales of RECs, assuming that the government has set a legally binding target for energy production. As noted earlier, for the case that the Peruvian

government does not set a legally binding renewables target, the renewable resource owner must recover its costs through sales of SFPFC energy, bilateral hedging arrangements, and short-term market sales.

5.3. Cost-Based LMP Market and Renewables Integration

The strength of a cost based LMP market design for RER integration is that all of the resources in the control area, including intermittent renewable resources, will be dispatched in a least cost manner using the variable costs determined by the market operator. How these resources are compensated for the energy they sold in the SFPFC auctions will not impact how the resource is ultimately used to produce energy. As noted earlier all suppliers have a strong financial incentive to supply their hourly allocation of SFPFC energy at the lowest possible cost, either by producing it or purchasing it from the short-term market.

A cost based short-term LMP market provides RER owners with a transparent short-term market to purchase energy from when their intermittent renewable units do not produce sufficient energy to meet their hourly SFPFC obligation and sell excess energy beyond this forward market obligation when their units produce more than this quantity of energy. This logic emphasizes the importance of a publicly disclosed process for clearing the day-ahead and real-time cost-based markets. The renewable resource owner can factor in how these imbalances will be settled in making offers to supply SFPFCs for energy.

Shifting renewable resource owners to fixed-price and fixed-quantity forward contract from fixed-price and quantity-produced contracts will also provide financial incentives for renewable resource owners to manage the intermittency of their production through storage investments and financial contracts that support investments in fast-ramping dispatchable generation resources to provide insurance against renewable energy shortfalls. Transitioning forward contracts for renewable energy to require the seller to manage the quantity risk associated with the energy it sells is a crucial step in increasing the amount intermittent renewable energy produced while maintaining high level of grid reliability.

In all LMP markets operating around the world there is an ongoing process of updating the set of constraints incorporated into the market mechanism to ensure that the match between how the market sets prices and dispatch levels agrees as closely as possible with how the grid is operated. This logic implies that as the share of intermittent renewable resources increases an LMP market can be easily adapted to deal with the new reliability challenges this creates.

For example, California has added several new operating reserves to account for the fact that the large share of solar RERs has created the need to manage a large daily ramp up of dispatchable resources at the end of the daylight hours and a slightly smaller ramp down in the early morning hours. The introduction of these new operating reserves required additional constraints in the day-ahead market-clearing mechanism and adding the offer prices times the offer quantities for these products to the objective function.

A multi-settlement LMP market can efficiently manage the sudden generation unit starts and stops that arise with a significant amount of intermittent renewable generation units and the need to configure combined cycle natural gas units to operate as either individual combustion turbines or as an integrated pair of combustion turbines and a steam turbine. A formal day-ahead market allows these generation units to obtain day-ahead schedules that are consistent with their physical operating constraints. The real-time market can then be used to account for unexpected changes in these day-ahead schedules because of changes in the operating characteristics of generation units such as a forced outage or limitations in the amount of available input fossil fuel, as well as changes in demand between the day-ahead and real-time markets.

6. Concluding Comments

Achieving the large shares of intermittent renewable energy necessary to reduce substantially the carbon content of a region's electricity supply is likely to be significantly less costly because of the recent reduction in the LCOE of wind and solar resources. However, ensuring that this transition occurs in a least cost manner requires efficient pricing in the short-term energy market and a long-term resource adequacy mechanism designed for an industry with a large share of intermittent renewables. Zonal pricing markets that do not account for all relevant operating constraints on dispatchable and intermittent renewable generation units in the day-ahead and real-time market, unnecessarily increase the cost of making this energy transition. The major system reliability challenge with a significant amount intermittent renewable resources changes from having sufficient generation capacity to meet annual system demand peaks to the ability to meet the hourly net demands (system demand less intermittent renewable output) for energy throughout the year. Particularly in an electricity supply industry with a summer annual peak demand and significant installed solar generation capacity, meeting daily system demand peaks is relatively straightforward because demand peaks when there is significant solar energy production. The new focus on meeting net demand peaks implies a system-wide focus on energy adequacy

where intermittent renewable resources have a financial incentive to hedge their short-term and production quantity risk with dispatchable generation resources to cover these net demand peaks.

