A Common Sense Guide to Wholesale Electric Markets

Collin Cain, M.Sc.
Jonathan Lesser, Ph.D.

Bates White
April 2007
Table of Contents

1. Introduction .............................................................................................................. 2
2. Competitive wholesale markets ............................................................................ 6
   2.1. Benefits of RTOs ............................................................................................... 6
   2.2. How Wholesale Markets Work .......................................................................... 11
   2.3. Security constrained economic dispatch ......................................................... 17
   2.4. Summary ........................................................................................................... 22
3. Is the system working as it should? ................................................................. 24
4. Conclusions .......................................................................................................... 37
Appendix – PJM ........................................................................................................ 38

Table of Figures

Figure 1: Least-Cost Dispatch and System Costs ..................................................... 10
Figure 2: A Simplified Power Pool ......................................................................... 18
Figure 3: Key Elements of a Centralized Wholesale Market .................................. 23
Figure 4: The PJM territory ...................................................................................... 38
Figure 5: NERC Regions Today ............................................................................... 40
Figure 6: Regional Transmission Organizations ....................................................... 41
1. INTRODUCTION

In the last year, there has been growing controversy regarding whether wholesale electric markets are working as they should. Much of this controversy stems from the impacts on retail ratepayers as price caps, which were often imposed as an initial condition of industry restructuring and deregulation, expired. Coupled with higher fossil fuel prices, retail consumers have seen the prices they pay for electricity increase significantly. These price increases have led some to question the competitiveness of wholesale markets and the mechanisms used to establish wholesale prices. These criticisms have also called into question whether wholesale markets are adding value to the electric industry, or are only exacerbating price increases at the retail level. And, with questions raised over the value of wholesale markets have come policy recommendations to reregulate retail electric prices and have local electric utilities resume their past responsibilities for owning and operating generating plants, and undertaking long-term integrated resource planning activities that were common in the 1980s and 1990s.

There is no doubt that wholesale electric markets are complex, in large part because electricity is different than most other goods and services. For example, electricity cannot be stored in a warehouse until it is needed. Moreover, when shipped over high voltage transmission lines – the electric equivalent of interstate highways – electricity doesn’t always follow simple or unique paths, especially when those highways become overcrowded (congested). Finally, unlike most other markets, wholesale electric markets require technical overseers, much like air traffic controllers, to ensure the entire transmission system operates safely and reliably.
From wholesale to retail: how wholesale prices become retail rates

How wholesale electric prices translate into retail rates is not always clear, especially for customers who continue to take services from their local electric distribution utility, even though their utility is no longer in the generation business.

In a number of jurisdictions, local electric utilities like BGE hold auctions to provide the electricity needed to serve their non-switching retail customers, typically residential and small commercial customers, who have not chosen another retail provider. The suppliers who bid the lowest prices win the auctions, and agree to provide what is called “full requirements” service at a fixed price.

Full requirements service includes all of the services needed to ensure that the lights always stay on. Some of these products, including on and off peak energy, capacity, reserves, voltage control, reactive power and other ancillary services, are available through the regional power pools, such as PJM. Wholesale suppliers are also responsible for “shaping” the energy to meet minute-by-minute changes in customer’s demand, meeting the requirements of any Renewable Portfolio Standards (RPS), and managing a variety of risks, from credit to weather to customer migration to other suppliers.

To the cost of supplying generation, which many states call “standard offer service” (SOS) or “provider of last resort” (POLR) service, are added the costs incurred by the local utility to transmit and distribute electricity to its customers, including maintaining poles and wires, building new substations, sending out bills, and so forth.

In many states that have restructured, all customers have the ability to choose a competitive supplier. However, because most residential rates have been capped, it has been difficult of competitors to beat the below market price offered by the utility. Many large
commercial and industrial customers, whose rates are no longer capped, have chosen their own retail provider. Some of these customers choose a “real time” product, which is priced at the cost of wholesale power each hour. Others negotiate fixed-price arrangements similar to homeowners who pre-buy their heating oil at a guaranteed price before winter. The scope and breadth of these arrangements is limited only by the creativity of willing buyers and sellers. Whatever prices those large customers negotiate, however, they too must still pay the cost of interstate transmission and to maintain the local distribution system, just like residential customers.

But electric markets are not so different that basic market principles of competition no longer apply. In fact, wholesale electric competition is working well. Our analysis has shown that wholesale electric competition has created significant value for customers in Maryland, Pennsylvania and elsewhere. Numerous studies show that competitive or “merchant” generation owners have improved operating efficiency and increased generating output relative to their still-regulated counterparts. In addition, it is well documented that dispatching power plants over a larger regional footprint allows the lowest cost supplies, which may not be the generation owned by the local utility, to be matched with load. Increased supply options and lower operating costs translate into benefits for consumers.

In our view, the concerns over the competitiveness of wholesale electric markets can be addressed through a better understanding of the different mechanisms that have been established to provide electricity to retail consumers at the lowest possible cost, and how those mechanisms work together. That is the primary purpose of this report. A secondary purpose is to address some specific concerns that have been raised by critics of the current wholesale markets.


2. COMPETITIVE WHOLESALE MARKETS

Wholesale competition in electricity was initiated and expanded through a succession of federal rules beginning in the late 1970s, and has been facilitated by the development of multi-utility power pools such as PJM\(^1\). The Federal Energy Regulatory Commission (FERC), which is responsible for regulating interstate transmission of electricity (as well as natural gas and oil), has pushed for transferring management and operation of the electric transmission system from individual utilities to independent regional organizations as a critical element in promoting electric competition. A number of these Regional Transmission Organizations (RTOs) currently operate in the United States, covering many of the country’s major load centers.\(^2\)

2.1. Benefits of RTOs

RTOs are, first and foremost, guarantors of system reliability. Whereas the level of control over individual generating plants differs among the different RTOs – some centrally dispatch all generators, while others just act as “traffic cops” – all RTOs must meet the same reliability standards, so as to prevent large scale blackouts. The secondary mission of RTOs is to operate the system as cost-effectively as possible. Some, like PJM, do this by centrally coordinating operations of all generating plants. Others RTOs, such as the Midwest Independent System Operator (MISO), act as scheduling coordinators between buyers and sellers, but do not control how individual generators are dispatched.

\(^1\) PJM, formally called PJM Interconnection, is the Regional Transmission Organization covering all or part of 13 states and the District of Columbia, stretching from the Mid-Atlantic States to the Midwest. Further background and description of PJM is provided in the appendix to this paper.

