Market Organization and Efficiency in Electricity Markets

Erin T. Mansur and Matthew W. White *

January 13, 2012

Abstract

Electricity markets exhibit two forms of organization: decentralized bilateral trading and centralized auction markets. Using detailed data on prices and quantities, we examine how market outcomes changed when a large region in the Eastern US rapidly switched from a bilateral system of trade to an auction market design in 2004. Although economic theory yields ambiguous predictions, the empirical evidence indicates that employing an organized market design substantially improved overall market efficiency, and that these efficiency gains far exceeded implementation costs. Our analysis suggests these gains arise from superior information aggregation about congestion externalities, enabling the organized market to support greater trade.

* *Mansur:* Associate Professor, Dept. of Economics, Dartmouth College and NBER. Email: erin.mansur@dartmouth.edu. *White:* Senior Economist, ISO New England. Email: matthew.white.phd@gmail.com. The authors gratefully acknowledge the University of California Energy Institute for facilitating access to proprietary transaction information, and to the PJM Interconnection, LLC., for network data. White thanks the University of Pennsylvania's Wharton School for its hospitality and support. We thank Severin Borenstein, William Hogan, John Kwoka, Michael Riordan, Frank Wolak, and Catherine Wolfram for valuable discussions and comments on earlier drafts. The opinions and conclusions expressed in this paper are solely those of the authors and should not be construed as representing the opinions and policies of the ISO New England.

1 Introduction

Since the seminal experiments of Smith (1962, 1964), economists have recognized that the process used to match buyers with sellers in a market can have substantial consequences for its efficiency. In recent years this line of inquiry has turned increasingly normative as a new field of market design emerged (Roth, 2002). It seeks to identify specific market rules and procedures that can speed information revelation, discover efficient prices, and improve market performance.

This paper examines how market organization affects performance, efficiency, and prices in competitive electricity markets. In many regions of the United States, wholesale electricity markets operate as a decentralized, bilateral trading system. In other regions, trade is mediated through centralized market designs. These markets aggregate offers to buy and sell, determine market-clearing prices, and handle settlements for several complementary services involved in power production and delivery.

The merits of these two forms of market organization have provoked significant debate. In 2001, the Federal Energy Regulatory Commission (FERC) initiated a formal proceeding to identify 'best practices' in market designs for electricity, and sought to promote their use in regions where bilateral trading practices prevail. The FERC's policy initiative has been vigorously challenged by market participants who expect to lose under a different trading system. Their central argument is that the benefit of expanding organized market designs into new regions remains speculative, and may not be worth the cost of implementation.

We address this question by examining how market outcomes changed when an organized electricity market in the Eastern US, known as the PJM Interconnection, expanded to serve a large region of the Midwest. Before the change, firms in the Midwest engaged in considerable trade with counterparts to the east, but all transactions were made bilaterally. After the change, buyers and sellers in both regions could be (anonymously) matched to one another through a central auction-based market. The empirical evidence indicates that total inter-regional trade promptly tripled, suggesting the auction-based market identified gains from trade that were not achieved with bilateral trading arrangements.

The evidence is appealing in its transparency. The organized market's expansion was implemented on a single day—creating a sharp 'before versus after' demarcation between two forms of inter-regional trade. There were no concurrent changes in the number—or even the identities—of the firms in these markets. Nor were there any changes in their technologies, nor the physical infrastructure of the (transmission) network that enables them to trade. Instead, what changed was the organization of trade: how firms' willingness to buy and sell was elicited and used to determine production and prices.

Although considerable economic theory guides electricity market design, persuasive arguments regarding bilaterally-based trade have proven elusive. The efficiency of unstructured bilateral markets depends how buyers and sellers are matched to one another. Without strong assumptions about how this matching occurs and how market participants' information sets form, economic theory (and laboratory experiments) admit a wide range of outcomes. In real markets, participants' information sets are difficult for researchers to observe and characterize, making the relative efficiency of decentralized versus organized markets difficult to establish. Thus it seems useful to assess empirically whether adopting an organized market design improves market efficiency—and to compare these gains with the cost of implementing it.

This objective differs from much of the burgeoning empirical literature on electricity markets. Previous contributions have focused, more or less exclusively, on whether organized electricity markets operate efficiently relative to a 'perfectly competitive' market benchmark (Mansur 2007, Borenstein Bushnell and Wolak 2002, Wolfram 1999). In contrast, our objective is to provide evidence on whether adopting an organized market design improves economic efficiency relative to the widely-used bilateral trading system. This compares the two workable market arrangements we observe in use, and seems the salient comparison to inform economic policy decisions.

To perform this comparison, we examine market-level data on demand, prices, and inputs costs. This detailed information enables us to evaluate the gains from trade under each form of market organization, and to measure how market performance changed along several dimensions. The efficiency gains arise from supply-side allocative efficiency improvements: increased trade reallocates production from higher-cost plants to lower-cost plants. We find that adopting the organized market design in this region produced efficiency gains of over $160 million annually, substantially exceeding the (one-time) $40 million implementation cost.

The next two sections explain the basic theoretical rationale for organized electricity market designs, a problem of incomplete information and network externalities. Section 3 summarizes our empirical strategy, and Sections 4 through 6 present empirical findings. A discussion of market implications follows the main findings.

2 Network Externalities and Information

A considerable body of economic theory informs electricity market design, and suggests why decentralized bilateral trading may yield different outcomes than centralized auction markets.[1] We articulate the main issues here, as these theoretical arguments play a central role in explaining our empirical results.

2.1 Background

The principal actors in electricity markets are producers (who own power plants) and retailers (local distribution utilities). Many, but not all, are vertically integrated. Retailers tend to build, or procure under long-term contract, sufficient capacity to serve their customers' annual peak demand. On a daily basis, this practice yields considerable excess capacity market-wide. That creates an opportunity for producers to trade among themselves, idling high-cost plants when other firms are able to deliver the same quantity at

[1] See Wilson (2002) for a survey. Schweppe *et al* (1988) spurred considerable research on electricity market design.

lower cost. These 'spot' wholesale markets can create significant gains from trade because producers have heterogeneous technologies, capital vintages, and factor prices.

An essential feature of electricity markets is that trade takes place over a network. Production can therefore create externalities, due to network congestion. The externalities are difficult for market participants to resolve—in Coasian fashion—because of incomplete information. Specifically, market participants have too little information about each others' production decisions to identify whose actions will alleviate congestion.

The central question is what form of market organization can aggregate enough information resolve these congestion externalities. We first explain precisely how these externalities arise, and why incomplete information hampers trade.

2.2 The Complementarity Problem

Most of the challenges to achieving efficient trade in electricity can be traced to a complementary goods problem. This complementarity arises when two production facilities are separated by a congested portion of the delivery network. In such circumstances, additional trade does not necessarily exacerbate the network's congestion—instead, it can often alleviate it. This makes it desirable to pair transactions that alleviate congestion with transactions that (otherwise) would create it. By doing so, market participants can 'internalize' congestion externalities and improve overall efficiency.

The efficiency gains from identifying complementary trades can be considerable, even in remarkably simple networks. An example illustrates the point. Consider a triangular network with three firms at the vertices, A, B, and C, as shown in Figure 1-a. The three firms have (constant) marginal valuations of $v_A = \$5$, $v_B = \$15$, and $v_C = \$16$ per unit. For concreteness, think of A, B, and C as the locations of three (separately-owned) plants, and the valuations as their marginal costs of production. All three network links are assumed to be identical, except that one has lower capacity (of 100) than the other two (500 apiece). For simplicity, we shall assume that all energy flows are real and ignore losses.[2]

Suppose a mass of consumers is currently served by firm B at that location, making B a potential buyer from A at the wholesale level. Imagine they agree to trade 900 units at a price (say) of \$12 per unit. Can this be implemented? Not by the two firms alone. Suppose A first increases production by 300 units, and B reduces production by 300 units. Because energy follows the path of least resistance, part of A's output will reach B's consumers on the direct $A \rightarrow B$ link and part will flow around the network $A \rightarrow C \rightarrow B$. In our triangular network, the latter path is twice as long as the former so the flows will spit in a 2:1 ratio. That is, 200 units flow directly from A to B, and the remaining 100 units reach B along the path through C.

At this point, their transaction hits a constraint: The link between A and C, which has a capacity of 100, is fully utilized. Like roadways and other networks, energy networks experience congestion in the sense that once a link reaches capacity no greater flow (in that direction) is possible. Unlike other networks, however, power cannot be re-routed around a congested path; once a link is congested, no greater production from the same source is feasible. Firm A and B's transaction is limited to 300 units, with a gain from trade of $300 \times (\$15 - \$5) = \$3000$.

This is the best they can do bilaterally, but it is not allocatively efficient. Suppose B now buys 300 units from C, in addition to the 300 units from A, withdrawing all 600 at B. Without considering network effects this seems plainly inefficient: C is producing for \$16 a product that B values at only \$15. There is a reason for B to pay for C's output, however: It alleviates congestion on the $A \rightarrow C$ link. In our example, A's 300 units flow directly to B on the $A \rightarrow B$ link and C's 300 units flow on the $C \rightarrow B$ link (see Figure 1-b). The flow between A and C becomes zero.[3]

[2]Incorporating complex (apparent) power and losses would not alter the economic insights here.

[3]Fixing output at 300 each, any flow of $\delta > 0$ from A to C would also increase the $C \rightarrow B$ flow to $300 + \delta$ and decrease the $A \rightarrow B$ flow to $300 - \delta$. If so, the flows would shift to follow the path of lower resistance, or $A \rightarrow B$, bringing δ to zero.

Firm C's production has an externality benefit for A, alleviating the constraint that previously limited A's production. To exploit it, suppose A now increases its production to 600 and B withdraws all 900 as originally proposed. Firm A's *incremental* production of 300 will again split in a 2:1 ratio between the $A \to B$ path and the $A \to C \to B$ path; see Figure 1-c. On the margin, it is efficient for B to pair the larger trade with A and the seemingly costly transaction with C: Each additional purchase from C (at a loss of \$1 per unit) enables B to acquire an additional unit from A (at a gain of \$10 between them). The total value of trade is now $\$10 \times 600 - \$1 \times 300 = \$5700$.

The main point to observe is that C's production benefits A. This is a classical congestion externality in the sense that the output C delivers to B affects the maximal output A can deliver to B. Unlike congestion in other settings, however, the externality is positive: increasing production at the receiving end of a congested link reduces the flow across it. In our example, reallocating production to exploit this positive externality increases total welfare by 90%, from \$3000 to \$5700.

Note further that the total volume of trade triples, from 300 to 900, when production is allocated efficiently. This has an important empirical prediction: If a decentralized trading system is not able to identify and implement all complementary transactions, the volume of trade may be sharply attenuated.

2.3 Informational Impediments to Trade

The basic difficulty these complementarities present is that single bilateral trades may be infeasible, but *sets* of bilateral trades may be simultaneously feasible and efficient. For instance, in the preceding example one trade is infeasible alone (600 units from $A \to B$), the other is inefficient alone (300 units from $C \to B$), but the combination of the two is both feasible and Pareto efficient.

It might seem simple enough to identify these complementarities if market participants know the structure of the network. This simplicity is a deceptive consequence of a three-node example. In a 'mesh' network—one with multiple links at each node—a trade between any two locations may create or alleviate congestion on (essentially) any link in the network. To evaluate whether or not this occurs, a firm must also know the quantities that all other firms are buying and selling at *every* location in the network. The concern is that without a formal mechanism that aggregates and reveals this information, the market may generate too little information for firms to determine which transactions are complements. If so, the market exhibits too little trade.

To explain precisely how this arises, we need a more general characterization of the network complementarity problem. This requires a bit of graph theory.

Consider a network represented by its graph: (V, E, K) where V is an (enumerated) list of network vertices, or nodes; E the set of edges, or links; and K the links' capacities. The ith element in E, denoted E_i, is a pair of connected nodes (u, v), $u < v$. It is useful to represent the network structure (V, E) by its *link matrix*, L, that indicates which links (rows) connect to a which nodes (columns). If i indexes links and j indexes nodes, the link matrix has (i, j)th element

$$l_{ij} = \begin{cases} 1 & \text{if } \; E_{i,1} = j \\ -1 & \text{if } \; E_{i,2} = j \\ 0 & \text{if } \; \text{otherwise}, \end{cases}$$

where $E_{i,1}, E_{i,2}$ are link i's first and second nodes, respectively. Signs merely preserve E's node-pair order.