A multi-settlement locational marginal pricing market design efficiently prices the system-wide and local reliability benefits provided by dispatchable resources relative to intermittent renewable resources. By co-optimizing the procurement of energy and ancillary services, this market design ensures that the demand for energy and ancillary services all locations in the transmission network are met at least cost. The standardized energy contracting approach to long-term resource adequacy described in this paper addresses the primarily reliability challenge in regions with significant intermittent renewables. It provides strong incentives for intermittent resources to cross-hedge their quantity and price risk associated with selling these standardized long-term contracts with dispatchable resources order to provide the revenue necessary to keep enough of this generation capacity available to meet hourly net demands throughout the year. The experience of Chile and Peru over the past 15 years, each of which has a market design that creates the same set of incentives for supplier behavior as the SFPFC mechanism, provides encouraging empirical evidence in favor of its adoption in regions with significant intermittent renewable energy goals.

Finally, if a region has a legal mandate to achieve a pre-specified renewable energy goal by a given date, such as 60 percent of energy consumed by 2040, then a renewable energy certificates market is the least cost approach to achieving this goal. If a region does not have a mandated renewable energy goal, then such a market is not necessary. The recent declines in the LCOE of wind and solar resources makes them a lower LCOE solution than natural gas and coal generation units in many regions.

References

Bjørndal, Endre, Mette Bjørndal, Hong Cai, and Evangelos Panos. "Hybrid pricing in a coupled European power market with more wind power." *European Journal of Operational Research* 264, no. 3 (2018): 919-931.

Bohn, Roger E., Michael C. Caramanis and Fred C. Schweppe (1984), 'Optimal Pricing in Electrical Networks over Space and Time', *RAND Journal of Economics*, **15** (5), 360-376.

Bushnell, James B., Benjamin F. Hobbs, and Frank A. Wolak (2008), "Final Opinion on 'The DEC Bidding Activity Rule under MRTU'", available at https://www.caiso.com/Documents/MSCFinalOpinionDECBiddingActivityRuleunder MRTU.pdf.

Carden, Kevin, Alex Krasny Dombrowsky, Chase Winkler, "2020 Joint IOU ELCC Study, Report 1," (2020), available at https://www.astrape.com/2020-joint-ca-iou-elcc-study-report-1/

EIA, Levelized Costs of New Generation Resources in the Annual Review Outlook 2021, United States Energy Information Administration, (2021) available at https://www.eia.gov/outlooks/aeo/pdf/electricity_generation.pdf

Galetovic, Alexander, Cristián M. Muñoz, and Frank A. Wolak (2015), 'Capacity Payments in a Cost-Based Wholesale Electricity Market: The Case of Chile', *The Electricity Journal*, **28** (10), 80-96.

Garver, Leonard L. "Effective load carrying capability of generating units." *IEEE Transactions on Power apparatus and Systems* 8 (1966): 910-919.

Graf, Christoph, Federico Quaglia, and Frank A. Wolak (2020) "Simplified Electricity Market Models with Significant Intermittent Renewable Capacity: Evidence from Italy," available at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/GrafQuagliaWolak_SimplifiedElectricityMarketModelsRenewables.pdf

IRENA, *Power Generation Costs in 2020*, International Renewable Energy Agency, Abu Dhabi, 2021.

Joskow, Paul L. (1997), 'Restructuring, Competition and Regulatory Reform in the U.S. Electricity Sector', *The Journal of Economic Perspectives*, **11** (3), 119-138.

Mansur, Erin T. and Matthew W. White (2012), 'Market Organization and Efficiency in Electricity Markets', accessed 26 September 2020 at http://www.dartmouth.edu/~mansur/papers/mansur_white_pjmaep.pdf.

McRae, Shaun D., and Frank A. Wolak. "Diagnosing the causes of the recent el nino event and recommendations for reform." (2016) available at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/diagnosing-el-nino_mcrae_wolak.pdf

Oren, Shmuel. S. (2001), 'Design of ancillary service markets', in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pp. 9-pp, IEEE.

Price, James E., and Mark Rothleder. "Recognition of extended dispatch horizons in California's energy markets." In 2011 *IEEE Power and Energy Society General Meeting*, pp. 1-5. IEEE, 2011.

Triolo, Ryan C. and Frank A. Wolak (2021) "Quantifying the Benefits of a Nodal Market Design in the Texas Electricity Market," available at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/BenefitsOfNodalDesignERCOT.pdf

Wolak, Frank A. "An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market." *International Economic Journal* 14, no. 2 (2000): 1-39.

