Market Organization and Efficiency in Electricity Markets

Erin T. Mansur[†] and Matthew W. White[‡]

Discussion Draft

October 3, 2008

Abstract

Electricity markets exhibit two forms of organization: decentralized bilateral trading and centralized auction markets. Using detailed data on prices, quantities, and production costs, 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 organized markets achieve these gains because of superior information aggregation about trading opportunities in markets with congestion externalities.

[†] Yale School of Management, 135 Prospect St., P.O. Box 208200, New Haven, CT 06520-8200, and NBER. Email: erin.mansur@yale.edu.

[‡] The Wharton School, University of Pennsylvania, Philadelphia PA 19104-6372, and NBER. Email: matthew.white@wharton.upenn.edu.

Acknowledgments. The authors gratefully acknowledge the University of California Energy Institute for facilitating access to proprietary transaction data, and to the PJM Interconnection, LLC., for network data. We thank Andrew Ott and Joe Bowring of PJM for helpful discussions and information about market procedures, and Frank Wolak for comments on an earlier draft.

v 11.00

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.[1] In recent years this line of inquiry has turned increasingly normative as a new field of market design has 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 most regions of the United States, wholesale electricity markets operate as a decentralized, bilateral trading system. In a handful of areas, however, 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 energy production and delivery.

The merits of these two forms of market organization have become a point of extensive 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.

The premise of such policy initiatives is that adopting a well-chosen market design would improve market efficiency in areas where decentralized, bilateral trading practices prevail. This paper provides direct empirical evidence on this hypothesis. We do so by examining an abrupt change in a regional energy market's organization. In 2004 the largest organized electricity market in the United States, known as the PJM Interconnection, expanded to serve a large region of the Midwest where buyers and sellers formerly transacted exclusively through bilateral trading arrangements. For technical reasons, this change in the organization of trade was implemented on a single day. These and related features convey exceptionally clear information on the impact of changing a market's organization, from a decentralized bilateral system to a centralized auction market design, in a setting of considerable economic consequence.

Although the merits of organized market designs for electricity have been examined in experimental laboratory settings (*e.g.*, Olson et al., 2003), persuasive evidence has remained elusive. In general, the efficiency of unstructured bilateral markets depends how buyers and

[1] Extensive surveys include Kagel and Roth (1995) and Davis and Holt (1996).

sellers are matched to one another. Without strong assumptions about how this matching occurs and how market participants' information sets form, theory (and laboratory experiments) admit a wide range of outcomes. In real decentralized markets, participants' information sets are difficult for researchers to observe and characterize, making the relative efficiency of decentralized versus organized markets difficult to establish outside the field. To make informed decisions, it is desirable to determine empirically whether the efficiency gains of an organized market design are realized, and compare this magnitude with the cost of implementing it.

This objective differs from much of the burgeoning theoretical and empirical literature on electricity markets. Previous empirical 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 the widely-used bilateral market arrangement operates efficiently or not—and if not, whether adopting an organized market design improves economic efficiency. This compares the two workable market arrangements we observe in use, and seems the salient comparison for market organization debates.

Our empirical strategy exploits several attractive features of our setting. First, we have both market-level and plant-level data on supply, demand, prices, quantities, and production costs. This detailed information allows us to assess how market organization affects market performance along several dimensions and to measure the economic efficiency gains. These efficiency gains ultimately arise from supply-side allocative efficiency improvements, as trade reallocates production from higher-cost plants to lower-cost plants.

Second, the synchronized market change date yields a clear 'before versus after' demarcation of the two forms of market organization. There were no concurrent changes in the number—or even the identities—of the firms operating in these markets, nor their technologies or capacities. Nor were there any changes in the physical infrastructure of the (transmission) network that enables them to trade. Instead, what changed were the rules of the market: most importantly, how the participants' willingness to buy and sell was elicited, aggregated, and used to determine prices and allocate production. We find that these changes enabled the organized market to direct production to the most efficient available resources, realizing significantly greater gains from trade than occurred under the bilateral trading system.

The next two sections addresses the trade-offs between bilateral and centralized markets from a theoretical standpoint, emphasizing the problem of price discovery with network externalities. Section 4 summarizes our empirical strategy, and Sections 5 through 7 present empirical findings. A discussion of market implications follows the main findings.

2 Network Externalities and Information

Economic theory provides considerable guidance regarding why decentralized bilateral trading may yield different outcomes than centralized auction markets. These insights have motivated much work on electricity market design.[2] We articulate the main issues here, as these theoretical arguments play a central role in explaining our empirical results.

2.1 Background

In wholesale electricity markets, the main actors are producers (who own power plants) and retailers (local distribution utilities). Nearly all producers are either vertically-integrated or have long-term supply contracts with retailers. The vertically-integrated utilities tend to build, and the stand-alone retailers procure, sufficient capacity to cover their consumers' 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: It is efficient to idle one's plant if a producer with lower costs is willing and able to deliver the same quantity instead. The potential gains from trade in these 'spot' wholesale markets are considerable if producers have different marginal costs, which is common in practice (due to heterogeneous technologies, factor prices, capital vintages, and so forth).