\(^2\) RTOs and Independent System Operators (ISOs) have essentially the same functions. Current usage distinguishes them on the basis of whether they cover a single state (ISO) or multiple states (RTO). See the appendix to this paper for a map of current RTOs.
**Ensuring Reliability**

RTOs and power pools like PJM provide a number of reliability benefits. First, because electricity cannot be warehoused like wheat, the amount of power generated must be able to meet customer demand every second. This can be difficult for several reasons. First, most power plants cannot adjust how much electricity they generate as fast as consumer demand changes. Second, power plants can break down unexpectedly (called a “forced outage”). If that happens, keeping the lights on requires that other plants can fill the void instantly.

Power pools greatly reduce the risk that these sorts of outages will result in customers not having electricity. They do this by coordinating, or “interconnecting,” the operations of many different generating plants. In this way, if one plant needs routine maintenance, or suddenly breaks down, other generating plants are available to meet customer demand. Power pools like PJM still need to have reserve generating capacity, or “operating reserves,” but such resources can be used far more efficiently than if individual utilities separately operate their power plants. RTO operators ensure that, at every moment of every day, there are generators ready to provide electricity at a moment’s notice if called on to do so.

RTOs also ensure that the transmission system operates safely. The “system operator” monitors how electricity is flowing throughout the transmission system at all times. The reason constant monitoring is necessary is that the flows of power associated with multiple generators and users across a wide power grid are highly complex and often unpredictable. While generators may have contracts with buyers, electrons are not bound by
contract, but instead obey the laws of physics, following the path of least resistance.\(^3\)

Without oversight, an individual generator’s sale to a single buyer could adversely affect the entire transmission system. In the extreme, the system could become overloaded and individual transmission lines could fail, causing a blackout.

### A Glossary of Common Reliability Terms

**Automatic generation control**: Equipment that allows the system operator to increase or decrease generation output automatically from a power plant in response to changing system conditions.

**Operating reserves**: Generation that is not running, but which can be brought on line within a specified timeframe, such as within 10 minutes or 30 minutes of being called on.

**Planning reserve**: The difference between a utility’s or region’s peak generating capability and its expected annual peak demand. For example, PJM requires a 15% planning reserve. (Also called capacity reserve.)

**Spinning reserves**: Generation that is not supplying power, but can be brought “on-line” instantly if needed.

### Reducing Costs

Another key benefit of power pools like PJM is that they determine which sources of electricity will be used to meet demand, selecting or “dispatching” the cheapest sources available at any time. Low-cost “baseload” plants are dispatched first, followed by higher-

---

\(^3\) People are discouraged from standing under trees during electrical storms, for example, because our bodies are excellent conductors of electricity.
cost resources, according to need, and independent of whether the resources are used to serve local or more distant needs. This is called “economic dispatch.” Before restructuring, “tight” power pools, such as PJM, NEPOOL, and NYPP controlled how all of the power plants in their areas were dispatched. This centralized dispatch was accomplished based on the estimated incremental (or marginal) costs to run each plant. Those incremental costs consist primarily of the cost of the fuel burned to produce each kilowatt-hour (kWh) of electricity, plus all of the incremental maintenance costs that increase as generation output increases. With that cost data, the system operator would then schedule all of those plants to meet overall system load every hour from the lowest to highest incremental cost units, as illustrated in Figure 1.

When the system uses economic dispatch, the cost of the very last generating plant needed to supply power in a given hour sets the system cost. For example, suppose the incremental cost of a coal plant is 4¢/kWh, the incremental cost of a combined-cycle plant is 6¢/kWh, and the cost of a peaking unit is 8¢/kWh. When the baseload coal plant can meet all electricity demand, the cost of meeting the power needs for the system will just be 4¢/kWh. As demand increases, and the higher cost combined-cycle plants are required, and so system costs increase to 6¢/kWh. Finally, on hot summer days when everyone switches on their air conditioner and peaking units are need to run to meet customer demand, system costs increases to 8¢/kWh.

Figure 1 shows a simplified supply-demand example with three types of generating plant – baseload, intermediate, and peaking. The vertical axis shows the variable cost (fuel

---

4 In contract to “tight” power pools, “loose” power pools do not have centralized dispatch.
5 These costs are referred to as variable operations and maintenance (O&M) costs.
6 This is called the “system lambda.”
cost plus other variable expenses) of each type of generation and the horizontal axis shows the quantity of generation available or needed at a particular point in time. Baseload generating plants, such as nuclear and coal plants, have the lowest fuel cost and run virtually at all times. These types of plants also tend to have very high startup and shutdown costs, so it makes economic sense to run them around the clock. So, when demand is low, such as in the middle of the night, baseload generation can meet then entire system demand, shown in Figure 1 as Demand (Night). As shown in the figure, the marginal cost of the last MWh of baseload generation used equals to Cost\textsubscript{BASE}.

**Figure 1: Least-Cost Dispatch and System Costs**

When customer demand peaks, such as during the middle of a hot summer day, baseload generation must be supplemented with intermediate and peaking generation. Intermediate generation tends to be gas- and oil-fired combustion turbines or combined-cycle plants. Peaking units are typically small combustion turbines, or even diesel engines.
Whereas baseload plants are expensive to build, but cheap to operate, peaking plants are cheap to build, but expensive to operate. Peaking plants aren’t designed to run all the time. They tend have higher fuel costs and higher maintenance costs for each kWh generated. When the system operator needs peaking units, the system cost rises to reflect the marginal cost of the last peaking unit needed, shown as Cost\textsubscript{PEAK} in Figure 1.

### 2.2. How Wholesale Markets Work

Most wholesale power is bought and sold through bilateral contracts, freely negotiated between individual buyers and sellers. In PJM, for example, about 85% of all market sales are bilateral. The remaining 15% are made into the spot market, which acts like a clearinghouse for buyers and sellers. However, even though it accounts for only a small percentage of total wholesale transactions, the spot market plays a crucial role in these bilateral sales, because it provides “visible” market prices, which economists sometimes call “price discovery,” just as organized markets like the Chicago Board of Trade and NYMEX provide price discovery for everything from natural gas to nickel.