Let q be an allocation: An n-vector of quantities at the nodes (positive for injections, negative for withdrawals), such that $\sum_j q_j = 0$ (aggregate supply equals demand). Let f be the link flows (positive for $u \to v$ flows, negative for $v \to u$). The central relation between allocations and flows is that the net flow at each node must sum to zero.[4] In matrix form, this implies

$$q = L'f$$

[4]This is Kirchhoff's law of conservation of charge; see, e.g., Howatson (1996).

as may be verified by expansion. Although this contains n equations, one is redundant; dropping the last (arbitrarily) yields an $n-1$ equation system we denote $\underline{q} = \underline{L}'f$. Solve for f:

$$f = \underline{L}(\underline{L}'\underline{L})^{-1}\underline{q} \equiv T\underline{q}.$$

T is known as the *transfer matrix.* The difference between any two columns j' and j in T indicates how a trade of one unit from node j' to node j changes the flow on each network link.[5]

Although the link matrix is sparse, the transfer matrix is not. Therein lies the basis for the informational difficulties confronting trade: energy flows across the network, in proportions given by T, across all possible network paths between j' and j. Specifically, consider a transaction of $\Delta q'$ units from a seller at j' to a buyer at j. This creates an incremental flow on link i of $\Delta q'(t_{ij'} - t_{ij})$. It results in a total flow on link i of

$$f_i' = \Delta q'(t_{ij'} - t_{ij}) + T_i\,\underline{q}, \tag{1}$$

for T_i the ith row of T. We say the transaction *congests link* i if the capacity constraint of the ith network link binds, or $|f_i'| = K_i$.

The central observation here is that the entire market allocation $\underline{q}$ enters (1). This means that, for a firm to evaluate whether a candidate transaction will create or alleviate congestion, it needs to know the quantities that every other firm in the network is buying or selling at *their* network locations.

The property of creating or alleviating congestion is central to whether two (or more) transactions are substitutes or complements. To characterize when transactions are complements, let us introduce a second bilateral transaction that involves $\Delta q''$ units from a new seller j'' to buyer j. Two transactions are *potential complements* if they are (i) jointly feasible but (ii) at least one transaction is infeasible individually. (If each transaction is

[5] We simplify: If links have unequal length (or impedance, generally), the same interpretation of T will apply; however, its form generalizes to $T \cong \underline{L}\Omega^{-1}(\underline{L}'^{-1}\underline{L})^{-1}$, where diagonal matrix Ω indicates the relative impedance of each network link. For clarity, we assume $\Omega = I$ (w.l.o.g.).

individually feasible, then they are mutual substitutes.) The first condition requires that

$$|f_i' + f_i''| \leq K_i \quad \forall\, i \tag{2}$$

and the second requires

$$\max\{\, |f_i'|, |f_i''|\,\} > K_i \quad \exists\, i \tag{3}$$

The final requirement of complementary transactions is that they create gains from trade. If v_j denotes the valuation of buyer j, and similarly for sellers j', j'', this requires:

$$\Delta q'(v_j - v_{j'}) + \Delta q''(v_j - v_{j''}) > 0. \tag{4}$$

The two transactions are complementary if they satisfy (2), (3), and (4). The extension to complementarities among sets of more than two bilateral transactions is straightforward.

It is useful to be clear about how these features affect the gains from trade. Congestion externalities arise because the feasibility of an individual bilateral transaction depends on production at all network locations, $\underline{q}$, which enters (1). Complementarities arise because flows in opposing directions alleviate congestion (rows of the transfer matrix T have positive and negative elements). Determining whether transactions are potential complements—that is, satisfy (2) and (3)—therefore requires knowledge of the network structure T and the market's current allocation, $\underline{q}$.

In a fully decentralized market, the nodal-level production and trading decisions of a market participant is its own private information. That makes it difficult for markets to exploit the benefits of these complementarities: Unless a firm can observe others' private information, it cannot determine the current allocation $\underline{q}$; and without $\underline{q}$, it cannot determine which transactions are complements.

In a broad sense, the difficulty here is that every market participant has 'small' bits of information—its valuation and quantity—but identifying complementary trades requires information in the *union* of their private information sets. The vexing economic question is how to structure market institutions so as to elicit this information and identify complementary transactions. This requires a brief discussion of market institutions.

In practice, bilateral electricity markets resolve this problem with an institutional arrangement known as a *transmission reservation system*. In brief, this system requires firms using the network to communicate (privately) to a system administrator a candidate bilateral transaction's quantity information (e.g., $\Delta q'$ units from $j' \to j$). The reservation system then uses the network graph and (1) to check whether the candidate transaction is feasible, given the previous transactions of all market participants. It privately reports back to the two parties whether it is feasible or not, and if yes, the parties confirm their transaction. The system appropriately updates the market allocation q, and the process continues.

This institutional arrangement effectively creates a 'first-come, first-serve' entitlement to the network's capacity. It leaves the problem of determining an efficient allocation of network capacity to the market, through re-trading among network users. This process of re-trading to reallocate scarce network capacity when the network is congested is where identifying complementary trades becomes essential to market efficiency.

The process of identifying complementary transactions in this environment suggests why this institutional arrangement may achieve less than full efficiency. Consider first the steps involved if we assume—counter to fact—that an individual firm j publicly observes the current allocation at all locations, q. In that case, it could identify the set of congested network links that render a candidate bilateral trade of $\Delta q'$ units from $j' \to j$ infeasible. It then needs to identify a change in the allocation that alleviates congestion on each such link, while not creating congestion along any other network path between j' and j. This requires identifying a perturbation of the allocation vector, $\Delta \in \mathbb{R}^n$, such that $\sum_{k=1}^{n} \Delta_k = 0$ and (omitting the last value of Δ in $\underline{\Delta}$ for conformity),

$$| \, \Delta q'(t_{ij'} - t_{ij}) + T_i \, (\underline{q} + \underline{\Delta}) \, | \leq K_i \quad \forall \, i. \tag{5}$$

If such a Δ exists, its non-zero elements indicate a set of potentially complementary trades with the bilateral transaction of $\Delta q'$ units from $j \to j'$. In a large network there may be many non-zero elements in any vector Δ that satisfies (5), so a set of complementary transactions may require multiple bilateral trades to implement it.

Yet without knowledge of q, a firm cannot by itself evaluate (5. Instead, the only way j can ascertain whether (5) holds is a two-step process: (i) enumerate a set of bilateral trades that implement a trial allocation perturbation Δ, and then (ii) submit this set of candidate bilateral transactions to the transmission reservation system. For firm j, this is an iterative, trial-and-error process: the reservation system provides no price signals—that is, no gradient information—that would enable j to determine in which direction to move Δ to find the set of complementary transactions that are defined by (5). Unlike finding a value of Δ that solves (5) when q is known, which effectively amounts to solving a simultaneous system of linear equations, finding a value of Δ when q is unknown to firm j is an extraordinary task.

Discussions with market participants point to a related problem that makes them hesitate to pursue the complementary trades they can identify (these are termed *redispatch* arrangements). The set of complementary transactions that satisfy (5) depends on the current market allocation, $\underline{q}$. Some of these transactions may be feasible individually, but have negative value alone. (For example, the transaction between firms B and C in Figure 1 has this property.) Because of this, sequentially arranging each transaction in a complementary set creates a problem of *execution risk.* This risk is that if the market moves (that is, q changes) while firm j is partway through the process of executing binding bilateral transactions with different counterparties, the complementary set may suddenly become infeasible. In that event, the value of the contracts that were executed first might not be zero—it may be negative.

The mere possibility of this event creates a disincentive to execute complementary transactions in which one or more component trades have negative stand-alone values. Thus, even if the market does not move adversely to render the set infeasible, known complementary trades might not be undertaken. This risk reduces trade and can lead a market not to implement transactions that are, in fact, efficient.[6]

[6]A closely-related exposure problem arises in auctions for complementary goods (a lucid treatment is

Fundamentally, there are two senses in which achieving market efficiency with congestion externalities can be regarded as an informational problem. As noted earlier, if there existed a market device that disseminated the prices and quantities generated by a process of sequential bilateral trades, then market participants could more readily search for efficient re-trading opportunities by solving (5). However, since no trader knows the aggregate market position (that is, q) in a bilateral market, there is no basis to expect that all complementarities will be realized. Distributing this information provides one possible avenue for an alternative market design that might speed the process of discovering efficient allocations.

The second sense in which it is informational is that a different way of aggregating market participants' private information provides a simpler way to solve this problem. Instead of accreting price and quantity information revealed through sequentially-arranged transactions, a market mechanism might elicit willingness to buy and sell offers from all participants simultaneously. The virtue of simultaneously-arranged transactions is that the trial-and-error process of finding complementary trades disappears, as does the markets need to evaluate $n!$ (worst case) potentially complementary transactions sequentially.

Imagine, for the moment, a market mechanism that induced all participants to reveal simultaneously their true valuations. For an auctioneer, the problem of determining an efficient allocation of production becomes an optimization problem subject to the network's feasibility constraints, $|T\underline{q}| \leq K$. The question of whether a market organized in this fashion will discover an efficient allocation reduces to whether there are enough market participants—and enough substitutability among them across locations—so that the prices at which firms actually offer to buy and sell in a simultaneous auction are driven to their true valuations. This logic, and procedure for (implicitly) matching buyers and sellers, lies at the core of nearly all organized market designs for electricity.[7]

Milgrom, 2004).

[7] An exception is the New Electricity Trading Arrangement (NETA) in the U.K., which uses a hybrid bilateral-and-centralized allocation system. See Green (200x).

2.4 The Market Design Controversy

In reality, what matters most is not whether one form of market organization or the other achieves a theoretically ideal market outcome, but whether the difference between them is economically significant. Industry participants in regions of the U.S. where bilateral trading prevails commonly argue that the cost of adopting (or joining an existing) organized market would exceed the benefit. This points to a central trade-off in market design: An organized market design might reduce inefficiencies that exist in an unstructured, decentralized market, allowing participants to realize gains from trade that would not otherwise be achieved. However, organized markets are costly to design and implement (particularly so for electricity). Thus, the value of shifting the venue of trade out of a decentralized bilateral system and into an organized market is ultimately an empirical matter.

This trade-off has emerged as a controversial policy issue recently, for two reasons. First, the industry's principal regulator (the FERC) retains an obligation to evaluate and approve changes in electricity market designs—a task not taken lightly in the wake of California's disastrous experience with an ill-designed market. Second, policy makers' goal of encouraging more efficient markets is not always aligned with the private incentives of market participants. A producer may have a strong private incentive to object to a new market design if it will result in a more competitive marketplace with lower prices; and a buyer that relies upon a constrained network path for delivery may not relish the prospect of increasing competition for this scarce resource. The practical consequence of these fundamental incentive problems is that modern regulatory policy makers face a panoply of conflicting claims about the costs and benefits of organized market designs.

3 The PJM Market Expansion

We bring new information to this problem by examining how market outcomes changed after an existing, organized electricity market expanded to serve a region where an (exclusively) bilateral trading system prevailed. The organized market is known as the PJM Interconnection (PJM). PJM is a non-profit, mutual-benefit corporation that operates several inter-related wholesale markets for electricity (energy), its delivery, and a variety of ancillary services. The five-hundred members of PJM comprise producers that own power plants, local utilities that buy electricity to distribute to homes and businesses, and third-party traders (financial institutions and commodities brokers) that participate in PJM's forward markets. PJM presently operates spot and forward markets for electricity production and delivery at thousands of delivery points from the East coast to Illinois. The nominal value of all transactions on PJM's spot and forward markets annually is approximately $22 billion (PJM, 2005b).

In contrast, utilities and power producers throughout most other regions of the United States engage in wholesale electricity trading through bilaterally-negotiated transactions.[8] Following several years of planning and regulatory approvals, in October 2004 nineteen Midwest-based firms that previously traded exclusively through bilateral market arrangements became members of PJM. Seven of these new members are affiliated subsidiaries of the American Electric Power Company (AEP), a holding company that, until joining PJM, was one of the largest participants in regional bilateral markets in the Midwest.

The decision of the new members to join PJM originates in an (unrelated) merger settlement with federal authorities half a decade earlier.[9] Whether that decision reflects forward-looking behavior by these new members about the value of participating in the organized market is an interesting question, and one that affects how we will interpret the results. It does not, however, alter our ability to identify whether PJM's expansion improved market efficiency overall. We discuss this issue next.

[8] As of 2004, the exceptions are the organized regional electricity markets in California, Texas, New York, and the New England states.

[9] *C.f.* 89 FERC ¶63,007 (1999) and 90 FERC ¶61,242 (2000).

3.1 Inference

In our setting, identifying how the organized market's expansion affected market efficiency entails two related, but conceptually distinct issues. The first issue is the question of cause and effect: Whether, and why, we may be confident that any changes in market outcomes we measure are attributable to the markets' expansion, and would not have occurred otherwise. This stems from the timing and nature of the changes we study.

The second issue is how we use the market outcomes we observe—prices and quantities, primarily—before and after the market's expansion to infer changes in market efficiency. This we describe next.

3.1.1 Efficiency

Although electricity markets can be complex, a simple analogy will clarify the main ideas. This analogy highlights the essential features we exploit to identify market efficiency changes using observable outcomes.