An essential feature of electricity markets is that trade takes place over a network. Production can therefore create externalities, due to network congestion. Although these externalities are technological in origin, they are potentially difficult for market participants to resolve (in Coasian fashion) due to imperfect information. Congestion externalities imply a buyer and seller's maximum gains from trade depend on the production decisions of others elsewhere in the network. But others' production is private information that is not observable by all parties it affects. This means a potential buyer and seller often do not know which other market participants can alleviate network congestion—nor what it might cost to do so.

In this setting, the central question is whether different forms of market organization are capable of aggregating enough valuation and production information to resolve these externalities and achieve an efficient allocation. We explain precisely how these externalities arise next, then return to their implications for market design.

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

[2] See Wilson (2002) for a survey. Schweppe *et al* (1984) initiated considerable work on optimal prices in electricity networks.

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 1a. 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 have identical electrical characteristics, 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.[3]

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 $A \to B$ link and part will flow around the network $A \to C \to B$. 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, flows in energy networks 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, as well as the 300 units from A, withdrawing all 600 at B. See Figure 1b. 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 $A \to C$ link. A key property of 'displacement'-type networks, such as energy networks, is that flows have direction and offsetting flows cancel.

[3] Modeling complex (harmonic) flows and network losses would not alter the economic implications here.

Consequently, only a link's net flow is constrained by a link's capacity. In our example, this implies A's 300 units flow directly to B on the $A \to B$ link and C's 300 units flow on the $C \to B$ link. The net flow on the link between A and C becomes zero.[4]

Firm C's production has an externality benefit on 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 it 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 1c. 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 here is that C's production directly 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 because increasing production at the receiving end of a congested link counteracts the flow across it. In our example, reallocating production to exploit this positive externality increases total welfare by 90%, from \$3000 to \$5700, and triples the total volume of trade.

2.3 Information

The challenge these network externalities pose for efficiency is to ensure that all complementary trades are identified and consummated. How well different forms of market organization accomplish this task depends on the information available to participants.

Consider first the information required to determine an efficient allocation as a mathematical exercise. The first requirement is the structure of the network: The nodes, links between them, and their capacities.[5] This information is necessary to determine which transactions will create or alleviate congestion (and how the network constrains the feasible set of allocations). In practice, the structure of the network is well known to the firms that use it.

The second informational requirement is each firm's willingness to buy and sell at its locations. This is necessary to evaluate the gains from trade associated with any reallocation of production. In contrast to the network structure, a firm's valuation is its own private information.

[4]The reason is that, fixing output at 300 each, any flow of $\delta > 0$ between A to C would also increase the $C \to B$ flow to $300 + \delta$ and decrease the $A \to B$ flow to $300 - \delta$. If so, the flows would then shift to follow the path of lower resistance, or $A \to B$, bringing δ to zero.

[5]In practice, this also requires information on the impedances and electrical characteristics of network components. We abstract from these issues for transparency.

With complete information, the efficient allocation is a well-understood network optimization problem (even when various engineering complications are introduced; see Mamoh, 2000). Prior to the advent of competitive electricity markets, this 'complete information' approach was used by vertically-integrated firms that owned both the (transmission) network and all plants connected to it. Today these firms trade with one another across larger, interconnected networks, and many plants are owned by non-integrated producers. In this environment, the question becomes whether markets are able to elicit enough private information to 'discover' an efficient allocation.

3 Impediments to Efficient Trade

Economic theory suggests the complementarities created by network effects may impede a decentralized market's ability to discover an efficient allocation and its supporting prices. There are several reasons this may occur, which we describe next. This discussion also identifies why some impediments may be overcome with an auction market design, while others present difficulties for both forms of market organization.

To frame the key ideas, consider the allocative adjustment process in a decentralized market with and without these complementarities. Imagine first a classic exchange economy with a large number of traders: All goods are substitutes, there are no externalities, and all trade is arranged via bilateral transactions. Without a device to aggregate information about market participants' valuations, the market seeks an efficient allocation through an iterative process. At each iteration, an individual (1) elicits a price offer from a potential counterparty, (2) evaluates whether or not the offer is privately profitable to accept, and (3) accepts if yes or declines if no. (There may be counter-offers between steps 2 and 3, a peripheral issue here.) In the absence of externalities, if accepting the offer is privately profitable for the individual then accepting the offer is Pareto improving for the market. Thus each successive transaction moves the market closer to an efficient allocation.[6]

In a market with complementarities of the sort described earlier, the basic mechanics of this process are the same but the information required of a market participant is considerably greater. These greater requirements are not only apt to impede the process of equilibrium price discovery, it may well move the market *further* from the efficient allocation as trading proceeds.

[6]The economic theory of non-*tatonnement* price adjustment addresses when this iterative process converges to an efficient allocation. Classic treatments are Arrow and Hahn (1971) and Fisher (1972).

3.1 Finding Complementary Trades

For an individual market participant, the informational problem is to identify which transactions are complements and which are not. The difficulty is that a single bilateral trade may be infeasible, but multiple bilateral trades may be simultaneously feasible and efficient. For instance, consider again the example in Figure 1a-c: There one bilateral trade is infeasible alone (600 units from $A \to B$), and the other is inefficient alone (300 units from $C \to B$). But the combination of the two bilateral transactions is both feasible and Pareto improving.