Where there is vigorous wholesale market competition, such as in PJM, centralized dispatch is no longer determined by a system operator using estimated incremental costs for each generating plant. Instead, PJM uses a bid-based system, in which individual generators submit their own price bids every day to the system operator to meet the expected customer demand for the following day.\textsuperscript{7} Once the system operator has all of the bids, he selects those generators whose bids are lowest and tells generators who are not selected that they will not be needed.

\textsuperscript{7} This is called the “day-ahead” market.
Day-ahead and real-time markets

Bid-based energy markets begin with generators submitting their bids to the system operator in the day-ahead (DA) market. As its name implies, the day-ahead market takes place the day before the actual operating day. The DA market is a financial market, rather than a physical one. Essentially, the DA market is like a commodities futures market that allows buyers and sellers to hedge their transactions. Selling generation in the DA market, like selling orange juice futures, doesn’t mean the seller is committed to physically delivering their product. Instead, generators whose bids in the DA market are accepted are bound into a financial obligation. A generator that cannot physically provide the power bid the previous day must obtain that power in the real-time (RT) market, which is a “physical” spot market. It is in the real-time market that generators provide the electricity needed to keep the lights on.

The reason for having both DA and RT markets is that a lot can happen in the 24 hours between when generators’ bids are accepted and when they are actually scheduled to run. For example, a generating unit can break down unexpectedly and not be available. Or, there may be a maintenance problem with a substation or a transmission line that prevents a generator from supplying the transmission grid with power. Thus, the RT market can be thought of as a “balancing” market, which responds to real-time events that affect supply and demand. (This is also why power pools require the three different types of reserves – operating, planning, and spinning – to ensure that when the unexpected happens, there is still enough capacity available to keep the lights on.)

All of these real-time operational risks mean that prices in the RT spot market tend to be more volatile than in the DA market. Moreover, because prices in the DA market are

---

8 This is called “injecting” power into the grid.
more stable than those in the RT market, buyers and sellers can use the results of the DA market to further hedge against factors that can drive prices unexpectedly up or down, such as a change in the weather than wasn’t forecast.

For example, day-ahead bilateral sales are often settled through contracts-for-differences. Using a highly simplified example, consider a case where a generator enters into a bilateral contract with a buyer for 100MW per hour at $30/MWh. However, suppose that the next day, the actual RT spot market price has increased to $35/MWh. When the generator physically sells electricity into the PJM market, he will receive $35/MWh from PJM and the buyer will pay $35/MWh to PJM. If the buyer and seller had signed a contract-for-differences, the generator would pay $5/MWh to the buyer, resulting in a net transaction price of $30/MWh on each side. If the RT spot price decreased to $25/MWh, the buyer would pay the seller $5/MWh. In either case, both the buyer and the seller have hedged their price risk.

What is most interesting is that the system operator doesn’t need to know anything about the terms of any bilateral contracts in order to optimize dispatch of the entire system.\(^9\) This is possible because the bids generators submit determine how the system is dispatched, and financial contracts between individual buyers and sellers ensure that the terms of their individual agreements are met.

**Prices in theory and practice**

In an ideal competitive market (the kind economists typically use for classroom examples), each generator’s bid would exactly equal its marginal cost, and the supply curve

---

\(^9\) Contracting may affect both supply and demand bids, but the system operator doesn’t need to know the contract details. In some cases, generators may submit operational schedules instead of bids to reflect contract arrangements.
would look just like the one in Figure 1. In practice, bids will in fact tend to approximate actual marginal costs – a generation owner stands to lose money by bidding above marginal cost and not being selected, or by bidding below marginal cost, being selected, and then being unable to recoup the costs to produce electricity.

There are several reasons why bids may not equal the instantaneous marginal cost for a generation unit. First, a generator who submits a bid must consider the costs of starting up and shutting down his plant. This is why most nuclear plant operators and other baseload plants submit $0 bids. By submitting a $0 bid, the system operator will select them first, ensuring they do not have to shut down and incur very large shutdown costs. Since the market-clearing price will always be positive, owners of baseload plants know they will still be paid for their generation, even though in many offpeak hours when electric demand is low, the market prices will be below their plants’ average costs.\(^\text{10}\)

The bid-based dispatch system also allows generation owners to factor in more complex considerations, such as the opportunity to sell output into other, higher-priced markets, and operational limits that might make it more profitable (and hence more valuable to the system) to operate later in the day rather than immediately. For example, this is often true for hydroelectric plants. It is also the case for plants that produce both electricity and steam for industrial purposes (co-generation plants) that factor their own need for steam into the wholesale market bids they submit.

Another difference between the competitive market “ideal” and reality is that generators not only have to cover all of their variable costs, like fuel, but they must earn sufficient revenues to pay their fixed costs, such as servicing debt borrowed to build the

\(^{10}\) It is actually possible for the market clearing price to equal zero, when demand is very low.
plant, wages and salaries, insurance, and so forth. Investors will include all of these costs as an opportunity cost of doing business, and will not enter a market if they believe they will not be able to recover all of their costs, including a return on their investment, through market prices.

**Market power concerns**

A final reason why bids could deviate from marginal cost is because of the exercise of market power. For instance, a generation owner with a portfolio of assets might be able increase market prices by submitting a high bid for one unit. Even if the unit doesn’t run, the strategy could raise prices by causing another high-cost unit to set the market price. Despite the lost revenue from one plant, it is possible for the increased revenue from other plants to result in higher profits.

There are substantial risks associated with market manipulation, however. First, the strategy may not work, and may result in losses. More important, getting caught will result in painful sanctions. In order to operate as an RTO, FERC requires that the markets be monitored by independent market monitors. These market monitors to scrutinize bidding behavior to prevent a range of potential manipulation strategies and report regularly to both the RTOs and to FERC. Market monitors also employ various FERC-approved mitigation techniques, including bid caps, cost-based bids, reliability must-run contracts and other out-of-merit dispatch in cases where the potential for market power is high – e.g. in a transmission-constrained area where the bidding behavior of a single plant could have a large

---

11 Fixed costs are costs that remain constant, no matter how much electricity a plant generates.
effect on prices. Additionally, one of FERC’s most important roles is to ensure that competitive wholesale markets are truly competitive. Generation owners who want to sell their output competitively must apply for “market-based rate” (MBR) authority with FERC every three years, showing that they either do not have, or have effectively mitigated, any market power. In addition, FERC has significantly boosted its enforcement capabilities, largely in response to new authority provided in the Energy Policy Act of 2005 that dramatically increased FERC civil and criminal penalty authority.

