Electricity Market Design and Renewables Integration in Developing Countries

by

Goran Strbac
Department of Energy Systems
Imperial College London
London, UK  SW7 2AZ
g.strbac@imperial.ac.uk

Frank A. Wolak
Department of Economics
Stanford University
Stanford, CA  94305-6072
wolak@zia.stanford.edu

Draft:  February 21, 2017
Abstract

This paper identifies the key features of successful electricity market designs that are particularly relevant to the experience of low-income countries. Important features include: (1) the match between the short-term market used to dispatch generation units and the physical operation of electricity network, (2) effective regulatory and market mechanisms to ensure long-term generation resource adequacy, (3) appropriate mechanisms to mitigate local market power, and (4) mechanisms to allow the active involvement of final demand in short-term market. The paper provides a recommended baseline market design that reflects the experience of the past 25 years with electricity restructuring processes. It then suggests a simplified version of this market design ideally suited to the proposed East and Western Sub-Saharan Africa regional wholesale market that is likely to realize a substantial amount of the economic benefits from a forming a regional market with minimal implementation cost and regulatory burden. Recommendations are also provided for modifying the Southern African Power Pool increase the economic benefits realized from its formation. How this market design supports the cost effective integration of renewables is discussed and future enhancements are proposed that support the integration of a greater share of intermittent renewables. The paper closes with proposes directions for future research in the area of electricity market design in developing countries.
1. Introduction

Electricity has been historically provided by vertically-integrated geographic monopolies where each of the four segments of electricity supply—generation, transmission, distribution and retailing—are contained in the same company. In most developing countries, these monopoly electricity suppliers were state-owned enterprises subject to government oversight, usually by the Ministry of Mines and Energy, with a generally poor record in providing a reliable supply of electricity at a reasonable price.

These entities typically experience high levels of technical and non-technical transmission and distribution network losses. Specifically, the amount of electricity billed for by the state-owned company is a significantly smaller percentage of the total amount of electricity generated, and the amount electricity ultimately paid for is an even smaller fraction of this amount. This is due in part to an inadequate and poorly maintained electricity delivery and metering infrastructure. But the vast majority of the difference between the amount of electricity generated and the amount that is paid for can be explained by theft and non-payment of electricity bills.

Because of high levels of technical and non-technical losses in developing countries, many of these state-owned companies have found it difficult to raise the capital needed for investments in new generation, transmission, and distribution assets necessary to meet a rapidly growing demand for electricity. This has led to long delays in the electrification, particularly in rural areas.

These shortcomings of the state-owned monopoly market structure have led many developing countries to consider re-structuring their electricity sector by isolating the monopoly segments of the industry to the transmission and distribution networks and allowing competition in electricity generation and retailing. Because privately-owned companies have a strong incentive to ensure that consumers receive and pay for as much of the electricity that is produced as is technically possible, this market structure has the potential to achieve the goals of a reliable supply of electricity to all consumers at a reasonable price. Moreover, privately-owned companies have an incentive to expand service to customers willing to pay at least the cost of supplying them.

Although these improvements in industry performance are not guaranteed by introducing wholesale and retail competition into a country’s electricity supply industry, international experience in the industrialized and developing world 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
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, and at least during the initial stages of the restructured industry, there is a small number of wholesale and retail market participants.

These two facts imply that wholesale and retail market rules can significantly impact the behaviour of these market participants, often to the detriment of electricity consumers. Consequently, designing the wholesale and retail market rules that govern an electricity market 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.

A market designer must therefore recognize that any wholesale or retail market rule will be exploited by all market participants to enhance their ability to pursue objectives such as maximizing profits from selling wholesale or retail electricity or minimizing retail electricity procurement costs. The most successful re-structured markets are those with market rules that account for the self-interested behaviour of all market participants. Much of this paper 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 electricity market for developing countries.

The experience of the past twenty-five years identifies the following necessary conditions for a successful market design. Section 2 first describes why electricity requires an explicit market design process. Section 3 demonstrates that a multi-settlement locational marginal pricing 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. Section 4, 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 capacity payment
Section 5 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, which exist in all US markets and most international markets, although the details of these mechanisms differ across markets. Section 6 emphasizes the need for active involvement of final demand in the wholesale market, particularly in regions that have deployed interval meters. A multi-settlement locational marginal pricing market is shown to provide the ideal platform for active participation of final demand.

Section 7 considers the prospects for a successful market design in Southern Africa and Sub-Saharan Africa. Particularly in East and West Africa, electricity industry restructuring has been extremely slow to deliver tangible economic benefits. The analysis of Sections 3-6 is used to formulate a market design these regions that is likely to capture a significant fraction of the potential benefits from reform with a minimal financial cost and regulatory burden. This market design avoids many of the more costly features of formal wholesale electricity markets in the industrialized world that also involve a significant regulatory burden. Although Southern African is further along in the reform process, Sections 3-6 also provide recommended changes to existing market mechanisms for this region so that they yield greater economic benefits with minimal cost and regulatory burden.