Wolak, Frank A. (2009), 'Report on Market Performance and Market Monitoring in the Colombian Electricity Supply Industry', accessed 26 September 2020 at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/files/sspd_report_wolak_july_30.pdf.

Wolak, Frank A. (2011b), 'Measuring the Benefits of Greater Spatial Granularity in Short-Term Pricing in Wholesale Electricity Markets', *American Economic Review*, **93** (2), 247-252.

Wolak, Frank A. "Economic and political constraints on the demand-side of electricity industry re-structuring processes." *Review of Economics and Institutions* 4, no. 1 (2013): 42.

Wolak, Frank A. (2014), 'Regulating Competition in Wholesale Electricity Supply', in Nancy L. Rose (ed.), *Economic Regulation and Its Reform: What Have We Learned*?, The University of Chicago Press, pp. 195-289.

Wolak, Frank A. "Level versus Variability Trade-offs in Wind and Solar Generation Investments: The Case of California." *The Energy Journal* 37, no. Bollino-Madlener Special Issue (2016).

Wolak, Frank A. Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables. No. c14586. National Bureau of Economic Research, 2021.

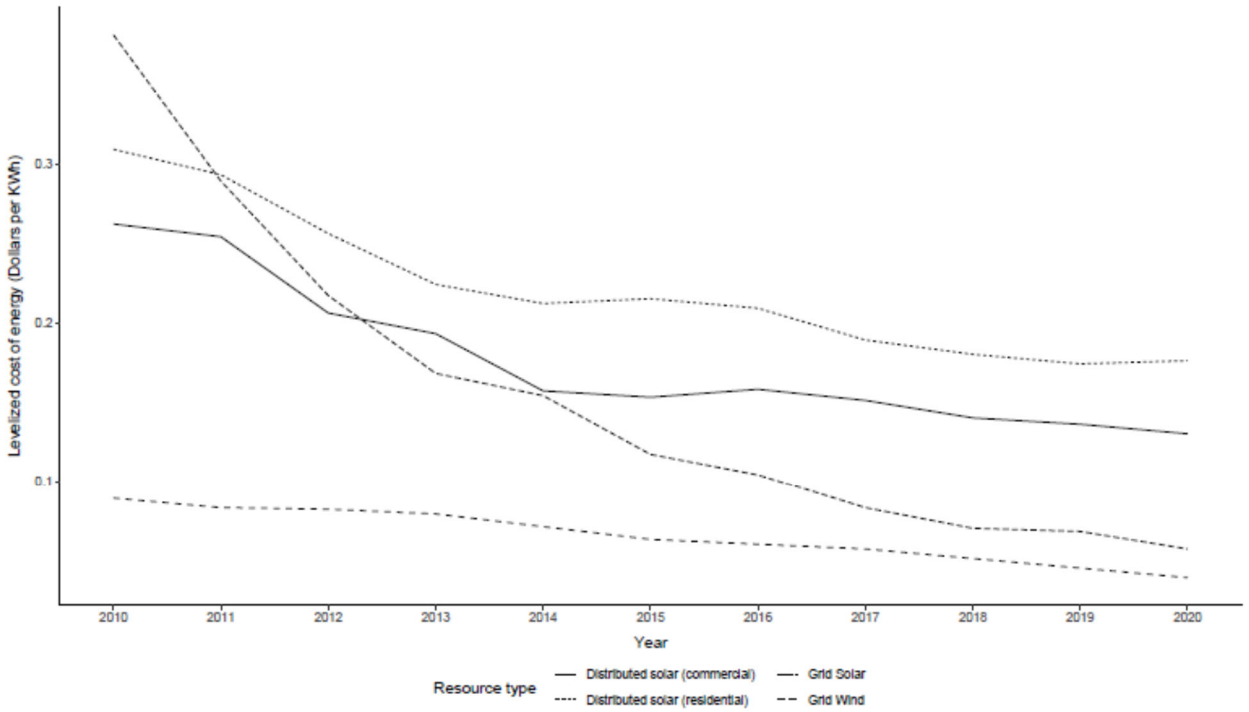

Figure 1: Levelized Cost of Energy from Grid Scale Wind and Solar and Distributed Solar Generation 2010 to 2020