So, why not force all generators to bid their costs? Although this would seem to eliminate market power, it is likely to do more harm than good. First, as we discussed above, it is not so easy to determine each generator’s exact marginal cost. The complexities of estimating start-up and shutdown costs, opportunity costs of selling elsewhere or deciding whether to generate today or tomorrow, significantly complicate our classroom example. Using a bid-based system with appropriate monitoring is easier, more efficient, and produces lower prices. A second reason is that forcing generators to bid their marginal costs, or instituting too onerous market mitigation, eliminates the important role of scarcity pricing. The fact that a generator bids more than their marginal cost is not prima facie evidence of market power. Instead, it may mean that there is just too little supply for a given level of demand. When that happens, all competitive markets “ration” the available supply using prices, which increase until quantity demanded equals the available supply. This is called scarcity pricing.

---

12 Of course, there is always a concern that these mitigation measures are disrupting legitimate scarcity prices need to incent investment in demand response, generation and transmission. This concern is discussed later.

13 Markets can still work with cost-based bids. There are some successful examples, mostly in South America, in which generators submit their bids, and the system operator assigns a marginal cost to them, but all these markets require additional capacity payments for generators to recover their investment and fixed operating costs.
Scarcity pricing sends an important signal to the market, encouraging demand to cut back when it is cost effective and telling potential market entrants that entering the market will be profitable. By encouraging appropriate response to price signals – by both demand and supply resources –, either load cuts back or supply increases and prices fall, benefitting consumers. If prices are never allowed to reflect scarcity, such as with price caps, then there won’t be appropriate conservation at peak time or needed market entry. As a result, and as has been observed when multi-year retail price caps have been lifted, prices can zoom upwards suddenly, harming the same customers whom the price caps were ostensibly designed to protect.

2.3. Security constrained economic dispatch

As explained above, the DA market is really just a financial market, while the RT market is a physical one. In a tight power pool with centralized dispatch, such as PJM, the system operator schedules generators to operate based on their real-time bids and physical availability. Thus, supplier A’s bilateral contract to sell 100 MW of power to buyer B doesn’t mean supplier A will necessarily be physically generating 100 MW in real time that will be routed to buyer B. In fact, supplier A may not be generating anything at all at the time the contract is exercised.

Instead, the system operator is constantly balancing the generation that is physically available with demand. Specifically, the system operator must “redispatch” generating plants to account for generation that is physically available, transmission system constraints, and real-time demand. This is called “security constrained economic dispatch” (SCED). Essentially, SCED is just least-cost dispatch (based on generator bids) that takes account of all of the physical constraints on the transmission system at a given moment. In some ways,
this is similar to planning a car trip by one route, but then hearing on the morning traffic report that the traffic is particularly heavy on that route. Drivers are encouraged to take alternate routes, which reduces the traffic pressure on the congested arteries.

A simple example can help explain SCED. Assume we have a transmission system with just two distinct areas, called Node 1 and Node 2 in Error! Reference source not found.. In each area, there is customer demand, or “load,” and there is also a generating plant. Node 1 has an 80-megawatt (MW) generator and 20 MW of load, while Node 2 has a 20-MW generator and 50 MW of load. So, the total generating capacity in the system, 100 MW, is enough to serve a total load of 70 MW. However, to meet the load at Node 2, some of the generation from Node 1 must be transmitted from Node 1 to Node 2.

![Figure 2: A Simplified Power Pool](image)

The question is how to meet that 70 MW of load at the lowest cost. To answer that question, we look at the marginal costs of each generator. The generator in Node 1 has marginal cost of just $10 per megawatt-hour (MWh). It is a large baseload plant that runs around the clock. The generator in Node 2 is a small peaking unit with a marginal cost of
$50 per MWh. If each generation owner bids his marginal cost and there are no transmission constraints, then the generator in Node 1 will supply the entire load in both areas. The market price will just equal $10 per MWh.\footnote{In this example, we are ignoring market power issues so as to provide a straightforward example of SCED.}

Things get a bit stickier if the transmission line in Figure 2 does not have sufficient capacity to transmit all of the power needed to serve load in Node 2. With transmission congestion, some of the generation in Node 2 has to be dispatched, increasing the cost of serving load in Node 2. Then, Node 2 becomes what is known as a “load pocket”—an area partially or fully isolated from the rest of the system where the cost of electricity is higher than generation from outside the load pocket—generation that would serve load in the pocket absent the congestion. In the example of Figure 2, let’s assume that in the summer the transmission line can only transfer 30 MW of power. Then, in order to serve a demand of 50 MW, the system operator will have to dispatch 20 MW of generation from Node 2, at a cost of $50/MWh.

As we start adding more nodes, the example in Figure 2 gets much more complicated. In PJM, for example, there are thousands of nodes. Transmission lines have physical and safety limits, and electrons flow according to the laws of physics, meaning that when operators intend to transfer power between two adjacent nodes A and B, they affect the whole system. Congestion might be increased in node C and relieved in a transmission line that serves node D, at the same time. The reasons for these effects stem from electrons’ responding to the laws of physics, rather than contracts drawn up by lawyers.

However, juggling all of these nodes and constraints is a problem for system operators. We will stick to our simplified example and ask a more relevant economic
question: what should the price of electricity be in Nodes 1 and 2? There are two main alternatives. The first one is to “socialize” the cost of serving load in Node 2. This is what traditional (pre-restructuring) vertically integrated utilities did for decades, and is how local distribution utilities like Baltimore Gas and Electric allocate costs among customers. For example, rural residential customers typically are not charged higher rates than customers in the city, even though the per-customer costs of poles and wires in rural areas is higher. In the same way, retail customers didn’t get charged different prices for generation; the cost of producing electricity was averaged out over all customers of a utility.

At the wholesale level, however, it is far more efficient to price generation in each node to reflect the cost of delivering the next MW of power to that specific node. This is called Locational Marginal Pricing (LMP), meaning setting the price at each location (and at each point in time) at the marginal cost of energy for that location. In our example in Figure 2, LMP prices (or LMPs) are $10/MWh in Node 1 and $50/MWh in node 2. The respective locational prices reflect the least-cost manner of providing an additional MW to each node. Again, the calculations get more complicated as the number of nodes increases, as generation in other nodes in the system may affect which generators are able to serve nodes 1 and 2, and their respective LMPs.

