
4. Wholesale electricity market design

Frank A. Wolak

1. INTRODUCTION

Although different jurisdictions around the world had different motivations for restructuring their electricity supply industries and introducing wholesale electricity markets, all had the goal of improving industry performance. Because electricity market design requires an explicit regulatory process to set market rules, initial conditions in the industry and political factors can drive the choice of a jurisdiction's initial market design. This choice can interfere with the ability of the subsequent regulatory process to improve industry performance and ultimately benefit electricity consumers.

Particularly in the United States (US) and Europe, there has been a convergence of market designs, although the standard market design in the US is quite different from the standard market design in Europe. This difference can be largely explained by differences in initial conditions and the political forces driving the restructuring process in each region. However, the desire of many jurisdictions to scale the amount of intermittent wind and solar generation resources is creating new market design challenges that are likely to lead to greater convergence between the market designs in the US and Europe.

International experience with more than 25 years of wholesale electricity market design has revealed several factors that are crucial to achieving lasting improvements in industry performance and tangible economic benefits to electricity consumers. These factors are: (1) the match between the short-term market used to set prices and dispatch generation units and how the actual electricity network is operated; (2) effective market and regulatory mechanisms to ensure long-term generation and transmission resource adequacy; (3) appropriate mechanisms to mitigate system-wide and local market power; and (4) mechanisms to allow active involvement of final demand in the short-term market.

As we discuss below, these factors must be addressed in any successful restructuring process because wholesale and retail market mechanisms decentralize many of the activities that formerly took place within the vertically integrated monopoly. In addition, at least during the initial stages of the restructured industry, there is a small number of wholesale and retail market participants, which implies that wholesale and retail market rules can significantly affect the behaviour of market participants, often to the detriment of electricity consumers.

Consequently, designing wholesale and retail market rules requires accounting for the impact each market rule has on the behaviour of individual market participants. These market rules must create economic incentives for actions by individual wholesale and retail market participants that enhance, or at least do not detract from, real-time system reliability or long-term supply adequacy.

The market designer must recognize that any wholesale or retail market rule will be exploited by all market participants to enhance their ability to maximize profits from

selling wholesale or retail electricity or minimizing retail electricity procurement costs. The most successful restructured markets are those with market rules that account for the self-interested behaviour of all market participants. Much of this chapter is devoted to analysing each of the above-mentioned four factors determining the performance of a restructured electricity supply industry from this perspective and using the lessons learned from international experience to provide recommendations for the design of a successful wholesale electricity market.

The remainder of this chapter proceeds as follows. Section 2 first describes why electricity requires an explicit market design process. Section 3 summarizes the factors leading to the different experiences with and outcomes of the wholesale market design process in the US versus Europe and other industrialized countries. Sections 4 to 7 present the four major lessons from international experience with wholesale market design. Section 4 outlines the importance of the match between the network model used to operate the forward market that sets prices and dispatch levels and the physical network used to operate the electricity supply industry in real time. This section argues that a multi-settlement locational marginal pricing (LMP) wholesale market is likely to achieve the best possible match between how the transmission network operates and how the wholesale market determines prices and dispatch levels for any feasible configuration of the transmission network. Section 5 explains why a liquid forward market for energy is likely to be the most efficient way to ensure both short-term and long-term resource adequacy. This section also discusses the properties of a capacity payment mechanism, an alternative approach to long-term resource adequacy. Section 6 describes why fixed-price long-term contracts are an effective mechanism for limiting the incentive of suppliers to exercise system-wide unilateral market power in the short-term market. This section also discusses local market power mitigation mechanisms that exist in all US markets and most international markets, although the details of these mechanisms differ across markets. Section 7 emphasizes the benefits of active involvement of final demand in the wholesale market, particularly in regions that have deployed hourly meters.¹ A multi-settlement LMP market is shown to provide the ideal platform for active participation of final demand.

Section 8 builds on the discussion in Sections 4 to 7 to propose a market design for developing countries that is likely to capture the major source of economic benefits possible from electricity industry restructuring, while recognizing the regulatory and institutional challenges facing developing countries. This market design avoids many of the more costly features of formal wholesale electricity markets in the industrialized world with a significantly reduced regulatory burden. This section also describes how this market design can be implemented in a small market with a vertically integrated utility that purchases a portion of its electricity from independent power producers.

Section 9 considers the question of what market design can successfully integrate a significant amount of intermittent renewable resources at both the transmission and distribution grid levels. A number of countries have policies to increase the amount of renewable energy they consume. I describe how a multi-settlement LMP market with a forward energy contracting-based long-term resource adequacy mechanism can support the efficient deployment of intermittent renewable resources in both the transmission and distribution grids.

2. WHY ELECTRICITY IS DIFFERENT

It is difficult to conceive of an industry where introducing market mechanisms at the wholesale and retail level is more challenging for a policymaker. Virtually every aspect of the technology of electricity delivery and how it has been historically priced to final electricity consumers enhances the ability of suppliers to raise the prices they are paid through their unilateral actions – what is typically referred to as exercising unilateral market power. Supply must equal demand at every instant in time and at each location in the transmission and distribution networks. If this does not occur then, these networks can become unstable, and brownouts and blackouts can happen. It is also very costly to store electricity. Constructing storage facilities typically requires substantial up-front investments and basic physics implies significantly more than 1 megawatt-hour (MWh) of energy must be produced to store 1 MWh of energy. Production of electricity is subject to extreme capacity constraints in the sense that it is physically impossible to get more than a pre-specified amount of energy from a generation unit in an hour. These capacity constraints limit the size of the supply response by competitors to the attempts of a generation unit owner to raise the price it is paid for electricity. Finally, delivery of the product consumed must take place through a potentially congested, looped transmission network, and how transmission capacity is allocated to different market participants exerts an enormous influence on their behaviour.

Historically, how electricity has been priced to final consumers makes wholesale demand extremely inelastic, if not perfectly inelastic, with respect to the hourly wholesale price. Customers are usually charged a single fixed price or according to a fixed non-linear price schedule for each kilowatt-hour (KWh) they consume during the billing cycle, regardless of the value of the wholesale price when each KWh is consumed. Paying according to a fixed-retail price schedule implies that these customers have hourly demands with zero price elasticity with respect to the hourly wholesale price, which significantly enhances the ability of suppliers to exercise unilateral market power in the short-term market.

The requirement to deliver electricity to final electricity consumers through a specialized transmission and distribution network that is too expensive to duplicate for a given geographic area precludes the usual approach to finding a market design that best meets the needs of consumers and producers. For most products, the market design process involves consumers deciding which products and locations to patronize and producers deciding which products and locations to serve. Some locations and products favour producers and others favour consumers. However, the existence of willing buyers and sellers of the same product at the same location is a necessary condition for trade to take place.

Coffee retailing is a recent example of the typical market design process. Historically, a customer interested in purchasing a cup of coffee would go to a diner or convenience store. However, specialized coffee retailers such as Starbucks entered and attracted customers, and, as a result, many traditional coffee outlets lost customers. The customers lost by traditional coffee shops and diners and gained by specialized coffee retailers reduced the profitability and increased the likelihood of exit by the former and increased the financial viability of the latter. This dynamic is continually taking place in all markets where consumers and producers can vote with their feet for their preferred market design.

This process of consumers and producers voting with their feet is not available to the electricity supply industry because the product must be injected and delivered to final consumers through the same transmission and distribution network, and even in developing countries customers have limited tolerance for an interruption of their supply of electricity. Moreover, reliable delivery of the electricity requires maintaining supply and demand balance at all locations in the grid at every instant in time. Consequently, any new supplier still delivers its electricity through this network and all customers still receive grid-supplied electricity from this network. As a consequence, the market design process must take place through explicit regulatory actions that set the rules for how market participants connect to the network and how they are paid for the electricity they inject and how they pay for the electricity they withdraw, rather than through the unilateral actions of producers and consumers of electricity.

3. WHOLESALE MARKET DESIGN IN THE US AND EUROPE

The electricity supply industries in the US and Europe and other industrialized countries started from different initial conditions that also led to different political motivations for electricity industry restructuring. Europe and other industrialized countries typically started with state-owned national or regional vertically integrated monopolies. The primary public policy concern was that these state-owned entities were over-capitalized and inefficiently operated. The political motivation was that privatizing them and introducing competition would create stronger incentives for efficient operation. For example, in England and Wales, the Central Electricity Generating Board (CEGB), a state-owned entity, was broken up and sold off by the Margaret Thatcher government to create three wholesale electricity suppliers – National Power, PowerGen and Nuclear Electric – and the National Grid Company that operated the transmission grid. The 12 regional electricity companies (RECs) that sold retail electricity and operated local distribution networks in their service territory were also privatized.² The Nordic countries, having both state-owned companies and local public authority-owned companies, did not choose to privatize them, but instead let them all compete in the same wholesale market (as Le Coq and Schwenen explain in Chapter 10 of this handbook). Australia and New Zealand pursued a similar strategy in restructuring their electricity supply industries, with both privately owned and government-owned companies competing to supply wholesale electricity.³

In the US, very few of the vertically integrated utilities in the regions that introduced a wholesale electricity market were state owned. The vast majority were shareholder owned, with their output prices set by state public utilities commissions for almost 100 years. Although far from perfect, this state-level regulatory process was significantly more effective at squeezing out productive inefficiencies than the state-owned utility model in Europe and in other industrialized countries.⁴ As Joskow (1997) notes, the major motivation for restructuring in the US was that competition for wholesale electricity would cause more efficient new capacity investment decisions than the vertically integrated regulated monopoly regime. According to Fabrizio, Rose and Wolfram (2007), there is also evidence that restructuring reduced plant labour and non-fuel expenses.

In contrast to the US, countries in Europe and other parts of the world had little experience with monopoly regulation given their history of state ownership of the electricity supply industry. Consequently, when the restructuring process began, these jurisdictions found their electricity supply industries with significantly different initial conditions relative to their privately owned, state-regulated counterparts in the US. Specifically, all US regions that restructured their electricity supply industries started with significantly less modern delivery infrastructure than many of their European counterparts. The efforts of state regulators in the US to limit retail price increases in response to the run-up in fossil fuel prices in the late 1970s led to very little investment in new transmission capacity until the early 2000s. Load growth over this time period in most parts of the country was met through generation capacity additions near major load centres rather than through large new capacity investments distant from the load centre served by new transmission capacity. Consequently, the wholesale markets in the US began operations with significantly less modern transmission networks and less transmission capacity than their counterparts in Europe and other industrialized countries, where the state-owned electricity supply industry was able to make significant investments in transmission capacity over this same time period.

