The Role of Efficient Pricing in Enabling A Low-Carbon Electricity Sector

by

Frank A. Wolak
Program on Energy and Sustainable Development (PESD)
and Department of Economics
Stanford University
Stanford, CA 94305-6072

March 31, 2019

Abstract

Producers and consumers will make the investments and innovations necessary to transition to a low carbon electricity supply industry only if they are compensated for their efforts. In the absence of explicit government support for these activities, this outcome will occur only if wholesale and retail prices provide this compensation. Efficient wholesale and retail pricing provides compensation for the cost-effective deployment these innovations. Multi-settlement locational marginal pricing markets set efficient short-term wholesale electricity prices. Marginal cost-based pricing of transmission and distribution networks is increasingly important in regions solar resources. More efficient wholesale and retail pricing implies significantly greater price volatility, particularly as the share of intermittent renewable generation increases, which requires implementing a number of competition and regulatory safeguards to protect consumers, while still providing the price signals necessary for a least cost transition to a low-carbon electricity supply industry.

1. Introduction

To meet their greenhouse gas emissions targets, many countries are forecasting massive increases in utility-scale and distributed renewable generation capacity, primarily utilizing wind and solar resources. This shift to intermittent renewable resources at both ends of the combined transmission and distribution network will require significant changes in the planning, operation, and regulation of transmission and distribution networks, the design of wholesale and retail electricity markets, and regulatory oversight of the entire industry.

Much more extensive and technologically sophisticated transmission and distribution networks will be required to manage this intermittency. More sensors must be deployed to improve real-time situational awareness throughout the transmission and distribution networks. Software and algorithms must be developed to compile this information and process it to provide signals to controllers embedded throughout the transmission and distribution networks to charge and discharge storage and deploy other devices to maintain a reliable supply of energy and voltage. Both grid-scale and distributed storage devices can to allow renewable energy produced during high-output hours of the day to be consumed during low-output hours. Automated demand-response technologies can reduce demand during hours with low renewable energy production and shift it into hours with high renewable energy production. All of these changes are unlikely to occur as rapidly or inexpensively as possible without efficient pricing of transmission and distribution network access and energy and ancillary services at the wholesale and retail levels.

This paper outlines how efficient wholesale, transmission and distribution network, and retail pricing can enable the cost-effective transition to a high penetration intermittent renewable generation electricity sector. Setting prices that reflect the marginal cost of withdrawing an additional megawatt-hour (MWh) of electricity at each location in the transmission network during each pricing interval provides the economically efficient wholesale price signal. Adding the marginal cost of supplying electricity to each location distribution grid to the relevant locational wholesale price yields the efficient retail price for that pricing interval.

Efficient wholesale and retail prices provide the revenue streams to generation unit owners, electricity retailers, and final consumers that support the upfront investments and ongoing operating costs necessary for this transformation to occur. Setting day-ahead and real-time market prices that reflect transmission network constraints and all other relevant operating constraints as well as the marginal losses associated with moving electricity from where it is produced to where it is consumed reduces the incentive of suppliers to take actions that reduce grid reliability and increase the cost to consumers of serving load in real-time. An increasing share of intermittent renewable generation capacity in a region increases the number and magnitude of reliability constraints that grid operator must address to maintain supply and demand balance in real-time. A day-ahead market that prices all transmission and other relevant operating constraints in the procuring both energy and ancillary services followed by a real-time market for energy and ancillary services that prices these same constraints will reduce cost of integrating an increasing amount of intermittent renewable generation capacity.

Wholesale electricity markets that do not price all transmission and other relevant operating constraints, such as those that exist in all European countries, are likely to become increasingly costly operate. That is because system operators will need to take more actions outside of formal market mechanisms to ensure that final generation and load schedules that emerge from the wholesale market mechanism are physically feasible for the configuration of the transmission network, the amount of available generation capacity, and the geographic location of demand.

The traditional approach to network pricing that recovers the cost of the transmission and distribution networks through a per-MWh charge, will continue to encourage inefficient bypass of grid-supplied electricity from distributed generation units, particularly in regions with significant solar resources, and even if utility-scale solar generation units are a lower cost source of renewable energy. As more customers substitute energy from rooftop solar systems for grid supplied electricity, the per-MWh charge for the transmission and distribution network will have to increase to recover the same fixed cost over a smaller number of MWhs. The resulting higher average retail price and the declining cost of installing a distributed solar system provides strong financial incentives for the utility's remaining customers to install rooftop solar systems.

Efficient retail pricing will face final consumers with greater price volatility, particularly as the share of intermittent renewable generation increases. This increased price volatility can finance investments in distributed storage and automated load-shifting technologies. Greater default retail price volatility increases the expected benefits consumers realize from hedging their exposure to short-term prices and increases the risk of bankruptcy for electricity retailers.

Retail price volatility also implies a greater role for the industry regulator to protect consumers from excessive prices. Regulatory safeguards such as a local market power mitigation mechanism for energy and ancillary services, a default retail tariff, hedging requirements for retailers, and competition among electricity retailers typically provide this protection.

The remainder of the paper proceeds as follows. Section 2 describes why efficient wholesale electricity pricing requires accounting for all relevant operating constraints in the transmission network and generation unit operation and marginal transmission and distribution network losses in setting day-ahead and real-time prices. This is section discusses the role of a multi-settlement locational marginal pricing market in achieving this goal and why regions that ignore transmission network and other reliability constraints in operating their day-ahead and real-time energy and ancillary services markets are likely to face a significantly higher cost of integrating are larger share of intermittent renewable resources. Section 3 illustrates why the availability of distributed solar photovoltaic (PV) generation capacity has increased the social cost of the historical approach to transmission and distribution network. A simple economic model is used to demonstrate a more efficient approach to network pricing when customers have the option to install distributed solar PV capacity. Section 4 explains why efficient retail pricing involves facing final consumers with greater price volatility, particularly as the share of intermittent renewable generation capacity increases. The need to manage this price volatility supports cost-effective adoption of new technologies necessary to maintain a reliable supply of electricity with significant larger share of intermittent renewable generation. Section 5 describes the major challenges facing competition and regulatory authorities under efficient wholesale and retail pricing and provides recommendations for addressing them.

2. More Efficient Wholesale Energy and Ancillary Services Pricing

Thirty years of experience with electricity industry re-structuring provides ample evidence that competition authorities and industry regulators face significant challenges in achieving competitive wholesale market outcomes during vast majority of hours of the year. Virtually all wholesale electricity markets have experienced periods of poor market performance that has required competition or regulatory authority intervention, or at least, resulted in calls for intervention.[1] There is a growing consensus around the world that a short-term wholesale market that sets prices that reflect all of the costs associated with withdrawing energy at each location in

[1] Wolak (2014) describes instances when wholesale electricity markets in United Kingdom, California, New Zealand and Colombia experienced periods of prices that reflected the exercise of significant unilateral market power.

the transmission grid during each pricing interval is necessary to limit the likelihood of periods of poor wholesale market performance.

2.1. Match Between Market Mechanism and Actual System Operation

A key determinant of wholesale market performance, particularly in regions with significant amount of intermittent renewable generation capacity, is the extent to which the market mechanism used to dispatch and operate generation units is consistent with how the grid is actually operated. In the early stages of electricity re-structuring, many regions attempted to operate wholesale markets using simplified versions of the transmission network. These markets either 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 created opportunities for market participants to increase their profits by taking advantage of the fact that in real-time the actual configuration transmission network and other operating constraints would need to be respected in order to serve demand at all locations in the transmission network.