Imagine a market with many participants who have heterogeneous, privately-known valuations. Participants trade with one another bilaterally, at prices determined in private negotiations. Suppose further that some of the market's participants are also members of an exchange, or clearinghouse, that matches offers to buy or sell among its members in an organized fashion. Exchange membership is open to any participant who pays a (fixed) membership fee. The exchange members are free to transact with non-members, but must do so outside the exchange in the bilateral market.

To complete the analogy, now suppose that a subset of the bilateral-market participants joins the organized exchange. Following our earlier terminology, we will refer to the two transaction venues in this analogy as the bilateral market and the organized (exchange-based) market.

At one level, the logic underlying our empirical strategy is straightforward. In this simple analogy—and in reality—any pair of market participants has the option to transact bilaterally outside the organized market. But the exchange has a membership cost. Thus, if we observe an increase in the quantities transacted by the new members after they join the organized market (*ceteris paribus*), we conclude that the new market participants realized gains from trade that they could not capture by transacting in the bilateral market.

This logic carries over to inference about market efficiency on the basis of price changes, although the argument is slightly more involved. In the absence of any trading frictions in the bilateral market—where all trade between exchange members and non-members must take place—arbitrage implies bilateral and exchange-based transactions should occur at the same price. Empirically this turns out not to be the case, so an alternative hypothesis about trading frictions is needed. Suppose now that contractual incompleteness, search costs, or some other trading imperfection exists in the bilateral market. In this case we expect a non-zero price spread between the bilateral and organized markets, and an incentive for some market participants to join the exchange.

We will draw conclusions about relative efficiency of the bilateral and organized markets not from the *fact* that some market participants joined the organized market, *per se.* Instead, we examine how prices and quantities *change* after they joined it. The logic for why the price spread between the two markets may shrink (in magnitude) after some firms join the organized market is that it shifts the distribution of valuations among each markets' participants. For example, suppose (without loss of generality) that prices in the bilateral market are lower than in the organized market. Then low-cost sellers have an incentive to join the exchange, withdrawing (or raising the offered price for) supply in the bilateral transaction market and expanding aggregate supply in the organized market. Such a shift narrows the price spread between the bilateral and the organized market, increasing the volume of trade overall.

In sum, after the new members join the organized market, efficiency-enhancing reallocations from low- to high-value market participants will reduce the (magnitude of the) price spread between bilateral- and exchange-based transactions. Thus, the first component of our empirical strategy will be to evaluate whether price spreads converged significantly after the organized market's expansion.

4 Evidence: Price Convergence

We now examine whether prices converged for similar transactions arranged in the bilateral market and in the organized market (PJM). Because the details of how prices are measured are important to our purposes, we first summarize the transactions they represent.

4.1 Price Data

To examine whether between-market arbitrage improved, we assembled detailed market price data at daily frequency covering a three-year span. There are two data sources for transaction prices in bilateral electricity markets: the Platt's daily price survey and the electronic 'over the counter' trading system operated by the Intercontinental Exchange, Inc. We have examined daily transaction data from both sources, and daily price indices for delivery points of interest are (essentially) identical. In the results below we have used the Platt's data due to its slightly broader coverage, unless indicated otherwise. The prices determined by PJM are public information (pjm.com).

Because electricity must be produced at precisely the moment it is used by consumers, trading in wholesale electricity markets is conducted on a forward basis. Our analysis centers on prices in the day-ahead forward markets. Day-ahead forwards are the highest-volume markets for wholesale electricity transactions, in both the bilateral and the organized market.

The bilateral market and exchange-based (PJM) day-ahead forward prices we compare represent identical commodities, up to delivery points. Each indicates the price for delivery of the same quantity of power, at the specified delivery location, for a pre-specified duration the following day. In bilateral markets, two standard contracts are traded: Peak and off-peak, in 50 megawatt units, for next-day delivery continuously from 6 AM to 10 PM or 10 PM to 6 AM. On PJM, separate prices are set for each hour of next-day delivery; we construct the equivalent prices for the industry-standard peak and off-peak delivery intervals, thereby matching exactly the delivery schedules for the contracts traded in the bilateral market.

These contracts differ in one respect: PJM's day-ahead markets use different delivery (pricing) points than bilateral market forward contracts. This will affect our analysis and interpretation, as discussed below. In terms of the data, we selected a set of delivery points in the mid-Atlantic and Midwestern states that are most likely to reveal any changes in market outcomes that result from PJM's expansion into the Midwest.[10] These delivery points are selected based on three criteria: (1) Proximity of each delivery point to one another (where proximity is with respect to structure of electric transmission network); (2) commonly-used delivery points, to ensure liquidity; and (3) for which complete location-specific day-ahead market price data exist. There are five delivery points that meet these criteria. Rather than select among them, we will report results for all five points and the price spreads between them. All of our results and their interpretations turn out to be highly robust to the choice of which delivery points to compare between PJM and the Midwestern bilateral markets, as will become clear presently.

There is a second, minor difference in the pricing of day-ahead forward contracts due to the timing of each market's close. Bids in the PJM forward market are due by noon the day prior to delivery, at which point the day-ahead market closes. Prices are posted by the market by 4 PM. Bilateral market price data include trades arranged up to close of the business day. Thus the information set of traders in bilateral markets is a superset of that incorporated into the organized market's day-ahead prices. Nonetheless, there is

[10] In this respect our analysis is a partial, rather than general, equilibrium analysis of the expansion's impacts. We have not included analysis of additional delivery point prices here primarily to reduce the volume of our analysis. More distant pricing points might also be affected by the expansion, ostensibly by lesser amounts.

no reason why any additional information would bias bilateral market forward prices one way or another, relative to PJM's forward prices.

4.2 Changes in Price Spreads

Tables 1 and 2 summarize the price levels and price spreads between the bilateral market and the organized market before and after the market's expansion on October 1, 2004. Panel A in each table presents average daily forward prices, by market type and delivery point, for six-month periods before and after expansion. Panel B summarizes the changes in price spreads between contrasting delivery point pairs.

The first numerical column in each table reveals that average prices differ at each delivery point, an empirical regularity in electricity markets generally. The standard explanation for these price differences is that they reflect occasional congestion on the transmission network used for delivery. That is, when the difference in prices between any two points creates excess demand for delivery (transmission capacity) from one point to the other, the market may not be able to close the price spread completely. In an efficient market, the price spread would be zero between any two delivery points when there is excess capacity and non-zero when there is not. The positive price spreads we see in Tables 1 and 2 reflect a mix of these two conditions that varies day to day.

The presence of non-zero price spreads due to network congestion between delivery points has an important implication for our analysis. We are not interested *per se* in testing whether arbitrage is 'perfect', in the sense of continuously equating prices between market-specific delivery points. Rather, we are interested in assessing whether arbitrage *improves* as a result of the market's expansion. That is, the central question is whether markets find better ways to use the existing network capacity to increase trade, thereby reducing price spreads.

In both Tables 1 and 2, the third column shows the changes in prices and spreads before and after the market change date. They indicate that price spreads changed at all locations in a striking way. For the peak-period contracts in Table 1 price spreads between markets converge at all six delivery point pairs. The magnitudes are similar at all six pairs, ranging from −\$2.67 to −\$3.49 per megawatt hour. In percentage terms, the decline in these price spreads ranges from 35 to 49 percent of the average pre-expansion price spread.[11]

The change in the price spreads between markets is even more dramatic for the off-peak delivery period in Table 2. Panel B shows the changes in off-peak price spreads for all six bilateral-PJM delivery point contrasts. Again, the price spreads between the two markets fall by similar magnitudes for all six pairs, ranging from −\$4.24 to −\$8.74 per megawatt hour. These correspond to 37-to-81 percentage point declines from average pre-expansion price spreads. Both the on- and off-peak reductions in average spreads are large relative to normal variation in daily spreads, and are highly statistically significant.[12]

Tables 1 and 2 are simple before-and-after comparisons, and do not account for any potentially confounding factors that may have also affected prices over the same period. In particular, the cost of natural gas and coal feedstocks in these two regions rose steadily (by about 20 percent) over the six months post-expansion. Because electricity prices are fairly sensitive to input fuel prices, these factor price increases will tend to (a) offset the price reduction in the mid-Atlantic region resulting from the organized market's expansion, and (b) amplify the price increase observed in the Midwest.

The potential confounding effects of fuel price increases are most apparent in Table 2: There, the price level in the PJM area *increases* after October 2004, presumably due to offsetting increases in fuel costs. We also observe an exceptionally steep increase in Midwestern off-peak price before versus after prices after October 2004, which is likely due

[11] After PJM's expansion, the organized market also set a price for delivery in central Ohio (AEP-Dayton). Although the precise set of network delivery points (nodes) comprising each venue's central Ohio hub differ slightly, the post-expansion PJM market price at AEP-Dayton is (essentially) the same as the daily bilateral-market transaction price.

[12] We use nonparametric (Newey-West) standard errors throughout, as there is slight persistence in the daily price spreads between most delivery points. This occurs because exogenous changes in network capacity that create congestion tend to last more than one day (e.g., weather disturbances and line deratings).

in (large) part to coal prices. (The thermal efficiency of a typical coal-fired power plant is roughly .3, so a 20 percent increase in fuel prices would increase output prices by about 60 percent—or about the size of the price increase at the Midwest delivery points.) In Section 7, we conduct an econometric analysis that separates the effects of fuel cost increases from the role of increased trade, and shows each region's supply curve to be considerably more elastic than Table 2 suggests.

If the change in market organization improved the efficiency of trade, a second prediction of price spread convergence is that we should see less dispersion, or volatility, in daily price spreads. Table 3 provides evidence on this. The first column shows the standard deviation of between-market price spreads for various delivery point pairs over the six months prior to the market change date. The second column shows the comparable data for the six months after it. The third column reports relative change, post versus pre.

Daily price spreads for power tend to be quite volatile: Standard deviations are roughly 1.5 times the mean spread for each pair. Yet the volatility of these daily spreads falls quite dramatically after the market change date. Panel A of Table 3 indicates that the standard deviation in daily between-market price spreads for peak period delivery fell by 25 to 37 percent. The decline is greater in the off peak periods, falling by 25 to 61 percent. All of these changes are far too large to be attributable to chance variation, as indicated by the F-statistics shown in the final column. Overall, it is clear that prices spreads converged substantially after the new market design was implemented—with far less volatility thereafter.

Tables 1 through 3 present price information using a six month 'window' pre- and post-expansion. This relatively long horizon is informative because the economic importance of a change in market outcomes depends whether it persists over time. We have replicated the analysis these tables using both shorter and longer pre- versus post-expansion windows. These yield quantitatively similar changes to those shown in Tables 1 through 3 for all six delivery point pairs, in both peak and off-peak periods.

Notably, the data indicate that price spreads fell quite quickly after the organized market's expansion. Table 4 shows the changes in price spreads for various time horizons centered on the market change date. The spreads are for PJM's Allegheny and the AEP-Dayton bilateral market delivery points, the physically closest pair of the six contrasting dyads in Tables 1 to 3. Comparing one day before and one day after integration, price spreads fall 16 percent on peak and 46 percent off peak. By the end of one week, the decline in average daily price spreads is 41 percent on peak and 65 percent off peak.

Regardless of the window length examined—from one day out to six months after the market expansion—we see peak period price spreads fall from pre-expansion levels. The decline in peak period spreads varies somewhat with the time 'window' employed, and—because spreads tend to be quite volatile—becomes statistically significant only with a full 12 months (that is, ±2 quarters) of data. Off-peak price spreads fall by substantially greater (percentage) amounts, and are uniformly smaller ex post at all time horizons. In sum, the price spreads between markets fell quickly after PJM expanded, and remained far smaller thereafter.

We next examine whether the same thing happened in prior years. A simple 'placebo analysis' is to replicate these calculations for a prior-year comparison period centered on October 1, 2003, when there was no change in the markets' organization. There we see no significant changes in average price spreads, whether we evaluate them with window lengths of six, three, or one month or one week.

The simplest interpretation of these data is that the organized market's expansion improved arbitrage between the new and existing members of PJM. Any firm that bought at PJM Western Hub or Allegheny prior to PJM's expansion faced systematically higher prices than at the bilateral-market delivery points. Their convergence suggests the organized market identified complementary trading opportunities, as in Figure 1, that enabled the network to accommodate greater trade between these regions.

The architecture of the organized market implies this increased arbitrage is taking place anonymously. Buyers and sellers are (implicitly) matched by PJM to reduce participants' total production costs post-expansion. Since producers' marginal costs increase with output (at the firm level), increased production by low-cost firms in the Midwest after joining PJM would raise the price at which these firms are willing to sell to trading partners that remain in the bilateral market. The result is higher prices in Midwestern bilateral markets after PJM's expansion, and the price spread convergence documented in Tables 1 through 4.

Of course, if this interpretation of the price data is correct, then the market's expansion should also be accompanied by an increase in the quantities traded between the new and existing members of the organized market. There is also the question of why price spreads converged substantially, but were not driven to zero, by the organized market. We examine these next.