Section 8 considers the question of integrating a significant amount of intermittent renewable resources into a formal wholesale market. A number of developing countries have policies to increase the amount renewable energy they consume. Consequently, an important issue is how to ensure that the market design a region adopts does not unnecessarily increase the cost of achieving its renewable energy goals. This section first introduces the three commonly employed policies used to foster renewable energy deployment: (1) feed-in tariffs, (2) renewables portfolio standards, and (3) fixed-price, fixed quantity forward contracts for energy. It then describes how our recommended market designs support the cost effective deployment of intermittent renewable resources.

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 occur. It is very costly to store electricity. Constructing significant storage facilities typically requires substantial up-front costs and basic physics implies more than 1 MWh of energy must be produced and consumed to store 1 MWh of energy. Production of electricity is subject to extreme capacity constraints in the sense that it is impossible to get more than a pre-specified amount of energy from a generation unit in an hour. This limits the size of 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 behavior.

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 nonlinear price schedule for each kilowatt-hour (KWh) they consume during the month, 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 a supplier 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 is involves consumers and producers deciding which products and locations to serve. Some locations and products favor producers and others favor consumers. But 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 this 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 and Peet's 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 customers have little 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 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 decisions of producers and consumers of electricity.

Although the market rules that are best suited to meet the needs of electricity producers and consumers for a specific region are continually changing (as a result of technological change and evolving consumer preferences for electricity), there are a number of common design challenges to all electricity markets, particularly in the developing country context. The next four sections discuss specific market design challenges and the lessons learned from both the industrialized and developing country experience in meeting these challenges.

Where the empirical evidence supports a firm recommendation on how to address a market design challenge, one is provided. When the empirical evidence does not support a clear recommendation, the trade-offs between the different choices are discussed.

3. 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. In the early stages of electricity supply industry reform processes, many regions attempted to operate wholesale markets that used simplified versions of the transmission network. These markets often assumed infinite transmission capacity between locations in the transmission grid or only recognized transmission constraints across large
geographic regions. The Southern African Power Pool (SAPP) is an example of such a market. Only transmission constraints between regions—primarily the member countries—are modelled in operating the market. 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 transmission network and other operating constraints would need to be respected.

Many early wholesale electricity 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 the 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 operating the unit and raise the total cost of electricity supplied to final consumers...

A similar set of circumstances can arise for constrained off generation units. Constrained-off suppliers are usually paid the difference between the market-clearing price and their offer price for not supplying electricity that it 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 in order to receive the highest possible payment for being constrained-off and raise the total cost of electricity supplied to final consumers. Bushnell, Hobbs and Wolak (2008) discuss this problem and the market efficiency consequences in the context of the California zonal market. However, it is not unique to industrialized country markets. Wolak (2009) discusses these same issues in the context of the Colombian single-price market.
3.1. Locational Marginal Pricing (LMP)

Almost any difference between the market model used to set dispatch levels and market prices and the actual operation of the generation units needed to serve demand creates an opportunity for market participants to take actions that raise their profits at the expense of overall market efficiency. Multi-settlement wholesale electricity markets that use locational marginal pricing (LMP), also referred to as nodal pricing, 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 prices in the wholesale market.

Generation unit owners and load serving entities submit their location-specific willingness-to-supply energy and willingness-to-purchase energy to the wholesale market operator, but locational prices and dispatch levels for generation units at each location in the transmission network are determined by minimizing the as-offered costs of meeting demand at all locations in the transmission network subject to all network operating constraints. No generation unit will be accepted to supply energy if doing so would violate a transmission or other operating constraint.

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

The nodal price at each location is the increase in the minimized value of the 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. Bohn, Caramanis, and Schweppe (1984) provide an accessible discussion of the properties of this 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 locational prices and dispatch levels. For example, suppose that reliability studies have shown that a minimum amount of energy must be produced by a group 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 LMPs. This property of the LMP markets is particularly relevant to
the cost-effective integration a significant amount of intermittent renewable generation capacity. Additional reliability constraints may need to be formulated and incorporated into LMP market to account for the fact that this energy supply can quickly disappear and re-appear.