Market mechanisms that assume simplified transmission networks typically produce prices and dispatch levels for generation units that are not feasible given the configuration of the transmission network or other constraints on how the system is operated. Between the close of the wholesale market and real-time system operation, the system operator must decrease the scheduled output levels of some units and increase the scheduled output of others in order to create a physically feasible configuration of operating levels for all generation units. The final outcome of this process yields generation units with offer prices (or variable costs) below the market-clearing price not producing electricity and units with offer prices (or variable costs) above the market-clearing price producing electricity.

Generation units operating "out of merit order" 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 instead.[2] The former units are typically called "constrained-off" units and the latter are called "constrained-on" or "must-run" units.

---

[2] Ordering generation units from the lowest to highest marginal cost of production (or lowest to highest offer price) is called the merit order. In the absence of transmission or other operating constraints, generation units would be

Constrained-on suppliers are typically paid their offer price or a regulator-determined price for this energy. Suppliers constrained-off typically sell their energy back to system operator at their offer price or at a regulated-determined price. The wholesale markets that use simplified transmission networks differ in terms of the sophistication of the mechanism used by the system operator to determine which generation units are constrained on and constrained off.

How frequently generation units are constrained on or constrained off will affect the offer prices they submit into the short-term 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 is highly probable the unit will be constrained-on, an expected profit-maximizing unit owner will submit an offer price far in excess of the variable cost of operating the unit. This action can increase the market-clearing price if the constrained on unit is not accepted to supply energy in the formal market mechanism.

Analogous circumstances can arise for constrained-off generation units. Constrained-off suppliers that are accepted to supply energy in the formal market mechanism are typically able to purchase the energy they are unable to supply because of the configuration of the transmission network at a price below the market-clearing price. 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 ensure that it will be accepted in the formal market. It will also receive the highest possible payment for being constrained-off if its offer price is the price at which it much purchase the energy it is unable supply because of the configuration of the transmission network.[3] Bushnell, Hobbs and Wolak (2008) discuss this problem and its market efficiency consequences in the context of the California zonal market. Similar problems occurred in the zonal markets that initially existed in other parts of the United States, such as New England and Texas. Australia operates a zonal market and is currently facing an increasing frequency of constrained-on and constrained-off generation units. This problem is not unique to industrialized country markets. Wolak (2009) discusses these same issues in the context of the Colombian single-zone market.

---

accepted to supply energy according to this merit order because that would minimize the total variable cost (or as-offered cost) of serving demand.

[3] If Q(M) is amount sold in the formal market at price P(M) and Q(A) (Q(A) < Q(M)) is amount the unit actually produced and P(O) is the supplier's offer price into the formal market, the net revenues is P(M)Q(M) – P(A)(Q(M) – Q(A)).

All wholesale markets in Europe ignore the configuration of the transmission network or assume a simplified version of the transmission network in setting prices and dispatch levels and are therefore increasingly plagued by this source of market inefficiencies. The final generation schedules that emerge from all of these market mechanisms must typically be adjusted to produce generation schedules that are physically feasible given the configuration of the transmission network. Each of these markets accomplishes this task using different mechanisms, but all produce outcomes where certain generation units are constrained on or required to produce more energy than was sold in the formal market mechanism and certain generation units are constrained off or required to produce less energy than was sold in the formal market mechanism. Moreover, because of the increased number and magnitude of reliability constraints that must be respected as the share of intermittent renewable energy in these markets have increased, the cost of making the dispatch levels that emerge from the short-term energy market physically feasible has increased over time.

Markets that use simplified transmission network models usually do not explicitly include transmission network losses in market prices in spite of the fact that energy losses occur between locations in the transmission network where energy is injected and locations where energy is withdrawal. These markets typically allocate losses using ad hoc rules, rather that according to how these losses occur in the operation of the transmission network. The physics of transmission networks imply that transmission losses increase with the square of energy flows in the network and the voltage capacity of the transmission line. Markets that explicitly account for the configuration of the transmission network and other operating constraints also account for transmission network losses between locations where the electricity is injected and locations where it is withdrawn.

An additional source of inefficiency in many of the early wholesale market designs in the United States and existing market designs in Europe is sequential procurement of energy and ancillary services. Under these designs, the market for purchasing ancillary services, such as automatic generation control or regulation reserve and operating reserves such as spinning, non-spinning and replacement reserves takes place after the close of the energy market. Sequential markets, which first procure of one product (e.g., energy) and then other products (e.g., reserves), limits the degree of substitutability between these products and create opportunities for suppliers to increase the prices they are paid to provide these services with no corresponding increase in system reliability. Oren (2001) provides several examples of this phenomenon in the initial market designs in California and ISO New England. Generation unit owners would understate the capability of units to provide one service with the goal of selling in a subsequent market for a higher price, which ultimately increased the total cost to consumers of the necessary ancillary services procurement. Simultaneously procuring all ancillary services and energy and allowing cheaper higher quantity services to substitute for more expensive lower quantity services chooses the least as-offered cost combination energy and ancillary services offers to meet the locational demands for energy and ancillary services.

All markets in the United States now employ a market mechanism used to set prices across locations and hours of the day that matches as closely as possible how the transmission network is operated and simultaneously procures energy and ancillary services and explicitly accounts for marginal transmission losses at all locations in the transmission network. This eliminates the need for constrained-on and constrained-off generation units, because the market mechanism sets potentially different prices at all locations in the transmission network. Each generation unit receives a price that reflects the cost of withdrawing an additional MWh of energy at that location in the transmission network during that pricing interval that includes the marginal losses associated with delivering energy to that location in the transmission network.

2.2. Co-Optimized Locational Marginal Pricing Market for Energy and Ancillary Services

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. Multi-settlement wholesale electricity markets that use locational marginal pricing (LMP), also referred to as nodal pricing, avoid these constrained-on and constrained-off problems, because all transmission constraints and other relevant operating constraints as well as marginal losses are respected in the process of determining dispatch levels and prices in the wholesale market.

If these markets simultaneously purchase energy and ancillary services, this limits the opportunities suppliers have to take advantage of the sequential procurement process to raise prices in these markets. In an LMP market that simultaneously procures energy and ancillary services, the market operator receives generation unit-level offers for energy and each ancillary service the unit is capable of providing and decides which combination of energy and ancillary services from each generation unit in the market is optimal on a market-wide basis for meeting the demand for energy and each ancillary service. In contrast, sequential energy and ancillary services markets provide generation unit owners with more opportunities to raise market prices (particularly in the ancillary services market) and sell their capacity into the market yields the highest expected profits because they are able to observe outcomes in previous markets before submitting their offers into subsequent markets. Market participants know that capacity sold in earlier markets is not available to compete with suppliers in later markets and this can allow them to raise market prices in both sequential markets.

In co-optimized LMP market for energy and ancillary services market, generation unit owners submit their location-specific willingness-to-supply energy and ancillary services and load serving entities submit their willingness-to-purchase energy and ancillary services to the wholesale market operator. Locational marginal prices (LMPs) and schedules for energy and ancillary services 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 ancillary services at all locations in the transmission network subject to all network operating constraints, including transmission network losses. No generation unit is accepted to supply energy or any ancillary service if doing so would violate a transmission or other operating constraint. Because the market operator minimizes the as-offered costs of meeting the demands for energy and all ancillary services, this process is called co-optimization of energy and ancillary services procurement.

A co-optimized LMP market sets potentially different prices for energy and ancillary services at all locations in the transmission network, depending the configuration of the transmission network and geographic location of demand and available generation units. Because the configuration of the transmission network and the location of generation units and demands is taken into account in operating the market, only generation units that can actually operate will be accepted to serve demand and they will be paid a higher price or lower price than the average LMP for that energy or ancillary service, depending whether the generation unit is in a generation-deficient or generation-rich region of the transmission network.