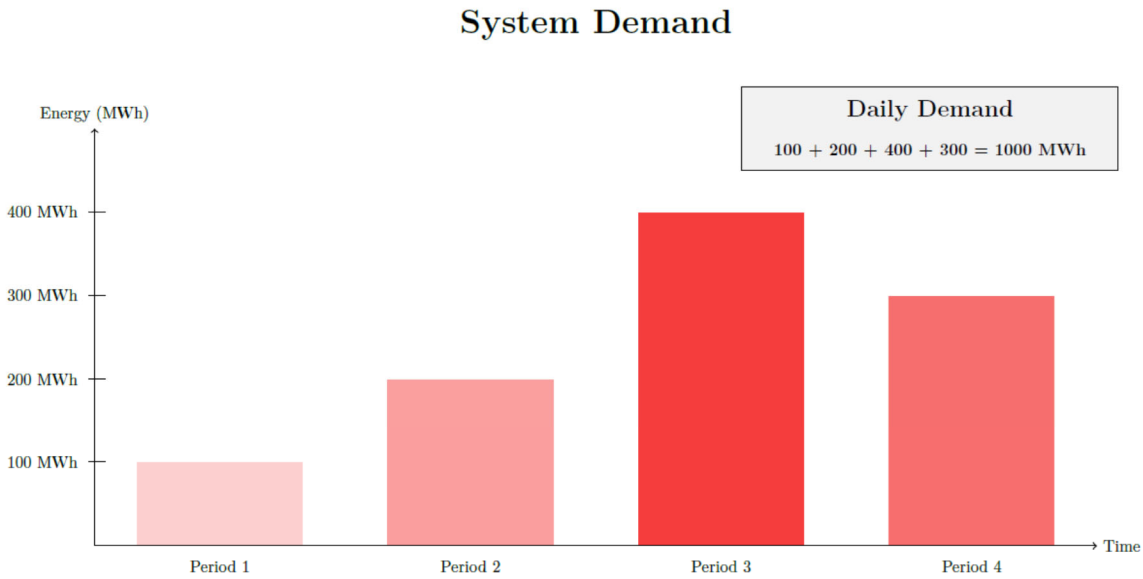

Figure 2: Hourly System Demands

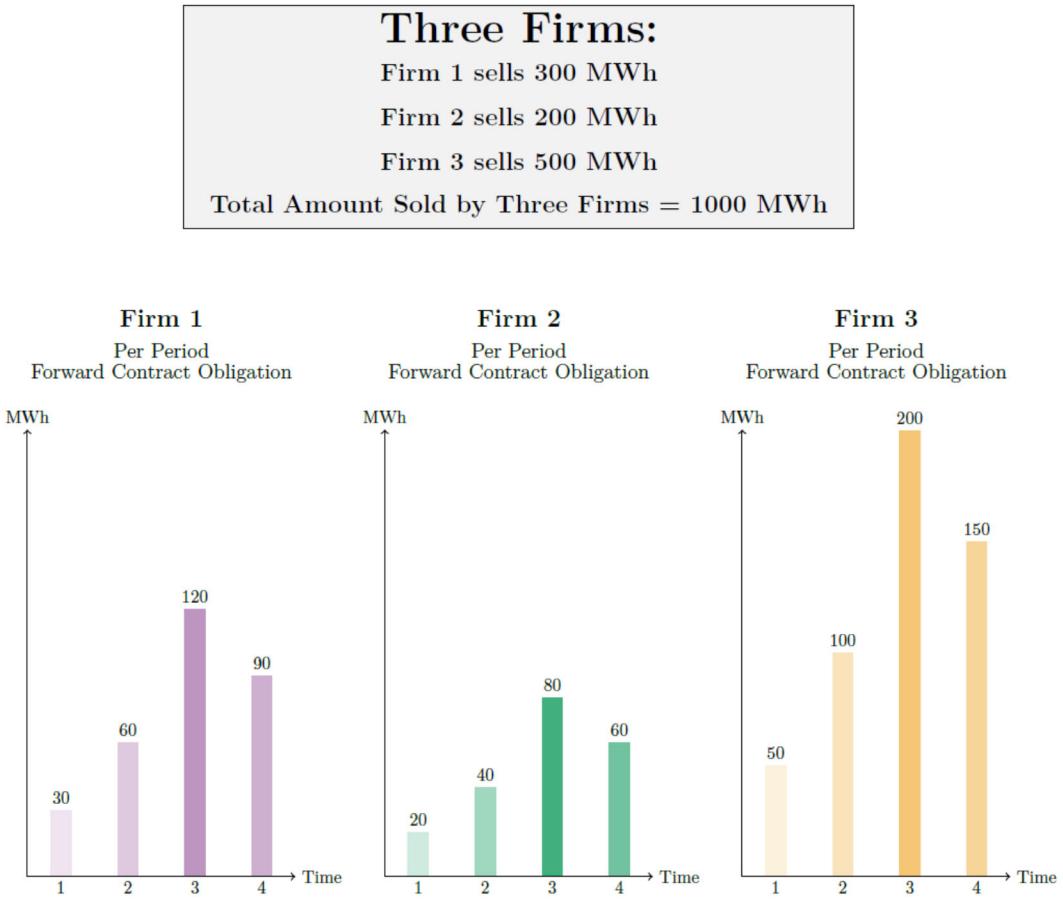

Figure 3: Hourly Forward