If firm B knows the structure of the network, it seems simple enough for B to identify this complementarity by eliciting price offers from A and C: Since B is willing to pay up to \$15 per unit, any pair of prices that satisfy $p_A + p_C < \$30$ will do.[7] This simplicity is a deceiving consequence of a three-node example. In a 'mesh' network (one with multiple links per node), evaluating which bilateral transactions are complements may require knowledge of the valuations and production quantity at *every* node in the network. To understand why, some graph theory is needed.

Networks. 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 all edges, or links; and K the links' capacities. The ith element in E, denoted E_i, is a pair of connected nodes (u, v), $u < v$. For present purposes, it is useful to represent the network structure (V, E) by its *link matrix*. If $i = 1, 2, \ldots, m$ indexes links and $j = 1, 2, \ldots, n$ indexes nodes, the link matrix L 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. A non-zero entry in L simply indicates a particular link (row) connects to a particular node (column). Signs indicate a link's node-pair order.

Since complementary trades alleviate network congestion, we next characterize when this occurs. Let q be an allocation: An n-vector of quantities at the nodes (positive for injections, negative for withdrawals). Let f be the link flows (positive for $u \to v$ flows, negative for $v \to u$). An allocation q is feasible if aggregate supply equals demand, $\sum_j q_j = 0$, and the flow on each link respects its capacity, $|f| \le K$.

The central relation between flows and allocations is that all flows at a node must sum

[7] This assumes B demands (up to) 900 units, as before. The amount B demands matters because it affects whether A and C are complements or substitutes.

to zero.[8] In matrix form, this implies

$$q = L'f$$

as may be verified by direct expansion. Although this contains n equations, one is redundant. Dropping the last (arbitrarily) with no loss of information, we have an $n-1$ equation system we write as $\underline{q} = \underline{L}'f$. Solve for f:

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

T is known as the *transfer matrix.* Each element t_{ij} has $|t_{ij}| \leq 1$, and the (column) difference between two elements $t_{ij'} - t_{ij}$ indicates how a trade of one unit from node j' to node j changes the flow on network link i.

Although the link matrix is sparse, the transfer matrix is not. Therein lies the mathematical basis for the informational difficulties confronting efficient bilateral trade. A bilateral transaction of $\Delta q'$ units from a seller at location j' to a buyer at j creates an incremental flow on link i of $\Delta q'(t_{ij'} - t_{ij})$, and a new total flow on i of

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

T_i the ith row of T. Given a feasible allocation q, the bilateral transaction of $\Delta q'$ units is feasible if $|f_i'| \leq K_i$ for all i. We say the transaction *congests link* i if the ith constraint binds, $|f_i'| = K_i$.

Quantity Information. Observe the market allocation $\underline{q}$ enters (1), and so affects the feasibility of bilateral trades. The physical basis for this interdependency is the fact that energy flows across a network, in proportions given by T, across all possible network paths between j and j'. The market position—that is, the quantities—of any other firm located along those paths is its own private information, however. That has considerable ramifications: It implies the ability to determine if a bilateral transaction is feasible—thus whether it has value at all—depends on information the two transacting parties do not possess.

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 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

[8] This follows from Kirchhoff's laws on conservation of charge. See Howatson (1996).

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.

Complementary Trades. To characterize precisely when two bilateral transactions are complements, let us introduce a second 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 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.

Valuing Complementary Trades. Return now to the iterative process by which a decentralized market identifies and implements mutually profitable trades. When a single bilateral transaction is feasible, a buyer or seller determines its profit from accepting an offer by comparing its private valuation to the offered price. What if a bilateral transaction is infeasible? In that event, a buyer or seller's valuation for it depends on whether or not there are complementary trades that make the transactions jointly feasible. The difficulty lies in determining which transactions are complementary to it—if any—if the current market allocation q is not public information.

To appreciate why, 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. Call this set $S(q)$. It then needs to identify a change in the allocation that alleviates congestion on each link $i \in S(q)$, while not creating congestion along any 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 $j \rightarrow j'$. In a large network with many binding constraints in $S(q)$, there may be many non-zero elements in any vector Δ that satisfies (5).

It remains to determine if executing a set of potentially complementary bilateral transactions that implement Δ is profitable for j. In a bilateral trading system, this involves eliciting a price offer p_k for the quantity Δ_k from each firm with a non-zero element in Δ. A set of bilateral transactions implementing Δ is profitable for j if

$$\Delta q'(v_j - p_{j'}) - \sum_{k=1}^{n} \Delta_k p_k > 0. \tag{6}$$

Note that if (6) holds, the set of complementary transactions that implement it create gains from trade for all participating firms that could not be achieved by individual bilateral trades alone.