The nodal price at each location is the increase in the minimized value of the total as-offered cost to meet demands at all locations (nodes) in the transmission grid as a result of a one unit increase in the amount of energy or ancillary service supplied at that location in the transmission network. There are four components of each locational marginal price: (1) the energy price, (2) the price of transmission congestion relative to the reference node, (3) marginal transmission losses relative to the reference node and (4) the shadow price of other relevant operating constraints such generation unit ramping constraints or local transmission network reliability constraints.[4]

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

An important implication of modelling all relevant operating constraints in setting LMPs is that withdrawing one more unit of energy or ancillary services at a location in a constrained region of the grid can require backing down generation units or demands at other locations in the grid, all of which can increase the LMP at that location. Consequently, an under-appreciated implication of LMP markets is that the price at constrained locations can be multiples of the maximum allowed offer price in the wholesale market. For example, in LMP markets with \$1,000/MWh caps on a generation unit owner offers, LMPs higher than \$3,000/MWh have been observed during heavily constrained pricing intervals. As more intermittent renewables are added to the generation mix, the potential is even greater for these system conditions to arise and extremely high and low LMPs to occur.

It is important to emphasize that these extreme LMPs reflect the cost of supplying an additional unit of energy or ancillary services at that location in the grid during that pricing interval. These prices provide the revenue streams to support investments in storage, automated load-

[4] The reference node is the (physical or virtual) location in the transmission network relative to which all congestion and transmission losses are defined. Bohn, Caramanis, and Schweppe (1984) provide an accessible discussion of the properties of this market mechanism.

shifting and other new technologies that can reduce the cost of managing the increased reliability challenges associated with a larger share of intermittent renewable generation capacity in a region.

2.3. Multi-Settlement Markets

A short-term forward market for energy and ancillary services that employs a co-optimized LMP market for energy and ancillary services in addition to a real-time time co-optimized LMP market can facilitate the cost-effective deployment of the new technologies necessary for a low-carbon electricity sector. The combination of short-term forward markets, typically the day-before and an hour-ahead of real-time system operation, along with a real-time market, is called a multi-settlement market. Multi-settlement nodal-pricing markets have been adopted by all US jurisdictions with a formal offer-based wholesale electricity market.

The day-ahead forward market sets firm financial schedules for all generation units and loads for all 24 hours of the following day. Suppliers submit generation unit-level offer curves for each hour of the following day and electricity retailers submit demand curves for each hour of the following day. The system operator then minimizes the as-offered cost to meet these demands for all 24 hours of the following day subject to the anticipated real-time configuration of the transmission network, other relevant operating constraints, and transmission network losses during all 24 hours of the following day. This gives rise to LMPs and firm financial commitments to buy and sell electricity and each ancillary service each hour of the following day for all generation unit and load locations.

The strength of this market design for regions with significant intermittent renewable generation capacity is that it accounts for all of the intertemporal reliability constraints in operating all generation resources. For generation units with start-up cost and ramping constraints, this market design can take into account the start-up cost and ramping constraints on individual generation units in determining whether to accept energy and ancillary services from the unit for any hours of the following day. Solving for the optimal schedules of all generation resources for all 24 hours of the following day simultaneously can significantly reduce the cost of serving load relative to a market design that clears each hour of the day independent of other hours of the day as is the case in most European wholesale electricity markets.

The day-ahead commitments that emerge from this market do not require a generation unit to supply the amount sold in the day-ahead market 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 much 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.

In all US 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 demand at all locations in the control area taking into account the current configuration of the transmission network, other relevant operating constraints, and marginal transmission losses at each location. This process gives rise to LMPs at all locations in the transmission network and 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 that 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 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 LMP market design is 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 amount of renewable resources can easily be incorporated into the day-ahead and real-time market models, because market prices and quantities are determined as the solution to a constrained optimization problem. Therefore, the economic benefits from implementing a multi-settlement LMP market relative to wholesale market designs that do not model all transmission and other operating constraint are likely to be greater the larger is the share of intermittent renewable resources because of the increasing number of operating constraints that must be accounted for in system (and market) operation.

An additional advantage of multi-settlement LMP markets for regions with significant renewable generation capacity is that this market design values of the dispatchability 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. 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 because of the unexpectedly large amount of wind energy produced. Consequently, besides rewarding the dispatchability of the thermal generation resource with a higher average price of energy, multi-settlement markets benefit intermittent resource owners that are better able to forecast, on a day-ahead basis, the real-time production of their generation units.

One complaint often leveled against LMP markets is that they increase the likelihood of political backlash from consumers because prices paid for wholesale electricity can differ significantly across locations within the same geographic region. For example, customers in urban areas that primarily import electricity over congested transmission lines will pay more than customers located in generation-rich rural regions that export electricity. Because more customers live in the urban areas than in the rural regions, charging final consumers in the urban areas a higher retail price to recover the LMP at their location may be politically challenging for regulators to implement.

Many regions with LMP pricing have overcome this concern by charging all customers in a given state or utility service territory a weighted average of the LMPs in the region. In the above example, this implies charging the urban and rural customers the weighted average of the LMPs in the urban and rural areas, where the weight assigned to each price is the share of system load that is withdrawn at that location during that pricing interval. Under this scheme, generation units continue to be paid the LMP at their location. For example, in Singapore all generation units are paid the LMP at their location, but all loads are charged the Uniform Singapore Electricity Price (USEP), which is the quantity-weighted average of the half-hourly LMPs for all load-withdrawal points in Singapore. California employs this approach to pricing to loads for each of the three large investor-owned utilities. All consumers in each utility service territory pays a wholesale price equal to the quantity-weighted average of all load-withdrawal nodes in the utility's service territory.

This approach to pricing to loads captures the reliability and operating efficiency benefits of an LMP market while addressing the equity concerns regulators often face with charging customers at different locations prices that reflect the configuration of the transmission network. Tangeräs and Wolak (2017) argue that imposing this requirement on retailers that are vertically-integrated into generation can improve the performance of short-term LMP wholesale markets.

The experience of all US wholesale electricity markets supports the argument that a co-optimized multi-settlement LMP market design is the most effective mechanism for achieving economically efficient wholesale prices. All US wholesale markets initially used simplified models of the grid in the dispatch of generation units and pricing of energy and many employed sequential energy and ancillary services markets. These designs created significant market performance problems, particularly in regions with limited transmission capacity. As a result, all these regions ultimately adopted multi-settlement LMP market designs with co-optimized energy and ancillary services procurement that explicitly price marginal transmission losses.

3. More Efficient Transmission and Distribution Network Pricing

Historically there were no financially viable substitutes for grid-supplied electricity. Consequently, the traditional approach to transmission and distribution network pricing of a fixed per-unit charge for withdrawals from the grid introduced a limited amount of economic inefficiencies. Specifically, charging a per unit price for use of the transmission and distribution grids that is higher that the marginal cost of withdrawing energy from the grid increases the per-unit retail price and reduces the demand for grid-supplied electricity relative to the level that would exist under marginal cost pricing of the transmission and distribution networks. Because customers had no financially viable alternative to grid-supplied electricity, the inefficiency in the historical approach to transmission and distribution network pricing only resulted in a lower level of electricity consumption relative to efficient pricing.