5 Quantity Evidence

The abrupt changes in price spreads shown above drew considerable attention from energy traders and power producers at the time. Although electricity trading is a specialized business, the *Wall Street Journal* ran a front-section article on the dramatic changes in power flows and prices in this area of the U.S. after PJM's expansion (Smith, 2005). One quantitative piece of information in the article is that shipments of eastbound power from the Midwest to PJM's pre-existing members tripled after the market's expansion.

To examine this we obtained information on the quantities traded between these areas before and after the market's expansion. Figure 2 shows the day-ahead scheduled transfers across the interface that separates the bilateral-market delivery points and the organized-market delivery points in Tables 1 and 2. (Flows are net, with positive values eastbound).[13] These day-ahead flows correspond exactly to the contracts whose prices are summarized

[13] Data are daily averages from http://pjm.com/markets-and-operations/ops-analysis/nts.aspx. It makes little difference whether gross or net transfers are used. The transfers in Figure 2 are eastbound approximately 98 percent of all hours.

in Tables 1 to 3 above. The time horizon is May 2003 through April 2005, with two years of data superimposed on the same twelve-month horizontal axis.

In Figure 2, the solid circles are the net transfer each day from April 2004 to April 2005. The solid line is their (locally-weighted) average before and after the the market change date on October 1, 2004. For comparison purposes, Figure 2 also shows the same data for the twelve-month period one year earlier, when there was no change in the market's organization. For May 2003 to April 2004, the open circles are the net transfer each day and the dashed near-horizontal line their (locally-weighted) average.

Figure 2 reveals a striking, abrupt increase in the quantity of power shipped between these two areas immediately after the market's expansion. The total flows from the Midwest increased nearly threefold, from 35 to 105 million kilowatt-hours per day. This increase is similar whether we compare the average post-expansion transfers to the same six (winter) months one year earlier, or to the six (summer) months immediately preceding the expansion.[14]

By any measure, the abrupt increase in east-bound power transfers after October 1, 2004, was an extraordinarily large change in where power is produced. To put the magnitudes in perspective, the *increase* in the average quantity transferred (80 million kilowatt-hours per day) is the amount of power typically consumed daily in a city of three million people. Abrupt changes of this magnitude in the quantities transferred across the transmission network are extraordinary events, and are otherwise precipitated only by large-scale plant or transmission network failures (large enough to affect millions of people, absent adequate reserve capacity). Yet no such events occurred in 2004. One is left with the seemingly indisputable conclusion that adopting PJM's market design in the Midwest increased trade by unprecedented magnitudes.

Two related pieces of evidence are informative here. The first is data regarding the frequency of network congestion across this interface. Publicly-available locational price

[14] Like the average price spreads, the average quantities transferred prior to the market's expansion were largely stable from month-to-month but quite volatile on a day-to-day basis.

data from PJM indicate that this interface was congested eastbound—that is, handling the maximum possible quantity—in 98% of all hours after October 1, 2004. By contrast, in a technical filing submitted to the FERC prior to the market expansion, the new market participants indicated there was little congestion across this interface during 2003 (AEP 2004, p. 36).

This is important because it indicates that congestion across this interface did not fall (for some exogenous reason) after the market expansion date, and thereby enable greater trade. Rather, the volume of trade increased up to the capacity of the network. That fact explains why there continues to be positive price spread between the delivery point areas in Tables 1 and 2 after the market's expansion: The network is evidently transferring the maximum quantity it can accommodate between these regions, but the capacity of the network is not high enough to drive the price spreads to zero.

Taken together, the price and aggregate quantities data point to a substantial increase in arbitrage in the day-ahead forward markets. Note the actual change in trading arrangements accompanying the organized market's expansion imply this arbitrage is taking place anonymously, through PJM's day-ahead forwards market. The new members joining PJM were matched by the system to buyers elsewhere in PJM (to the east), increasing the new members' production from generating assets physically located in the Midwest and decreasing production from generating assets in the mid-Atlantic and Eastern seaboard. The result of this reallocation in production is the enormous change in power shipments between regions in Figure 2.

6 Gains from Trade

The magnitude of the quantity changes and price spread reductions that followed the organized market's expansion suggest that the gains from increased trade may be substantial. Our next task is to refine this analysis, and provide quantitative evidence on the magnitude of these economic efficiency gains. This requires information on the elasticity of supply in each region, and how it varies with factor prices that changed during the period of our study.

Our methods are based on an econometric model of supply and inter-regional trade. The economic logic of our approach is illustrated in Figure 3. Let $S_1(Q)$ be the (inverse) supply function of producers in the Midwest, who are not initially members of the organized market. Here Q_1^d is the total electricity consumption of final consumers in region 1. Because retail electricity prices are regulated and change (typically) only on an annual basis, electricity consumption is insensitive to day-to-day changes in wholesale electricity market prices. Consequently, their aggregate consumption is indicated by the dashed vertical line.

Similarly, let $S_2(Q)$ be the (inverse) supply function of producers in the mid-Atlantic region. The distance Q_2^d represents the electricity consumption of final consumers in region 2. Note we have reversed the horizonal axis direction for region 2, so that Q_2^d and S_2 increase to the left. As shown, the mid-Atlantic supply curve S_2 is effectively a residual demand curve for wholesale market purchases from the lower-cost Midwest suppliers.

Under the bilateral trading system, we observe a level of inter-regional trade that reduces production in region 2 by Δq^b and shifts it to less expensive sellers in region 1. This produces gains from trade represented by the trapezoidal region W^b in Figure 3. The greater volume of inter-regional trade after the organized market's expansion, Δq^o, produces greater gains from trade of $W^b + \Delta W$. The increase following the markets' integration, ΔW, is the shaded trapezoidal area in Figure 3.

A simple calculation suggests the magnitude of the increased gains from trade. Suppose supply is approximately linear in output over the range of production affected by trade. Then

$$\Delta W \approx \frac{1}{2}(\Delta q^o - \Delta q^b)(\Delta p^o + \Delta p^b)$$

where Δp^o and Δp^b are the price spreads under the organized and bilateral trading systems, respectively. Provided the supply functions do not shift over time, we can approximate

ΔW by inserting the before-and-after average price spreads from Tables 1 and 2 and the quantity data from Figure 2. These imply efficiency gains from increased trade of $\Delta W \approx \$500$ thousand per day, or \$175 million annually.

This order-of-magnitude calculation is subject to bias if there are other, confounding factors that differed before-versus-after the organized market's expansion. Our primary concern is factor prices: The price of Appalachian coal rose 30 percent over the 12 months after October 2004 from the previous year, and natural gas prices rose 20 percent. If these factor price changes shifted regional supply curves further apart—thus increasing the marginal benefit of incremental trade—then perhaps the benefits of inter-regional trade might have increased under the bilateral trading system even in the absence of the organized market's expansion.

Addressing this possibility quantitatively requires an estimate of how much inter-regional trade would have occurred under the bilateral system if, counter to fact, the organized market's expansion had never occurred. For this we require a model of trade.

6.1 Methods

The first step is to evaluate how changes in input factor prices affect each region's supply schedule. Extending the ideas in Figure 3, assume each region's price on day t is given by

$$p_{1t} = S_1(Q_{1t}^d + \Delta q_t, F_{1t}) + \varepsilon_{1t} \tag{6a}$$

$$p_{2t} = S_2(Q_{2t}^d - \Delta q_t, F_{2t}) + \varepsilon_{2t} \tag{6b}$$

Here the marginal willingness-to-sell (inverse supply) curve S_i in region i varies with regional electricity consumption Q_{it}^d, the net exports Δq_t from region $1 \rightarrow 2$, and a vector of region-specific factor prices F_{it}. Supply shocks ε_{it} arise from maintenance shutdowns and transmission network contingencies that affect which plants operate each day.

We use data on market prices, consumption, quantities traded, and producers' input factor prices to estimate the supply functions S_1, S_2.[15] This serves two purposes. First, it enables us to distinguish variation in market prices attributable to *shifts* of regional supply curves (due to changes in input factor prices) from movement *along* a supply curve (due to changes in consumer demand). Second—as a consequence—it provides estimates of the elasticity of supply in each region with respect to changes in the quantity traded. Estimation of S_1, S_2 poses three econometric issues: (a) Specification of a functional form for S_i; (b) potential endogeneity of Δq_t; and (c) the correlation of supply shocks between regions and over time. We address estimation further below.

As depicted in Figure 3, the gain from trade attributable to integration is the difference in sellers' valuations for the incremental quantities traded:

$$\Delta W_t = \int_{\Delta q_t^o}^{\Delta q_t^b} S_2(Q_{2t}^d - \theta, F_{2t})\, d\theta - \int_{\Delta q_t^b}^{\Delta q_t^o} S_1(Q_{1t}^d + \theta, F_{1t})\, d\theta \tag{7}$$

We observe Δq_t^0 directly beginning October 1, 2004. However, Δq_t^b is not observed: It is the counterfactual quantity that would have been traded after October 2004 'but for' the organized market's expansion.

We estimate Δq_t^b for the factor prices and demand conditions after October 1, 2004, as follows. Let p_{it}^* be the *autarky prices* that would prevail in each region in the absence of any trade between them (see Figure 3). Under the model, these are

$$p_{it}^* = S_i(Q_{it}^d, F_{it}) + \varepsilon_{it}\ .$$

Note there is no Δq_t here. We assume that the volume of inter-regional trade realized with the bilateral trading system is an (increasing) function g of the autarky spread:

$$\Delta q_t^b = g(\Delta p_t^*) + \eta_t \tag{8}$$

[15]PJM reports actual load, day ahead scheduled net flows, and day ahead prices (www.pjm.com). We use the day ahead prices for PJM's Western Hub and APS Zone. ECAR load data are from FERC form 714. Platts provides daily block price data for PJM Western Hub, Into AEP, Cinergy, North ECAR, and Northern Illinois. Fuel prices are from Platts. Natural gas prices are the daily Texas Eastern, M-3, price. Coal data are the daily Central Appalachia (CAP) prices for Ohio and Northern Appalachia (NAP) prices for Pennsylvania. Monthly pollution prices are from Cantor Fitzgerald for the SO_2 market and from the EPA for the NO_x market. Average daily temperature data, which we use to measure heating and cooling degree days for Pittsburgh and Cleveland, are from NOAA.

where $\Delta p_{it}^* = p_{2t}^* - p_{1t}^*$. We call g the *bilateral arbitrage efficiency* function. If, for example, supply functions are linear and arbitrage eliminates a constant proportion $0 < \alpha < 1$ of the autarky price spread, then $g = \alpha\gamma\, \Delta p_t^*$ for $\gamma = 1/(S_2' + S_1')$. Alternatively, if arbitrage drives the quantity traded to the lesser of a network capacity limit κ_t (when binding) or the efficient level (when not), the g is kinked: $g = \min\{\kappa_t, \frac{\gamma}{2}(\Delta p_t^*)^2\}$. Other assumptions are possible, with more complex specifications for g.

Since the true map $\Delta p_t^* \mapsto \Delta q_t^b$ depends on market participants' information sets and network complementarities described in Sections 2 and 3, it seems difficult to specify theoretically. We take an empirical approach. Although autarky prices are not directly observable, they can be estimated from the fitted marginal willingness-to-sell functions S_1, S_2 using

$$\hat{p}_{it}^* = \hat{S}_i(Q_{it}^d, F_{it}) + \hat{\varepsilon}_{it} \tag{9}$$

where $\hat{\varepsilon}_{it}$ is the residual from fitting (6).[16] We first estimate g by projecting the observed daily net trade volume Δq_t^b before October 2004 onto a set of orthogonal polynomial functions of the estimated autarky price spread $\Delta \hat{p}_t^*$. Applying this fitted bilateral arbitrage efficiency function $\hat{g}$ during the post-expansion period gives

$$\Delta \hat{q}_t^b = \hat{g}(\Delta \hat{p}_t^*), \qquad t \geq \text{October 2004}.$$

This yields a counterfactual estimate of the volume of trade that would have been achieved bilaterally after October 2004 'but for' the market's new organization. Changes in input factor prices and demand conditions affect the volume of trade by changing the supply curves and autarky prices in (9). The central assumption here is that the effectiveness of the bilateral trading system in arbitraging inter-regional price spreads—that is, the mapping g—would have continued unchanged 'but for' the organized market's expansion.

[16] Note that here we are (implicitly) assuming the same disturbance term ε_{it} that actually occurred on date t post-expansion would have also applied on that date had the expansion not occurred. This seems sensible, as the main random factors that we cannot account for in the model (6) that might affect prices (network line failures, generator forced outages large enough to move prices, and the like) should not be assumed away in the counterfactual case of no market expansion.