This difference in initial conditions eventually led to a different evolution of wholesale market designs, despite the fact that most wholesale market designs in the US were initially very similar to those in Europe. The United Kingdom was the first European country to restructure in the early 1990s and it chose a market design that assumed a transmission network where all generation units were equally effective at meeting demand at any location in England and Wales. Specifically, this market set a single national price for electricity each half-hour of the day. The other early wholesale market in Europe, the Nord Pool, set potentially different prices in a small number of geographic zones in Norway and Sweden, the two initial members of the Nord Pool, which later expanded to include Denmark and Finland and now allows energy trading in 13 European countries.⁵

All the European markets have evolved to zonal or single-zone pricing models where the day-ahead forward market clears at this level of spatial granularity. These markets determine market-wide or zonal generation and load schedules and market prices. This is followed by a redispatch process where intra-zonal transmission constraints and other real-time operating constraints are managed by the system operator for that region or country.

Virtually all the US wholesale electricity markets began as either a single pricing zone market, as was the case in the New England ISO, or a zonal market, as was the case in the PJM Interconnection, New York ISO, California ISO and Electricity Reliability Council of Texas (ERCOT), which followed a similar design to the standard European market. These markets had a day-ahead forward market where generation and load schedules and prices were determined at the zonal or single-pricing zone level. This was followed by a congestion management or redispatch process run by the system operator to produce final schedules that were physically feasible given the configuration of the transmission network.

Starting with the PJM Interconnection in the Eastern US, market operators found that the incidence and magnitude of transmission congestion was so large and the cost of managing it increasingly expensive that a market design with a more spatially granular pricing was adopted. This process led the PJM Interconnection to adopt an LMP market

design with a day-ahead forward market and a real-time market – what is referred to as a multi-settlement LMP market design. All the other regions of the US had similar experiences with the operational challenges and cost of managing transmission congestion within the pricing zones in their single or multiple zone pricing markets. Each eventually implemented a multi-settlement LMP market design, with ERCOT being the last to do so in December of 2010.

Because of their history of state-owned utilities and significantly greater investments in transmission capacity from the 1970s up to the start of the restructuring process, wholesale markets in Europe and most other industrialized countries have managed to maintain their zonal market designs. However, the costs of making final schedules physically feasible has grown significantly as the share of intermittent renewable energy in many European countries has increased. In 2017, these costs were over 1 billion euros in Germany, more than 400 million euros in Great Britain, over 80 million euros in Spain and approximately 50 million euros in Italy.⁶ These costs have led a number of European countries to consider adopting more granular approaches to electricity pricing.

4. MATCH BETWEEN MARKET MECHANISM AND ACTUAL SYSTEM OPERATION

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 actually operated. As noted in the previous section, 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 can create 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, despite the fact that 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 a unit’s owner knows that it will be constrained-on, a profit-maximizing unit owner will submit an offer price far in excess of 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 can arise for constrained-off generation units. Constrained-off generation units are usually paid the difference between the market-clearing price and their offer price for not supplying electricity that the units would have supplied 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 to receive the highest possible payment for being constrained-off and raise the total cost of electricity supplied to final consumers. Bushnell et al. (2008a) 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 behaviour created by the divergence between the day-ahead zonal market model and the 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 reconciliations payment mechanism.

A zonal or single zone pricing market design can address the requirement that the market model used to set prices and dispatch levels and the model used to operate network in real time match if there is sufficient investment in transmission capacity to make the actual network match for as many hours of the year as possible the network assumed in the forward markets used to set schedules and prices. For example, a single zone pricing model can be a viable market design if the transmission planning process commits to transmission investments that make transmission congestion and the need for redispatch of generation units sufficiently infrequent and unpredictable as to not impact the behaviour of generation unit owners. As noted in Wolak (2015a), the wholesale electricity market for the province of Alberta in Canada sets a single price for the entire province and has adopted a transmission policy that ensures sufficient transmission capacity to make this energy market assumption a reality for actual system operation for the vast majority of hours of the year. The wholesale market in Australia operates a zonal market based on state boundaries. Because most of the major load centres are on the east coast of the country and the major generation units tend to be further inland near the coal and natural gas resources, historically there was a limited amount of intrazonal congestion. However, this situation is changing with the increasing amount of intermittent renewable generation capacity installed in the Australian electricity market.

4.1 Locational Marginal Pricing

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 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 prices in the wholesale market.

All LMP markets in the US co-optimize the procurement of energy and ancillary services.⁷ This means that all suppliers submit to the wholesale market operator their generation unit-specific willingness-to-supply schedules for energy and any ancillary service the generation unit is capable of providing. 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 ancillary services 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 ancillary service if doing so would violate a transmission or other operating constraint.

An important distinction between an LMP market design and the standard European market design is the centralized commitment of generation units to provide energy and ancillary services. European markets do not typically require all 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. 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 ancillary services and instead rely on sequential procurement of ancillary services after energy schedules have been determined. As Oren (2001) demonstrates, sequential clearing of energy and ancillary services markets increases the opportunities for generation unit owners to exercise unilateral market power in the ancillary services market, because they know that units committed to supply energy cannot compete in the subsequent ancillary services market.

In contrast, LMP markets that co-optimize the procurement of energy and ancillary services ensure that each generation unit is used in the most cost-effective manner based on the energy and ancillary services offers of all generation units, not just those owned by a single market participant. Specifically, the opportunity cost of supplying any ancillary service a unit is capable of providing will be explicitly taken into account 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 price for the only ancillary service the unit can supply is 5 \$/MW, then the unit will not be accepted to supply the ancillary service. At a \$0/MW offer price, it would be accepted to supply the ancillary service only if the price was greater than or equal to 10 \$/MW because of the 10 \$/MWh opportunity cost of energy for that unit. In contrast, self-commitment markets such as those that exist in Europe and other industrialized countries must rely on individual market participants to make the choice between supplying energy or ancillary services from each generation unit.

The LMP process sets potentially different prices at all locations in the transmission network, depending on the configuration of the transmission network and geographic location of demand and the availability of generation units. Because the configuration of the transmission network and the location of generation units and demand is taken into account in operating the market, only generation unit dispatch levels that are expected to be feasible in real time given the expected configuration of the transmission network will be accepted to serve demand and they will be paid a higher or lower LMP than other

units, depending on whether the generation unit is in a generation-deficient or a generation-rich region of the transmission network.

The nodal price at each location is the increase in the minimized value of the 'as-offered costs' objective function as a result of a one unit increase in the amount of energy withdrawn at that location in the transmission network. The price of each ancillary service is defined as the increase in the optimized value of the objective function as a result of a one unit increase in the demand for that ancillary service. In most LMP markets, ancillary services are procured at a coarser level of spatial granularity than energy. For example, energy is typically priced at the nodal level and ancillary services are priced over larger geographic regions. Bohn, Caramanis and Scheweppe (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 takes into account 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 the LMP market to account for the fact that this energy can quickly disappear and reappear.

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 LMP 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.

4.2 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 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 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 locational marginal prices (LMPs) and firm financial commitments to buy and sell electricity 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 no-load cost offers as well as energy offer curves, and both of these costs enter the objective function used to compute hourly generation schedules and LMPs 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. As noted earlier, to the extent that generation unit owners submit start-up, no-load and energy offer curves that are representative of their actual costs, the total cost of committing and dispatching the generation units that arises from this centralized unit commitment process is likely to be less than the total commitment and dispatch costs that result from a self-commitment market, such as those that exist in Europe and other industrialized countries.

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.

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 geographic region that contains the LMP market, taking into account 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 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 \$3000 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 \$2300, the \$3000 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 (what Europeans call a supplier) 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.

By this same logic, a multi-settlement nodal pricing market is well-suited to countries 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 (2011b) quantifies the magnitude of the economic benefits associated with the transition to nodal pricing from a zonal-pricing market that was very similar to the standard market design in Europe and other industrialized countries. On 1 April 2009, the California market transitioned to a multi-settlement nodal pricing market design from a multi-settlement zonal-pricing market. Wolak (2011b) compares the hourly conditional means of the total amount of input fossil fuel energy in millions of British thermal units (BTUs), the total hourly variable cost of production from fossil fuel generation units, and the total hourly number of starts from fossil fuel units before versus after the implementation of nodal pricing, controlling non-parametrically for the total hourly output of the fossil fuel units in California and the daily prices of the major input fossil fuels. Total hourly BTUs of fossil fuel energy consumed to produce electricity is 2.5 per cent lower, the total hourly variable cost of production for fossil fuels units is 2.1 per cent lower, and the total number of hourly starts is 0.17 higher after the implementation of nodal pricing. This 2.1 per cent 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.

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

A multi-settlement LMP market also values 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 \text{ MWh} \times 30 \text{ $/MWh} - 20 \text{ MWh} \times 60 \text{ $/MWh})/30 \text{ MWh}$ for the 30 MWh it produces, while the dispatchable thermal unit is paid an average price of 40 \$/MWh = $(40 \text{ MWh} \times 30 \text{ $/MWh} + 20 \text{ MWh} \times 60 \text{ $/MWh})/60 \text{ MWh}$ 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 \text{ MWh} \times 30 \text{ $/MWh} + 20 \text{ MWh} \times 10 \text{ $/MWh})/50 \text{ MWh}$ and the average price received by the thermal unit is 40 \$/MWh = $(60 \text{ MWh} \times 30 \text{ $/MWh} - 20 \text{ MWh} \times 10 \text{ $/MWh})/40 \text{ MWh}$. 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.

One complaint often levelled 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 to those regions. Because more customers live in urban areas than in rural regions, charging final consumers in urban areas a higher retail price to recover the LMPs at their location may be politically challenging for the regulator to implement.

Most regions with LMP markets have addressed this issue by charging all customers in a state, region or utility service territory a weighted average of the LMPs at all load withdrawal points in the geographic region. In the above example, this implies charging the urban and rural customers the weighted average of the LMPs in urban and rural areas, where the weight assigned to each price is the share of system load that is withdrawn at that location. Under this scheme, generation units continue to be paid the LMPs at their location, but all loads pay a geographically aggregated hourly price. For example, in Singapore all generation units are paid the LMPs 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 generation nodes in Singapore. This approach to pricing 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. Tangerås and Wolak (2018) present evidence that this market rule can result in more competitive behaviour in the wholesale market by vertically integrated suppliers with the ability to exercise unilateral market power.