**What purpose do LMPs serve?**

LMP pricing is both a natural outgrowth of a centrally-dispatched market and a key element in creating incentives for efficient behavior by suppliers and consumers. In order to optimize the dispatch of generation, the system operator needs to know the value of relieving congestion in different areas. That means he must be able to calculate the marginal cost of generation at each node and in each hour. Yet LMPs are much more than an artifact
of the optimization system. LMPs correspond to the actual incremental cost of power for each location – i.e., the cost of producing a tiny bit more power. This incremental cost serves as a signal for both consumers and producers of energy.

Although LMPs in PJM are similar in most areas and most of the time, since the system typically is not seriously congested, LMPs will be higher for areas where the least-cost dispatch is limited by transmission constraints. These higher LMPs can be temporary, occurring only when there is an unexpected transmission line failure or a generator experiences a sudden maintenance problem, or can reflect areas in the system that are chronically congested.

When the transmission system is chronically congested in a particular location, LMPs provide both consumers and suppliers with important price signals. On the demand side, higher LMPs can encourage consumers to use energy efficiently, i.e., to use less power when prices are high. Such demand response can lower market prices by reducing the need for high-cost generation, and can produce large system benefits, even if only a small portion of load reacts to price. By making prices local and transparent, LMPs provide a strong incentive for cost-effective demand response programs that reduce electric use during peak hours.

On the supply side, persistently high LMPs signal to potential investors where additional energy supplies are most needed. LMPs serve to induce investment – in new generation supplies, in transmission capacity to reduce congestion, and in demand response measures.
2.4. Summary

A centralized wholesale market such as PJM can be characterized by three key interrelated elements:

1. **Auction Market**: a free-entry market where generation suppliers bid to meet the expected demand of local utilities and other retailers;

2. **Security Constrained Economic Dispatch**: a system operator ensures that the system works safely and reliably by directing traffic to avoid congestion, making sure everyone’s electrons get to where they need to be at the lowest possible cost; and,

3. **Locational Marginal Prices**: visible hourly prices that vary by location (when there is congestion), providing market signals to consumers and producers in the near term, and to potential investors in new generation, transmission or demand reduction in the long term.
Figure 3: Key Elements of a Centralized Wholesale Market

- **Auction Market**
  - Security Constraints
  - Local Market Scarcity

- **Locational Marginal Prices**
  - Signals to Consumers, Producers and Investors
  - Local Incremental Costs

- **Economic Dispatch**
  - Least-Cost Bids
  - Critically Congested Nodes
3. IS THE SYSTEM WORKING AS IT SHOULD?

Higher wholesale market prices that have led to increases in retail electric rates have caused some to question whether competitive wholesale generating markets are functioning as they should, and whether those “competitive” wholesale markets are truly competitive. These questions have also led to recommendations that electric markets be re-regulated, abandoning wholesale and retail competition in favor of the pre-restructuring market in which vertically integrated electric utilities owned and operated generating plants, and rates were determined using traditional cost-of-service principles.

Although it is critical to determine whether wholesale electric markets are functioning as planned, much of the recent criticism of those markets, and electric restructuring in general, arises from fundamental misunderstandings of how the market operates. Specifically, critics of restructuring have raised the following questions about competitive wholesale markets:

1. Why is so little generation being built where LMPs are the highest?

2. Why not pay generation suppliers what they bid, rather than market-clearing prices. Won’t that save customers money?

3. Can competition work in the electric industry, or is the industry so different from others that it must continue to be fully regulated?

**Does LMP provide a sufficient price signal to attract new investment where and when needed?**

Some areas of PJM experience congestion, and associated higher locational prices, on a persistent basis. That market incentives have not prompted investment in new generation or transmission sufficient to eliminate such price differentials has led critics to challenge the current LMP market design. While there is some research documenting patterns of price
difference and investment, there is no evidence to demonstrate that the patterns reflect market failure. First, there has been significant investment in both transmission and generation in PJM. Second, there are many reasons why there has not been sufficient investment to eliminate the LMP differences.

**Transmission investment**

To date, over $683 million has been investment in the PJM transmission system since the first Regional Transmission Expansion Plan (RTEP) was adopted in 2000. Currently, over $638 million in transmission upgrades are under construction and transmission projects totaling over $4.2 billion are currently planned.

**Generation investment**

Since 1999, 157 new generation projects have been developed in PJM, totaling more than 19,000MW of capacity, with 3,500 MW currently under construction.\(^\text{15}\) Over 1,100MW of wind and other renewable generation has also been added in the PJM footprint. This new generation represents 14% of the existing capacity in PJM’s expanded footprint. Moreover, PJM has itself expanded significantly. By doing so, generation dispatch has become more efficient, as the pool of available generating plants has increased.

**The role of LMPs**

As we discussed above, one of the most important questions in the case of electricity is whether the overall wholesale markets are competitive. With competitive markets, when prices remain at levels above those necessary to finance baseload plants, new developments take place and the increased supply brings prices back to normal levels. Similarly, projects do not obtain financing when prices are too low (nobody builds a nuclear plant to charge $10/MWh, even if it covers the fuel cost), and eventually, prices increase. The best way to

---

\(^\text{15}\) PJM Regional Transmission Expansion Plan (RTEP), February 27, 2007.
lower market prices is to either build new generation or to reduce demand. Markets do that when prices are allowed to reflect scarcity. That is what LMPs do—they signal where to invest in new generation supply or in demand response programs. Of course, a single hour's LMP does not make a long-term investment signal, nor would it be cost-effective to reduce all congestion on the system. Developers base their expectations of future market prices on longer-term price data, including past LMP trends, forward market prices, and so forth, to determine the expected returns from new investment. This is true not only of firms that own and operate generation, but also firms that sell demand response services to customers.

LMPs exist everywhere. They are the market-clearing prices at each zone or node. LMPs are higher in areas where transmission capacity and local generation are not sufficient to meet local demand—typically the more densely populated areas of PJM along the east coast. These are also the areas where it is difficult to site new generation and transmission. In some cases, this is the result of NIMBY reactions, whether because of concerns that generating plants and transmission lines can reduce property values, fears about environmental impacts, or even fears that those facilities will be targets for terrorists.

In other cases, however, locally higher LMPs may not be high enough to attract new investment. If the costs of building new transmission or generation in an area with a higher LMP are greater than the expected revenues over the long-term, then it will not make economic sense to build new generation. That is true whether purchasing power in a competitive market or having a local utility own and operate a new generating plant. In those cases, higher local LMPs can best be addressed by demand-response programs targeting peak LMP hours.