Efficiency. Market efficiency requires that participants ultimately identify all sets of complementary trades. This points to a fundamental reason why achieving an efficient allocation through sequential bilateral trade is considerably more difficult with complementary goods than if all goods are substitutes, as in a classical exchange market. The latter requires traders to determine no reallocation would create gains from trade between any *pair* of them. With complementary goods, however, they must also determine that no reallocation would create gains from trade among any *subset* of them. The difference between these requirements is that between n^2 tasks and $n!$ tasks—a difference that grows rather rapidly. In our application, the number of network nodes (or n) is of the order 10^4, and most node pairs are connected by multiple network paths. Identifying all possible sets of complementary trades necessary for an efficient allocation using a sequential bilateral trading system seems a heroic task to ask of market participants.

There is yet a further problem. The preceding discussion assumed firm j observed the current market allocation vector q. In reality, the components of q at locations other than its own are the private information of other firms. Consider now how that informational limitation affects the previous procedure for identifying potentially complementary trades.

Without knowledge of q, firm j 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 a set of candidate bilateral transactions that implement it to the transmission reservation system for a simultaneous feasibility check. For firm j, this is an iterative, trial-and-error process: the reservation system provides no gradient information that would enable j to determine in which direction to move Δ to find the convex polytope 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 amounts to a shot in the dark. For market participants to find all complementary trades using this trial-and-error method is effectively an insurmountable task.

Discussions with market participants pointed to a related problem that makes them hesitate to pursue the complementary trades they can identify (these are termed *redispatch* agreements). The set of complementary transactions that satisfy (5) depends on the current market allocation 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(a)-(c) 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 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.[9]

Information Aggregation. Fundamentally, there are two senses in which achieving market efficiency with congestion externalities can be regarded as an informational problem. At a broad level, the difficulty is that every market participant has 'small' bits of information—its valuation and quantity—but identifying some complementary trades requires the union of everyone's information sets. 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 suggests one possible avenue for an alternative market design that might speed the process of discovering efficient allocations.

[9] A closely-related exposure problem arises in auctions for complementary goods (a lucid treatment is Milgrom, 2004). In electricity markets, this problem is mitigated somewhat using options.

The second sense in which it is informational is that a different way of aggregating market participants' private information provides a rather 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. Mathematically, the problem of determining an efficient allocation of production becomes an optimization problem subject to the network's feasibility constraints, $|T\underline{q}| \leq K$. Provided valuation functions are (weakly) convex, this is a familiar problem solved in polynomial time. Thus, 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.[10]

3.2 Local Non-Improvement: The Multiplier Problem

The preceding analysis characterized the problem that network complementarities pose as impeding the market's ability to discover to an efficient allocation. There is another related, but conceptually distinct, concern. With externalities, implementing profitable trades may not be Pareto improving. That is, as sequential bilateral trades are executed the market may move *away* from the efficient allocation.

An example illustrates the concern. Imagine that a particular network link has a capacity of 80 units. A buyer and seller arrange transaction A for 80 units, of which 75% will flow between them across this link and the other 25% will flow between them on other network paths. Suppose the difference in their valuations is \$2 per unit. Now imagine two other parties arrange a higher-value transaction B for 1200 units, of which only 5% will flow across this link. The difference in their valuations is \$10 per unit.

Note that each transaction would flow 60 units across the network link in question, which exceeds its capacity of 80. What happens? Suppose that the lower-value transaction A was arranged first. When the second transaction B is later submitted to the reservation system, they would learn that the link has only 20 units of capacity available given all previously scheduled trades (that is, transaction A). Since 5% of transaction B flows across the link,

[10] 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).

the parties to transaction B are limited to a total trade of 400 units (of which 20 flow on the link, congesting it, and 380 flow through other network paths). Total gains from trade are $80 \times \$2 + 400 \times \$10 = \$4160$, on total trade of 480 units.

This is inefficient, because the partially-foreclosed transaction B creates greater gains from trade on the margin. The efficient allocation allows all of transaction B to occur, flowing 60 units on the link, and rations transaction A to the remaining 20 units (thus A transacts 20 / 75% = 26.7 units in total). Total gains from trade are $1200 \times \$10 + 26.7 \times \$2 = \$12,053$, on total trade of 1226 units.

There are several points to note here. The first is that transaction A has to be 'undone'—or at least $80 - 26.7 = 53.3$ units of it—by the firms in B. In this sense, a mutually profitable transaction (for the A firms) can move the market as a whole farther from the efficient allocation of production than if it had not occurred at all.

Getting to the efficient allocation when A arrives first entails three complementary trades: Paying A's seller to produce only 26.7 instead of 80 units, paying A's buyer to consume only 26.7, and implementing transaction B. With complete information there is little doubt this would be achieved. However, with imperfect information it may be difficult to do so for the reasons described earlier. The two parties to transaction B may need to know the entire market allocation vector q to determine which firms' transaction can be reduced to reallocate scarce network capacity to their higher value trade.

Let's suppose for the moment that the B firms did not manage to implement these complementary (re)trades. In that case, B would have been rationed to 20 units of this link's fixed capacity for arriving second. Normally, the welfare loss from rationing an inelastically supplied good (i.e., a link of fixed capacity) is strictly allocative: It distorts who receives the good, but not the total quantity of trade. That logic would imply the welfare loss from the inefficient initial allocation equals the number of units originally mis-allocated on the link, or 53.3, times the difference in the two transactions' gains from trade: $53.3 \times \$8 = \$426$. Yet the actual welfare loss is $\$12,053 - \$4160 = \$7893$, which is greater by a factor of approximately 18.