5. MECHANISMS TO ENSURE LONG-TERM RESOURCE ADEQUACY

Why do wholesale electricity markets require a regulatory intervention to ensure long-term resource adequacy? Consumers want to be able to withdraw electricity from the network when they need it, just like other goods and services. But it is unclear why electricity is so fundamentally different from other products that it requires paying suppliers for production capacity to exist. For example, consumers want cars, but they do not pay for automobile assembly plants. They want point-to-point air travel, but they do not pay for airplanes. They want a loaf of bread, but do not pay for the existence of a bakery. All these industries are high fixed cost, relatively low marginal cost production processes, similar to electricity supply. Nevertheless, all these firms earn their return on capital invested by selling the good that consumers want at a price above the variable cost of producing it. Clearly, cars, air travel and bread are in many ways essential commodities, yet there is no regulatory intervention that ensures there is sufficient production capacity for these products to meet demand.

So, what is different about electricity markets that necessitates the need for a long-term resource adequacy mechanism? The answer lies in how short-term markets for these products operate relative to that for wholesale electricity. This difference is the result of the regulatory history of the electricity supply industry and the technology historically used to meter electricity. The limitation on the level of short-term prices and the way that supply shortfalls are dealt with in wholesale electricity markets creates what has been termed a ‘reliability externality’ that requires a regulatory intervention to internalize (Wolak 2013).

In the market for automobiles, air travel and even bread, there is no explicit prohibition on the short-term price of the good rising to the level necessary to clear the market. Take the example of air travel. Airlines adjust the prices for seats on a flight over time in an attempt to ensure that the number of customers travelling on that flight equals the number of seats flying. This can result in very different prices for a seat on the same flight, depending on when the customer purchases the seat. A customer that waits too long to purchase a seat faces the risk of an infinite price in the sense that all of the seats on the flight are sold out. This ability to use prices to allocate the available seats is also what allows the airline the flexibility to recover its total production costs. Airlines can set low prices to fill flights with low demand and extremely high prices on other flights, or at other times for the same flight, when demand is high.

The ability to use the short-term price to manage the supply and demand balance in the electricity supply industry is limited first by the fact that many wholesale electricity markets have offer caps that limit a supplier’s offer price into the wholesale market and/or a price cap that limits the magnitude of the eventual market-clearing price. In addition, historically virtually all electricity supply industries did not have hourly meters that could record a customer’s consumption during each hour of the month. Even today, most regions only have mechanical meters that compute the customer’s consumption for the entire month as the difference between two consecutive meter readings. With monthly or bi-monthly reading of mechanical meters, it is impossible for the utility to know how much electricity a customer consumed within a given hour of the month.⁸

Although these 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 a reliability externality, it is not the cause.

This externality exists because offer caps limit the potential downside to electricity retailers and large consumers (able to purchase from the short-term market) delaying their purchases of electricity until real-time operation. Specifically, if a 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 more than the amount suppliers are willing to offer at or below the offer cap. This outcome implies that the system operator must be forced to either abandon the market mechanism or curtail load until the available supply offered at or below the offer cap equals the reduced level of demand, as was the case a number of times during the period January 2001 to March 2001, and more recently in August 2020, in California.

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

Particularly for markets with very low offer caps, retailers have little incentive to engage in sufficient fixed-price forward contracting with generation unit owners to ensure a reliable supply of electricity for all possible realizations of future real-time demand. For example, a 200 MW generation unit owner that expects to run 100 hours during the year with a variable cost of 80 \$/MWh would be willing to sign a fixed-price forward contract to provide up to 200 MWh of energy for up to 100 hours of the year to a retailer. Because this generation unit owner is essentially selling its expected annual output to the retailer, it would want a \$/MWh price that at least exceeds its average total cost of supplying energy during that year. This price can be significantly above the average price in the short-term wholesale market during the hours in which this generation unit operates because of the offer cap on the short-term market and other market power mitigation mechanisms. This fact implies that the retailer with the option purchase energy from a short-term market with a low price cap would find it profit-maximizing not to sign a forward contract that allows the generation unit owner full cost recovery but instead wait until the short-term market to purchase the necessary energy at prices that are limited by the offer cap.

Although this incentive for retailers to rely on a price-capped short-term market is most likely to impact generation units that run infrequently, if the level of demand relative to the amount of available supply is sufficiently large, it can even impact intermediate and baseload units. Because of the expectation of very low prices in the short-term market

and the limited prospect of very high prices because of offer caps, retailers may decide not to sign fixed-price forward contracts with these generation unit owners and purchase their energy in the short-term market. By this logic, a short-term energy market with an offer cap always creates an incentive for retailers to delay purchasing some of their energy needs until real time, when these caps can be used to obtain the energy at a lower price than the supplier would be willing to sell it in the forward market.

The lower the offer cap, the greater is the likelihood that the retailer will delay its electricity purchases to the short-term market. Delaying more purchases to the short-term market increases the likelihood of the event that insufficient supply will offer into the short-term market at or below the offer cap to meet demand. If a retailer knows that part of the cost of its failure to purchase sufficient fixed-price forward contracts will be borne by other retailers and large consumers because of random curtailment, then it has an incentive to engage in less fixed-price forward contracts than it would in a world where all customers had hourly meters and all customers could be charged hourly prices high enough to cause them to reduce their demand to equal the amount of supply available at that price.

Because externalities are generally caused by a missing market, another way of characterizing this reliability externality is as a missing market for long-term contracts for energy. In this case, because retailers do not bear the full cost of failing to procure sufficient energy to meet their real-time needs in the future, there is a missing market for long-term contracts for long enough delivery horizons into the future to allow new generation units to be financed and constructed to serve demand under all possible future conditions in the short-term market.

The above discussion implies that unless the regulator is willing to eliminate the offer cap on the short-term market so that the short-term price can be used to equate available supply to demand under all possible future short-term market conditions, some form of regulatory intervention is necessary to internalize the resulting reliability externality. However, if customers do not have hourly meters that can record their consumption during each hour of the billing cycle, they have a limited ability to benefit from shifting their consumption away from high-priced hours, so having no offer cap on the short-term market would not be advisable.

As the above discussion makes clear, relying on a capped short-term energy market price to ensure long-term resource adequacy does not address the reliability externality and leaves both a 'missing money' and a 'missing market' problem. Capacity mechanisms are one approach to addressing this reliability externality designed primarily for thermal generation-dominated markets, where the major concern is insufficient generation capacity to meet future demand peaks. In intermittent renewable-dominated markets the major reliability concern is more likely to be insufficient renewable energy (that is, not enough water, wind or sunshine) to meet a sustained demand for energy, which implies that other approaches to addressing the reliability externality may dominate a capacity-based approach for these electricity markets.

As the share of intermittent renewable generation in a wholesale electricity market increases, the magnitude of the reliability externality is also likely to increase. The uncertain availability of wind and solar resources increases the magnitude and duration of potential future energy supply shortfalls that must be managed, which implies many more instances when a capped short-term energy market may not yield a sufficient

energy supply increase or demand decrease to maintain real-time supply and demand balance.

Two general approaches have been developed to address this reliability externality. The first is based on fixed-price and fixed-quantity long-term contracts for energy signed between generation unit owners and load-serving entities at various horizons to delivery. The second approach is a regulator-mandated capacity mechanism. Typically, the regulator requires that load-serving entities purchase sufficient firm generation capacity, a magnitude defined by the regulator, to cover their annual peak demand. Generation unit owners receive a regulator-determined payment for the capacity they provide to the load-serving entity. Differing degrees of regulatory invention are used to determine this \$/KW per year payment across the existing capacity mechanisms.

5.1 Fixed-price Forward Contract Approach to Long-term Resource Adequacy

The fixed-price forward contract solution is the standard approach used to ensure a real-time supply and demand balance in markets for products with high fixed costs of production. The prospect of a high real-time price for the product provides incentives for customers to hedge this real-time price risk through a fixed-price forward contract. A supplier benefits from signing such a contract because it has greater quantity and revenue certainty as result.

The airline industry is a familiar example of this phenomenon. There is a substantial fixed cost associated with operating a flight between a given origin and destination pair. Regardless of how many passengers board the flight, the airplane, pilot and co-pilot, flight attendants and fuel must be paid for. Moreover, there is a finite number of seats on the flight, so passengers wanting to travel face the risk that if they show up at the airport one hour before the flight and attempt to purchase a ticket, they may find that it is sold out or tickets are extremely expensive because of the high real-time demand for seats. Customers hedge this short-term price risk by purchasing their tickets in advance, which is a fixed-price, fixed-quantity (one seat) forward contract for travel on the flight. These forward market purchases allow the airline to better plan the types of aircraft and flight staff it will use to serve each route and how much fuel is needed for each flight.

Similar arguments apply to wholesale electricity markets to the extent that real-time prices can rise to very high levels. For example, in Australia the offer cap on the short-term market is currently A\$15 000 per MWh, yet annual average wholesale prices are less than 100 A\$/MWh. The potential for short-term prices at or near the price cap provides a very strong incentive for electricity retailers and large customers to purchase their electricity through fixed-price forward contracts, rather than face the risk of these extreme short-term prices. However, even at this level of the offer cap on the short-term market in Australia there have been a small number of half-hour periods when supply shortfalls occur, consistent with the reliability externality argument.

Purchasing fixed-price and fixed-quantity forward contracts far enough in advance of delivery for new entrants to compete to provide this energy ensures that retailers will receive a competitive forward market price for their purchases. These forward market purchases far in advance of delivery also ensure that the seller of the contract has sufficient time to construct the new generation capacity needed to meet the demand secured through the fixed-price forward contracts. Consequently, in the same sense that

fixed-price forward contracts for air travel allow an airline to better match airplanes and flight staff to routes, fixed-price forward contracts for electricity allow generation unit owners to choose the least-cost mix of capacity to serve the demand that has purchased the fixed-price forward contracts for energy.