A technical complication arises when using $\Delta\hat{q}_t^b$ to evaluate the gains from trade in (7). As in Figure 3, the gain from trading Δq_t^b is

$$W_t^b(\Delta q_t^b) = \int_{\Delta q_t^b}^{0} S_2(Q_{2t}^d - \theta, F_{2t})\, d\theta - \int_0^{\Delta q_t^b} S_1(Q_{1t}^d + \theta, F_{1t})\, d\theta \ .$$

If $S_i' > 0$, W_t^b is nonlinear in Δq_t^b and the error η_t in (8) creates an (upward) bias in the naive 'plug-in' estimator $W_t^b(\Delta\hat{q}_t^b)$. Specifically,

$$W_t^b(\Delta q_t^b) = W_t^b(g(\Delta p_t^*)) - \psi(\Delta p_t^*)$$

where $\psi(\Delta p_t^*) > 0$. If S_1, S_2 are linear over the range of production affected by trade, the bias correction has the exact form

$$\psi(\Delta p_t^*) = v(\Delta p_t^*)\frac{(S_{1t}' + S_{2t}')}{2}$$

where $v(\Delta p_t^*) = E[\eta_t^2|\Delta p_t^*]$ is the conditional variance of the error in (8). To incorporate this, we project the squared residuals from estimation of (8) onto the autarky spreads during the bilateral regime to estimate the conditional variance function $\hat{v}(\cdot)$. We then evaluate, for $t \geq$ October 1, 2004,

$$\hat{\psi}_t = \hat{v}(\Delta\hat{p}_t^*)\frac{(\hat{S}_{1t}' + \hat{S}_{2t}')}{2}.$$

The error-corrected estimate of the bilateral gains from trade that would have occurred on day t 'but for' the organized market's expansion is therefore

$$\widehat{W}_t^b = \int_{\Delta\hat{q}_t^b}^{0} \hat{S}_2(Q_{2t}^d - \theta, F_{2t})\, d\theta - \int_0^{\Delta\hat{q}_t^b} \hat{S}_1(Q_{1t}^d + \theta, F_{1t})\, d\theta \ - \hat{\psi}_t \ .$$

Empirically, the bias correction term $\hat{\psi}_t$ is small (about 1 percent of $\widehat{W}_t^b$).[17] We evaluate $\widehat{W}_t^b$ separately for each day and each price block (peak and off peak), then sum these values to estimate the total gains from trade under the bilateral system on an annual basis.

[17] We ignore additional the errors-in-variables bias in $\Delta\hat{q}_t^b$ that arises from using $\Delta\hat{p}_t^*$ in lieu of Δp_t^* when estimating (8). The variance of the observed bilateral spreads relative to predictions (from fitting (6)) suggests the resulting bias in $\widehat{W}_t^b$ is apt to be be small.

Estimating the gains from trade under the organized market design is considerably simpler, because we can apply the observed quantity of trade Δq_t^o directly. Relative to autarky, the gains from trade after the organized market's expansion are estimated with

$$\widehat{W}_t^o = \int_{q_t^o}^{0} \hat{S}_2(Q_{2t}^d - \theta, F_{2t})\, d\theta - \int_0^{q_t^o} \hat{S}_1(Q_{1t}^d + \theta, F_{1t})\, d\theta \ .$$

Our estimate of the gains from increased trade due to the organized market's expansion is the difference,

$$\Delta\widehat{W}_t = \widehat{W}_t^o - \widehat{W}_t^b \ , \qquad t \geq \text{October 2004.}$$

These three statistics, summed to an annual level, are the gains from trade measures that we discuss below.

6.2 Specification and Estimation of $S_{it}(Q)$

Implementation requires of a specification for market-level supply. At the plant level, electricity production is fixed-proportions (Leontief) in two variable factors: fuel and emissions permits. The marginal cost of plant k is

$$m_k = h_k \cdot (p^f + p^e \cdot e_k^f)$$

where h_k is the plant's efficiency, p^f the price of fuel, p_t^e the price of emissions permits, and e_k^f the plant's emissions (NO_x and SO_2 production) per unit input. If all plants of type f in a region have the same emissions rate and factor prices, their aggregate marginal cost function has the form

$$m^f(Q, F) = h^f(Q)(p^f + p^e \cdot e^f) \tag{10}$$

for factor prices $F = \{p^f, p^e\}$. The function h^f is an aggregate heat rate (inverse thermodynamic efficiency) curve. It increases with Q because less efficient units operate as demand rises. In practice, the assumption that all plants in a region with the same fuel have the same emissions rate is tenuous (some have scrubbers, some do not). However, the empirical consequence of this assumption is likely to be minor: permit costs ($p^e e^f$)

are small relative to fuel costs, and the overwhelming determinants of the marginal cost structure are the first two terms, $h^f(Q)p^f$.[18]

Two production technologies set the market price at different times: gas- and coal-fired generation. To aggregate to a market-level supply curve, we are helped by the fact that one technology (coal) dominates the other (gas) during our study period. The marginal cost of supplying Q units is then

$$c(Q, F) = m^c(Q, F)I^c(Q) + m^g(Q - \kappa, F)I^g(Q) \tag{11}$$

where c, g denote the two technologies. The indicator $I^c = 1$ iff the lower-cost technology's available capacity κ exceeds Q so type c operates at the margin, and $I^g = 1 - I^c$ if otherwise.

Equation (11) is best thought of as an instantaneous marginal cost; in contrast, prices are quoted for a (8 or 16 hour) peak or off-peak delivery period. Since there are no inventories, we shall assume the observed price for a given delivery period on day t is

$$p_t = \bar{c}_t + \varepsilon_t, \tag{12}$$

where $\bar{c}_t$ is the average region-level marginal cost during the delivery period (inclusive of net exports). The cost shock ε_t is assumed to be orthogonal to the factor prices F and regional retail consumption Q, but possibly correlated with net exports.

Complications arise in specifying $\bar{c}_t$ because demand varies during the delivery period. Averaging $c(Q, F)$ over a delivery period for day t implies

$$\bar{c}_t = E_t[m^c(Q + \Delta q, F) \,|\, I^c]\, \phi_t + E_t[m^g(Q + \Delta q, F) \,|\, 1 - I^c\,]\, (1 - \phi_t). \tag{13}$$

E_t is against the frequency distribution P_t of $Q + \Delta q$ during the day t delivery period, and $\phi_t = P_t(Q + \Delta q \leq \kappa)$. To approximate (13), we treat each technology's aggregate efficiency curve as linear in output:

$$m^f(Q, F) = (\alpha^f + \beta^f Q)(p^f + \gamma^f p^e),$$

[18] Fuel expenses are 80 to 90 percent of plants' marginal costs (PJM 2005, p. 106).

for technology-specific efficiency parameters $\alpha^f, \beta^f, \gamma^f$. The expectations in (13) then satisfy

$$E_t[m^f(Q + \Delta q, F)|I^f] = m^f(\mu_t^f(\kappa), F)$$

where $\mu_t^f(\kappa) = E_t[Q + \Delta q|I^f]$ is the truncated mean of total output when technology f is on the margin (during t's delivery period). For the present analysis, however, we make the simplifying assumption that[19]

$$E_t[Q + \Delta q|I^f] = \lambda^f \, E_t[Q + \Delta q].$$

The final supply model can then be expressed in terms of average output during the delivery period as

$$S_i(\overline{Q}_{it}^d + \Delta\overline{q}_t, F_{it}) = \delta' \left[F_{it} \otimes I_2\right] \begin{bmatrix} 1 \\ \overline{Q}_{it}^d + \Delta\overline{q}_t \end{bmatrix} \tag{14}$$

with δ a vector of reduced-form parameters and I_2 a (2×2) identity matrix. This means price p_{it} is a linear-in-the-parameters function of the factor prices in F_{it}, and their cross-product with output plus net exports. (To match the depiction in Figure 3, we reverse the sign convention on Δq in region 2.) The structural parameters that characterize costs $(\alpha^f, \beta^f, \gamma^f)$, demand (λ^f), and their interaction $(\overline{\phi}_t)$ can be (partially) recovered from the elements in δ. We estimate the supply specification using daily data from May 2003 to October 2005. This covers two years prior and one year after the market's organization change date in October 2004.

Note that in aggregating from marginal costs to market supply using (12) and (14), we implicitly assume that producers' marginal willingness to sell is their marginal cost. Is this sensible? Regulatory reports from PJM indicate that the marginal seller's markup over its marginal cost is small in the organized market, both pre- and post-expansion (averaging approximately 3.4 percent; PJM 2005, p. 68).

[19] These truncation functions can also be estimated using a separate sample of our market data that contains (high-frequency) observations on I^c and $Q + \Delta q$ during 2004. The results are not appreciably different from those reported below.

In our model, any differences between price and marginal cost will be absorbed into the estimated efficiency parameters, β^f. Because of this, the fitted model would still provide estimates of the gains from trade even if firms' offers to sell incorporate a (proportional) markup over marginal cost. The interpretation of these gains from trade as comprising production cost savings relies on the evidence that these markups are negligible, however. For the Midwest region, no such external regulatory reports on seller's markups are available (during our study period). If Midwest prices exceeded marginal costs, however, then procedures outlined here will tend to underestimate the true cost savings from reallocating production to Midwestern producers. We elaborate further on these market power issues in Section 8.

A second estimation concern is the possibility that the quantity traded between regions may be correlated with the unobserved supply shifters in ε_{it} in (6). This could arise if there are multi-day network contingencies or transmission line failures with a region that directly raise price and simultaneously curtail net transfers between regions. To address this possibility, we instrument for Δq_t in each region's supply specification using putatively exogenous data that is highly correlated with daily inter-regional electricity trade. Specifically, we use differences in weather conditions (cooling and heating degree-day differences) between Midwestern and mid-Atlantic cities, at daily frequencies. There is no reason weather conditions would otherwise affect supply, except through its (substantial) effect on retail electricity consumption (Q_{it}).

6.3 Results

Tables 5 and 6 summarize the supply function parameter estimates. Because we observe wide variation in demand and factor prices during this period, the reduced form parameters can be estimated quite precisely. Columns (1) and (4) report parameter estimates fitted via ordinary least squares; columns (2) and (5) instrument for net trade using weather information. Although the first-stage F-statistics are high, the OLS and IV parameter estimates are substantively similar. Since we regard weather as a valid instrument for supply, we interpret this to suggest the potential endogeneity bias in the OLS estimates is a minor issue. Columns (3) and (6) perform a seemingly-unrelated regression that accounts for the high (approximately .7) contemporaneous correlation in the supply shocks between regions.

Tables 5 and 6 assume there is a single supply function in each region, and use the price data from all delivery points in the region to estimate it. We have also estimated the supply models separately for each delivery pricing point; the results are not appreciably different from the pooled estimates. The interpretation that follows are based on the IV/SUR estimates in Tables 5 and 6.

Although the reduced form coefficients are difficult to interpret directly, the marginal effects of fuel prices (the β^f's) are consistent with expectations. In the mid-Atlantic region on peak, when gas tends to be on the margin, a one unit (in \$/MMBtu) increase in the cost of gas raises marginal willingness to sell by approximately \$8/MWh. This implies a marginal technical (thermal) efficiency of approximately 40 percent, which is in line with engineering estimates.

Table 7 converts the raw parameter estimates into supply functions' slopes and elasticities. These we report by pricing point, as well as for the combined (regional) supply function estimates. The slope information in the first and third numerical columns are in megawatts per dollar. They indicate that in the mid-Atlantic region peak period prices rise by \$1 per MWh as consumption increases by (approximately) 700 megawatts; in the Midwest region, this requires a consumption increase of 1400-to-1700 megawatts. These figures imply (inverse) supply functions are rather flat, and particularly so in the Midwest.

The second and fourth columns in Table 7 report the elasticity of net imports/exports with respect to a region's price. The elasticity value of 48 in the first row for PJM Western Hub means that a 48% increase in net imports (in MW) into the mid-Atlantic region decreases the market price at Western Hub by 1%. (There is no decimal point missing in Table 7's elasticities—but note they are elasticities of net exports, not of aggregate supply). Overall, these results imply that doubling net imports into the mid-Atlantic region decreases mid-Atlantic prices on peak by about 2 percent, and off peak about six percent. Doubling net exports from the Midwest increases Midwest prices on peak by about one percent, and off peak about three percent.

These figures provide useful information on how observed changes in the level of trade affected each region's prices, absent the confounding effects of fuel price increases that occurred during the same period. In the data shown in Figure 2, we observe an average increase in the quantities traded of 194 percent on peak and 221 percent off peak. Absent fuel cost increases, these results suggest the organized market's expansion would have reduced prices in PJM's mid-Atlantic region by 194/47 = 4 percent on peak, 221/14 = 16 percent off peak; and increased wholesale prices in the Midwest by 194/85 = 2 percent on peak, and 221/45 = 5 percent off peak. In contrast to the simple before-and-after comparisons provided in Tables 1 and 2, the estimated supply elasticities imply that prices fell considerably more in the mid-Atlantic than they rose in the Midwest due to the increase in inter-regional trade.