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

These day-ahead commitments do not require a generation unit to supply the amount sold in the day-ahead market or a load to consume the amount purchased in the day-ahead market. The only requirement is that any shortfall in a day-ahead commitment to supply energy 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 control area 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 $3,000 in revenues from this sale. However, if the generation unit owner fails to inject 50 MWh of energy into grid during that hour of the following day, it must purchase the energy it fails to inject at the real-time price at that location. Suppose that the real-time price at that location is $70/MWh and generator only injects 40/MWh of energy during the hour in question. In this case, the unit owner must purchase the 10 MWh shortfall at $70/MWh. Consequently, the net revenues the generation unit owner earns from selling 50 MWh in the day-ahead market and only injecting 40/MWh is $2,300, the $3,000 of revenues earned in the day-ahead market less the $700 paid for the 10 MWh real-time deviation from the unit’s day-ahead schedule.

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

All US markets started the electricity re-structuring process with significantly less extensive transmission networks relative to their counterparts in other industrialized countries, and initially all of them attempted to operate a wholesale market that did not fully account for the configuration of the transmission network and all relevant operating constraints in the market mechanism. All of them eventually switched to the LMP market design.

By this same logic, a multi-settlement nodal-pricing market is well-suited to developing 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, currently a popular market design outside of the US. On April 1, 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 BTUs, the total hourly variable cost of production from fossil fuel 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 percent lower, the total hourly variable cost of production for fossil fuel units is 2.1 percent lower, and the total number of hourly starts is 0.17 higher after the implementation of nodal pricing. This 2.1 percent cost reduction implies a roughly $105 million reduction in the total annual variable costs of producing fossil fuel energy in California is 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 model transmission and other operating constraint are likely to be greater the larger is the share of intermittent renewable resources. Consequently, any region with significant renewable energy goals would benefit from implementing a multi-settlement LMP market.

A multi-settlement LMP market also values of the dispatchability of generation units even though it pays all resources at the same location in the grid the same price in the day-ahead and real-time markets. Suppose that a wind unit sells 50 MWh and a thermal resource sells 40 MWh in the day-ahead market at $30/MWh. If in real-time, not as much wind energy is produced, the dispatchable thermal unit must make up the difference. Suppose that the wind unit produces only 30 MWh, so that the thermal unit must produce an additional 20 MWh. Because of this wind generation shortfall, the real-time price is now $60/MWh. Under this scenario, the wind unit is paid an average price of $10/MWh = \( \frac{50 \text{ MWh} \times 30 \text{$/MWh} - 20 \text{ MWh} \times 60 \text{$/MWh}}{30 \text{ MWh}} \) for the 30 MWh it produces, whereas the dispatchable thermal unit is paid an average price of $40/MWh = \( \frac{40 \text{ MWh} \times 30 \text{$/MWh} + 20 \text{ MWh} \times 60 \text{$/MWh}}{60 \text{ MWh}} \) for the 60 MWh it produces. Similar logic applies to the case that the wind resource produces more than expected and the thermal resource reduces its output because the real-time price is lower than the day-ahead price.
because of the unexpectedly large amount of wind energy produced.

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

Many regions with LMP pricing have overcome this potential problem by charging all customers in a given state or utility service territory a weighted average of the LMPs in the region. In the above example, this implies charging the urban and rural customers the weighted average of the LMPs 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 LMP at their location, but all loads pay a geographically aggregated hourly-price. For example, in Singapore all generation units are paid the LMP at their location, but all loads are charged the Uniform Singapore Electricity Price (USEP), which is the quantity-weighted average of the half-hourly LMPs for all load-withdrawal points in Singapore. This approach to pricing captures that 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.

On this market design feature, the accumulated experience supports a multi-settlement LMP market design as the recommended wholesale market mechanism. Attempts to run wholesale markets that use simplified models of the grid, typically create significant market performance problems, particularly in regions with limited transmission capacity. As noted above, equity concerns with charging different prices to different customers can be addressed through having loads pay geographically averaged LMPs.

4. Mechanisms to Ensure Long-Term Resource Adequacy

A major challenge to all electricity markets is how to ensure there is sufficient generation capacity to meet demand. Achieving this goal can be even more difficult for developing countries because of the rapid load growth they often experience and the difficulty of obtaining sufficient
revenues from customers to recover the full cost of the electricity produced. The typical solution to a supply shortfall in other markets is to allow short-term prices to rise to the level necessary to make the available supply equal demand. Even in many industrialized countries, this solution is not politically feasible for electricity given its essential role in the modern life. This fact has necessitated the design of regulatory mechanisms to ensure long-term resource adequacy.

Two general approaches have been developed. The first is based on fixed-price and fixed-quantity long-term contracts for energy signed between electricity suppliers and load-serving entities at various horizons to delivery. The second approach is a regulator-mandated capacity payment 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. Suppliers 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-year payment across the existing capacity payment mechanisms.