Key to the success of this strategy for obtaining sufficient generation capacity to meet future demand is the threat of very high short-term prices, which provides the incentive for load-serving entities to sign fixed-price forward contracts for their expected future demand far enough in advance of delivery to allow new entrants to compete with existing generation unit owners in the provision of these forward contracts for energy. However, most regions with restructured electricity markets are unwilling to allow short-term prices to rise to the level allowed in Australia. Until 17 November 2016, all wholesale markets in the US that were under the oversight of the Federal Energy Regulatory Commission (US FERC) had offer caps equal to 1000 \$/MWh. FERC Order 831 revised that policy, allowing suppliers to submit, after that date, the higher of 1000 \$/MWh or the unit's verified cost-based incremental energy offer. This order also capped verified cost-based incremental energy offers at 2000 \$/MWh.⁹ The ERCOT, the only non-FERC jurisdictional wholesale market in the US, currently caps the offer price that suppliers can submit at 9000 \$/MWh, which in US dollars is very close to Australia's offer cap. Again, this level of the offer cap significantly limits but does not eliminate the likelihood of supply shortfalls in the real-time market due to the reliability externality.

Many wholesale electricity markets outside the US, particularly those in the developing countries, have offer caps far below 1000 \$/MWh. Low offer caps do not create a strong enough incentive for load-serving entities to purchase enough fixed-price forward contracts far enough in advance of delivery to ensure sufficient generation capacity to meet future demand. Consequently, in a number of Latin American countries, there are regulator-mandated requirements for load-serving entities to purchase certain percentages of their final demand in fixed-price forward contracts in advance of delivery. For example, it could be 90 per cent of forecast demand one year in advance, 85 per cent two years in advance and so forth. This regulatory mandate provides sufficient demand for long-term contracts far enough in advance of delivery to ensure enough generation capacity to meet future demand.

It is important to emphasize that mandating these contracting levels is unlikely to impose a financial hardship on retailers that lose customers to competing retailers. If a retailer purchased more fixed-price forward contract coverage than it ultimately needs because it lost customers to a competitor, it can sell this obligation in the secondary market. Unless the market demand for energy in the future is unexpectedly low, this retailer is just as likely to make a profit on this sale as it is to make a loss, because one of the retailers that gained customers needs forward contracts to meet its regulatory requirements for coverage of its final demand. Only in the unlikely case that the aggregate amount of forward contracts purchased is greater than the realized final demand for the entire market will there be a potential for stranded forward contracts held by retailers that lose customers.

Fixed-price forward contract obligations also significantly limit the incentive of generation unit owners to exercise unilateral market power in the short-term market. To understand this logic, let PC equal the fixed price at which the generation unit owner agrees to sell energy to an electricity retailer in a forward contract and QC equal the

agreed upon quantity of energy sold. This contract is negotiated in advance of the date in which the generation unit owner will supply the energy, so the value of PC and QC are predetermined from the perspective of the supplier's behaviour in the short-term wholesale market.

Wolak (2000) demonstrates that the quantity of fixed-price forward contract obligations held by the generation unit owner determines what short-term market price the firm finds ex post profit-maximizing given its marginal cost of producing energy, the supply offers of its competitors and the level of aggregate demand. Incorporating the payment stream a generation unit owner receives from its forward contract obligations, its variable profit function for a given hour of the day is:

$$\pi(PS) = (PC - C)QC + (QS - QC)(PS - C) \quad (4.1)$$

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

The first term in (4.1) is the variable profit from the forward contract sales and the second term is the additional profit or loss from selling more or less energy in the short-term market than the generation unit owner's forward contract quantity. Because the forward contract price and quantity are negotiated in advance of the delivery date, the first term, $(PC - C)QC$, is a fixed profit stream to the generation unit owner before it offers into the short-term market. The second term depends on the price in the short-term market, but in a way that can significantly limit the incentive for the generation unit owner to raise prices in the short-term market.

For example, if the generation unit owner attempts to raise prices by withholding output, it could end up selling less in the short-term market than its forward contract quantity ($QC > QS$), and if the resulting market-clearing price is greater than the firm's marginal cost ($PS > C$), the second term in (4.1) will be negative. Consequently, only in the case that the generation unit owner is confident that it will produce more than its forward contract quantity in the short-term market does it have an incentive to withhold output to raise short-term prices.

The quantity of forward contract obligations held by a firm's competitors also limits its incentive to exercise unilateral market power in the short-term market. If a producer knows that all of its competitors have substantial fixed-price forward contract obligations, then this producer knows these firms will submit offer curves into the short-term market close to their marginal cost curves. Therefore, attempts by this generation unit owner to raise prices in the short-term market by withholding output are likely to be unsuccessful because the aggressiveness of the offers into the short-term market by its competitors with substantial fixed-price forward contract obligations limits the price increase a producer can expect from these actions.

This dynamic creates the following virtuous cycle of forward contracting. If a producer knows all its competitors have a substantial amount of their expected energy sales covered by fixed-price forward contracts, then it has an incentive to sign fixed-price forward contracts for a substantial fraction of its expected energy sales. When all generation unit owners have a substantial fraction of their expected energy sales covered by

fixed-price forward contracts, then all of them have a common interest in reducing the cost of meeting these fixed-price forward contract obligations.

The resulting reductions in the short-term prices creates another virtuous cycle from fixed-price forward contracts. Generation unit owners that have sold forward contracts have a strong incentive to make the cost of supplying these contracts as low as possible. This should result in low short-term prices for energy, which will then be factored into subsequent negotiations for the next round of fixed-price forward contracts. This dynamic for reducing short-term wholesale prices results from a persistent high level of coverage of final demand by fixed-price forward contracts that creates the incentive for all generation unit owners to reduce the cost of meeting their forward contract obligations.

5.2 Capacity Mechanism Approach to Long-term Resource Adequacy

Particularly in the US, capacity mechanisms appear to be a holdover from the vertically integrated regulated regime with regional power pools where capacity payments compensated generation units for their capital costs, because the regulated power pool typically only paid unit owners their variable operating costs for the electricity they produced. Therefore, all fixed costs had to be recovered through other mechanisms besides the sale of wholesale electricity.

Capacity payments typically involve a dollar per kilowatt per year (\$/kW/year) payment to individual generation units based on some measure of the amount of their capacity that is available to produce electricity during stressed system conditions, what is often referred to as the unit's 'firm capacity'. This value depends, among other things, on the technology of the generating unit. For example, a baseload coal-fired unit would have a firm capacity value very close to its nameplate capacity. Indeed, the firm capacity of a thermal unit is usually equal to the unit's capacity in megawatts times its availability factor.¹⁰

In hydroelectric-dominated markets, determining the firm capacity of a generation unit is an extremely challenging task. The firm capacity of a hydroelectric generation unit is typically based on the amount of energy the unit is capable of providing under the worst possible hydrological conditions. However, it is difficult, if not impossible, to determine the maximum amount of capacity or energy a hydroelectric unit can provide under these conditions, so there is a significant degree of arbitrariness in setting the firm capacity of a hydroelectric unit. Second, because every hydroelectric unit owner would like a larger capacity value for its generation unit, to avoid accusations of arbitrary firm capacity values for individual generation units, the entity making this decision typically bases the figure on the amount of energy the unit produced during the historically worst hydrological conditions even though the system operator may have sound reasons for believing that this firm capacity value is set too high. As consequence, particularly in Latin America, there are numerous examples of capacity mechanisms that failed to ensure an adequate supply of energy and rationing conditions have been declared. Virtually all the restructured markets in Latin America that have capacity mechanisms – specifically, Brazil, Chile and Colombia – have experienced supply shortfalls that have required rationing. McRae and Wolak (2019) present an analysis of the most recent supply short-fall period in Colombia and conclude that the perverse incentives created by the capacity payment mechanism there was a major contributing factor to this outcome.

Wind and solar generation units have a firm capacity value significantly below their nameplate capacity but substantially higher than the amount of energy these units are able to produce during stressed system conditions, which suggests that the capacity market construct is poorly suited to an electricity supply industry with significant intermittent renewable generation capacity. For example, on an extremely hot night solar generation units are not likely to produce any energy, while on a hot sunny day very little wind energy will be produced. Consequently, the process of computing the firm capacity values for wind and solar units involves a significant number of unverifiable assumptions often aimed at increasing the resulting firm capacity values.

Capacity mechanisms differ along a number of dimensions. In some regions, the payment is made to all generation unit owners regardless of how much total generation capacity is needed to operate the system. In other regions, the independent system operator (ISO) specifies a system-wide demand for capacity equal to peak system demand plus some planning reserve, typically between 15 to 20 per cent, and only makes capacity payments to enough generation units to meet this demand.

There have been attempts to use market mechanisms to set the value of the \$/kW/year payment to the generation units needed to meet the total demand for capacity. However, these capacity markets have been subject to almost continuous revision because they are extremely susceptible to the exercise of unilateral market power. The nature of the product sold – installed generation capacity – and a publicly disclosed perfectly inelastic demand for the product creates extreme opportunities for suppliers to exercise unilateral market power.

In early versions of Eastern US capacity markets, there were instances of the exercise of enormous amounts of unilateral market power. During the off-peak months of the year when no single supplier is pivotal in the capacity market, the price paid for capacity was very close to zero, which is the marginal cost of a supplier providing an additional megawatt of available capacity from existing generation capacity.¹¹ During the peak and shoulder months when one or more suppliers are pivotal in the capacity market, there was no limit on the price a supplier could charge.

This market power problem leaves open the question of how to set the value of the \$/kW/year price cap on the capacity payment. In all regions of the US with capacity payment mechanisms, there is an administratively set process for determining this price. The value of the maximum capacity payment is based on the regulator's estimate of the annual \$/kW fixed cost of a peaking generation unit. This maximum price is typically backed by the argument that because of the offer cap on the short-term market and other market power mitigation mechanisms this peaking unit could only set an energy price slightly higher than its variable operating costs. Because this generation unit and all other generation units are missing the hours when the market price would rise above their variable operating costs, the annual \$/kW cost of the peaking unit is needed to compensate all generation units for the revenues they do not receive because of the offer cap and market power mitigation mechanisms.¹²

This logic for setting the value of the \$/kW/year capacity payment explicitly assumes that the real-time demand for electricity is completely price inelastic and that suppliers are unable to exercise significant amounts of unilateral market power in the short-term market for energy. Both of these assumptions are clearly false. An increasing number of jurisdictions around the world are installing hourly meters that allow dynamic pricing

plans to be implemented. Wolak (2013) discusses these technologies and the pricing plans they enable.