Table 8 summarizes the gains from trade analysis. Values labeled 'post 10/2004' are calculated daily over the 12 month period beginning October 1, 2004, and 'pre 10/2004' values over the preceding 12 months. Note the last two columns separate off-peak periods into weeknights (10pm to 6am) and weekends (all hours). Because these are annual averages and off-peak is split into two periods, the average price spreads differ here from the six-month pre-and-post values in Tables 1 and 2.

The top panel reports average price spreads between regions under three market arrangements: autarky, the bilateral trading system, and the organized market design. Autarky price spreads are calculated daily using (9); the counterfactual bilateral spreads are calculated similarly, but assume trade of $\Delta \hat{q}_t^b$ reported separately below. The autarky spreads differ pre and post because they are evaluated at different input factor prices and demand conditions (although the difference in demand conditions is slight). While the estimated autarky price *levels* are significantly higher after October 2004 than before (approximately 30 percent higher in the mid-Atlantic region and 45 percent higher in the Midwest), the price *spreads* between regions are substantively the same. This reveals that the changes in factor prices after October 2004 did not result in significant asymmetric supply shifts across the two regions; rather, the increase in fuel costs (in particular) shifted both regions supply curves upward, and by similar amounts. The implication is that for most comparisons of interest, the observed actual averages under the bilateral system before October 2004 are apt to be good measures of what would have occurred under the same system after October 2004.

The final row in Panel A indicates that had the bilateral trading system continued, we estimate price spreads between regions would be between 2 and over 3 dollars per MWh higher than actually occurred after October 2004. Because the supply curves are quite elastic, however, even modest reductions in price spreads correspond to large increases in quantities traded. Panel B reveals that on an annual basis, the quantities traded increased by about 2000 MW per hour (197%) on peak and about 2600 MW per hour (221%) off peak.

The final panel indicates the estimated gains from trade under the two market trading arrangements. Under the bilateral trading system, we estimate total gains from trade of approximately \$150 million annually. This is distributed (in aggregate, not hourly) roughly equally across the peak, weeknight off-peak, and weekend off-peak periods. It differs little whether we evaluate it at the factor prices and demand conditions observed prior to October 2004, or under the higher cost conditions the following year.

After adopting the organized market design, we estimate the 2.5-fold increase in quantities traded between regions increased the total gains from trade to \$313 million annually. The difference, or $\Delta\widehat{W}$, is $\$313 - \$150 = \$163$ million per year. Exclusive of implementation costs noted below, this is our estimate of the annual efficiency gains—aggregate production cost savings—due to the expansion of the organized market design. In proportionate terms, the relative efficiency of the bilateral trading system is 150/313 = 48 percent. Thus, it appears that the decentralized bilateral trading system is able to realize slightly less than half the gains from trade achieved with a more efficient market design.

We also evaluate the gains from increased trade using supply models for each delivery point. Aggregated to an annual basis, the efficiency gains we estimate range from $162 million to $181 million across the four contrasting delivery points. These should not be summed together; rather, because we have evaluated the gains from trade pairwise (as opposed to solving for the implied flows between all delivery points simultaneously), these should be interpreted as providing different estimates of the total gains from improved trade between all of these delivery regions.

By any measure, these are large efficiency gains following the adoption of the organized market's design. As noted in the introduction, however, there are costs to implementing a new system of market organization. These costs can be compared to the efficiency gains reported above, providing a better assessment of the net benefits of expanding the organized market design.

The costs of implementing the new market design were incurred by two sets of market participants: The market operator itself (PJM), and the individual firms that joined the market. Regulatory accounting filings prepared by PJM for its members and the FERC report total expansion expenses of $18 million, through 2005. These are one-time, non-recurring expenses due to the expansion of the market. For the new members, accounting data filed with the SEC by American Electric Power indicate internals costs of re-organizing its wholesale market operations due the PJM expansion of $17 million; forward-looking statements characterize this as a one-time expense. Other market participants' expenses are more difficult to obtain, but based on volume-of-production and trading data, and the fact that all other new members relied upon AEP's regional transmission network prior to the market's expansion, we believe are likely on the order of $4 to 5 million. In total, this amounts to approximately $40 million in one-time implementation costs of expanding the market's design.

Combining these benefits and costs, the picture that emerges is that for an initial investment of approximately $40 million the participants in these markets realized *increased* efficiency gains of $163 million over the first year alone. At the usual risk of extrapolation, if gains of this magnitude in subsequent years are of similar magnitude, the present value to society of expanding this organized market's design is remarkably large.

7 Discussion

One perspective that merits brief discussion relates to pricing changes by the firms new to the organized market. Specifically, perhaps the efficiency improvements we have pointed to here arose because the expansion of the organized market led the new market participants to change their willingness to supply. We alluded to this possibility when discussing our interpretation of the supply specification model in section 6.2. Stated in other words, perhaps the firms that joined PJM simply decided to offer their production at lower prices (that is, by bidding more aggressively) into the organized market, relative to their previous supply behavior bilateral market.

We are skeptical of this possibility, for several reasons. First, from a theoretical perspective, it is difficult to conceive why such a change in willingness-to-sell would be profit-maximizing behavior. The identities and number of firms operating in these markets was the same throughout the period we study, and—*if* the bilateral markets were not subject to trading imperfections—*then* the new exchange members would have faced the same set of trading opportunities before and after the organized market's expansion. Second, there is the empirical fact that prices in the delivery region where the new members physical production assets are located increased sharply following the market's expansion. This fact is inconsistent with firms offering to sell their production at lower prices, but consistent with an increase in demand from buyers to the east.

Third, while our results (in Figure 2) indicate that the quantities delivered to the two main PJM delivery regions the Midwest nearly tripled post-expansion, these two PJM delivery regions' price levels fell only about ten percent. This is an extraordinarily large elasticity response, although perhaps that is to be expected in an homogeneous-good market. Empirically, it would not have been profitable for the new exchange members to produce as little as they actually did before the market's expansion—unless bilateral market imperfections obscured the trading possibilities subsequently identified by the organized market's design.

8 Conclusion

The motivation for this paper arose from a vigorous policy debate about the merits of organized market designs in electricity markets. This debate reflects two distinct, but related difficulties that frequently confront policy makers. First, the potential for a more efficient market design to reallocate production from high-cost firms to lower-cost competitors will create a political incentive for market participants that stand to lose to oppose it. Second, there is a technocratic challenge that the theoretical appeal of a different market design must be balanced against the cost of implementing it. Given these challenges, it is not surprising that consensus has proved elusive on the merits of expanding organized market designs to regions where they are not yet present.

The central contribution of this paper is to provide a detailed empirical assessment of this question. The expansion of the organized market design used by the PJM Interconnection in 2004 created a particular informative opportunity, and exceptionally rich data, with which to evaluate its consequences. As industry participants rapidly discovered, there were dramatic changes in market outcomes after the expansion: price differences between Midwestern and mid-Atlantic regions converged, the quantities of energy traded between them increased substantially, and production shifted from higher to lower-cost facilities.

We are led to the seemingly inexorable conclusion that the organized market design identified new trading opportunities that were not realized by the bilateral trading system that preceded it. These findings are consistent with the theoretical concern that decentralized bilateral markets may have difficulty achieving efficient allocations of the complementary services—viz., generation and transmission—required in these markets. Moreover, the magnitude of these gains calls into question the assertion that organized market designs are not worth their costs of implementation.

References

[1] AEP. 2004. "Market Power Analysis of the AEP Power Marketing Companies." Compliance Filing of American Electric Power Corporation in Federal Energy Regulatory Commission Docket ER97-4143-008 (July 18).

[2] Arrow, K., and F. Hahn (1971). *General Competitive Analysis.* San Francisco: Holden-Day.

[3] Borenstein, Severin, James B. Bushnell, and Frank A. Wolak (2002). "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review* 92(5): 1376–1405.

[4] Bushnell, James B., Erin T. Mansur, and Celeste Saravia (2008). "Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured U.S. Electricity Markets." *American Economic Review* 98(1): 237–266.

[5] FERC (2005). *FERC State of the Markets Report 2004.* Federal Energy Regulatory Commission, Office of Market Oversight and Investigation (June). Available at http://www.ferc.gov/market-oversight/st-mkt-ovr/som-rpt-2004.pdf

[6] Fisher, Frank M. (1972). "On Price Adjustment Without an Auctioneer." *Review of Economic Studies,* 29(1): 1–15.

[7] Howatson, A. M. (1996). *Electrical Circuits and Systems.* New York: Oxford University Press.

[8] Joskow Paul L., and Jean Tirole (2000). "Transmission Rights and Market Power on Electric Power Networks." *RAND Journal of Economics* 31(3): 450–487.

[9] Kleit, Andrew N., and James D. Reitzes (2008). "The Effectiveness of FERC's Transmission Policy: Is Transmission Used Efficiently and When Is It Scarce?" *Journal of Regulatory Economics* 34(1): 1–26.

[10] Olson, Mark, Stephen Rassenti, Mary Rigdon, and Vernon Smith (2003). "Market Design and Human Trading Behavior in Electricity Markets." *IEEE Transactions* 35: 833-849.

[11] Mamoh, James A. (2000). *Electric Power System Applications of Optimization.* New York: Marcel Dekker.

[12] Mansur, Erin T. (2007). "Upstream Competition and Vertical Integration in Electricity Markets." *Journal of Law and Economics* 50(1): 125-156.

[13] Mansur, Erin T. (2008). "Measuring Welfare in Restructured Electricity Markets." *Review of Economics and Statistics* 90(2): 369-386.

[14] Milgrom, Paul (2004). *Putting Auction Theory to Work.* Cambridge University Press.

[15] PJM (2005). *2004 State of the Market Report.* Valley Forge, PA: PJM Interconnection LLC., Market Monitoring Unit. Available at http://www.pjm.com/markets/market-monitor/som-reports.html

[16] PJM (2005b). *PJM 2005 Financial Report.* Valley Forge, PA: PJM Interconnection, LLC. Available at http://www.pjm.com/about/downloads/2005-financial-finalprint.pdf

[17] Reiss, Peter C., and Matthew W. White (2008). "What Changes Energy Consumption? Prices and Public Pressures." *Rand Journal of Economics*, 39(3): 636–663.

[18] Roth, Alvin E. (2002). "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70(4): 1341–1378.

[19] Rothschild, Michael (1973) "Models of Market Organization with Imperfect Information: A Survey." *Journal of Political Economy* 81(6): 1283-1308.

[20] Smith, Vernon L. (1962). "An Experimental Study of Competitive Market Behavior." *Journal of Political Economy* 70(2): 111–137.

[21] Smith, Vernon L. (1964). "Effect of Market Organization on Competitive Equilibrium." *Quarterly Journal of Economics* 78(94): 181–201.

[22] Smith, Rebecca (2005). "Eastern Power Is Getting Cheaper As Midwest Utilities Join Market." *Wall Street Journal* (Eastern Edition). New York, N.Y. (Jan 26, 2005), p. A.2.

[23] Schweppe, Fred C., Michael C. Caramanis, Richard D. Tabors, and Roger E. Bohn (1988). *Spot Pricing of Electricity.* Boston, MA: Kluwer Academic Publishers.

[24] Wilson, Robert A. (2002). "Architecture of Power Markets." *Econometrica* 70(4): 1299–1340.

[25] Wolfram, Catherine D. (1999). "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review* 89(4): 805–826.

Tables and Figures

TABLE 1

PRICE SPREADS BETWEEN MARKETS — PEAK DELIVERY

Contract Delivery Point[a] (and approximate location)		Pre-Expansion (Apr.-Sep. 2004)	Post-Expansion (Oct. '04-Mar. '05)	Post – Pre Percent Δ	Post – Pre Difference	Std. Error of Difference[b]	
		Average Prices for Day-Ahead Forwards ($ per MWh)					
		Panel A: Price Levels					
Exchange-based Prices							
PJM Western Hub (Pa.)		50.98	50.71	-1%	-0.27	(1.79)	
PJM Allegheny (Pa. and W. Va.)		50.41	49.80	-1%	-0.60	(1.92)	
Bilateral Market Prices							
AEP-Dayton (C. Ohio Valley)		43.41	45.81	6%	2.40	(1.87)	
Cinergy (S. Ohio Valley)		43.59	46.33	6%	2.75	(1.80)	
NI Hub (N. Illinois)		42.10	44.99	7%	2.88	(1.93)	
		Panel B: Price Spreads Between Markets					
PJM Western Hub	v. AEP-Dayton	7.57	4.90	-35%	-2.67	(1.06)	**
	v. Cinergy	7.40	4.38	-41%	-3.02	(1.02)	***
	v. NI Hub	8.88	5.73	-36%	-3.15	(1.11)	***
PJM Allegheny	v. AEP-Dayton	7.00	3.99	-43%	-3.01	(1.17)	**
	v. Cinergy	6.82	3.47	-49%	-3.35	(1.15)	***
	v. NI Hub	8.30	4.82	-42%	-3.49	(1.20)	***

Notes. Price spreads are average price differences between delivery points. Separate contracts are traded for peak (6 A.M. to 10 P.M.) and off-peak (10 P.M. to 6 A.M.) delivery; for off-peak, see Table 2. (a) Delivery points for electricity transactions are defined by area of the high-voltage transmission grid, not single points on a map. The locations above correspond to contiguous geographic regions, as follows (approximately): PJM Western Hub is central and western Pa.; PJM Allegheny is southwestern Pa. and northern W. Virgina; AEP-Dayton is central Ohio and southern W. Virginia; MichFE is northern Ohio and lower Michigan; Cinergy is southern Indiana and southwestern Ohio; and the NI Hub is Northern Illinois. (b) Newey-West standard errors assuming a five-day lag structure. Significance indicated for 1% (***), 5% (**), and 10% (*) levels.