It is also difficult to prove a negative – that potential developers are not building new generation or transmission in certain areas despite sufficient economic incentive (and
practical capability) to do so. So we have to turn elsewhere to examine the question of whether market incentives are working.

There is every indication that the PJM energy markets are highly competitive – i.e., that bids and prices are based on true costs rather than wholesale suppliers exercising market power. Studies to the contrary have typically ignored real-world costs, particularly opportunity costs of selling into other markets, and operational constraints, such as start-up and shutdown costs, so-called “ramp-rates” – varying levels of plant performance at different output points – and mandated operational patterns imposed by system operators, called “reliability must-run” requirements. When these factors are properly assessed, as they are in analyses performed by the PJM Market Monitoring Unit, the result indicates that the market is functioning competitively.¹⁶

**Energy and capacity markets**

The PJM market monitor has concluded that, since 1999, the revenues available to generators have been insufficient to allow the recovery of the costs associated with new investment in PJM. There are several reasons for this. First, PJM has lots of excess generation. Even with concerns about shrinking reserve margins, PJM has well over the 15 to 18 percent reserve margins generally expected for reliability. Of course, every megawatt in PJM is not the same as every other megawatt. Because of transmission constraints, excess capacity in Chicago may not help keep the lights on in New Jersey. That is why PJM’s new Reliability Pricing Model (RPM) has a locational component. This will allow price signals for capacity to be set in the location where that capacity is needed. Second, wholesale energy markets have price caps and other forms of out-of-merit dispatch such as not letting reserves and imports set the clearing price, must run contracts and cost-capping in constrained areas.

This affects primarily developers of peaking plants, who rely on revenues earned in a small fraction of hours to earn sufficient revenues. If prices are capped during those hours when demand is higher, peaking unit owners will be unable to earn enough revenues to recoup their investment. Third, the price of building new generating plants has increased, owing to increases in the prices of labor, raw materials, and key equipment, such as turbines.

Another reason stems from the fact that reliability – the primary purpose of every RTO – is what economists call a public good. As discussed in Section 2, reliability is provided by having different types of reserves, including capacity reserves. However, reliability is provided either to everybody or nobody – the electric system cannot practically be made “reliable” for one customer without also making it reliable for all of the other nearby customers. As a result, like all public goods, too little reliability will be provided by the market alone. Therefore, there must be intervention to provide the necessary incentives to do so.

This is not a problem in the abstract. PJM is facing shrinking reserve margins and there are growing concerns that some areas within PJM could face reliability challenges in the years ahead. To address the concerns outlined above, PJM has developed, and FERC has approved, a significantly revised capacity market known as the Reliability Pricing Model or RPM, which takes effect in June 2007. The RPM is designed to send market price signals to maintain current generation, transmission and demand response investment and promote new investment when and where it is needed.

While the evidence indicates a robust market environment in PJM, one recent analysis indicates that more new generation has been built in PJM areas with low LMPs than
in high-LMP areas.\textsuperscript{17} This, the author claims, indicates a market failure, since higher LMPs have not resulted in new capacity investments in those areas. It indicates no such thing.

As we have noted above, the areas with the highest locational prices are those where it is also most difficult to site new projects. In order to identify a market failure, one must be able to demonstrate that current and expected future LMPs exceed the necessary investment threshold, i.e., are expected to return a developer’s investment including a return on investment commensurate with the business and financial risk of development, but that still nothing is being built.

The cost threshold is higher in more populated areas, a fact that is not captured in a simple comparison of LMPs and construction trends. Siting and permitting issues raise the investment cost threshold in highly populated areas by reducing the number of potential sites for development, limiting the possible technologies that can be employed because of environmental quality issues, and creating a more challenging permitting process. Local opposition to new plants or transmission in highly populated areas can also raise the effective cost threshold, because it increases uncertainty. If we add that wind and hydro can only be built where wind and water are, the result can be that coal, nuclear, wind, and hydro power are not viable options in load pockets.\textsuperscript{18} This is also a reason for the increasing level of investment in demand reduction programs, which are most valuable in load pockets.

Finally, regulatory uncertainty about the electric industry can further delay generation developments. It takes years to build a generating plant or a transmission line. Investors do


\textsuperscript{18} Renewable technologies like wind and solar also depend on the back-up of other generation in PJM. Specifically, because wind and solar cannot be scheduled the way fossil and nuclear plants are, they must be “firmed up” with other types of generation. For example, wind energy will be of no value on a hot, windless summer day when demand peaks. Thus, there must be sufficient capacity available when the wind is not blowing to meet demand.
not want to commit large amounts of capital for the long-term if they think policy makers will “pull the rug out” from them later on or otherwise make significant market and policy changes that may adversely affect returns to committed capital. This continues to be a crucial risk factor in today’s wholesale electric markets in light of recent wholesale and retail price increases. With calls to re-regulate the industry, restore price caps, or even levy “windfall” profits taxes, investors will be less likely to commit funds, exacerbating problems of insufficient new investment and higher market prices.

These factors help to explain the observation that more capacity has been built in lower LMP regions of PJM; none of those factors implies that the market is broken or that lowering prices where investment is most needed would contribute to a solution.

**Is a Single Market Clearing Price Fair?**

Rising electricity rates have led some regulators and policymakers to challenge the basic market clearing mechanisms of centralized power markets. Opponents of these auctions argue that generators should be paid the price they bid, rather than the market-clearing price, for the power they generate. In particular, opponents question whether baseload units, such as nuclear power plants, with very low variable operating costs, should benefit from natural gas price increases that raise market prices when gas-fired plants are on the margin. Critics see the resulting increase in profits for baseload plants as unjustified and unfair.

One proposed solution is “pay-as-bid” pricing, in which suppliers are paid the prices they bid individually, rather than the price set by the most expensive unit dispatched. Whereas pay-as-bid may seem superficially attractive, it ignores key economic facts. Virtually every competitive market operates the way centralized electricity markets do – that is, on a
single-price basis. Some people are uncomfortable when this concept is applied in electricity markets, but this discomfort largely reflects a poor understanding of why the same mechanism that serves so many other markets is appropriate for electricity. Opponents of single-price markets in electricity also tend to confuse foresight and hindsight when assessing a particular market outcome.