When a customer has the option to install distributed solar generation capacity, recovering all of the costs of the transmission and distribution networks through a per-unit charge can introduce two significant distortions. First, a customer considering a distributed solar PV unit compares the levelized cost of electricity from this facility to the average per unit retail price, not the average wholesale price implicit in the retail price it faces. That is because each kilowatt-hour (KWh) reduction in consumption of grid-supplied electricity reduces the customer's bill for grid-supplied electricity by this retail price. Second, the customers that install distributed solar now consume significantly less grid-supplied energy, but the sunk costs of the grid remain the same. This means that the per unit price must be increased for all customers to recover the sunk costs of the transmission and distribution grids, which encourages more customers to install distributed solar. Therefore, recovering the sunk cost of the transmission and distribution grids through a per unit charge favors investments in distributed solar generation over grid-scale solar generation, even if grid scale solar is cheaper on a levelized cost of energy basis, what is referred to as economically inefficient bypass of grid-supplied electricity.[5]

3.1. Economically Inefficient Bypass of Grid-Supplied Electricity

In regions with LMP wholesale markets, each pricing interval there is a marginal cost of withdrawing an additional unit of electricity from any location in transmission grid. The marginal cost of transmission losses are included in the LMP at each node in the transmission network. Adding the marginal cost of distribution network losses between the point of withdrawal from the transmission grid to the customer's premises, yields the marginal cost of supplying that customer with an additional KWh of grid-supplied electricity during that hour.

[5] Wolak (2018) provides empirical support for this mechanism for the case of the three investor-owned utilities in California.

Because the current retail tariff in virtually all jurisdictions recovers the vast majority of the cost of the transmission and distribution grid in the per unit price, during the vast majority of hours of the year the fixed price charged for grid-supplied electricity is vastly in excess of marginal cost supplying the customer with an additional KWh of grid-supplied electricity during that hour. To take the example of Pacific Gas and Electric in California, until very recently a residential customer consuming on the highest step of the increasing block price schedule paid a marginal price of 36 cents/KWh for grid-supplied electricity that typically sold in the short-term wholesale market for between 3 and 4 cents/KWh.

If the levelized cost of energy (net of any government support to the consumer) from a distributed solar PV system is less than the cost to the consumer of purchasing this electricity from the grid under the current retail tariff, the customer will find it privately cost-effective to install a distributed solar system. However, if the average incremental cost of providing this customer with the electricity produced by its distributed solar system is less than incremental cost of obtaining this energy from the grid, then the total cost of supplying the customer with energy (grid-supplied plus distributed-solar-supplied) has risen as a result installing the distributed solar system and the customer has made a socially inefficient investment in a distributed solar system (even excluding the cost of the government support).

A numerical example based on Pacific Gas and Electric illustrates this point. Suppose the levelized cost of a solar PV system net of the government support is 25 cents/KWh and the customer typically ends the month consuming 200 KWh on the 36 cents/KWh step of the nonlinear price schedule before installing a solar system. Suppose the average wholesale price for energy is 4 cents/KWh and marginal cost of distribution losses is 1 cent/KWh, for a delivered marginal cost of energy of 5 cents/KWh. Although the customer is saving 11 cents/KWh = 36 cents/KWh – 25 cents/KWh from investing in a solar PV system that reduces monthly consumption on the 36 cents/KWh step to zero, the average cost of supplying him with this 200 KWh energy increased by 20 cents/KWh = 25 cents/KWh – 5 cents/KWh as a result of investing in the distributed solar system.

3.2. Toward More Efficient Transmission and Distribution Network Pricing

This section presents an economic model that illustrates several pathways for improving the efficiency of transmission and distribution network pricing. As noted above, the efficient price of energy for a customer is the marginal cost of delivering an additional KWh to their premises. Pricing in this manner is unlikely to yield sufficient revenues to recover all of the sunk costs of the transmission and distribution networks. The specific question addressed in this section is how to allocate the remaining sunk costs of delivering electricity to final consumers not recovered from marginal cost pricing of energy. I first consider the case that customers have meters that can record their consumption at a frequency that matches the frequency that the marginal cost of electricity delivered to the customer changes. Then I consider the case of customers with mechanical meters that can only record their monthly consumption.

Let C(h) equal the marginal cost of retail electricity facing the customer during hour of the year h, for h=1,2,…,H, where H is the total number of hours in the year. To a first approximation, $C(h) = PW(h)(1 + \eta(h))$, where PW(h) is the hourly wholesale price where the transmission network connects with the customer's distribution network and $\eta(h) > 0$ is the hourly marginal loss factor for delivery to the customer's premises through this distribution network. Define the customer's hourly demand curve for electricity to be $Q(h) = A(h) – P(h)$. A(h) is the customer's willingness to pay for the first unit of consumption. Suppose that both C(h) and A(h) are random variables with compact support and a joint density, F(C,A). The support of C(h) is $[C_L,C_H]$ and $[A_L,A_H]$ where $0 < C_L < C_H < \infty$, $0 < A_L < A_H < \infty$, and $A_L > C_H$. The last inequality imposes the reasonable assumption that it is socially optimal for the customer to consume a non-zero amount of electricity because A_L is greater than the highest possible marginal cost realization, C_H.

Economically efficient pricing implies that the hourly retail price of electricity should be set equal to the hourly marginal cost of grid supplied electricity, so that P(h) = C(h). Using the logic of two-part tariff pricing, the maximum fixed charge that the consumer is willing to pay for grid-supplied electricity during this hour is the area below the demand curve above the hourly price, which is CS(P(h)), the consumer's surplus at price P(h) associated with hourly demand curve Q(h) = A(h) – P(h). This is the shaded area in Figure 1 and is equal to $½(A(h) – C(h))^2$.

Suppose before setting the fixed charge for the year or month, that the regulator only knows that the (A(h),C(h))′ are independent, identically distributed draws across the H hours of the year. Figure 2 shows the value of hourly consumer surplus (CS) for the extreme case of when $A(h) = A_L$ and $C(h) = P(h) = C_H$, and the hourly value of consumer surplus is extremely small. Figure 3 shows the other extreme of $A(h) = A_H$ and $C(h) = P(h) = C_L$, and the hourly value of consumer surplus is extremely large. Note that the definition of Q(h) implies that $CS(C(h)) = ½ (A(h) – C(h))^2 = ½ (Q(h))^2$.

Suppose that the consumer is risk neutral with respect to his electricity consuming decisions and will remain connected to the grid for the year if the expected annual fixed charge is less than the expected value of the annual consumer surplus obtained from consuming at P(h) equal to the hourly marginal cost of grid-supplied electricity, C(h), each hour of the year. Taking the expected value of $\frac{1}{2}(A(h) - C(h))^2$, yields the following result:

$$\text{Annual Expected Hourly CS} = \tfrac{1}{2} E[(Q(h))^2] = 1/2\{Var(Q(h)) + [E(Q(h))]^2\}.$$

This expression provides guidance for setting the value of the fixed charge for each customer. This simple model of demand implies that customers with a higher mean and higher variance of their hourly consumption have a greater willingness to pay to consume at the hourly marginal cost.

Although this simple model does not specify the absolute magnitude of fixed charges for each customer, it does provide clear guidance for setting relative values of these fixed charges across customers. High average demand customers with a volatile hourly demand should pay the highest fixed charges. Low average demand customers with stable demands should pay the lowest fixed charges.

This analysis yields two recommendations for more efficient transmission and distribution network pricing. First, the per-unit charge should only reflect the marginal cost of withdrawing energy from the transmission and distribution grid at the customer's location during each hour of the day. LMP markets already set prices at each location in the transmission network that reflect the marginal. cost of transmission losses at that location. Consequently, the marginal cost of losses from moving the energy in the distribution grid to the customer's premises should be added to this LMP. This will produce a substantially smaller per-unit hourly price of grid-supplied electricity than is currently the case.

Second, because transmission and distribution network pricing is fundamentally a sunk cost recovery problem, the burden of paying for these sunk costs should be allocated based on the willingness to pay of customers without causing disconnection of any customer from the electricity delivery network. Wolak (2018) proposes a methodology for allocating these fixed costs to individual customers based on the first two moments of the annual distribution of Q(h), the customer's hourly consumption.