TABLE 2

PRICE SPREADS BETWEEN MARKETS — OFF-PEAK DELIVERY

Contract Delivery Point[a] (and approximate location)		Pre-Expansion (Apr.-Sep. 2004)	Post-Expansion (Oct. '04-Mar. '05)	Post – Pre Percent Δ	Post – Pre Difference	Std. Error of Difference[b]	
		Average Prices for Day-Ahead Forwards ($ per MWh)					
		Panel A: Price Levels					
Exchange-based Prices							
PJM Western Hub (Pa.)		27.71	31.88	15%	4.17	(1.55)	***
PJM Allegheny (Pa. and W. Va.)		27.83	30.51	10%	2.68	(1.36)	*
Bilateral Market Prices							
AEP-Dayton (C. Ohio Valley)		17.32	27.98	62%	10.66	(1.20)	***
Cinergy (S. Ohio Valley)		16.99	28.41	67%	11.42	(1.17)	***
NI Hub (N. Illinois)		16.35	24.77	51%	8.42	(1.34)	***
		Panel B: Price Spreads Between Markets					
PJM Western Hub	v. AEP-Dayton	10.39	3.90	-62%	-6.49	(0.91)	***
	v. Cinergy	10.72	3.47	-68%	-7.25	(0.88)	***
	v. NI Hub	11.35	7.11	-37%	-4.24	(1.06)	***
PJM Allegheny	v. AEP-Dayton	10.51	2.53	-76%	-7.98	(0.77)	***
	v. Cinergy	10.84	2.10	-81%	-8.74	(0.76)	***
	v. NI Hub	11.47	5.74	-50%	-5.73	(0.92)	***

Notes. Price spreads are average price differences between delivery points. Separate contracts are traded for peak (6 A.M. to 10 P.M.) and off-peak (10 P.M. to 6 A.M.) delivery. (a) For delivery point regions, see Table 1 notes and text. (b) Newey-West standard errors assuming a five-day lag structure. Significance indicated for 1% (***), 5% (**), and 10% (*) levels.

TABLE 3
VOLATILITY OF PRICE SPREADS, PRE- AND POST-EXPANSION

Exchange vs. Bilateral Market Delivery Point Pairs[a]		Pre-Expansion (Apr.-Sep. 2004)	Post-Expansion (Oct. '04-Mar. '05)	Post / Pre Ratio	F Statistic of Ratio[b]	
		Standard Deviation of Day-Ahead Forward Price Spreads ($ per MWh)				
		Panel A: Peak Delivery				
PJM Western Hub	v. AEP-Dayton	10.58	6.63	0.63	38	***
	v. Cinergy	9.75	6.55	0.67	62	***
	v. NI Hub	11.77	7.39	0.63	78	***
PJM Allegheny	v. AEP-Dayton	9.95	6.56	0.66	26	***
	v. Cinergy	9.06	6.67	0.74	39	***
	v. NI Hub	11.10	7.24	0.65	49	***
		Panel B: Off-Peak Delivery				
PJM Western Hub	v. AEP-Dayton	11.36	6.36	0.56	42	***
	v. Cinergy	11.54	6.08	0.53	47	***
	v. NI Hub	12.44	9.28	0.75	69	***
PJM Allegheny	v. AEP-Dayton	11.50	4.73	0.41	48	***
	v. Cinergy	11.66	4.56	0.39	55	***
	v. NI Hub	12.56	7.57	0.60	82	***

Notes. Separate contracts are traded for peak (6 A.M. to 10 P.M.) and off-peak (10 P.M. to 6 A.M.) delivery. (a) For delivery point regions, see Table 1 notes and text. (b) Newey-West standard errors assuming a five-day lag structure, using delta-method for ratios. Significance indicated for 1% (***), 5% (**), and 10% (*) levels.

TABLE 4

CHANGES IN PRICE SPREADS BETWEEN MARKETS OVER TIME

Spreads are for PJM Allegheny vs. the AEP-Dayton bilateral market price

	Post – Pre Change in Average Daily Price Spread					
	Peak Delivery			Off-Peak Delivery		
Time Window	Percent	$/MWh		Percent	$/MWh	
–1 day to +1 day	-16%	-2.43		-46%	-5.12	
–4 days to +4 days	-25%	-3.93		-60%	-7.60	
–1 week to +1 week	-41%	-5.32		-65%	-8.18	
–2 weeks to +2 weeks	-52%	-5.55	*	-85%	-9.92	***
–1 month to +1 month	-10%	-0.77		-72%	-7.66	***
–2 months to +2 months	-9%	-0.51		-86%	-9.22	***
–1 quarter to +1 quarter	-15%	-0.68		-80%	-8.98	***
–2 quarters to +2 quarters	-43%	-3.01	**	-76%	-7.98	***

Notes. Table entries report the change in between-market average daily price differences (basis spreads) between the PJM Allegheny pricing zone and the AEP-Dayton bilateral market hub, for various pre- versus post-expansion time windows on either side of the market change date. Percent changes are relative to the average pre-expansion price spread for each window span. Statistical significance of price changes ($/MWh) indicated for 1% (***), 5% (**),and 10%(*) levels for windows exceeding 7 days, using Newey-West standard errors assuming a 5-day lag and Gaussian p-values.

SUPPLY FUNCTION PARAMETER ESTIMATES -- PEAK PERIOD

Standard errors in parentheses, except as noted.

		Mid-Atlantic Region						*Mid-West Region*					
		OLS		IV		IV/SUR		OLS		IV		IV/SUR	
Variable	Mnemonic	(1)		(2)		(3)		(4)		(5)		(6)	
Natural gas price, \$/MMBtu	PGAS	5.87		6.59		5.20		4.68		5.94		5.30	
		(2.48)	**	(2.53)	***	(2.99)	*	(4.93)		(5.02)		(5.80)	
Coal price, \$/MMBtu	PCOAL	59.26		64.69		66.44		78.11		77.16		52.84	
		(11.93)	***	(12.13)	***	(14.62)	***	(12.58)	***	(12.74)	***	(15.19)	***
SO_2 price, \$100/ton	PSO2	-6.71		-7.88		-6.37		-10.05		-10.14		-3.52	
		(2.11)	***	(2.15)	***	(2.58)	**	(2.60)	***	(2.65)	***	(3.13)	
NO_x price, \$100/ton	PNOX	1.13		1.11		0.36		-1.60		-1.56		-1.29	
		(0.19)	***	(0.19)	***	(0.23)		(0.18)	***	(0.18)	***	(0.22)	***
Net demand, GW	Q	4.00		4.21		3.45		1.51		1.60		1.46	
		(0.54)	***	(0.56)	***	(0.66)	***	(0.42)	***	(0.43)	***	(0.50)	***
Net demand × gas price	QxPGAS	-0.10		-0.12		-0.08		0.04		0.02		0.00	
		(0.06)	*	(0.06)	*	(0.07)		(0.06)		(0.06)		(0.07)	
Net demand × coal price	QxPCOAL	-1.90		-2.05		-2.03		-1.03		-1.02		-0.70	
		(0.31)	***	(0.32)	***	(0.38)	***	(0.15)	***	(0.15)	***	(0.18)	***
Net demand × SO_2 price	QxPSO2	0.34		0.37		0.31		0.17		0.17		0.09	
		(0.05)	***	(0.06)	***	(0.07)	***	(0.03)	***	(0.03)	***	(0.04)	**
Net demand × NO_x price	QxPNOX	-0.03		-0.03		-0.01		0.02		0.02		0.02	
		(0.00)	***	(0.01)	***	(0.01)		(0.00)	***	(0.00)	***	(0.00)	***
Constant	C	-118.05		-126.06		-100.43		-133.94		-141.67		-117.06	
		(21.35)	***	(21.76)	***	(26.18)	***	(34.17)	***	(34.77)	***	(40.87)	***
R-square		0.57		0.57		0.53		0.70		0.70		0.68	
First-stage F-stat (p-value)		n/a		15.20		18.93		n/a		15.20		18.93	
				(0.00)	***	(0.00)	***			(0.00)	***	(0.00)	***
N. Observations		1,244		1,242		1,242		1,242		1,242		1,242	

Notes. Reduced-form parameter estimates of text equation [14]. Net demand is regional demand plus net exports (positive for Midwest, negative for Mid-Atlantic). The IV estimates instrument for net exports with daily net weather differences (measured by degree day difference between Pittsburgh and Cleveland); the SUR estimates perform GLS with contemporaneous error correlation between regions. The Mid-Atlantic model pools PJM Western Hub and APS pricing points; Mid-West model pools Cinergy and NI Hub pricing points. Peak delivery period is 6am to 10pm. Significance indicated for 1% (***), 5% (**), and 10% (*) levels.

TABLE 6

SUPPLY FUNCTION PARAMETER ESTIMATES -- OFF-PEAK PERIOD

Standard errors in parentheses, except as noted.

		Mid-Atlantic Region						Mid-West Region					
		OLS		IV		IV/SUR		OLS		IV		IV/SUR	
Variable	Mnemonic	(1)		(2)		(3)		(4)		(5)		(6)	
Natural gas price, \$/MMBtu	PGAS	1.06		1.02		0.56		8.60		9.77		7.79	
		(0.54)	*	(0.55)	*	(0.70)		(2.79)	***	(2.87)	***	(3.45)	**
Coal price, \$/MMBtu	PCOAL	0.21		4.21		11.54		28.14		29.79		17.11	
		(4.18)		(4.28)		(5.44)	**	(8.28)	***	(8.45)	***	(10.31)	*
SO_2 price, \$100/ton	PSO2	2.79		2.09		1.56		-6.23		-6.88		-3.66	
		(0.70)	***	(0.72)	***	(0.91)	*	(1.65)	***	(1.69)	***	(2.06)	*
NO_x price, \$100/ton	PNOX	0.14		0.17		0.02		-0.44		-0.45		-0.45	
		(0.06)	**	(0.06)	***	(0.08)		(0.10)	***	(0.10)	***	(0.13)	***
Net demand, GW	Q	1.34		1.50		1.62		1.12		1.24		0.92	
		(0.22)	***	(0.23)	***	(0.29)	***	(0.30)	***	(0.31)	***	(0.38)	**
Net demand × gas price	QxPGAS	0.00		0.00		0.01		-0.09		-0.10		-0.08	
		(0.02)		(0.02)		(0.02)		(0.04)	**	(0.04)	**	(0.05)	
Net demand × coal price	QxPCOAL	0.15		0.00		-0.27		-0.45		-0.48		-0.28	
		(0.16)		(0.16)		(0.20)		(0.12)	***	(0.13)	***	(0.15)	*
Net demand × SO_2 price	QxPSO2	-0.05		-0.02		-0.01		0.12		0.13		0.08	
		(0.03)	*	(0.03)		(0.03)		(0.03)	***	(0.03)	***	(0.03)	**
Net demand × NO_x price	QxPNOX	-0.01		-0.01		0.00		0.00		0.00		0.00	
		(0.00)	***	(0.00)	***	(0.00)		(0.00)	***	(0.00)	***	(0.00)	**
Constant	C	-27.72		-31.55		-33.79		-72.98		-81.47		-57.76	
		(6.20)	***	(6.31)	***	(8.12)	***	(20.23)	***	(20.73)	***	(25.12)	**
R-square		0.76		0.76		0.75		0.60		0.59		0.59	
First-stage F-stat (p-value)		n/a		6.10		12.56		n/a		6.10		12.56	
				(0.00)	***	(0.00)	***			(0.00)	***	(0.00)	***
N. Observations		1,246		1,244		1,244		1,244		1,244		1,244	

Notes. Reduced-form parameter estimates of text equation [14]. Net demand is regional demand plus net exports (positive for Midwest, negative for Mid-Atlantic). The IV estimates instrument for net exports with daily net weather differences (measured by degree day difference between Pittsburgh and Cleveland); the SUR estimates perform GLS with contemporaneous error correlation between regions. The Mid-Atlantic model pools PJM Western Hub and APS pricing points; Mid-West model pools Cinergy and NI Hub pricing points. Off-Peak delivery period is 10pm to 6am. Statistical significance indicated for 1% (***), 5% (**), and 10% (*) levels.