This phenomenon is known as the *multiplier problem.* It occurs because distorting the allocation of one element in a set of strictly complementary goods distorts the choices of the other goods. For the high-value transaction, one complementary good (the constrained link) is required in a fixed 1:20 ratio with another other complementary good (the other network links). These ratios are an attribute of the network structure, as determined by the elements in the transfer matrix T. Since all elements of T have modulus of one or less, any mis-allocation of a congested network link's capacity—regardless of how small—has a multiplier effect that magnifies the efficiency consequences.

The multiplier problem also operates on total quantities. This has the empirical prediction that, if a decentralized bilateral trading system is not able to re-allocate all congested network capacity efficiently, the volume of trade on the network as a whole may be sharply diminished.

The bottom line is that in electricity markets, a failure of a system of bilateral trade to converge exactly to the efficient allocation of network capacity can have large welfare costs. This motivates considerable attention to organized market designs that, if prices are close to valuations, will get the allocation of congested network links correct the first time. This seems a useful improvement over assigning the allocation of scarce network capacity as transactions occur, and hoping for the best in a secondary re-trading process subject to considerable informational limitations.

3.3 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.

4 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.[11] 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.[12] 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.

4.1 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.

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

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

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.

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, a central component of our empirical strategy will be to evaluate whether price spreads converged significantly after the organized market's expansion. When combined with the quantity data, this supports empirical inference about whether or not market efficiency improves after the organized market expands.

5 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.

5.1 Price Data

To examine whether between-market arbitrage improved, we assembled detailed market price data at daily frequency covering a two-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 the daily price indices for the 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 and were obtained directly from PJM.

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.[13] These delivery points are selected based on three simple 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 no reason why any additional information would bias bilateral market forward prices one way or another, relative to PJM's forward prices.

5.2 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

[13] 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 expansion, ostensibly by lesser amounts.

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 in 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.

Price Spread Convergence. The third column in Tables 1 and 2 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.[14]

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.[15]

[14] 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.

[15] 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).

[*Note here why price levels increased in PJM post instead of falling – issue is cost of fuels rose (more about which later).*]

Volatility of Price Spreads. 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.

Alternative Time Horizons. 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.

Prior Years. Did the same thing happen in prior years? The answer is no, unequivocally. A simple 'placebo analysis' is to replicate these calculations for a comparison period centered on October 1, 2003. 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. A two-year (difference-in-difference) comparison is provided in Appendix Table A-1, with related information (on quantities) in Section 5.2. In general, prior years' data indicate little if any seasonal differences in price spreads between delivery points.

Interpretation. 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 Figures 1a-c, 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.

6 Quantity Evidence

6.1 Quantity Changes from Transfers Data

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).[16] These day-ahead flows correspond exactly to the contracts whose prices are summarized in tables 1-3 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-horizonal 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.[17]

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 data from PJM indicate that this interface was congested eastbound—that is, handling the

[16] It makes little difference whether gross or net transfers are used. The transfers in Figure 2 are eastbound approximately 98 percent of all hours.

[17] 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.

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.

6.2 Quantity Changes in Plant-level Production Data

[*Preliminary*]

At a microeconomic level, there are efficiency gains from trade in wholesale electricity markets when trade reallocates production from high-cost plants to lower-cost plants. While the aggregate transfers results in the previous subsection indicate a major quantity reallocation took place post-expansion, it is useful to examine whether there is corresponding evidence for changes in production at the individual plant level.

To do this, we turn to a new, additional data source. To track compliance with air emissions regulations, the US Environmental Protection Agency continuously monitors the operating performance of major fossil-fueled power plants in the United States. Their *Continuous Emissions Monitoring System* (CEMS) database contains the hourly fuel consumption, emissions production, and gross electricity output for most power plants in our markets, nuclear plants excepted. We use the CEMS data to assess changes in output after the organized market's expansion, at the individual generating-facility level.

There were more than 600 generating units owned by members of the organized market at the time of its expansion, with an aggregate capacity exceeding 100,000 megawatts.

Consequently, detecting which plants operations were affected by the marginal change of slightly over 1000 megawatts in power shipments (on an average hourly basis) is a nontrivial task. However, the natural place to look for changes in production is among the subset of plants whose owners joined the organized market on October 1, 2004. These are the plants most likely to have realized new opportunities for trade under the organized market's design.

There are 66 such generating units in the CEMS data. The large changes in price spreads during the off-peak period, as shown in Panel B of Table 2, suggest we should see quantity changes for the largest, relatively efficient coal-fired power plants that are typically on the margin during off-peak periods. In fact, in our 66-unit group the six largest generating units saw a 22% increase in off-peak production over the six months post-integration relative to the six months pre-expansion. This is an enormous output increase, equal to approximately 1100 megawatts per hour (on an average hourly basis). By comparison, over the same calendar-month periods one year earlier these six plants' total off-peak production was virtually unchanged, at -0.2%. All six of these generating units are physically situated in the delivery region indexed by the AEP-Dayton pricing point. A change in production of this magnitude by such a large portfolio of generating stations is difficult to explain by any means other than that the owners of these plants found new opportunities to sell to distant buyers—starting in October 2004.