As discussed in Section 2, in the PJM bid-based system, generators submit their own price bids and the system operator dispatches those generators whose bids are lowest. All suppliers get the same market clearing price, based on the marginal cost of the last generating unit needed to supply power (the system lambda in Figure 1). The market price is, thus, equal to the cost of the system producing a little more power. This is appropriate, because it signals to users the true cost associated with their consuming a little more power, and it signals to producers the potential profit from making a little more output available. When market prices are high, the signal creates short-term incentives for users to consume less (if they see real-time price signals) and for producers to generate more. In the longer-term, the pattern of market prices creates the incentive for investments in new supply and, increasingly, in load management and programs offered by competitive firms.

This is the same as what occurs in other competitive markets. The difference is that in centralized electricity markets, this single price mechanism is explicit and rule-based, while in other markets it is implicit and an outgrowth of competition itself. In a generalized competitive market, sellers (and buyers) are considered to be price-takers, which means that no single participant can act to change the market price. When the market equilibrates, each seller’s output is at a point where trying to charge less and sell more, or trying to charge more and sell less, reduces profits; this leads to a single market-clearing price. However, it does not mean that all sellers make the same amount of profit. Factors such as talent,
innovation, patents, proximity to resources, timing of market entry, financing, luck, etc., can give some producers a cost advantage over others, resulting in high profits for some players and low profits for others in the same single-price market.

With electricity, there is the added complexity that the different ways of creating the same commodity have distinct cost structures. As noted earlier, baseload generators such as nuclear and coal-fired plants are expensive to build and maintain, but relatively inexpensive to operate. The reverse is true for simple gas-fired turbines. And, there are a multitude of technologies in between. In centralized markets, every bidder has to recoup their variable and fixed costs. This is analogous to household expenses – not only do you have to pay every month for water, garbage, gas bills, etc., but you also have to pay the mortgage.

Does this mean a windfall for baseload generators when peaking units set the market price based on the highest cost fuel price? The answer is clearly no. Like any business, generators need to recover their fixed costs, including a return on their investment, otherwise they will eventually go out of business. They need to recover their multi-million dollar investments by running at all times and by means of prices well above their own operating costs. In fact, baseload plants have generally recovered fixed costs over 20-30 years of operation. Most nuclear plant owners in PJM also took a substantial risk when acquiring plants when natural gas prices were very low – i.e., when nuclear generation was

---

19 Another issue that can be confusing is the recovery of depreciation costs. Under the traditional cost-of-service regulatory framework, a utility recovers an annual depreciation expense for all capital investment. This is done to account for the need to eventually replace that equipment. Without depreciation, a utility with fully-depreciated assets would have no ratebase (and thus earn no return for its investors). In essence, depreciation is charged so that a utility can recover the cost of its capital investments (separate and apart from earning a return on those investments.) Without charging for depreciation, a utility would be unable to make new capital investments when equipment reached the end of its useful life. So, while an older generating plant may be fully depreciated, its owner will expect to replace that plant much sooner than the owner of a new plant. For additional discussion, see, J. Lesser and L. Giacchino, *Fundamentals of Energy Regulation*, (Alexandria, VA: Public Utilities Reports, Inc.), forthcoming July 2007, Chapter 5.
far less competitive relative to gas-fired generation. If market prices had gone down, the owners of some nuclear plants might very well be out of business. Or, if they had suffered an extended outage—not so infrequent with nuclear plants when many changed hands, their investment cost recovery would also have been severely compromised.

Policies that attempt to limit short-term profits because they seem too high—especially after market prices equilibrate after years of artificial price caps—are shortsighted. The problem such policies create is that potential investors are less likely to invest in projects for the long term if they believe the rules of the market will change. High-risk investments require the probability of high payoff. If, instead, investors expect that profits will simply be taken away from them (i.e., a “heads I win, tails you lose” framework), the incentive to invest, take risks, and innovate will be stifled.

But the main reason why the pay-as-bid concept is only superficially attractive is that the proposal would not result in any consumer savings, not even in the short term. Pay-as-bid proponents base their arguments on the step-shape of the supply curve shown in Figure 1, in which generators bid their variable (“marginal”) operating costs, such as fuel costs. But, in a pay-as-bid auction, the supply curve will never look like that. Instead, generators would radically change their behavior and bid at the market price they anticipate in each hour (baseload plants would no longer submit $0 bids). If generators knew they would only be paid what they bid, they would have to bid higher to cover their fixed costs.

This is exactly what happens when a distribution utility procures power through bilateral contracts. Although the utility pays to each supplier a different price, the negotiated price includes a capacity payment (to cover fixed costs) and an energy payment (covering fuel costs), or both payments are included in an all-in price. Each supplier negotiates a price
that covers all its costs, including a return on investment, without any free lunch for the utility.

A pay-as-bid auction will also suppress market price signals, for several reasons. First, it discourages consumers from using electricity wisely and investing in energy efficiency. While a single price auction sets prices at the level needed to balance supply and demand, with a pay-as-bid auction, consumers do not see the true cost of generating the last kilowatt-hour demanded. Second, the uncertainty that pay-as-bid auctions create also leads to higher overall generation costs, because plants won’t always be operated based on their generating costs. Finally, abuse of market power and collusion is more difficult to detect under the pay-as-bid mechanism, because bid prices are less transparent (generators are not expected to bid their variable operating costs) than under the single market-clearing price auction.20

At best, pay-as-bid auctions would bring no noticeable difference in energy costs. At worst, economic dispatch would be compromised when market prices do not turn out as plant operators expect and some low-cost units are outbid by higher-cost units.

Some critics of the single clearing price auction have acknowledged that pay-as-bid is not an improvement and have simply argued for a return to cost-of-service or an average pricing approach. While it is easy to forget that the “good old days” where not as good as some remember, it simply isn’t possible, in any electric system, to provide all customers, all the time, with power only at baseload prices. In any regulatory model, the debate over whether residential customers should subsidize industrial users, or the other way around,

---

20 Transparent market prices are those that clearly reflect market demand and supply conditions. If suppliers bid their marginal costs, then knowing what demand is in a particular hour and the marginal costs of all available suppliers, allows the market price to be estimated accurately. That is a transparent market price.
consumes endless time and resources, since getting it “just right” is virtually impossible. At the end of the day, someone has to pay for the intermediate and peaking resources when those are need to meet system demand and the price signals provided by a single clearing price auction do the best possible job of reflecting the actual cost of meeting the last increment of demand on the system.