A number of studies document the exercise of unilateral market power in wholesale electricity markets. Wolfram (1999) applies a structural model of imperfect competition to measure the extent of unilateral market power exercised in the early England and Wales electricity market. Wolak and Patrick (2001) document the strategic use of market rules to raise prices by two large generation unit owners in this same electricity market. Wolak (2003a) documents changes in the unilateral ability of the five largest generation unit owners in the California electricity market to exercise unilateral market power from 1998 to 2000. Borenstein, Bushnell and Wolak (2002) document the magnitude of the rent transfers and economic inefficiencies caused by these actions. McRae and Wolak (2012) demonstrate that the four largest generation unit owners in New Zealand have both a substantial ability and incentive to exercise unilateral market power.

Capacity mechanisms make it extremely difficult for consumers to benefit from electricity industry restructuring relative to the case of a market without a capacity payment mechanism and active demand-side participation in the wholesale market. Recall that the capacity payment is made to either all generation units in the system or all generation units needed to meet the ISO's demand for capacity. On top of this, all generation unit owners typically receive the same market-clearing price for capacity. Thus, to the extent that producers are able to exercise unilateral market power in the short-term energy market, they can raise market-clearing energy prices significantly above the variable cost of the highest cost unit operating within the hour for all hours of the year, on top of receiving a market-clearing capacity price set by the highest offer price needed to meet the system demand for capacity.

As noted above, capacity payment mechanisms are typically accompanied by offer caps on the short-term energy market that significantly limit the incentive for final consumers to become active participants in the short-term wholesale market. For example, if the maximum wholesale price in an hour is 250 \$/MWh because of an offer cap at this level, then a 1 KWh reduction in demand for a residential customer (a very large demand reduction) during an hour only saves the customer 25 cents, which is likely to be insufficient to cause that consumer to reduce its demand. This lack of an active demand-side in the wholesale market impacts how generation unit owners offer their units into the market, because all producers know that system demand will be the same regardless of the hourly wholesale price.

Active participation by final demand substantially increases the competitiveness of the short-term wholesale market because all producers know that higher offer prices will result in less of their generation capacity being called upon to produce because the offers of final consumers to reduce their demand are accepted instead. Without an active demand-side of the wholesale market, generation unit owners know that they can submit offers that are further above their variable cost of supplying electricity and not have these offers rejected. Consequently, a market with a capacity payment mechanism can charge consumers for the \$/kW/year fixed cost of a peaker unit for their entire capacity needs and then give producers greater opportunities to exercise unilateral market power in the short-term market, which clearly reduces the likelihood that consumers will realize economic benefits from the restructuring of the electricity sector.

Another argument given for capacity payments is that they reduce the likelihood of long-term capacity inadequacy problems because the promise of a capacity payment provides incentives for new generation units to enter the market. However, until very recently, in most markets around the world, capacity payments were only promised for at most a single year and only paid to existing generation units. Both of these features substantially dulled the incentive for new generation units to enter the market, because a generation unit that entered the market only had a guarantee of receiving the capacity payment for one year and no guarantee that if it received the payment the first year the unit owner would continue to receive it later. This led the Eastern US wholesale markets to develop a long-term capacity product that is sold two to three years in advance of delivery to provide a sufficient lead time for new generation units to participate and capacity payments beyond a single year. This is a positive development for capacity markets, but it also raises the question of why not simply transition to a forward energy purchase requirement rather than a forward capacity requirement, given that most consumers do not want more generation capacity built but they do want their future energy needs met.

Capacity markets are also poorly suited to regions with a significant share of intermittent renewables. In these markets it is rarely, if ever, the case that there is a capacity shortfall in the sense that there is insufficient installed generation capacity to meet peak demand. The more common problem is insufficient energy, typically in the form of water stored behind a dam, to meet anticipated demand. With wind and solar photovoltaic generation units, capacity shortfalls are also extremely unlikely. It is more likely that the sun does not shine or the wind does not blow for a sustained period of time. In both these cases, the problem is not a capacity shortfall but an energy shortfall. Consequently, a capacity payment mechanism that focuses on ensuring adequate installed capacity is unlikely to deliver the most cost-effective solution for consumers to the problem of long-term energy adequacy in regions with a significant amount of intermittent renewable resources.

The argument for a capacity market is strongest in a region with all dispatchable thermal generation units and no potential for active participation of final consumers in the wholesale market, particularly if the capacity procurement decision is made far enough in advance of delivery and for a long enough period of time to support new investment. The similarity between this forward capacity market solution and the long-term energy contracting solution argues in favour of the latter solution to the long-term resource adequacy problem. Galetovic, Muñoz and Wolak (2015) use the example of the Chilean market design to demonstrate the market efficiency improvements from transitioning from a capacity payment-based market to an energy contracting-based market.

5.3 The Role of a Liquid Forward Market for Energy

The previous two subsections emphasize that short-term energy and capacity markets are extremely susceptible to the exercise of unilateral market power and the key to long-term resource adequacy at a reasonable price is the purchase of sufficient energy or capacity far enough in advance of delivery by electricity retailers and large customers for new entrants to compete with existing generation unit owners to provide the product.

Signing a fixed-price forward contract for energy or capacity a day, month, or even a year ahead of delivery limits the number of firms and technologies that are able to provide this energy or capacity. For example, a contract negotiated one day in advance limits the sources of supply to existing generation units able to produce energy the following day. Even a year in advance limits the sources that can compete to the generation units already in place because it takes longer than 18 months to site and build a substantial new generation unit in virtually all wholesale electricity markets. Therefore, to obtain the most competitive prices, at a minimum, the vast majority of the fixed-price forward contracts should be negotiated far enough in advance of delivery to allow new entrants to compete with existing generation unit owners.

This logic argues for regulatory intervention to internalize the reliability externality through a long-term resource adequacy process that ensures a liquid forward market for energy for delivery two to three years into the future. If a liquid forward market for energy exists at this time horizon and there is adequate demand for energy at this delivery horizon, a restructured market will achieve long-term resource adequacy. A liquid forward market at the two- to three-year delivery horizon implies less need for regulatory intervention into shorter-term forward markets. The regulator can raise the offer cap on the short-term market and this will stimulate the demand for retailers and large consumers to hedge their wholesale energy purchases at delivery horizons less than two years into the future. By purchasing a hedge against the short-term price risk at the locations in the network where it is going to withdraw energy, the buyer can rely on the financial incentives that the seller of the contracts faces to procure or produce this energy at the lowest possible cost.

Focusing the long-term resource adequacy process on the construction of new generation units misses the important point that there is an increasing number of ways for markets to achieve long-term resource adequacy besides building additional capacity. For example, by the appropriate choice of the mix of generation units, the same pattern of hourly demand throughout the year can be met with less total generation capacity and that can also cost electricity consumers less on an annual basis. Investment in distributed generation and storage, and active participation of demand in the wholesale market can also allow the same number of customers to be served with less grid-connected generation capacity.

It is important to emphasize that most capacity mechanisms do little to limit the ability or incentive of generation unit owners to exercise unilateral market power in the short-term energy market. Capacity mechanisms typically have the requirement that the generation unit owner must offer their capacity into the short-term market at or below the offer cap on the short-term energy market. This requirement does little to limit the ability or incentive of generation unit owners to raise short-term prices by exploiting the distribution of residual demand curves they face, as long as the resulting short-term prices are below the offer cap. Even in capacity markets with a scarcity of payment refunds, the capacity payment mechanism does not limit the ability or incentive of producers to exercise unilateral market power below the scarcity price.¹³ This is different from the case that a producer has a fixed-price forward contract obligation to supply energy, which endows the producer with a strong incentive to reduce the cost of meeting this forward contract obligation for energy.

Another advantage of focusing on the development of a liquid forward market for energy instead of capacity is that an active forward market for energy has other hedging

instruments besides so-called ‘swap contracts’, where a generation unit owner and a retailer agree to a fixed price at a location in the transmission network for a fixed quantity of energy. Cap contracts are also very effective instruments for guarding against price spikes in the short-term market and for funding peaking generation capacity. For example, a generation unit owner might sell a retailer a cap contract that says that if the short-term price at a specific location exceeds the cap contract exercise price, the seller of the contract pays the buyer of the contract the difference between the spot price and the cap exercise price times the number of MWh of the cap contract sold. For example, suppose the cap exercise price is 300 \$/MWh and market price is 400 \$/MWh, then the payoff to the buyer from the cap contract is $100 \text{ \$/MWh} = 400 \text{ \$/MWh} - 300 \text{ \$/MWh} \times$ the number of MWh sold. If the spot price is less than 300 \$/MWh, then the buyer of the cap contract does not receive any payment.

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

Cross-hedging between generation resource owners is likely to become even more important in ensuring long-term resource adequacy in energy systems with significant amounts of wind and solar resources. A wind or solar resource owner that sells a fixed-price and fixed-quantity forward contract for energy to a retailer is going to need to reinsure the quantity risk associated with such a contract. The wind or solar resource owner can sign a contract with a thermal resource owner that provides insurance against the quantity risk it faces from selling a fixed-price and fixed-quantity contract. For example, the wind resource owner could purchase a cap contract for the quantity of energy sold in the fixed-price, fixed-quantity contract to the retailer at a certain strike price and in this way have insurance against having to purchase energy from the short-term market at an extremely high price when the wind or solar resource is not producing energy. The up-front payment to the thermal resource for the price spike insurance would help to finance the fixed costs of the thermal resource that operates significantly less frequently because of the large amount of intermittent renewable generation capacity. The wind or solar resource owner would then factor in the cost of this quantity risk insurance in the price it is willing to sell any fixed-price, fixed-quantity forward contract for energy.

One question often asked about an approach that focuses on the development of an active forward market for energy is whether sufficient generation resources will be built to meet demand if consumers only buy forward financial hedges against the spot price at their location in the network. On this point, it is important to bear in mind the incentives faced by a seller of the forward financial contract once this contract has been sold. The generation unit owner maximizes its expected profits by ensuring that the forward contract quantity of energy can be purchased at the agreed-upon location in the spot market (or whatever market the forward contract clears against) at the lowest possible short-term price. The seller of the contract bears all the risk associated with higher spot prices at that

location. To prudently hedge this risk, the seller has a very strong incentive to ensure that sufficient generation capacity is available to set the lowest possible price in the short-term market at that location in the network for the quantity of energy sold in the fixed-price forward contract.