This net increase in production has an important additional implication for interpreting how improved efficiency arises post-expansion. The buyers and sellers in these markets are interconnected to a large number of other regional power markets throughout the Eastern U.S. As a result, the increase in quantities shipped eastward across the (transmission) network noted earlier could, in principle, reflect two different forms of arbitrage. One is an increase in production by low-cost sellers that joined the organized market, displacing higher-cost production elsewhere. Alternatively, the changes in (transmission) networks flows could reflect a shift in the destination market for the output from some other, more distant low-cost seller who did not join the organized market at all.

The distinction between these two possibilities relates to our understanding of how the organized market's expansion improved efficiency. Recall that since only members of PJM can trade in its forward markets, prior to the market's expansion the exchange's initial members and its members-to-be transacted in the bilateral market. One interpretation of the data presented so far is that the organized market is a more efficient venue than decentralized bilateral trade, resulting in increased production by low-cost firms that joined the organized market post-expansion. However, if the changes in trading patterns observed in the network power shipments in Section 5.1 were the result of increase production by other, more distant producers that never joined the organized market—thus traded bilaterally both before and after the expansion—then this interpretation would require some amendment.

For this reason, it is reassuring to observe in actual, metered-at-the-plant generation data that the price spread convergence after October 1, 2004 is matched by large production changes at the firms that joined the organized market. These are the firms and power plants that must have changed behavior to support the theory that overall market efficiency improved after they joined the market. Here we find that the increase in quantities transferred between delivery areas (in Figure 2) is due primarily to an increase in physical production by sellers who joined the organized market, and whose plants previously operated at lower levels before the market's expansion.[18]

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.

7 An Empirical Model of Imperfect Trade

We now present an empirical model of imperfect trade suited to our setting. Its primary purpose is to estimate the economic significance of the changes in price spreads that occurred after the organized market's expansion. This necessitates information on the elasticity of supply, and how it varies pre- and post-expansion, for the various delivery points affected.

In addition, we have a second objective. Over a very short (one day to one week) horizon, it is difficult to conceive that actual plant-level marginal costs in these markets change appreciably. However, over longer horizons this assumption may become tenuous. In particular, during the winter of 2004-05, the price of the fossil fuels and emissions permits that comprise the primary variable factors of production for producers rose significantly. In an environment were network congestion occasionally serves as a barrier to trade, it becomes conceivable that changes in these factor prices could account for a portion of the changes in price spreads over time. Consequently, before we base economic efficiency conclusions on the observed changes in price spreads and quantities, it is desirable to determine quantitatively if any portion of these changes can be attributable to exogenous changes in factor prices.[19]

[18] One interesting piece of anecdotal evidence comes from AEP's 2005 Annual Report to shareholders, which reports on the increase in output and profitability of this firm's low-cost, high-capacity coal generating capacity after AEP joined the PJM market (AEP, 2006).

[19] A similar argument can be made with respect to retail electricity consumption, changes in which we handle here as well.

7.1 A Model

To model the effect of integration on price spreads, it is useful to imagine a simple two-sector trade model with imperfect trade between them. Here the two sectors correspond to the organized market and the bilateral market. Consider Figure 3. Let $S_1(Q)$ be the inverse supply function of all producers who are members of the organized market, and let Q_1^d be the consumption of final consumers directly served by the members of the organized market (e.g., local distribution utilities). 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, aggregate consumption is indicated by the dashed vertical line.

Similarly, let $S_2(Q)$ be the inverse supply of all producers who are not (initially) members of the organized market and who trade at one of the bilateral-market delivery points. Here Q_2^d represents the electricity consumption of final consumers served by the firms that are not part of the organized market. In Figure 3, we have reversed the horizontal axis for market 2, so S_2 increases to the left.

In the absence of any trade between the organized market's members and non-members, the prices that would prevail in each market are indicated by p_1^* and p_2^*. Since there is always some trade between them in practice, we do not observe p_1^* and p_2^* directly. Instead, we observe prices p_1 and p_2 in each market, corresponding to a quantity of trade equal to (the width of) the shaded region in Figure 3. Note the prices that we observe in each market (p_1, p_2) are not assumed to capture all gains from trade, although they admit this possibility.

In this two-sector model, the effects of improved arbitrage are manifest through changes in quantities supplied, not through shifts in the supply curves themselves nor through demand changes. By contrast, changes in factor prices or (putatively exogenous) retail electricity consumption (e.g., due to weather) will shift supply or shift demand. Distinguishing these potential sources of price spread variation from improvements in arbitrage activity *per se* requires being able to separate out shifts of supply from movement *along* supply.