For customers with mechanical meters that read on a monthly or bi-monthly basis, determining the efficient retail price is considerably more complicated because only the customer's monthly consumption can be measured. Utilities typically use assumed hourly load profiles for customers to allocate their monthly consumption to different hours of the billing cycle. This same mechanism can be used to set both the monthly energy price and the monthly fixed charge for each customer. The rapidly declining cost of interval meters that can record a customer's consumption on 15-minute or finer level of temporal granularity implies that the share jurisdictions with interval meters will continue to grow, so that the more efficient approach to retail pricing described above will be possible for an increasing number of customers and regions.

4. Efficient Retail Pricing for a Consumer-Friendly Electricity Sector

The efficient retail price is the marginal cost of delivering grid-supplied electricity to the customer's premises during each pricing interval. This marginal cost equals the real-time LMP at that customer's location in the transmission network for that pricing interval plus the marginal cost of losses from delivering the energy through the distribution network to the customer's premises during that pricing interval.

The remaining sunk and fixed costs of the transmission and distribution network must recovered through a monthly fixed-charge based on the customer's willingness to pay to receive grid-supplied electricity at the efficient energy price. As shown above, this willingness to pay is proportional to the annual mean and variance of the price interval level real-time consumption of the customer.

Because the LMP at each location in the transmission grid includes marginal losses and these losses scale with the square of the amount of energy transmitted, there will be over-recovery of the cost of total losses in the transmission network. This additional revenue from including marginal losses in the computation of LMPs can be used to reduce the amount of the sunk cost and other non-volume variable costs of operating the transmission grid that must be recovered from monthly fixed charges. Pricing marginal losses in the distribution grid will also over-recover the cost of distribution network losses. These revenues can be used to reduce the amount of sunk and non-volume variable costs of operating the distribution network that must be recovered by monthly fixed charges.

As discussed in previous sections, the hourly marginal cost of energy to each customer is likely to be increasingly volatile as the share of intermittent renewable generation resources connected to the grid increases. However, these prices reflect real-time conditions at each location in the transmission and distribution grid during each hour of the day and in that sense provide economically efficient signals for investments in storage devices, automated response technologies, and sensors and control systems that can reduce the cost of serving demand at all locations in the grid.

Investments in storage, automated response technologies and sensors and control systems are financially viable because they enable the owner to profit from the price swings caused by the intermittency of renewable production. The larger the average price difference between the discharging hours and charging hours of the day, the larger the revenues from a battery that stores energy during the low-price periods and discharges during the high price periods. A persistently high or low price of energy to a customer provides no revenue for a storage resource.

Setting the monthly fixed charge a customer faces using the mean and variance of a customer's consumption provides an additional source of benefits from installing distributed solar and storage capacity. A rooftop solar system will reduce the customer's average hourly consumption of grid-supplied electricity and a battery can reduce the volatility of the customer's hourly consumption of grid-supplied electricity, both will reduce the customer's monthly fixed charge.

It important to emphasize that all consumers with interval meters that can record their consumption during each pricing interval must face the real-time marginal cost of delivering energy to their location as their default price. This does not mean that any customers will ultimately pay for all of their electricity at this retail price, only that customers must face this price as their default option. Customers can purchase fixed-price and/or fixed quantity hedges against this short-term price risk. As noted in Section 4.3 of Wolak (2013), unless customers face this short-term price risk, they will have not the financial incentive to become actively involved in the wholesale market and make the up-front investments in storage and load-shifting technologies necessary to manage this wholesale price risk.

4.1 Active Involvement of Final Demand in the Wholesale Market

The active involvement of final consumers in the wholesale market can also reduce the amount of installed generation capacity needed to serve them and can reduce the cost of integrating an increasing amount of intermittent renewable generation. If enough consumers reduce their demand as a result of a high marginal cost of delivered energy each time a peaker unit would need to be turned on, this generation unit is no longer required to serve annual demand and the revenues necessary to pay for it would not need to be recovered from wholesale prices. If consumers respond to the real-time price and increase their demand when the marginal cost of grid-supplied electricity is low and increase their demand when the marginal cost of grid-supplied electricity is high these actions can allow the system operator to manage a system with significant intermittent renewable capacity with less operating reserves.[6]

A multi-settlement market with a day-ahead forward market and real-time market facilitates active participation by final demand in the wholesale market. This mechanism allows loads to purchase energy in the day-ahead market that they can subsequently sell in the real-time market. As discussed above, a customer that purchases 3 KWh in the day-ahead market for a given hour of the day and only consumes 2 KWh during that hour because of an extremely high real-time price, effectively sells the 1 KWh not consumed at the high real-time price. This revenue reduces the customer's monthly electricity bill.

Without the ability to purchase demand in the day-ahead market that is not consumed in real-time, demand reduction programs require the regulator to set an administrative baseline relative to which the customer can sell demand reductions. Bushnell, Hobbs, and Wolak (2009) argue that this approach to demand response can significantly reduce the system-wide benefits of active demand-side participation.

4.2. Informed Customers with Interval Meters Can Respond to Dynamic Retail Prices

To understand why it is essential that all customers face the real-time marginal cost of delivering electricity to their premises during each pricing interval as their default option, consider the three necessary conditions for active involvement of final consumers. First, customers must have the necessary technology to record their consumption on an hourly basis. Second, they must receive actionable information that tells them when to alter their consumption.[7] Third, they must pay according to a price that provides an economic incentive consistent with the actionable information to alter their consumption.

There is growing empirical evidence that all classes of customers can respond to short-term wholesale price signals if they have the metering technology to do so. Patrick and Wolak (1999) estimate the price-responsiveness of large industrial and commercial customers in the United Kingdom to half-hourly wholesale prices and find significant differences in the average half-

[6] Anderson, L.M., Hansen, L.G., Jensen, C.L., and Wolak, F.A. (2019) give an example of this phenomenon for the case of Denmark, as country with significant intermittent renewable generation capacity.

[7] McRae and Meeks (2016) resents the results of a field experiment in Central Asia that demonstrates the importance of actionable information for facilitating active demand-side participation.

hourly demand elasticities across types of customers and half-hours of the day. Wolak (2006) estimates the price-responsiveness of residential customers to a form of real-time pricing that shares the risk of responding to hourly prices between the retailer and the final customer. The California Statewide Pricing Pilot (SPP) selected samples of residential, commercial, and industrial customers and subjected them to various forms of real-time pricing plans in order to estimate their price responsiveness. Charles River Associates (2004) analyzed the results of the SPP experiments and found precisely estimated price responses for all three types of customers. More recently, Wolak (2011a) reports on the results of a field experiment comparing the price-responsiveness of households on a variety of dynamic pricing plans. For all of pricing plans, Wolak found large demand reductions in response to increases in hourly retail electricity prices across all income classes.

Although all of these studies find statistically significant demand reductions in response various forms of short-term price signals, none is able to assess the long-run impacts of requiring customers to manage short-time wholesale price risk. Section 2.2 of Wolak (2013) describes the increasing range of technologies available to increase the responsiveness of a customer to short-term price signals. However, customers have little incentive to adopt these technologies unless regulators are willing to install hourly meters and require customers to manage short-term price risk by setting each customer's default price equal to the real-time marginal cost of delivering electricity to their premises during each pricing interval.

4.3. Managing Bill Risk with Dynamic Pricing

Politicians and policymakers often express the concern that the subjecting consumers to real-time price risk will introduce too much volatility into their monthly bill. These concerns are, for the most part, unfounded as well as misplaced. Wolak (2013) suggests a scheme for facing a consumer with the hourly marginal cost of delivered energy for her consumption above or below a pre-determined load shape so that the consumer faces a monthly average price risk similar to a peak/off-peak time-of-use tariff.