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

An additional source of economic benefits from the energy-based resource adequacy process is that the energy contract adequacy approach does not require the regulator to set the total amount of firm capacity needed to meet system demand. Instead, the regulator only ensures that retailers and large customers have adequate fixed-price forward contract coverage of their demand at various delivery horizons into the future and then relies on the incentives that the sellers of these contracts face to construct sufficient generation capacity or procure other resources to meet these forward contract obligations for energy. The sellers of these forward contracts for energy have a strong incentive to find the least-cost mix of generation and demand-side resources necessary to meet their contractual obligations.

6. MANAGING AND MITIGATING SYSTEM-WIDE AND LOCAL MARKET POWER

The configuration of the transmission network, the level and location of demand, as well as the level of output of other generation units can endow certain generation units with a significant ability to exercise unilateral market power in a wholesale market. 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 (LMPM) mechanism, there is no limit to what offer price that generation unit owner 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 electricity producers and retailers and large consumers. The logic of Wolak (2000) described in Section 5.1 demonstrates how fixed-price, fixed-quantity forward contracts limit the incentives of generation unit owners to exercise system-wide unilateral market power in the short-term market.

6.1 Solutions to the Local Market Power Problem

In an offer-based market, the regulator must design and implement an automatic LMPM mechanism that is built into the market-clearing software. 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 temporarily suspend market operations. All these tasks require a substantial amount of subjective judgment on the part of the regulatory process, which can be extremely challenging for countries and regions with limited regulatory experience.

In all offer-based electricity markets, an LMPM mechanism is necessary to limit the offers a generation unit owner submits when it faces insufficient competition to serve a local energy need because of a combination of the configuration of the transmission network, the levels and geographic distribution of demands, and the concentration of ownership of generation units. One lesson from the experience of US markets (in particular) is that system conditions can arise when virtually any generation unit owner has a substantial ability and incentive to exercise unilateral market power. That is why prospective market power mitigation mechanisms where certain units are designated in advance as having the ability to exercise unilateral market power, a practice typically used in Europe and other industrialized countries and initially employed in the US as well, are likely to miss many instances where the exercise of substantial unilateral market power is possible.

An LMPM mechanism built into the market software that relies on actual system conditions to determine whether a generation unit has a substantial ability and incentive to exercise unilateral market power is likely to be significantly more effective than the prospective approaches used in Europe and initially in the US. This logic explains why all US markets currently have such a mechanism built into their market software and running automatically during each pricing interval.

An LMPM mechanism is a pre-specified administrative procedure (written into the market rules) that determines: (1) when a producer has local market power worthy of mitigation; (2) what the mitigated producer will be paid; and (3) how the amount the producer 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 generation unit owner can exercise substantial unilateral market power. It is increasingly clear to regulators around the world, particularly those that operate markets with limited amounts of transmission capacity, that these automatic regulatory interventions 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 the public at large. This is often termed 'smart sunshine regulation' and implies that the regulatory process gathers a comprehensive set of information about market outcomes, analyses it and makes it available to the public in a manner and form that increases the likelihood of market participants' 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 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 tool a regulator has in managing local and system-wide market power in an offer-based market is determining the configuration of the transmission network. Because the configuration of the transmission network can often determine the extent of

competition that individual generation unit owners face, the regulator must take a more active role in the transmission planning and expansion process to ensure that competition-enhancing upgrades that improve market efficiency are built. Welsh (2020) describes the distinctly different roles of transmission network planning in the vertically integrated monopoly regime versus the wholesale market regime. Wolak (2015a) presents a framework for measuring the competitiveness benefits of transmission expansions in an offer-based wholesale market and applies it to the wholesale electricity market of the province of Alberta in Canada.

6.2 Cost-based Short-term Market

An alternative approach to market power mitigation is 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.

This mechanism avoids the need for an LMPM mechanism but is not without its challenges. For example, it does not completely close off opportunities for generation unit owners 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.

6.3 Solutions to System-wide Market Power

As discussed in Section 5 and in detail in Wolak (2000), 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 generation unit owner 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.

This logic argues in favour 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 et al. (2008b), 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. Consequently, to protect against periods when one or more generation unit owners have a strong incentive to exercise unilateral market power, the regulator should, at a minimum, monitor the forward contracting levels of the retailers they oversee as the primary mechanism to protect against the exercise of system-wide unilateral market power.

7. ACTIVE INVOLVEMENT OF FINAL DEMAND IN THE WHOLESALE MARKET

The active involvement of final consumers in the wholesale market can 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. An important market design feature that facilitates active participation by final demand is a multi-settlement market with a day-ahead forward market and a real-time market. This mechanism allows loads to purchase energy in the day-ahead market that they can subsequently sell in the real-time market. Without the ability to purchase demand in the day-ahead market that is not consumed in real-time, demand reduction programmes require the regulator to set an administrative baseline relative to which demand reductions are measured, which can significantly reduce the system-wide benefits of active demand-side participation. This issue is discussed in Bushnell, Hobbs and Wolak (2009).

7.1 Customers Can Respond to Dynamic Retail Prices

There are three necessary conditions for active involvement of final consumers. First, customers must have the necessary technology to record their consumption on time granularity/time scale similar to that of the wholesale market products. Second, they must receive actionable information that tells them when to alter their consumption.¹⁴ Third, they must pay according to a price that provides an economic incentive consistent with the actionable information to alter their consumption.

Mechanical meters that record a customer's consumption during a monthly or bi-monthly billing cycle as the difference of two consecutive meter readings cannot be used in a dynamic pricing plan because the retailer has no way of knowing when during the billing cycle the customer consumed electricity. Even if the customer is billed for wholesale electricity using a standardized load shape, a customer's monthly bill falls by the same amount regardless of when the customer reduces its consumption within the month. For example, if $w(h)$ is the load shape weight for hour h in the month ($\sum_{h=1}^H w(h) = 1$), $p(h)$ is the hourly wholesale price, and $Q(m)$ the household's monthly consumption, the wholesale electricity cost for the monthly billing cycle is $\sum_{h=1}^H p(h) w(h) Q(m)$, where H is the total number of hours in the billing cycle.

The above equation makes it clear that with mechanical meters read on a monthly or bi-monthly basis, dynamic pricing is impossible to implement because the customer's monthly bill falls by the same amount regardless of when consumption is reduced within the month. Therefore, a household faced with a higher average monthly price would reduce consumption when it is least costly for it to do so, not when the hourly price of electricity is the highest.

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-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 to estimate their price responsiveness. Charles River Associates (2004) analysed 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 the pricing plans, Wolak found large demand reductions in response to increases in hourly retail electricity prices across all income classes.

Although all these studies find statistically significant demand reductions in response to various forms of short-term price signals, none are able to assess the long-run impacts of requiring customers to manage short-term wholesale price risk. 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. Although most dynamic pricing experiments and programmes have relied on day-ahead price signals, recent work by Andersen et al. (2019) has shown that customers can respond to within-day price signals.

7.2 Managing Bill Risk with Dynamic Pricing

Politicians and policymakers often express the concern that 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 (described below) for facing a consumer with the hourly wholesale price for its 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.

If a state regulatory commission sets a fixed retail price or fixed pattern of retail prices throughout the day (time-of-use prices), it must still ensure that over the course of the month or year, the retailer's total revenues less its transmission, distribution and retailing costs, cover its total wholesale energy costs. If the regulator sets this fixed price too low relative to the current wholesale price, then either the retailer or the government must pay the difference.

Charging final consumers the same hourly default price as generation units owners provides strong incentive 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 increase a 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.

7.3 Fostering Investments in Automated Response Technologies

Perhaps the most important, but most often ignored, lesson from electricity restructuring processes in industrialized countries is the necessity of treating load and generation symmetrically. Symmetric treatment of load and generation means that unless a retail

consumer signs a forward contract with an electricity retailer, the default wholesale price the consumer pays is the hourly wholesale price. This is precisely the same risk that a generation unit owner faces unless it has signed a fixed-price forward contract with a load-serving entity or some other market participant. The default price a generation unit owner receives for any short-term energy sales is the hourly short-term price. Just as very few generation unit owners are willing to risk selling all of their output in the short-term market, consumers are likely to have similar preferences against too much reliance on the short-term market and would therefore be willing to sign long-term contracts for a large fraction of their expected hourly consumption during each hour of the month. Consistent with the above logic, a residential consumer might purchase a right to buy a fixed load shape for each day at a fixed price for the next 12 months. This consumer 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 that same price.¹⁵

This type of pricing arrangement would result in a significantly less volatile monthly electricity bill than if the consumer made all of its purchases at the hourly wholesale price. If all customers purchased according to this sort of pricing plan, then there would be no residual short-term price risk that the government needs to manage using tax revenues. All consumers manage the risk of high wholesale prices and supply shortfalls according to their preferences for taking on short-term price risk. Moreover, because all consumers have an incentive to reduce their consumption during high-priced periods, wholesale prices are likely to be less volatile. Symmetric treatment of load and generation does not mean that a consumer is prohibited from purchasing a fixed-price full requirement contract for all the electricity it might consume in a month, only that the consumer must pay the full cost of the retailer supplying this product.

Customers facing the risk of paying the real-time price for their electricity consumption is what creates the business case for investments in automated demand-response and storage technologies. If a customer can avoid consumption when the real-time price is high and consume more when this price is low through an investment in one of these devices, the customer is very likely to do so if the avoided wholesale energy purchase costs that this technology avoids more than covers the investment cost. A single fixed retail price or single fixed price schedule regardless of real-time system conditions cannot provide the revenue stream needed to finance investments in these technologies. Consequently, without exposing customers to the real-time price risk, investments in these technologies will not occur, unless some explicit support mechanism is introduced. Wolak and Hardman (2021) provide a comprehensive analysis of these issues and others related to the future of electricity retailing.

8. MARKET DESIGN LESSONS FOR DEVELOPING COUNTRIES

Electricity market design in developing countries and small control areas has the additional challenge of delivering significant economic benefits with a low implementation cost and limited regulatory burden. This section proposes such a market design using the insights from Sections 4 to 7.

8.1 A Simplified Market Design for Developing and Small Countries

The transition to formal market mechanisms in a number of developing and small countries has been slow. A number of 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 LMP. 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.