To do this, we relate observed prices to supply and demand fundamentals in each market. Since some trade occurs both pre and post, observed price spreads are bounded by the autarky spreads: $|p_{1t} - p_{2t}| \le |p_{1t}^* - p_{2t}^*|$. Equivalently, we can view arbitrage as reducing the markets' no-trade price spread by a (random) proportion ν_t each period:

$$|p_{1t} - p_{2t}| = \nu_t \, |p_{1t}^* - p_{2t}^*| \tag{7}$$

where $\nu_t \in [0, 1]$. At the extremes, $\nu_t = 1$ implies no gains from trade are being realized and $\nu_t = 0$ implies an efficient, integrated market. The exact (marginal) distribution of ν_t is unrestricted.

Changes in exogenous market conditions that shift supply and demand—factor prices, weather conditions, or the like—alter observed prices by changing p_1^* and p_2^*. Gains from trade that are due to improved trading efficiency between markets *given* these fundamentals are reflected in the random variable ν_t. Consequently, we are interested in assessing whether the average value of ν_t fell as a result of the market's expansion.

Although the no-trade prices p_1^* and p_2^* are unobserved, they characterize willingness-to-sell in each market: $p_i^* = S_i(Q_i^d)$, where S_i is market i's (inverse) supply curve. See Figure 3. To obtain an estimable model, we take logs of (7) to get

$$\ln|p_{1t} - p_{2t}| = \ln|S_{1t}(Q_{1t}^d) - S_{2t}(Q_{2t}^d)| + \ln\nu_t. \tag{8}$$

The first terms on the right-hand side strip the variation attributable to changes in factor prices and consumer demand from the total variation in observed (log) price spreads. It remains to specify a model for aggregate supply in each market, $S_{it}(Q_{it}^d)$.

The appeal of this model is that is yields a convenient decomposition and interpretation of changes in the efficiency of trade when there are confounding shifts in supply and demand curves. Let α_1 be the difference in mean (log) spreads post- versus pre-expansion:

$$\ln\nu_t = \alpha_0 + \alpha_1 I_t^{post} + \xi_t. \tag{9}$$

Let '$\%\Delta$' denote the change in the (absolute) average price spread post- versus pre-expansion, expressed as a percentage of the (absolute) average pre-expansion spread. With a little algebra this total percent change in average spreads can be decomposed into three components:

$$\%\Delta = m\,V\,\Omega - 1 \tag{10}$$

where $m = \exp(\alpha_1)$, V is (loosely speaking) the unexplained post-pre "variance" ratio

$$V = \frac{E[\exp(\xi_t)|I_t^{post} = 1]}{E[\exp(\xi_t)|I_t^{post} = 0]}\ ,$$

and

$$\Omega = \frac{E[\,|S_{1t}(Q_{1t}^d) - S_{2t}(Q_{2t}^d)|\,|I_t^{post} = 1]}{E[\,|S_{1t}(Q_{1t}^d) - S_{2t}(Q_{2t}^d)|\,|I_t^{post} = 0]}\ .$$

We can interpret Ω as the portion of the total percent change in average spreads that is attributable to changes in the (exogenous) observables, namely, factor prices and consumer electricity use, each day. In constrast, we can interpret $m\,V$ as the portion that is 'unexplained' by the observables. Intuitively, the total percent change in average price spreads due to integration is composed of a 'level' effect, $m = \exp(\alpha_1)$, and a reduction in the dispersion of price spreads given by the variance ratio term, V. This provides a convenient way to decompose the observed price spread changes in the separate components of interest. Note that, if the residual log-spread error ξ_t is normally distributed, then V simplifies to $\exp(\sigma^2_{\xi,post}/2 - \sigma^2_{\xi,pre}/2)$. An estimate of the percent change in average (absolute) spreads due to the markets' integration, holding changes in the exogenous observed factors constant, is then

$$\exp(\alpha_1 + (\sigma^2_{\xi,post} - \sigma^2_{\xi,pre})/2) - 1. \tag{11}$$

As written, the model in (8) presents a partial identification issue. To see why, note that we could subtract a constant c from the error term $\ln(\nu_t)$ and multiply each supply function by e^c and leave the left-hand side of (8) unchanged. Thus, a normalization is required. In estimation we impose (the sample analog to) $c = E[\ln \nu_t]$, which centers the disturbance term on zero. The unknown value of c can be recovered post-estimation using additional information contained in the model, viz., that aggregate supply and aggregate demand must balance. We discuss this step in the following section. Note that this normalization has no effect on the estimate of α_1 using the difference in the average fitted residuals from (8) post and pre, since the unknown value of c is differenced out.