Charging final consumers the same hourly default wholesale price as generation units owners, provides strong incentives for them to become active participants in the wholesale market or purchase the appropriate short-term price hedging instruments from retailers to eliminate their exposure to short-term price risk. These purchases of short-term price hedging instruments by final consumers increases the retailer's demand for fixed-price forward contracts from generation unit owners, which reduces the amount of energy that is actually sold at the short-term wholesale price.

4.4. Efficient Retail Pricing and Automated Response Technologies

Efficient retail pricing can provide the business case for investment in automated response technologies. A customer facing a default price equal to the hourly marginal cost of delivering energy to their premises may want to hedge this price risk by purchasing a fixed load shape for each day at a fixed price for the next 12 months from a retailer. This customer would then be able to sell energy it does not consume during any hour at the hourly wholesale price or purchase any power it needs beyond this baseline level at the retail price.[8] This type of pricing arrangement would result in a significantly less volatile monthly electricity bill than if the customer made all purchases at the hourly wholesale price.

If automated response devices that reduce the amount of energy flowing to certain appliances in the customer's house based on real-time price signals save more money on an annual basis in reduced energy costs than the annual cost of the devices, customers or their retailers can be expected to make these investments. A default retail price equal real-time price for delivering electricity to the customer's premises creates the business case for these investments. A single fixed retail price or single fixed price schedule regardless of real-time system conditions can never provide the revenue stream needed to finance investments in these technologies.

It important to emphasize that the customer does not even need to know they are responding to real-time price signals in order to benefit from facing the real-time price as their default price. Their retailer could offer them a fixed price for all the electricity they consume during the month based on their load shape and the condition that the customer installs certain devices in their home that allow their retailer to control the amount electricity flowing to certain appliances. For example, the customer would pay say $100 per month for their electricity as long as they do not over-ride the devices that their retailer installed and maintain a total consumption for the month less than a certain amount. Each time the customer over-rides a device the retailer could assess a penalty. The retailer could also assess a penalty for consuming more than the customer's monthly

[8] Wolak (2013) draws analogy between this pricing plan for electricity and how cellphone minutes are typically sold. Consumers purchase a fixed number of minutes per month and typically companies allow customers to rollover unused minutes to the next month or purchase additional minutes beyond these advance-purchase minutes at some penalty price. In the case of electricity, the price for unused KWhs and additional KWhs during a given hour is the real-time wholesale price.

purchase of energy. The retailer takes over the job of managing the customer's consumption within the month to keep the sum of the monthly fixed charge plus total real-time energy costs less than the $100 per month it charges the customer plus any penalties for over-riding the automated response devices or consuming too much energy within the month.

This "prices and devices" approach to electricity retailing significantly simplifies the active involvement of final consumes in the wholesale market. Setting the retailer's default cost of serving each customer equal to the customer's monthly fixed charge plus the customer's hourly marginal cost of energy delivered to their premises times customer's hourly consumption summed over all hours of the month provides the economic incentive for customers and retailers to manage real-time price risk in this manner.

5. Regulatory Safeguards to Support Efficient Pricing

Setting the efficient retail price as the default price for all customers entails exposing them to significant real-time price risk. The need to manage this price risk is what provides the economic incentive for customers to shift their consumption and make investments in devices that reduce their monthly electricity bill. This section discusses regulatory safeguards at the wholesale and retail market level aimed at maximizing the economic benefits consumers realize from efficient retail pricing.

5.1. Managing and Mitigating System-wide and Local Wholesale Market Power

There is general agreement among regulators that consumers should be protected against price volatility caused by the extreme exercise of unilateral market power. The configuration of the transmission network, the level and location of demand, as well as the level of output of other generation units can endow certain generation units with a significant ability to exercise unilateral market power. A prime example of this phenomenon is the constrained-on generation problem described earlier. The owner of a constrained-on generation unit knows that regardless of the unit's offer price, it must be accepted to supply energy. Without a local market power mitigation mechanism, there is no limit to what offer price that supplier could submit and be accepted to provide energy.

The system-wide market power problem is typically addressed through sufficient fixed-price and fixed-quantity long term contracts between suppliers and electricity retailers and large consumers. As discussed in Wolak (2000) and McRae and Wolak (2012), fixed-price forward contract obligations limit the incentive of suppliers to exercise system-wide unilateral market power in the short-term market.

5.1.1. Solutions to the Local Market Power Problem

There are a variety of regulatory mechanisms that exist around the world to address the local market power problem. In an offer-based wholesale market, the regulator must design and implement a local market power mitigation mechanism. In general, the regulator must determine when any type of market outcome causes enough harm to some market participants to merit explicit regulatory intervention. Finally, if the market outcomes become too harmful, the regulator must have the ability to suspend temporarily market operations. All of these tasks require a substantial amount of subjective judgment on the part of the regulatory process.

In all offer-based electricity markets a local market power mitigation (LMPM) mechanism is necessary to limit the offers a supplier submits when it faces is insufficient competition to serve a local energy or ancillary services need because of the combination of the configuration of the transmission network and concentration of ownership of generation units. As the share of intermittent renewable resources grows, the need for an LMPM mechanism increases, because the market operator must account for more reliability constraints in the dispatch process, which creates more opportunities for dispatchable generation units to exercise local market power.

A LMPM mechanism is a pre-specified administrative procedure (usually written into the market rules) that determines: (1) when a supplier possess the ability and incentive to exercise local market power worthy of mitigation, (2) what the mitigated supplier will be paid, and (3) how the amount the supplier is paid will impact the payments received by other market participants. Without a prospective market power mitigation mechanism, system conditions are likely to arise in all wholesale markets when almost any supplier can exercise substantial unilateral market power. It is increasingly clear to regulators around the world, particularly those that oversee markets with a significant amount of intermittent renewable generation capacity, that formal regulatory mechanisms are necessary to deal with the problem of insufficient competition to serve certain local energy needs.

An important component of any local and system-wide market power mitigation mechanism is the provision of information to market participants and public at large, is often termed, “smart sunshine regulation.” This means that the regulatory process gathers a comprehensive set of information about market outcomes, analyzes it, and make it available to the public in a manner and form that ensures compliance with all market rules and allows the regulatory and political process to detect and correct market design flaws in a timely manner. Smart sunshine regulation is the foundation for all of the tasks the regulatory process must undertake in the wholesale market regime. Wolak (2014) discusses the benefits of smart sunshine regulation and public data release on wholesale market performance.

Another regulatory tool for managing local and system-wide market power in an offer-based market is the configuration of the transmission network, which can determine the extent of competition that individual suppliers face in the short-term market. For this reason, the regulator must take a more active role in the transmission planning and expansion process, particularly as the share of intermittent renewables increases, to ensure that competition-enhancing upgrades that improve market performance are built. Wolak (2015) presents a framework for measuring the competitiveness benefits of transmission expansions in an offer-based wholesale market and applies it to the Alberta, Canada wholesale electricity market.

5.1.2. Cost-Based Short-Term Markets

An alternative approach to limiting system-wide and local market power used in a number of Latin American markets is a cost-based market. Under this mechanism generation unit owners do not submit offers to the market operator. Instead, the market operator takes the technical characteristics of generation units and input fuel prices to compute the variable cost of operating each generation unit. These variable cost estimates are used by the market operator to dispatch generation units and set market prices, which are typically equal to the highest variable cost necessary to meet demand. Galetovic, Munoz, and Wolak (2015) describe the operation of the cost-based market in Chile.

This mechanism avoids the need for a local market power mitigation mechanism, but is not without its challenges. For example, it does not completely close off opportunities for suppliers to exercise unilateral market power because they can still withhold their output from the cost-based dispatch as a way to increase short-term prices. They can also take actions to raise their regulated variable cost that enters the cost-based dispatch process. Wolak (2014) discusses the market efficiency trade-offs between offer-based versus cost-based markets.