4.1. 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 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 the 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 price cap on the short-term market is 12,500 Australia dollars ($AU) per megawatt-hour (MWh), yet annual average wholesale prices range from $AU 30/MWh to $AU 40/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. Purchasing these 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 purchase. 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 purchased through 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 electricity suppliers to match the mix of generation capacity to the demand that has purchased 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 demands far enough in advance of delivery to allow new entrants to compete with existing suppliers 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. For example, all US markets except the Electricity Reliability Council of Texas (ERCOT) have caps on the offer price that suppliers can submit at $1,000/MWh. ERCOT’s offer price cap is currently $9,000/MWh, which very close to Australia’s offer cap in US dollars.

Many wholesale electricity markets outside of the US, particularly those in the developing countries, have offer caps far below $1,000/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, 90 percent of forecast demand one year in advance,
85 percent 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 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 is going to need a forward contract to meet its regulatory requirements for coverage of its final demand. Only in the very unlikely case that the aggregate amount of forward contracts purchased is greater than the realized final demand for the system, 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 suppliers to exercise unilateral market power in the short-term market. To understand this logic, let \( PC \) equal the fixed price at which the supplier agrees to sell energy to an electricity retailer in a forward contract and \( QC \) equal to the quantity of energy sold. This contract is negotiated in advance of the date that the generation unit owner will supply the energy, so the value of \( PC \) and \( QC \) are predetermined from the perspective of the supplier’s behaviour in a short-term wholesale market.

Wolak (2000) demonstrates that the quantity of fixed-price forward contract obligations held by the supplier 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:

\[
\pi(PS) = (PC - C)QC + (QS - QC)(PS - C)
\]  

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 (1) is the variable profit from the forward contract sales and the second term is the additional profit or loss from selling more or less energy in the short-term market than the supplier’s forward contract quantity. Because the forward contract price and quantity are
negotiated in advance of the delivery date, the first term, \((PC - C)QC\), is a fixed profit stream to the supplier 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 supplier to raise prices in the short-term market.

For example, if the supplier 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 (1) will be negative. Consequently, only in the case that the supplier is confident it will produce more than its forward contract quantity in the short-term market does it have an incentive to withhold output in order to raise short-term prices.

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

### 4.2. The Capacity Payment Approach to Long-Term Resource Adequacy

Particularly in the US, capacity payment 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 electricity.

Capacity payments typically involve a dollar per kilowatt year \(($/kW\cdot\text{year})\) payment to individual generation units based on some measure of the amount of their capacity that is available to produce electricity at peak demand times during the year, what is often referred to as the unit’s “firm capacity.” For example, a base load coal-fired unit would have a firm capacity value very close to its nameplate capacity. Usually, the firm capacity of a thermal unit is equal to the unit’s
capacity in MWs times its availability factor. A wind generation unit would have a capacity value significantly below its nameplate capacity, but likely more than the amount of energy it is able to produce during peak demand periods.

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 owner is typically based on the amount of energy that 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 supplier can provide under these conditions, so there is a significant degree of arbitrariness in setting hydroelectric unit’s firm capacity value. Second, because every hydroelectric supplier would like a larger capacity value for their generation unit, in order 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 payment mechanisms that failed to ensure an adequate supply of energy and rationing conditions have been declared. Virtually all of the restructured markets in Latin America that have capacity payment mechanisms--Brazil, Chile, and Colombia have experienced supply shortfalls that have required rationing.

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

There have been attempts to use market mechanisms to set the value of the $/kW-year payment to the generation units needed to meet the total demand for capacity. However, these 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.
In the eastern US capacity markets, there have been numerous instances of the exercise of the enormous unilateral market power. During the off-peak months of the year when no single supplier is pivotal in the capacity market, the price of paid for capacity was very close to zero, which is the marginal cost of a supplier providing an additional MW of available capacity from existing generation capacity.1 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 annual $/kW fixed cost of a peaking generation unit. This is 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 its 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.

This logic for setting this value of $/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. Both of these assumptions are clearly false. An increasing number of jurisdictions around the world are installing interval meters that allow dynamic pricing plans to be implemented. As noted in McRae and Wolak (2012), exercising all available unilateral market power is equivalent to a privately-owned firm serving its fiduciary duty to its shareholders or a publicly-owned firm serving its fiduciary responsibility to its ratepayers. For these reasons, it seems highly unlikely that any market power mitigation mechanism could prevent the exercise of all unilateral market power.