TABLE 7

SUPPLY SLOPES AND NET EXPORT ELASTICITIES BY REGION

IV Method. Standard errors in parentheses.

	Peak Delivery Periods		*Off-Peak Periods*	
	Slope (MW/$)	Elasticity	Slope (MW/$)	Elasticity
Mid-Atlantic Region				
PJM Western Hub	726	48	796	14
PJM Allegheny	709	47	866	15
Pooled	715	47	826	14
Mid-West Region				
AEP-Dayton	1736	96	2671	33
Cinergy	1401	78	2631	33
N. IL Hub	1607	89	5296	66
Pooled	1526	85	3593	45

Notes. Slope figures report the derivative of each regions estimated supply function with repect to the market price, evaluated at the average factor prices during the first year post-expansion. Elasticity figures are the percent change in net exports (Mid-West) or net imports (Mid-Atlantic) with repect to a percent change in the regional market price. Pooled figures are based on the IV/SUR method estimates in Tables 6 and 7; individual delivery point figures based on the IV method estimates in Tables 5 and 6.

TABLE 8

GAINS FROM TRADE UNDER BILATERAL AND ORGANIZED MARKETS

		Peak Period	*Off-Peak Periods*	
	Total	Weekdays	Weeknights	Weekends
Price Spread Between Regions ($ per MWh) [a]				
Autarky price spreads				
Pre 10/2004 average (estimate)		10.78	12.37	14.39
Post 10/2004 average (estimate)		11.48	12.36	13.62
Bilateral market				
Pre 10/2004 actual average		9.79	9.60	12.01
Post 10/2004 counterfactual average		10.36	9.75	11.37
Organized market				
Post 10/2004 actual average		8.38	6.58	8.12
Difference post 10/2004 (Organized – Bilateral)		-1.97	-3.18	-3.24
Quantity Traded Between Regions (MW)				
Bilateral market				
Pre 10/2004 actual average	1455	978	2070	1741
Post 10/2004 counterfactual average	1485	984	2093	1816
Organized market				
Post 10/2004 actual average	3761	2923	4644	4425
Difference post 10/2004 (Organized – Bilateral)	2276	1939	2550	2610
Gains from Trade (million $ / year)				
Bilateral market				
Pre 10/2004 estimate	151.6	52.3	45.0	55.0
Post 10/2004 counterfactual	150.1	49.7	46.2	54.9
Organized market				
Post 10/2004 estimate	312.9	107.4	90.9	116.3
Change in Gains from Trade (million $ / year)	162.8	57.7	44.6	61.4
	(26.6)	(58.0)	(11.2)	(23.0)
Relative Efficiency of Bilateral Trade (in percent)	48%	46%	51%	47%

Notes. Peak periods are 6am-10pm weekdays; off-peak weekday periods are 10pm-6am, and off-peak weekend periods are all day. (a) Estimated price spreads based on the IV/SUR model results in Tables 6 and 7; actual price spreads are averages across the pricing points in Tables 1 and 2. See text.

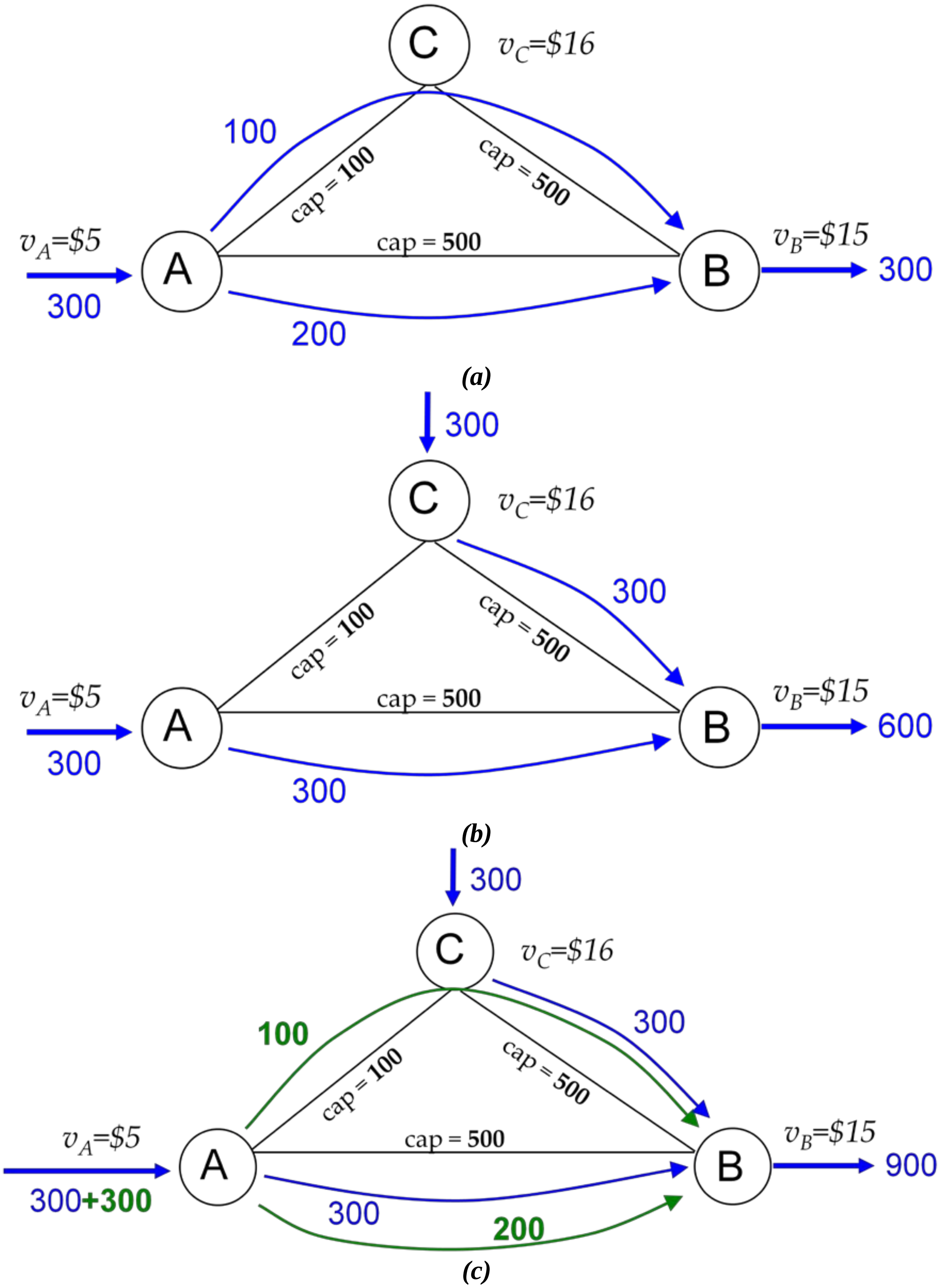


Figure 1: An Example of Counterflow Externalities

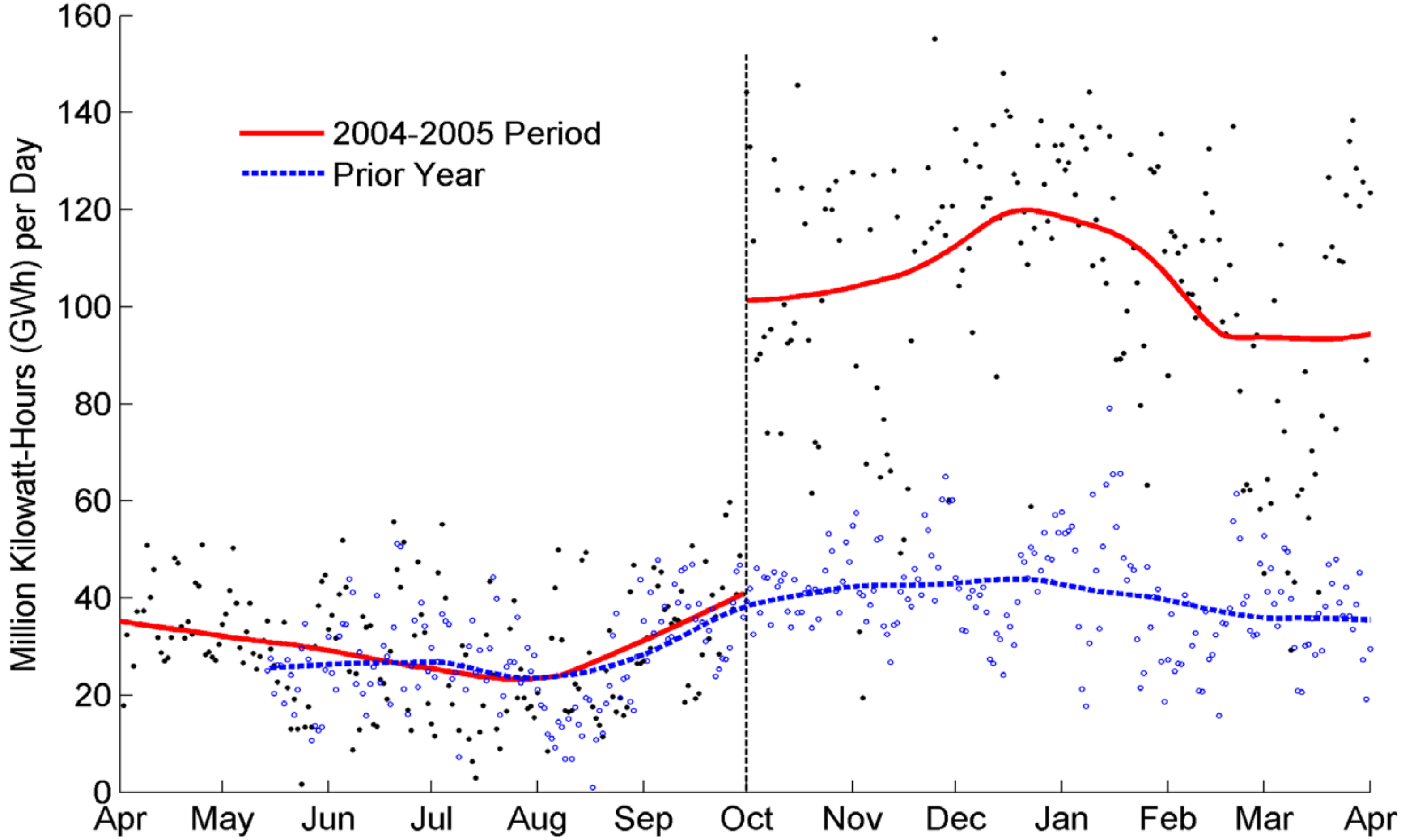
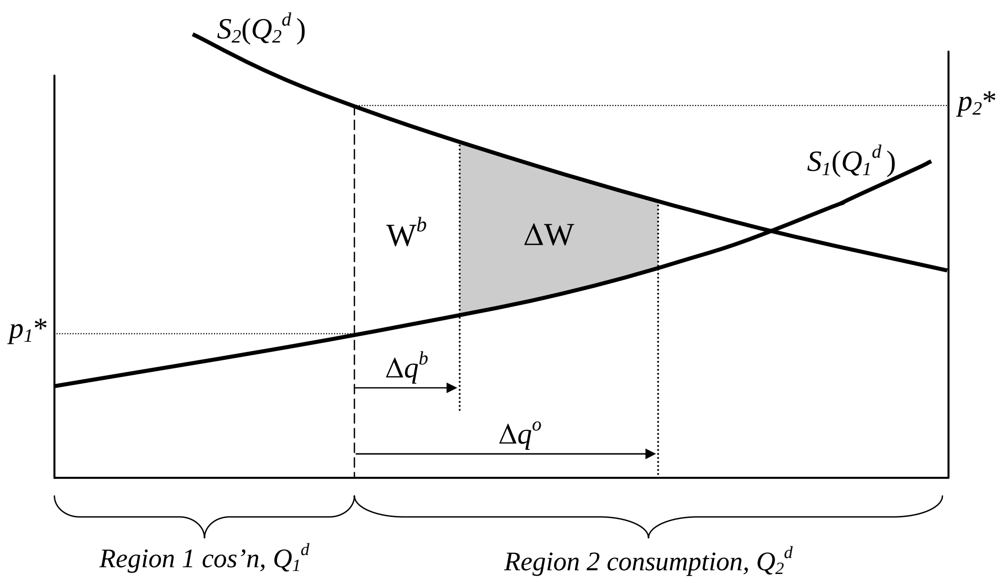


Figure 2: Day-Ahead Net Exports from Midwest to East (PJM)

Figure 3


$S_2(Q_2^d)$
p_2*
$S_1(Q_1^d)$
W^b
ΔW
p_1*
Δq^b
Δq^o
Region 1 cos'n, Q_1^d
Region 2 consumption, Q_2^d