Because it is cost based rather than offer based, this market design also eliminates the need for a local or system-wide market power mitigation mechanism, which typically involves a significant regulatory burden. Because it uses the LMP market-clearing mechanism to set locational prices and generation unit dispatch levels, the resulting market outcomes optimize the use of the limited transmission network within and across regions. This cost-based market could be run as a multi-settlement market with day-ahead prices and schedules and real-time pricing and settlement or with a single real-time market and settlement.

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. 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 LMP 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.

To ensure long-term resource adequacy in this market, retailers would be required to purchase forward contracts for energy at various delivery horizons equal to pre-specified

fractions of their actual realized demand or face a financial penalty for under-procurement. For example, retailers could be required to purchase 100 per cent of their actual demand in a forward contract purchased before the short-term market operates for that day, 95 per cent of their demand one year in advance, 92 per cent two years in advance and 90 per cent three years in advance. The financial penalty for under-compliance should be sufficiently high to ensure compliance with the mandated level of contracting.

These contracting mandates for all retailers are necessary to establish a liquid forward market for energy in a region with a cost-based short-term market. As discussed in Section 5, without the risk of high short-term prices, retailers have a financial incentive to purchase all their energy from the short-term market, which could quickly lead to inadequate generation resources to serve demand. The contracting mandate on retailers described above ensures that adequate generation capacity will always be available to serve demand because there is a generation unit owner that has sold each MWh of energy the retailer's customers consumes in a fixed-price forward contract.

The role of the short-term cost-based market is simply to provide a transparent mechanism for buyers and sellers of these forward contracts to clear their imbalances, because a generation unit owner rarely produces the exact quantity sold in a fixed-price forward contract during any given hour of the day. Retailers also rarely consume exactly their hourly fixed-price forward contract quantity. The cost-based short-term market provides a transparent mechanism for differences between forward energy sales and actual production and forward energy purchases and actual consumption to be settled. For example, if the generation unit owner sold 400 MWh each hour of the day in a forward contract and its unit failed to operate during certain hours of the day, it needs a mechanism for purchasing replacement energy during these hours. The cost-based short-term market provides that mechanism. The seller knows it can purchase the replacement energy at the price set in the cost-based short-term market during those hours. It is likely that the seller would pay a high price for this replacement energy because units with higher costs than its own units would be required to operate. This provides an incentive for the unit owner to maximize the availability of its own units to avoid this set of circumstances.

It is important to emphasize that this short-term market is only for settling imbalances. That is the purpose of the requirement for retailers to procure 100 per cent of their realized demand as of the actual delivery date in a fixed-price forward contract. Because of these contracting mandates on retailers and large consumers, retailers are purchasing no net energy from the short-term market.

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 is a low-cost and low-regulatory burden approach to realizing more efficient investments in new generation capacity.

The counterparties to the fixed-price forward contracts sold to the electricity retailers have a strong financial incentive to find the least-cost mix of new generation capacity to supply the energy they have sold in these forward contracts. The cost-based short-term market tells them what they will be paid or pay for differences between the hourly production of their generation units and the amount of energy they have sold for that hour in a fixed-price forward contract. Electricity retailers can use this short-term market to clear

hourly imbalances between the amount of energy they withdraw from the transmission network and their fixed-price forward contract purchases.

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 LMPM mechanism. The LMP market is already in 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.

To address concerns that different consumers purchase their imbalances at different locational prices, all consumers in the region could be charged the quantity-weighted average of all LMPs at all load withdrawal nodes in the region. This would allow all fixed-price forward contracts to clear against this quantity-weighted average price, which enhances liquidity in the forward market for energy.

8.2 Improving Performance in Small Markets

Even in regions that have a vertically integrated monopoly industry structure but have or would like to have independent power producers sell energy to this monopoly, the cost-based market has the potential to reduce the cost of serving demand. The cost-based short-term market is used to dispatch all generation units in the region – those owned by the vertically integrated monopoly and those owned by independent power producers. The vertically integrated firm can still be subject to cost-of-service regulation for the retail price it is allowed to charge and the independent power producers can continue to be compensated according to their power purchase agreements. During the periods when an independent power producer's generation unit does not operate because it is not dispatched in the cost-based market, the unit owner can purchase the energy required under its power purchase agreement from the cost-based short-term market.

This cost-based market mechanism ensures that all of the resources in the control area are used in the least-cost manner to serve demand, rather than allowing certain generation units to be dispatched in preference to lower variable cost units because of pre-existing contractual requirements. This mechanism should also reduce the cost to the vertically integrated utility of purchasing long-term contracts for wholesale energy because all independent power producers that sell energy to the vertically integrated monopoly know that they have the option to purchase energy from the cost-based market and sell additional energy from their generation units when it is economic to do so.

9. INTEGRATING RENEWABLES

An increasingly important consideration in formulating any wholesale market mechanism is the extent to which it can accommodate a significant amount of intermittent renewable generation in both the transmission and distribution networks. A growing number of jurisdictions in the developing world have significant renewable energy targets

in place. A number of small control areas also have significant renewable energy goals.¹⁶ Consequently, any market design adopted by these regions should support cost-effective integration of renewable resources. As we discuss below, the cost-based LMP market design is well-suited for integrating any amount of intermittent renewables into a national or regional electricity market. However, as we also discuss, the integration of an increasing share of renewables is likely to require incorporating additional constraints into the region's LMP market and the introduction of additional products to deal with the increasing share of intermittent renewable resources.¹⁷

9.1 Cost-based Market and Renewables Integration

The strength of a cost-based LMP market design 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 sold in the forward market will not depend on how the resource is used to produce energy. The existence of a cost-based LMP market will allow renewable resource owners to sell fixed-price and fixed-quantity contracts for energy because they have the short-term market to purchase energy from when their renewable units do not produce sufficient energy to meet their forward market obligation and can sell excess energy beyond their forward market obligation when their units produce more than the forward contract quantity.

The renewable resource owner can factor in how these imbalances will be settled in making offers to supply fixed-price and fixed-quantity long-term contracts 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 short-falls. Transitioning forward contracts for renewable energy to require the seller to manage the quantity risk associated with the energy it sells is an important step in the process of increasing the amount intermittent renewable energy is produced in a region.

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 actually operated. This logic implies that as the share of intermittent renewable resources increases, the LMP market can be easily adapted to deal with the new reliability challenges this creates.

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. As discussed in Section 4, this multi-settlement market also rewards dispatchable resources for their ability to supply more or less energy, depending on the instructions of the market operator.

Markets with significant intermittent renewables such as California and Australia have seen significant investments in battery storage, mostly driven by regulatory mandates. Battery storage systems can be connected to either the high-voltage transmission system or the local distribution network, but decisions about where to install battery systems are currently driven by regulatory mandates or inefficient distribution network pricing, as discussed in Wolak (2018).¹⁸ An energy-contracting approach to long-term resource adequacy and the increase in short-term price volatility that it allows should bring about more market-based investments in battery storage.

10. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The experience of the past 25 years identifies the following fundamental ingredients for a successful market design. First, a multi-settlement LMP wholesale market is most likely to achieve the best possible match between how the transmission network operates and how the wholesale market determines prices and dispatch levels. Second, a liquid forward market for energy appears to be the most efficient way to ensure both short-term and long-term resource adequacy, although capacity payment mechanisms continue to be employed in many regions of the world. Capacity mechanisms have also begun to emphasize the development of a liquid forward market for capacity, which would make it easier to transition to a forward market for energy. Third, fixed-price long-term contracts are an effective mechanism for limiting the incentive of generation unit owners to exercise system-wide unilateral market power in the short-term market. All US markets and most international markets have LMPM mechanisms, although the details of these mechanisms differ across markets. Fourth, there is increasing recognition of the need for active involvement of final demand in the wholesale market, particularly in regions that have deployed interval meters. A multi-settlement LMP market provides the ideal wholesale market platform for this to occur. The need for active involvement of final demand is even greater in regions with significant intermittent renewable energy.

Nonetheless, there are several issues that still deserve additional research. The two major drivers of future research on electricity market design are: (1) outstanding issues in markets with conventional generation resources; and (2) new issues created by the increasing penetration of renewables both at the transmission and distribution level. On the first topic, the chapter identifies challenges for future research. A regulatory mechanism for developing an active forward market for energy to ensure long-term resource adequacy is perhaps the most important topic for future research. There are many different regulatory mechanisms for developing an active forward market to ensure long-term resource adequacy that exist around the world. A comparative quantitative analysis of the performance of these mechanisms could be extremely informative. There are also many different approaches to LMPM that exist around the world. A comparative quantitative study of the performance of these mechanisms could help all regions improve their mechanisms. An understanding of the advantages and disadvantages of cost-based versus

offer-based markets as a function of the initial conditions in the country and the electricity industry could provide important guidance to developing countries and small regions that are considering reforming their electricity industries. A number of economic experiments with information provision and dynamic pricing programmes could inform how to achieve the greatest amount of customer acceptance and participation in active load management.

On the second topic, the engineering studies of how demand for ancillary services is likely to scale with different scenarios for renewables deployment in the transmission and distribution grids could be extremely helpful for developing countries wanting to expand the contribution of renewable resources to their electricity mix. Another important area for economic and engineering studies is on the design of new wholesale market products to reward fast-ramping and fast-starting dispatchable generation resources. It is also likely that new paradigms for transmission and distribution system operation will need to be developed to deal with increasing intermittency at the customer level because of investments in distributed generation and at the transmission level because of investments in grid-scale renewables.