**Does competition work in the electric market?**

In some ways, electricity is different from other commodities. It cannot be stored cost-effectively and it is crucial to all of our daily lives. But it is not so different that basic market principles no longer apply. Wholesale competition was introduced by Congress in 1978 and solidified in 1992. It achieved momentum after FERC issued its Open Access Policy in 1996. Since then, the restructuring of electricity markets has provided significant benefits to electricity customers. In a market economy, the main economic rationale for applying traditional cost-of-service regulation to any industry is in the case of a “natural monopoly,” in which a good or service is provided most efficiently by a single firm. This characterization may apply to certain aspects of electricity transmission and distribution, but certainly does not apply to electricity generation, where technological improvements have made it possible for relatively small suppliers to achieve efficient scale economies. This doesn’t mean there is no need for regulation whatsoever. All competitive markets are regulated to some extent, whether to ensure the food we eat is safe, the planes we fly in have trained pilots, and so forth (and in each of these markets one could argue that the commodity is unique). In electricity markets, FERC determines who can charge market prices and sets transmission access rules for generators and utilities. In addition, the
behavior of generators and other market participants is closely monitored by system
operators.

The rationale for competition in electricity is being attacked today because of recent
headline increases in retail electric rates. The rise in prices reflects substantial increases in
underlying fuel costs over the past several years, the costs of complying with regulatory
mandates, and the impact on consumers has been magnified by the expiration of artificially
capped rates in place for years in some jurisdictions.

Our analyses, and others, have shown significant benefits of wholesale electric
competition. Nuclear plant operations have improved dramatically, market prices have
fallen since the centralized market was created, when fuel prices, supply balance and inflation
are accounted for, and, in states where retail price caps have expired, the mechanisms for
drawing through the benefits of wholesale competition to customers appear to be highly
efficient.

Impact of Regulatory Restructuring on U.S. Electric Generation Efficiency,” Massachusetts Institute
of Technology, Center for Energy and Environmental Policy Research, UC Energy Institute, CSEM
http://faculty.haas.berkeley.edu/wolfram/Papers/frw.041306.pdf; Bushnell, J., and C. Wolfram,
“Ownership Change, Incentives and Plant Efficiency: The Divestiture of U.S. Electric Generation
Plants,” California Energy Institute, Center for the Study of Energy Markets,CSEM-WP-140, March
2005 (“Bushnell and Wolfram”). Available at: http://www.ucei.berkeley.edu/. See also, C. Cain and
Application to the PJM Region,” Center for the Advancement of Energy Markets, September 2003;
“Electricity Prices in PJM: A Comparison of Wholesale Power Costs in the PJM Market to Indexed
Power Markets to the Test – The Benefits of Competition in America’s Electrical Grid: Cost Savings
4. CONCLUSIONS

There is no doubt that electric markets are complex. But electric markets are not so different that basic market principles of competition no longer apply. In fact, wholesale electric competition is working well and has created significant value for customers in Maryland, Pennsylvania and elsewhere. Retail rates in the initial PJM states have increased less than in the Southern states that have not restructured their electricity markets. Excluding the effects of fuel costs, wholesale electricity prices have actually decreased significantly since the inception of the PJM wholesale market. Competition has also motivated power suppliers to build and operate power plants more efficiently.

Power pools like PJM provide a number of benefits, relative to the traditional isolated utility model. Power pools greatly reduce the risk of black outs, ensure that the transmission system operates safely, and dispatch electricity from the cheapest sources available in the pool. LMPs further induce consumers to use energy efficiently and signal potential investors where energy is needed the most. Persistently high LMPs in load pockets reflect the reality that congestion tends to occur in highly populated areas, just the places where it is difficult and expensive to build new generation and transmission.

The proposed pay-as-bid auctions will not result in any consumer savings, because generators would have to bid higher to cover their fixed costs. They also have other undesired consequences. They discourage consumers from using electricity wisely and investing in energy efficiency, they create uncertainty that raises generation costs, and they make it more difficult to detect abuse of market power.
APPENDIX – PJM

PJM, the acronym for the Pennsylvania – New Jersey – Maryland power pool, is the largest regional power pool in the United States. Although PJM originally covered just portions of these three states, it has since grown to encompass Delaware, Virginia, West Virginia and the District of Columbia, as well as parts of Illinois, Indiana, Michigan, Ohio, and North Carolina, as shown in Figure 4.

Figure 4: The PJM territory

Source: PJM Interconnection

Before the 1960s, PJM was the only integrated power pool in the United States, and there was limited coordination among the various power systems in North America. That changed with the huge blackout that affected the east coast in 1965. In response, coordinating councils were established by region to promote power system reliability, and were subsequently linked under the North American Electric Reliability Council (NERC) in
1968. With the passage of the Electric Policy Act of 2005, NERC became the Electric Reliability Organization (ERO) and was tasked with developing and enforcing specific reliability standards. Those reliability standards are reviewed by the Federal Energy Regulatory Commission (FERC), which also oversees the competitiveness of all wholesale electric markets. The regional coordinating councils monitor and develop guidelines for transmission system operations that ensure the security and reliability of the system. A map of the current regional councils is shown in Figure 5.\(^{23}\)

\(^{22}\) See, Docket No. RM06-16-000, Mandatory Reliability Standards for the Bulk-Power System, Notice of Proposed Rulemaking, 117 FERC ¶61,084, October 20, 2006.

\(^{23}\) Note that four of the eight regions incorporate parts of Canada.
The 1965 blackout also spurred formation of other integrated power pools. The New York Power Pool (NYPP) was established in 1966, and the New England Power Pool (NEPOOL) was created in 1971. While the coordinating councils establish rules to ensure the integrity and reliability of the bulk transmission system, the power pools provide the day-to-day operation and coordination of the electric system. For example, PJM coordinates the operations of every generating plant within its borders to minimize overall generating costs and ensure that transmission facilities are not overloaded.
With restructuring of the industry, FERC separated transmission system ownership and transmission system operation. FERC did this to ensure competitive electric markets, in much the same way that it required natural gas pipeline owners to divest their ownership of natural gas supplies. As a result, power pools like PJM have come to be called “regional transmission organizations” (RTOs), whose mission is to operate the high voltage transmission system safely and economically, and ensure that all generators have equal access to the transmission system to sell electricity.

**Figure 6: Regional Transmission Organizations**

Source: FERC