Capacity payment mechanisms make it extremely difficult for consumers to benefit from electricity industry restructuring relative to market without a capacity payment mechanism and

---

1 A supplier is said to be pivotal in a market with an inelastic demand if some of their supply is needed to meet this demand regardless of the offers of other market participants.
active demand-side participation in the wholesale market. Recall that the capacity payment is made to either all generation units in the system or all generation units needed to meet the ISO’s demand for capacity. On top of this, all suppliers typically receive the same market-clearing price for capacity. Thus, to the extent that suppliers are able to exercise unilateral market power in the short-term energy market, they can raise energy prices significantly above the variable cost of the highest cost unit operating within the hour for all hours of the year, on top of receiving a capacity payment 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 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 may 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 suppliers 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 suppliers know that higher offer prices will result in less of their generation capacity being called upon to produce because the offers of final consumers to reduce their demand are accepted instead. Without an active demand-side of the wholesale market suppliers know that they can submit offers that are farther above their variable cost of supplying electricity and not have these offers rejected. 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 suppliers greater opportunities to exercise unilateral market power in the short-term market, which clearly reduces the likelihood that consumers will realize net benefits from electricity restructuring.

Another argument given for capacity payments is that they reduce the likelihood of long-term capacity inadequacy problems because of the promise of a capacity payment provides incentives for new generation units to enter the market. However, until very recently capacity payments in most markets around the world were only promised for at most a single year and only paid to existing generation units. Both these features substantially dulled the incentive for new generation units to enter the market, because a generation unit that entered the market had no
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. This has led the eastern US wholesale markets to develop of 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. This is 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 do want their future energy needs met.

Capacity markets are also poorly suited for regions with a significant share of 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 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 of 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 efficient 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 done far enough in advance of delivery and for a long enough period of time to support new investment. The similarity between this solution and the long-term energy contracting solution argues in favour the long-term energy solution to the long-term resource adequacy problem. Galetovic, Munoz 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-only market.

4.3. 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 reasonable price is purchasing sufficient energy or capacity far enough in advance of
delivery by electricity retailers and large customers for new entrants to compete with existing suppliers 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. For example, a contract negotiated one day in advance limits the sources of supply to existing generation unit owners able to produce energy the following day. Even a year in advance limits the sources that can compete with existing generation unit owners, because it takes longer than eighteen months to site and build a substantial new generation unit in virtually wholesale electricity markets. To obtain the most competitive prices, at a minimum, the vast majority of the fixed-price forward contracts should be negotiated far enough in advance of delivery to allow new entrants to compete with existing suppliers.

This logic argues for regulatory intervention in the long-term resource adequacy process to develop a forward market for energy or capacity for delivery 2 to 3 years into the future. If a liquid forward market for energy exists at this time horizon to delivery and there is adequate demand for energy at this a horizon to delivery, a restructured market will achieve long-term resource adequacy. A liquid forward market at the 2 to 3 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 for their wholesale energy purchases at delivery horizons less than 2 years into the future. By purchasing a hedge against the spot price risk at the locations in the network where the retailer or large consumer withdraws energy, the buyer can rely on the financial incentives that the seller of the contracts faces to procure this energy at the lowest possible cost.

Focusing the long-term resource adequacy process on the construction of generation units misses the important point that there is an increasing number of ways for markets to achieve long-term resource adequacy besides building generation units. For example, by the appropriate choice of the mix of generation units, the same pattern of hourly demands throughout the year can be met with less total generation capacity that can also cost electricity consumers less. Distributed generation and storage investments, active demand-side participation in the wholesale market can also allow the same number of customers to be served with less grid-connected generation capacity.
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 supplier and a retailer agree to a fixed price at a location in the transmission network for a fixed quantity of energy. Cap contracts are also very effective instruments for guarding against price spikes in the short-term market and for funding for peaking generation capacity. For example, a supplier 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/MWh = $400/MWh – $300/MWh times the number of MWh sold. If the spot price is less than $300/MWh, then the buyer of the cap contract does not receive a payment.

Because the seller of a cap contract is providing insurance against price spikes, it must make payments when the price exceeds the cap exercise price. This price spike insurance obligation implies that the buyer must make a fixed up-front payment to the seller in order for the seller to be willing to take on this obligation. This up-front payment can then be used by the seller of the cap contract to fund 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.

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 supplier has an obligation to ensure 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 of the risk associated with higher spot prices at that location. In order to prudently hedge this risk, the seller has a very strong incentive to 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 supplier signs a forward contract guaranteeing the price for 500 MWh of energy for 24 hours a day and 7 days per week at a specific location in the network, it will construct or contract for more than 500 MWh of generation capacity to hedge this short-term price risk. Building only a 500 MW facility to hedge this risk would be extremely imprudent and expose the supplier to significant risk, because if this 500 MW facility is unavailable to provide electricity, the supplier must 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 suppliers 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 energy contracts have a strong incentive to find the least-cost mix of generation and demand-side resources necessary to meet their contractual obligations.

5. 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 mechanism, there is no limit to what offer price that supplier could submit and be accepted to provide energy.