Contract Quantities for Three Suppliers

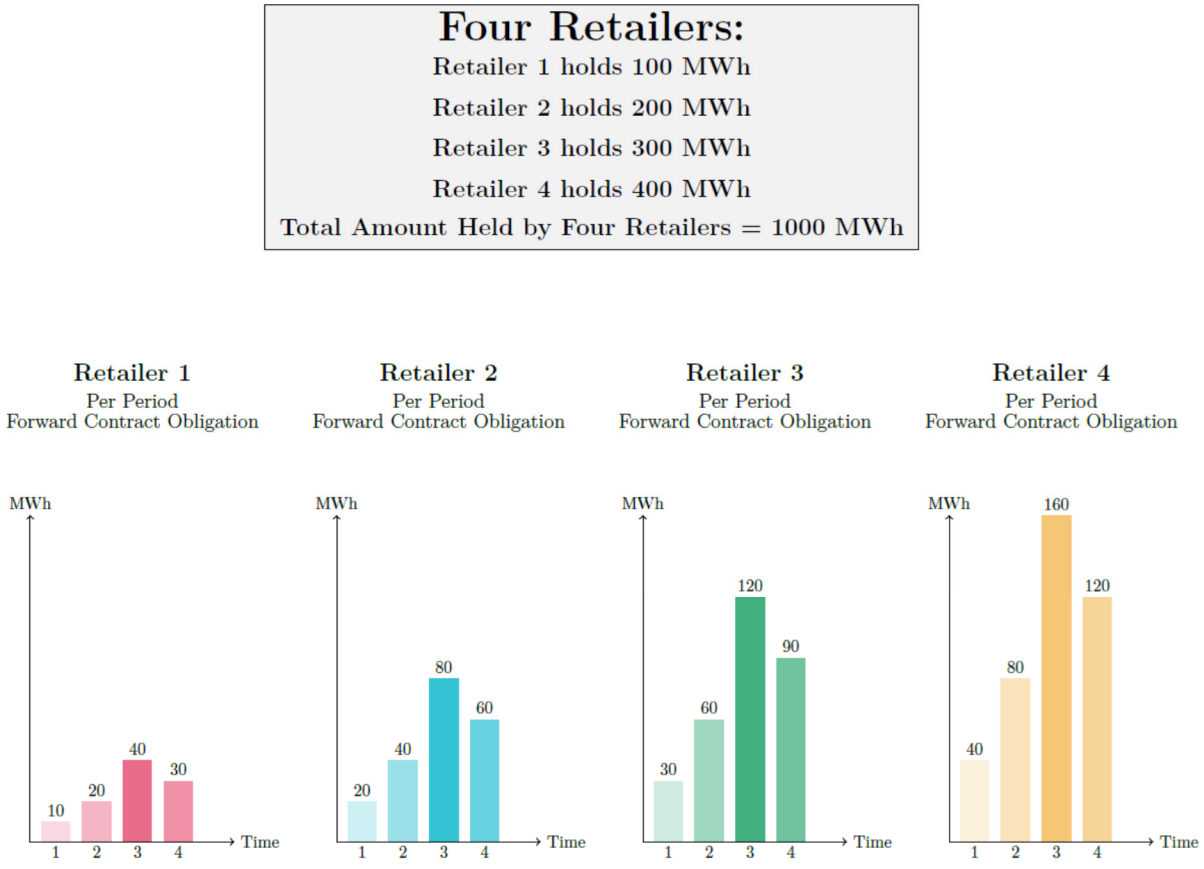

Figure 4: Hourly Forward Contract Quantities for Four Retailers