5.2. Protecting Retail Consumers from Economic Harm

There are a number of regulatory safeguards to protect retail consumers such as: (1) retail competition, (2) monitoring retailer hedging activity to ensure that they are not imprudently exposed to short-term prices and (3) providing the necessary technology, information and retail pricing plans to provide consumers with low-risk ways to participate in the wholesale market.

5.2.1. Customer Choice

Customer choice is a crucial driver of the adoption of new technologies in the electricity sector. Efficient pricing without market participants having the freedom to respond to these economic signals will not produce the intended economic benefits to consumers. There are a number of competition and regulatory policies that facilitate customer choice. These are: (1) interval meters to record a customer's half-hourly or hourly consumption, (2) the ability of customers to share their consumption data with competitive energy service providers, (3) actionable information about available technologies and pricing plans, and (4) clearly specified default provider obligations.

5.2.2. Interval Meters

In order to set the dynamic retail prices that provide the business case for investments in storage, load-shifting and control technologies, customers must have interval meters. Without the ability to measure a customer's consumption within a pricing interval, it is impossible to provide the full economic benefits to the customer from shifting their consumption into or away from that time interval.

With mechanical meters it is only possible to measure at customer's consumption between two meter readings. As discussed in Wolak (2013), with monthly reading of a mechanical meter, the customer's monthly bill is reduced by the same amount of regardless of which hour in the month the customer reduces consumption.

Consequently, in order to unlock the full economic benefits of efficient pricing, all customer must have interval meters. In most industrialized countries this has been accomplished by making meter deployment and reading a regulated service provided by the distribution utility. Given the substantial economic benefits of that these meters enable, the declining cost of these meters and the need to eventually replace all mechanical meters, this approach offers customers the maximum flexibility to participate in the wholesale market.

It important to emphasize that installing meters that simply record consumption at a 15-minute level or finer level of temporal granularity is sufficient to accomplish this task. There is little need for meters with enhancements beyond these basic functions to be part of a regulated distribution service. Enhancements to this basic service can be provided by third-party energy service firms.

5.2 3. Sharing Customer Data

In order for retailers to compete for customers they must have information on the load shape of the customer. A customer with a daily load shape that is highest during low-priced hours of the day is cheaper to serve than one with a load shape that is highest during the high-priced hours of the day. The most straightforward approach to addressing this issue is to allow customers to opt into providing their historical hourly consumption data to competing retailers and other third-parties.

An alternative approach would be to provide all data in an anonymized manner to all potential retailers and demand-response providers and then allow the retailer to solicit customers based on their permission to be contacted. It is important to recognize that without granting third-party access to a customer's data, the diffusion of new storage, load-shifting and control technologies will likely be slowed as well as more costly.

5.1.4. Actionable Information

Under retail competition, the price-setting function of the regulator is no longer relevant. However, this does not mean that the regulator should no longer protect consumers from the exercise of market power. Instead, the regulator must now transition to assisting consumers with becoming more able market participants.

Regulators should inform customers of what is likely to be lowest cost retail pricing plan for them through a web-site or other customer engagement mechanism. Similar information could be provided about new storage, load shifting and control technologies that could be benefit the consumer. The regulatory could serve as the "honest broker" in introducing these technologies to electricity consumers by providing informational web-sites that assist them in deciding whether investments in these technologies make economic sense.

5.2.1. Monitoring Forward Contract Positions of Retailers

As noted above, fixed-price forward contract commitments sold by generation unit owners reduce their incentive to exercise unilateral market power in the short-term energy market because the supplier only earns the short-term price on any energy it sells in excess of its forward contract commitment and pays the short-term price for any production shortfall relative to these forward contract commitments. Consequently, fixed-price forward contracts also provide the buyer and seller of a contract with protection against short-term price risk for the quantity of energy traded in the forward contract.

Because short-term wholesale prices are likely to become more volatile with a larger renewable energy share, this role for fixed-price forward contracts is likely to become even more important. Moreover, because the failure of a retailer to adequately hedge their exposure to short-term prices can impose costs on all of its customers as well as the customers of other retailers if it goes bankrupt, there is regulatory rationale for ensuring that retailers are not imprudently exposed to short-term wholesale price risk.

This logic argues in favor of the regulator monitoring the forward contract positions of retailers as part of its regulatory oversight process to ensure that there is adequate fixed-price forward contract coverage of final demand. As discussed in Wolak (2003b) and reinforced by the simulation results of Bushnell, Mansur and Saravia (2008), the California electricity crisis is very unlikely to have occurred if there had been adequate coverage of California's retail electricity demand with fixed-price and fixed-quantity forward contracts. High levels of fixed-price forward contract coverage of final demand would have protected retailers selling to final consumers at a fixed price from having purchase significant amounts of energy in the short-term market at extremely high and prices, which eventually caused at least one large retailer to declare bankruptcy.

The regulatory process would require retailers to make regular filings of the their fixed-price retail load obligations and the fixed-price forward contracts they have to hedge the wholesale price risk associated with serving these fixed-price retail load obligations. To the extent that final consumers are willing to manage short-term price risk through dynamic pricing plans for some their hourly demand, retailers can reduce their fixed-price forward contract purchases.

5.3.4. Default Provider Obligation

The regulator must set a clear rules for determining the default provider obligation. Specifically, what is the retail price that a customer must pay if their retailer goes bankrupt or exits the industry. During the early stages of retail competition, when more entry and exit is likely to occur, it is important to have clear rules for determining this default provider obligation.

As retail competition matures, there is less need for a formal regulatory process to address this issue. If a customer's retailer exits, there should may competitors willing to provide service at a reasonable price.

5. Conclusions

This paper argues that efficient pricing at the wholesale and retail level is a key driver of the investments and innovations necessary to achieve a significant intermittent renewable energy share at least cost to consumers. Multi-settlement locational marginal pricing markets that co-optimize energy and ancillary services procurement are the consensus choice internationally for a market design that sets efficient short-term wholesale electricity prices. The cost of not adopting the most efficient wholesale pricing mechanism are likely to increase at the share of intermittent renewables increases.

The conventional approach to distribution network pricing is increasingly costly in regions where distributed solar PV can be easily deployed. Two suggestions are provided for improving the efficiency of distribution network pricing.

Efficient wholesale and retail pricing implies significantly greater price volatility, which requires a number of competition and regulatory safeguards to protect consumers that include a local and system-wide market power mitigation mechanism, transmission planning process that recognizes the competitiveness benefits of upgrades, dynamic pricing that protects consumers and retailers from imprudent exposure to short-term prices. Regulatory policies that encourage customer choice such as widespread deployment of interval meters, mechanisms for customers to share their data with third-party providers of energy services, and clearly defined default provider prices and obligations provide further protections for consumers while still providing the price signals necessary for consumers to adopt beneficial technologies and pricing structures that will facilitate the least cost transition to significantly more intermittent renewable generation capacity.

References

Anderson, Laura M., Lars G. Hansen, Carsten Lynge Jensen, and Frank A. Wolak. "Can Incentives to Increase Electricity Use Reduce the Cost of Integrating Renewable Resources." No. w25615. National Bureau of Economic Research, 2019.

Bohn, Roger E., Caramanis, Michael C. and Schweppe, Fred C. (1984) "Optimal Pricing in Electrical Networks over Space and Time," *RAND Journal of Economics*, Vol. 15, No. 5, 360-376.

Borenstein, Severin (2015) "The Private Net Benefits of Residential Solar PV: The Role of Electricity Tariffs, Tax Incentives, and Rebates, National Bureau of Economic Research Working Paper Number 21342, July.