The system-wide market power problem is typically addressed through sufficient fixed-price and fixed-quantity long term contracts between suppliers and electricity retailers and large consumers. The logic described in Section 4.1 demonstrates how forward contracts limits the incentives of suppliers to exercise system-wide unilateral market power in the short-term market.
5.1. Solutions to the Local Market Power Problem

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

In all offer-based electricity markets a local market power mitigation (LMPM) mechanism is necessary to limit the offers a supplier submits when it faces is insufficient competition to serve a local energy need because of combination of the configuration of the transmission network and concentration of ownership of generation units. A LMPM mechanism is a pre-specified administrative procedure (usually written into the market rules) that determines: (1) when a supplier has local market power worthy of mitigation, (2) what the mitigated supplier will be paid, and (3) how the amount the supplier is paid will impact the payments received by other market participants. Without a prospective market power mitigation mechanism system conditions are likely to arise in all wholesale markets when almost any supplier can exercise substantial unilateral market power. It is increasingly clear to regulators around the world, particularly those that operate markets with limited amounts of transmission capacity, that formal regulatory mechanisms are necessary to deal with the problem of insufficient competition to serve certain local energy needs.

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

Another 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 suppliers, face the regulator must take a more active role in the transmission planning and expansion process to ensure that competition-enhancing upgrade that improve market efficiency are built. Wolak (2015) presents a framework for measuring the competitiveness benefits of transmission expansions in an offer-based wholesale market and applies it to the Alberta, Canada wholesale electricity market.

5.2. Cost-Based Short-Term Markets

An alternative approach that 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, which are typically equal to the highest variable cost necessary to meet demand.

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

5.3. Solutions to System-Wide Market Power

As discussed in Section 4 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 supplier only earns the short-term price on any energy it sells in excess of its forward contract commitment and pays the short-term price for any production shortfall relative to these forward contract commitments.

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, Mansur and Saravia (2008), the California electricity crisis is very unlikely to have occurred if there had been adequate coverage of California’s retail electricity
demand with fixed-price and fixed-quantity forward contracts. Consequently, in order to protect against periods when one or more suppliers has 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.

6. 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 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 programs require the regulator to set an administrative baseline, which can significantly reduce the system-wide benefits of active demand-side participation. This issue is discussed in Bushnell, Hobbs, and Wolak (2009).

6.1. 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 an hourly basis. Second, they must receive actionable information that tells them when to alter their consumption. Third, they must pay according to a price that provides an economic incentive consistent with actionable information to alter their consumption. A major challenge to active involvement of final consumers in the wholesale market in a developing country is the availability of the technology to record the customer’s consumption on an hourly basis. This can require the installation of new interval meters which can put additional financial stressed on any already stressed electricity retailer.

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).

---

2 McRae and Meeks (2016) resents the results of a field experiment in Central Asia that demonstrates the importance of actionable information for facilitating active demand-side participation.
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 in order to estimate their price responsiveness. Charles River Associates (2004) analyzed the results of the SPP experiments and found precisely estimated price responses for all three types of customers. More recently, Wolak (2011a) reports on the results of a field experiment comparing the price-responsiveness of households on a variety of dynamic pricing plans. For all of pricing plans, Wolak found large demand reductions in response to increases in hourly retail electricity prices across all income classes.

Although all of these studies find statistically significant demand reductions in response to various forms of short-term price signals, none are able to assess the long-run impacts of requiring customers to manage short-time 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.

This evidence suggests that for the developing country context, the key question is: For what classes of customers does installing interval meters make economic sense? However, the price differential between a conventional mechanical meter and an interval meter is currently so small that for regions where grid-supplied electricity is being introduced it makes very little economic sense not to install interval meters.

6.2. Managing Bill Risk with Dynamic Pricing

Politicians and policymakers often express the concern that the subjecting consumers to real-time price risk will introduce too much volatility into their monthly bill. These concerns are, for the most part, unfounded as well as misplaced. Wolak (2013) suggests a scheme for facing a consumer with the hourly wholesale price for her consumption above or below a pre-determined load shape so that the consumer faces a monthly average price risk similar to a peak/off-peak time-of-use tariff.
It is important emphasize that 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 the over the course of the month or year, the retailer’s total revenues less its transmission, distribution and retailing costs, must 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 increases the retailer’s demand for fixed-price forward contracts from generation unit owners, which reduces the amount of energy that is actually sold at the short-term wholesale price.