NOTES

1. In this chapter, I use the term hourly meter to refer to any meter that can record consumption in a fixed time interval such as every 15 minutes, half-hour or hour.
2. Vickers and Yarrow (1988) provide a comprehensive treatment of this process. In Chapter 6 of this handbook, Newbery discusses more recent experience with the British electricity supply industry.
3. In Chapter 9 of this handbook, Simshauser describes recent experience with the Australian electricity market.
4. Different from government-owned utility oversight, US state public utilities commissions are explicitly tasked with protecting consumers from excessive prices and only allowing the utilities they regulate to recover prudently incurred costs.
5. Wolak (1999) provides a description of the wholesale market design and regulatory oversight processes in England and Wales, Norway and Sweden, Australia, and New Zealand as of 1 January 1998.
6. See Figure 90 in European Network of Transmission System Operators – Electricity (ENTSO-E) (2018).
7. Ancillary services are composed of different operating reserves required by the system operator to maintain real-time supply and demand balance. For the California ISO market, these are composed of regulation up and regulation down automatic generation control (AGC) reserves, spinning reserves and non-spinning reserves.
8. Many regions have eliminated or are eliminating this technological barrier to allow price to manage the real-time supply and demand balance by installing interval meters for all customers and offering real-time meter reading as a regulated service.
9. A copy of FERC Order 831 is available at <https://www.ferc.gov/sites/default/files/2020-06/RM16-5-000.pdf>, accessed 26 September 2020.
10. Thermal generation units convert heat energy into electricity. These include coal-fired, natural gas-fired, oil-fired, nuclear and geothermal power plants. The availability factor of a generation unit is equal to the fraction of hours of the year such unit is available to produce electricity.
11. A supplier is said to be pivotal in a market with an inelastic demand if some of its supply is needed to meet this demand regardless of the offers of other market participants.
12. In Section 2 of Chapter 3 of this handbook, Joskow and Léautaud discuss this process in detail.
13. A scarcity price refund requires the seller of firm capacity to pay $\max(0, P(\text{market}) - P(\text{scarcity})) \times Q_{\text{Firm}}$, where $P(\text{market})$ is the market price, $P(\text{scarcity})$ is an administratively set scarcity price, and Q_{Firm} is the firm energy sold by the supplier. The function $g = \max(0, y)$ means that $g = y$ if $y > 0$ and zero otherwise.
14. McRae and Meeks (2016) present the results of a field experiment in Central Asia that demonstrates the importance of actionable information for facilitating active demand-side participation. Kahn and Wolak (2013) find that once customers understand non-linear pricing, they subsequently make energy-consuming

- decisions consistent with responding to the marginal price. Wolak (2015b) presents evidence consistent with real-time consumption feedback producing energy conservation efforts by households in Singapore.
15. This pricing plan requires the customer to have an interval meter. Wolak (2013) draws analogy between this pricing plan for electricity and how cell phone 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 the minutes purchased in advance at some penalty price. In the case of electricity, the price for unused KWh and additional KWh during a given hour is the real-time wholesale price during that hour.
 16. The state of Hawaii has a 100 per cent renewable energy target by 2045 and already has the largest share of distributed solar capacity as a percentage of peak demand of any control area in the world: more than 400 MW out of a peak demand of 1200 MW.
 17. Interested readers may also look at Chapter 15 by Richard Green for a general introduction to the economics of wholesale electricity markets with significant intermittent renewables.
 18. Relations between the central power system and off-grid or mini grid systems are addressed by Pérez-Arriaga et al. in Chapter 20 and by Foster et al. in Chapter 22 of this handbook.

REFERENCES

- Andersen, L. M., L. G. Hansen, C. L. Jensen and F. A. Wolak (2019), 'Can incentives to increase electricity use reduce the cost of integrating renewable resources?', accessed 26 September 2020 at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/Into_versus_Away_January_2019.pdf.
- Bohn, R. E., M. C. Caramanis and F. C. Scheppe (1984), 'Optimal pricing in electrical networks over space and time', *RAND Journal of Economics*, **15** (5), 360–76.
- Borenstein, S., J. B. Bushnell and F. A. Wolak (2002), 'Measuring market inefficiencies in California's restructured wholesale electricity market', *American Economic Review*, **92** (5), 1367–405.
- Bushnell, J. B., B. F. Hobbs and F. A. Wolak (2008a), 'Final opinion on "The DEC Bidding Activity Rule under MRTU"', accessed 26 September 2020 at <https://www.aiso.com/Documents/MSCFinalOpiniononDECbiddingActivityRuleunderMRTU.pdf>.
- Bushnell, J. B., B. F. Hobbs and F. A. Wolak (2009), 'When it comes to demand response, is FERC its own worst enemy?', *The Electricity Journal*, **22** (8), 9–18.
- Bushnell, J. B., E. T. Mansur and C. Saravia (2008b), 'Vertical arrangements, market structure, and competition: an analysis of restructured US electricity markets', *American Economic Review*, **98** (1), 237–66.
- Charles River Associates (2004), *Statewide Pricing Pilot Summer 2003 Impact Analysis*, Oakland, CA: Charles River Associates.
- European Network of Transmission System Operators – Electricity (ENTSO-E) (2018), *Bidding Zone Configuration Technical Report 2018*, October, accessed 26 September 2020 at https://docstore.entsoe.eu/Documents/Events/2018/BZ_report/20181015_BZ_TR_FINAL.pdf.
- Fabrizio, K. R., N. L. Rose and C. D. Wolfram (2007), 'Do markets reduce costs? Assessing the impact of regulatory restructuring on US electric generation efficiency', *American Economic Review*, **97** (4), 1250–77.
- Galetovic, A., C. M. Muñoz and F. A. Wolak (2015), 'Capacity payments in a cost-based wholesale electricity market: the case of Chile', *The Electricity Journal*, **28** (10), 80–96.
- Graf, C., F. Quaglia and F. A. Wolak (2020), 'Simplified electricity market models with significant intermittent renewable capacity: evidence from Italy', accessed 26 September 2020 at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/GrafQuagliaWolak_SimplifiedElectricityMarketModelsRenewables.pdf.
- Joskow, P. L. (1997), 'Restructuring, competition and regulatory reform in the U.S. electricity sector', *The Journal of Economic Perspectives*, **11** (3), 119–38.
- Kahn, M. E. and F. A. Wolak (2013), 'Using information to improve the effectiveness of non-linear pricing: evidence from a field experiment', accessed 26 September 2020 at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/files/kahn_wolak_July_2_2013.pdf.
- Mansur, E. T. and M. W. White (2012), 'Market organization and efficiency in electricity markets', accessed 26 September 2020 at https://mansur.host.dartmouth.edu/papers/mansur_white_pjmaep.pdf.
- McRae, S. D. and R. Meeks (2016), 'Price perceptions and electricity demand with non-linear prices', accessed 26 September 2020 at <https://www.sdmcrae.com/publication/price-perception-and-electricity-demand/>.
- McRae, S. D. and F. A. Wolak (2012), 'How do firms exercise unilateral market power? Evidence from a bid-based wholesale electricity market', in J.-M. Glachant and E. Brousseau (eds), *Manufacturing Markets: Legal, Political and Economic Dynamics*, Cambridge, UK: Cambridge University Press.

- McRae, S. D. and F. A. Wolak (2019), 'Market power and incentive-based capacity payment mechanisms', accessed 26 September 2020 at web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/2019-03-mcrae-wolak-capacity.pdf.
- Oren, S. S. (2001), 'Design of ancillary service markets', in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, <https://doi.org/10.1109/HICSS.2001.926283>.
- Patrick, R. H. and F. A. Wolak (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 M. A. Crew (ed.), *Regulation Under Increasing Competition*, Boston, MA: Kluwer Academic Publishers, pp. 155–82.
- Tangerås, T. P. and F. A. Wolak (2018), 'The competitive effects of linking electricity markets across space and time', accessed 26 September 2020 at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/The_Competitive_Effects_of_Linking_Electricity_Markets_Across_Space_April_2021.pdf.
- Vickers, J. and G. K. Yarrow (1988), *Privatization: An Economic Analysis*, Cambridge, MA: MIT Press.
- Wolak, F. A. (1999), 'Market design and price behavior in restructured electricity markets: an international comparison', in T. Ito and A. Krueger (eds), *Competition Policy in the Asia Pacific Region*, Chicago, IL: The University of Chicago Press, pp. 79–134.
- Wolak, F. A. (2000), 'An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market', *International Economic Journal*, **14** (2), 1–40.
- Wolak, F. A. (2003a), 'Measuring unilateral market power in wholesale electricity markets: the California market, 1998–2000', *American Economic Review*, **93** (2), 425–30.
- Wolak, F. A. (2003b), 'Diagnosing the California electricity crisis', *The Electricity Journal*, **16** (7), 11–37.
- Wolak, F. A. (2006), 'Residential customer response to real-time pricing: the Anaheim critical-peak pricing experiment', *CSEM WP 151*, Center for the Study of Energy Markets, accessed 26 September 2020 at <https://haas.berkeley.edu/wp-content/uploads/csemwp151.pdf>.
- Wolak, F. 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, F. A. (2011a), 'Do residential customers respond to hourly prices? Evidence from a dynamic pricing experiment', *American Economic Review*, **101** (3), 83–7.
- Wolak, F. A. (2011b), 'Measuring the benefits of greater spatial granularity in short-term pricing in wholesale electricity markets', *American Economic Review*, **101** (3), 247–52.
- Wolak, F. A. (2013), 'Economic and political constraints on the demand-side of electricity industry restructuring processes', *Review of Economics and Institutions*, **4** (1), 1–42.
- Wolak, F. A. (2014), 'Regulating competition in wholesale electricity supply', in N. L. Rose (ed.), *Economic Regulation and Its Reform: What Have We Learned?*, Chicago, IL: University of Chicago Press, pp. 195–289.
- Wolak, F. A. (2015a), 'Measuring the competitiveness benefits of a transmission investment policy: the case of the Alberta electricity market', *Energy Policy*, **85**, 426–44.
- Wolak, F. A. (2015b), 'Do customers respond to real-time usage feedback? Evidence from Singapore', accessed 26 September 2020 at http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/Singapore%20residential_nov_2015.pdf.
- Wolak, F. A. (2018), 'The evidence from California on the economic impact of inefficient distribution network pricing', *NBER Working Paper 25087*, National Bureau of Economic Research.
- Wolak, F. A. (2020), 'Transmission planning and operation in the wholesale market regime', *Transmission Network Investment in Liberalized Power Markets*, pp. 101–33. Springer, Cham.
- Wolak, F. A. and Hardman, I. (2021) *The Future of Electricity Retailing and How We Get There, Lecture Notes in Energy Volume 41*, Springer International Publishing.
- Wolak, F. A. and R. H. Patrick (2001), 'The impact of market rules and market structure on the price determination process in the England and Wales electricity market', *NBER Working Paper 8248*, National Bureau of Economic Research.
- Wolfram, C. D. (1999), 'Measuring duopoly power in the British electricity spot market', *American Economic Review*, **89** (4), 805–26.