Borenstein, Severin, Bushnell, James, and Wolak, Frank A. (2002) "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market," *American Economic Review*, December, 1367-1405.

Bushnell, J. B., Mansur, E. T., & Saravia, C. (2008). Vertical arrangements, market structure, and competition: An analysis of restructured US electricity markets. *The American Economic Review*, 98(1), 237-266.

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

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

Charles River Associates (2004) "Statewide Pricing Pilot Summer 2003 Impact Analysis," Charles River Associates, 5335 College Avenue, Suite 26, Oakland, California 94618.

Galetovic, A., Muñoz, C. M., & Wolak, F. A. (2015). Capacity payments in a cost-based wholesale electricity market: the case of Chile. *The Electricity Journal*, 28(10), 80-96.

McRae, Shaun D. and Meeks, Robyn (2016) "Price Perceptions Price Perceptions and Electricity Demand with Nonlinear Prices," available at http://www.sdmcrae.com/research/

McRae, Shaun D. and Wolak, Frank A. (2012) "How Do Firms Exercise Unilateral Market Power? Evidence from a Bid-Based Wholesale Electricity Market," *in Manufacturing Markets: legal, political and economic dynamics*," Jean-Michel Glachant and Eric

Brousseau (editors), Cambridge University Press.

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

Patrick, Robert H. and Wolak, Frank A. (1999) "Customer Response to Real-Time Prices in the England and Wales Electricity Market: Implications for Demand-Side Bidding and Pricing Options Design Under Competition," in *Regulation Under Increasing Competition*, Michael A. Crew (editor) Kluwer Academic Publishers, 155-182.

Tangeräs, Thomas, P. and Wolak, Frank A. (2017) "The Competitive Effects of Linking Electricity Markets Across Space and Time, available at http:// http://www.stanford.edu/~wolak.

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

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

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

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

Wolak, Frank A. (2011a) "Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment," *American Economic Review*, May, 83-87.

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

Wolak, Frank A. (2015). Measuring the competitiveness benefits of a transmission investment policy: The case of the Alberta electricity market. *Energy Policy*, 85, 426-444.

Wolak, Frank A., (2013) "Economic and Political Constraints on the Demand-Side of Electricity Industry Re-structuring Processes," *Review of Economics and Institutions*, Vol. 4, No. 1, Winter, Article 1.

Wolak, Frank A. (2014). Regulating competition in wholesale electricity supply. In *Economic Regulation and Its Reform: What Have We Learned*? (pp. 195-289). University of Chicago Press.

Wolak, Frank A. (2018) The Evidence from California on the Economic Impact of Inefficient Distribution Network Pricing, National Bureau of Economic Research Working paper number 25087, (September) available at https://www.nber.org/papers/w25087.

Figure 1: Two-Part Tariff Pricing

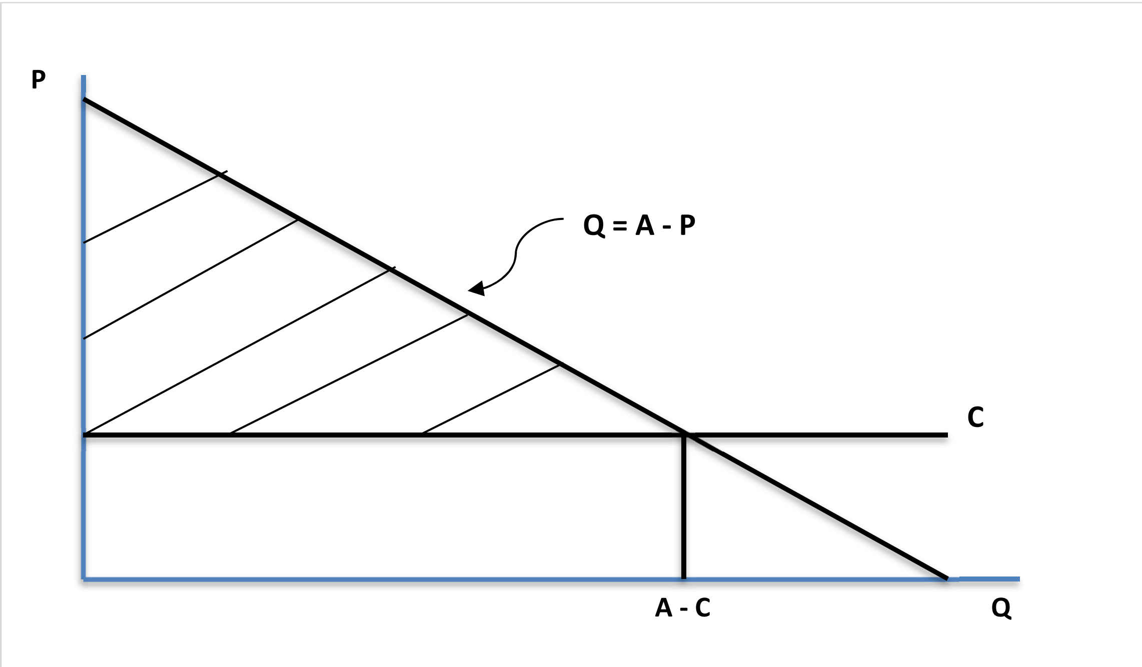

Figure 2: Worst-Case Fixed Fee Determination

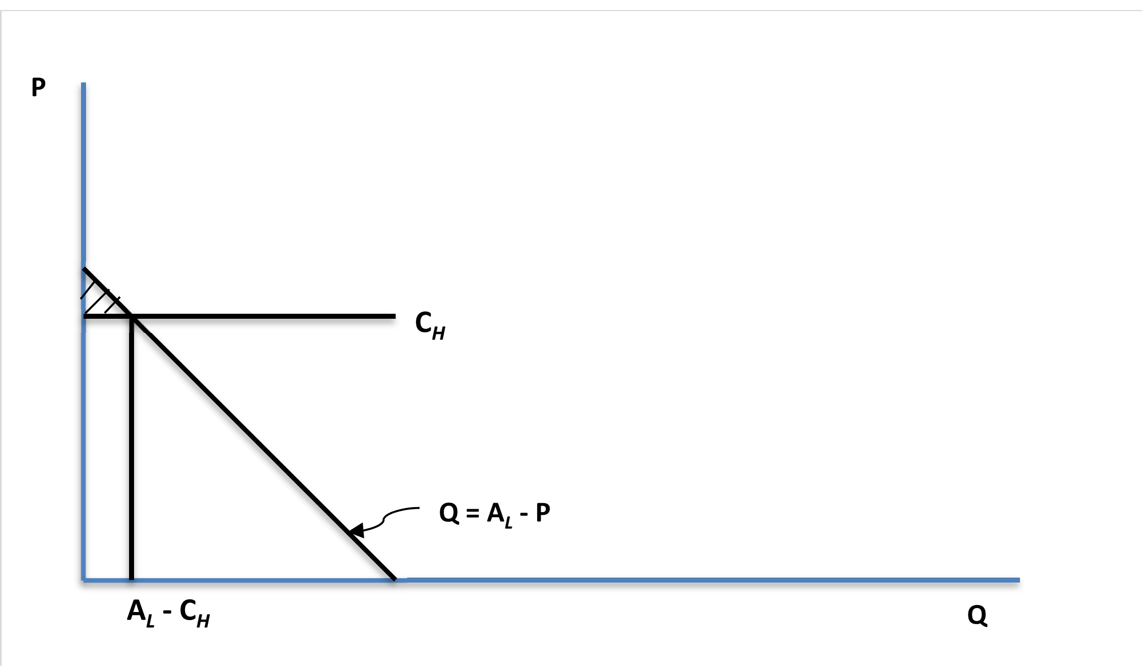

Figure 3: Best Case Fixed Fee Determination