6.3. Fostering Investments in Automated Response Technologies

Perhaps the most important, but most often ignored, lesson from electricity re-structuring 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 it receives for any short-term energy sales is the hourly short-term price. Just as very few suppliers are willing to risk selling all of their output in the short-term market, consumers should have similar preferences against too much reliance on the short-term market and would therefore be willing to sign a long-term contract 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.3 This type of pricing

---

3 Wolak (2013) draws analogy between this pricing plan for electricity and how cellphone minutes are typically sold. Consumers
arrangement would result in a significantly less volatile monthly electricity bill than if the consumer made all of his 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 requirements contract for all of the electricity they might consume in a month, only that the consumer must pay the full cost of the retailer supplying this product.

The risk of paying the real-time price for their electricity is what creates the business case for investments automated response technologies and storage technologies. If the customer can avoid consumption when the real-time price is high and consume more when the price is low, through an investment in one of these devices, they are very likely to do so if the avoided wholesale energy purchase costs this technology avoids more than covers the cost of this investment. A single fixed retail price or single fixed price schedule regardless of real-time system conditions can never provide the revenue stream needed to finance investments in these technologies. Consequently, without exposing customers to the risk of the real-time price in the same way that generation unit owners face this price as their default price for electricity sales, investments in these technologies will not occur without explicit support mechanisms.

7. Market Design Lessons for Sub-Sahara and Southern Africa

The analysis of Sections 3 to 6 provides recommendations for electricity market designs in Eastern and Western Sub-Sahara Africa that can deliver significant economic benefits with a low implementation cost and limited regulatory burden. This analysis also yields suggestions for improving the current market design in Southern Africa.
7.1. *A Simplified Market Design for Eastern and Western Africa*

The transition to market mechanisms in both Western and Eastern Africa has been very slow. Both regions proposed formal regional wholesale markets in the early 2000’s, but both regions have yet to begin operating a formal market mechanism. Both regions face significant challenges because of limited transmission capacity between and within their member countries. Consequently, any attempt to operate an offer-based market for either region is likely to run into severe local and system-wide market power problems. In addition, virtually no deployment of interval 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 East and West Africa is a cost-based short-term market that uses locational marginal pricing (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. The market operator would only need each market participant to declare the available capacity of each of the units it owns. Then the market operator could compute the LMPs and dispatch levels for each generation unit given the realized demand at each point of withdrawal from the transmission network for each hour of the following day.

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 optimizes 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 suppliers would submit the technical characteristics of their generation units to the regional market operator and it would determine 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 regional market operator could use a globally traded price for coal and a benchmark delivery cost to the generation unit to determine the fuel variable cost of the unit. This
would be multiplied by the heat of the unit to compute the variable fuel cost. An estimate of the variable operating and maintenance cost for the unit could be added to this variable fuel cost to arrive at the total variable cost of the unit. In order to provide incentives to minimize their total actual variable cost of producing electricity, the values of the components of the total variable cost should be based on benchmark values for the technology used by the generation unit owner, whether 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 a day-ahead schedules and prices for each location in a multi-settlement version of this market design. In real time the dispatch and locational pricing process would be completed using the actual system demand and actual configuration of the transmission network.

To ensure long-term resource adequacy in this market, retailers would be required to purchase forward contracts for energy at various horizons to delivery equal to pre-specified fractions of their realized demand or face a financial penalty for under procurement. For example, retailers could be required to purchase 100% percent of their actual demand in a forward contract purchased before the short-term market operates for that day, 95% percent of their demand one year in advance, 92% two years in advance, and 90% 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. As discussed in Section 4, 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 generation unit owner that has sold each MWh of energy the retailer’s customers consume in a fixed-price forward contract so that all wholesale energy sold to final consumers will be purchased through these fixed-price forward contracts.

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. A generation unit owner rarely produces the exact quantity sold in fixed-price forward contracts during any given hour of
the day. Retailers rarely consume 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, 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 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 units would be required to operate. This provides an incentive for the unit owner to maximize the availability of their unit to avoid this set of circumstances.

It is important to emphasize that this short-term market is only for settling imbalances. That is purpose of the requirement for retailers to eventually procure 100% 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 assures 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 is a fixed-price forward contract during that hour. 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 obligation.

The market design also have the advantage that it can easily transition to an offer-based market once the transmission network in the region is expanded, interval meters are deployed and the regulator is able to design an effective local market power mitigation mechanism. The LMP
market is in place and suppliers costs as computed by the market operator can easily be replaced by the offers of suppliers. 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 following 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.

7.2. Improving the Southern Africa Power Pool

The Southern Africa Power Pool currently operates a zonal market despite limited transmission capacity both between and within the regions. As discussed in Section 4, a zonal market design does not make optimal use of the limited transmission capacity available within the Southern Africa Region. This market design can also create incentives for suppliers to take advantage of difference between the zonal market used to set market prices and generation unit dispatch levels and the actual operation of the regional transmission network.

As 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 regions. Mansur and White (2012) dramatically demonstrate this point by comparing the trades between regions of the eastern US before and after these regions are integrated into a single locational marginal pricing market that accounts for the configuration of the transmission network throughout the entire regions. Hourly energy flows between the two regions increase by almost 1000 MWh immediately following the integration of the two regions into an LMP market. There was not change in the physical configuration of the transmission network for the two regions. This increase in flows was purely due the incorporating the two regions into a formal LMP market that recognizes the configuration of the transmissions network for the two regions.

This result suggests that trading between regions in Southern Africa could be significantly increased by operating a LMP market over this region. Given the challenges of managing the potential local market power problems that are likely to emerge, this market should initially be run as a cost-based market with the mandated contacting levels described above. Once sufficient transmission upgrades, interval meters are deployed and effective local market power mitigation mechanism is implemented, the market could transition to be offer-based.
8. 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. A growing number of jurisdictions in the developing world have significant renewable energy goals. Consequently, any market design adopted by these regions should support cost effective integration of renewable resources. As we discuss below, the LMP market design is ideally suited for integrating any amount of intermittent renewables into a national or regional electricity supply industry. 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 additional products to deal with the increasing share of intermittent renewable resources.

8.1. Locational Marginal Pricing Markets and Renewables integration

A strength of the LMP market design for a national or regional wholesale electricity market is that any constraint that is relevant to operating the transmission network can be incorporated into the pricing and dispatch mechanism and if this constraint is binding it will be reflected in the prices that generators receive and consumers are paid. Consequently, if certain operating constraints become relevant as the amount of intermittent renewable generation in a region is increased and these constraints can be expressed mathematically with an acceptable degree of precision, they can be built into the LMP pricing mechanism.

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 3, 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.

8.2. Enhancing the LMP market to Support Renewables Integration

Intermittent renewable energy sources such as solar and wind represent non-synchronous power sources that do not contribute to system inertia. As their share increases, the overall system inertia will decrease which will increase requirements for primary frequency regulation. Overall, unpredictability and lack of inertia associated with these resources will impose a very considerable demand for additional flexibility, particularly for the ancillary services to maintain the second-to-second real-time supply/demand balance. Strbac et.al. (2012) argues that increased requirements for real time ancillary services, if provided by conventional generation running part-loaded, can not only reduce the efficiency of system operation but may undermine the ability of the system to accommodate production of variable renewable generation.

8.3. Development of Ancillary Services Market to Support Renewables Integration

As the value of flexibility increases, the demand for ancillary services may increase order of magnitude, as noted in Strbac et.al. (2012), Sturt and Strbac (2012) and Strbac et.al. (2015). Introducing products that provide flexibility will be critical for the cost-effective operation of the transmission network, as argued in Teng et.al. (2016), Samuel et.al. (2015) and Ela et.al (2014). Adjusting the demand for ancillary services on an hourly basis and setting trading interval level prices will become increasingly necessary. The LMP market is again ideally suited to deal with this challenge because it can allow co-optimisation of the energy market with the ancillary service markets, setting locational price for energy and prices for a range of ancillary service products in each trading interval.
9. Directions for Future Research

The experience of the past twenty-five years identifies the following necessary conditions for a successful market design. First, a multi-settlement locational marginal pricing 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. Capacity payment mechanisms have also begun to emphasize the development of a liquid forward market. Third, 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. All US markets and most international market have local market power mitigation 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, and a multi-settlement locational marginal pricing market provides the idea wholesale market platform for this to occur. The need for active involvement of final demand is even greater in regions with significant renewable energy goals. Finally, three commonly employed approach to financing renewables investments: (1) feed-in tariffs for renewable power, (2) renewables portfolio standards, and (3) fixed-price, fixed quantity forward contracts for energy. The first two approaches are the most common, but they can have adverse impacts on short-term market performance. The third is the least common, but is the least likely to degrade short-term market efficiency.

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 distributed renewables and grid-scale renewables.

On the first topic, the paper identifies several directions for future research. The least cost regulatory mechanism for a developing an active forward market to ensure long-term resource adequacy is perhaps the most important issue. 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 local market power mitigation 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 function of initial conditions in the country and the electricity supply industry could provide important guidance to developing countries considering reforming their electricity supply industries. A number of economic experiments with information provision and dynamic pricing programs 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 ancillary services demands are likely to scale with different scenarios for renewables deployment in the transmission and distribution grid could be extremely helpful for developing countries wanting the expand the contribution of the 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 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 distributed generation investments and at the transmission grid scale because for grid-scale renewables investments.
References


Volume 8, Takatoshi Ito and Anne Krueger (editors) University of Chicago Press, 79-134.(*