7.2 Specifying Supply $S_{it}(Q)$

We now turn to an empirical model for market-level supply. We start with our knowledge of the underlying production technology: At the level of an individual plant (generating unit), electricity production is fixed-proportions (Leontief) in two variable factors of production, fuel and emissions permits. This implies the marginal cost of an individual plant k is

$$MC_{kt} = HR_k\,(p^f_t + ER_k \cdot p^e_t)$$

where HR_k is the plant's heat rate (inverse thermodynamic efficiency), p^f_t is the price of fuel, ER_k is the plant's emissions rate (NO_x and SO_2 production), and p^e_t is the price of emissions permits. The price of fuel and permits varies over time but does not vary (in our data) across firms or plants in the same region. If, in addition, the emissions rate is constant across plants, then the market level marginal cost function for plants of fuel type f with aggregate output Q can be written as

$$MC^f_t(Q) = g^f(Q)\,(p^f_t + ER^f p^e_t)$$

where $g^f(Q)$ is a monotonically increasing step function. Empirically, the assumption that all plants in a region that use 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 negligible, as a plant's total permit costs are small relative to its fuel costs and the overwhelming determinant of the marginal cost structure is the first two terms, $g^f(\cdot)p_t^f$.[20]

In these markets, there are two production technologies that set the markets' prices: gas- and coal-fired generation.[21] We can combine the two to achieve a market-level marginal cost function using an indicator variable I_t^f for the marginal technology type at time t, giving

$$MC_t(Q) = \textstyle\sum_f \; I_t^f \, g^f(Q) \, (p_t^f + ER^f p_t^e)$$

where f indexes fuel type (coal or gas) and $I_t^f = 1$ if and only if type f is the marginal (price setting) technology at t. We observe the marginal fuel type indicator variable I_t^f directly in the organized market.[22]

To complete the specification, we need to match the frequency of the price data. These cover either 16 hour (peak) or 8 hour (off peak) delivery durations. Since we observe retail electricity consumption in each region at an hourly frequency, we aggregate up to daily (16 or 8 hour) market supply functions explicitly. If we let h index hours and now set t to index the day (price block), and a new subscript i to index each markets, the market-level supply (per megawatt hour) in market i on day t becomes

$$S_{it}(Q_{it}^d) = \gamma_{i0} + \frac{1}{H} \sum_{h \in t} \sum_{f=c,g} I_{ih}^f \, p_{ih}^{total} \, g^f(q_{ih}^d) \tag{12}$$

where the factor price and emissions-related terms are condensed to

$$p_t^{total} = (p_t^f + ER^f p_t^e) \ ,$$

and Q_{it}^d is now a vector of (either 16 or 8) hourly consumption observations $\{q_{ih}^d\}$ during day t.

In aggregating from marginal costs to market supply in (12), we implicitly assume that producers' willingness to sell is a (constant) proportional markup over their marginal cost. 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 approxi-

[20]In the markets we study, fuel expenses are 80 to 90 percent of plants' marginal costs (see PJM 2005, p. 106).

[21]In eastern Pennsylvania, there are a small number of oil-fired plants still in operation. Data provided by PJM indicate these units are not relevant for explaining prices at PJM West and Allegheny, which are in central and southwestern Pennsylvania, except under unusual circumstances.

[22] These data are courtesy of the PJM Market Monitoring Unit. For the Midwestern bilateral delivery points, we assume coal is always the marginal fuel. This is consistent with data from system operators in these areas and fuel prices during the period studied (viz., the 2003 and 2004 EIA-714 system-λ data for the AEP, Cinergy, and FirstEnergy transmission control areas).

mately 3.4 percent; PJM 2005, p. 68). In our model, any differences between marginal cost and willingness to sell that affect a market's price will be absorbed into the estimated productive efficiency function, g^f. As a result, our assumption that this technology-dependent function g^f is time invariant is not entirely innocuous; one possible concern is that, coincident with (or as a result of) of the market's expansion, suppliers may have changed their behavior and that in turn may affect price spreads. The plant-level data help inform this, and we address it further below.

Supply Model Data. We briefly summarize a few features of the data that enter the supply function here. For prices set by the organized market, the variable q^d_{ih} represents the total retail electricity consumption in hour h of consumers served by local distribution utilities that are (pre-existing) members of PJM. For prices set by the bilateral market, the same applies for utilities that are not initial members of PJM. These were obtained from PJM and the major utilities serving load in the Midwestern regions; a complete list and technical details are given in the Appendix.

Second, although the productive efficiency function g^f is technically a monotone step function, there are hundreds of individual generating units in our markets. This makes the differences between g^f and a smooth polynomial approximant to it small, especially over the relevant range of quantities we observe. In estimation, we find no evidence that third or higher-order terms have any effect on our empirical results; thus, the results reported below are based on second-order polynomials for g^f.

Last, there is the possibility that the marginal fuel indicator I^f_h might be endogenous in the price spreads equation (8). The reason is that successful arbitrage (between Midwest producers and buyers on the East coast) may tend to displace gas-fired production and make coal more likely to be the marginal fuel (in PJM). Although we think the extent to which this occurs in minor, we nevertheless instrumented for I^f_t in the PJM supply specification. For instruments we projected observed I^f_t onto an array of weather variables and time dummies (by hour of day, month, season, and their interactions), giving us predicted marginal fuel indicators that are functions of time and weather alone.[23]

For the Midwestern bilateral market prices this problem does not arise, as system-operator data for those areas indicate the marginal fuel is always coal (*c.f.* note 22). Further details about fuel data sources and their construction are provided in the Appendix, including fuel prices, point sources, delivery costs, and plant emissions rate information.

[23] We are also missing marginal fuel data for the organized market in 2003, and use these predictions in their place.

7.3 Estimated Effects of Expansion on Price Spreads

Table 6 summarizes the estimated model's results with respect to price spread changes. The raw parameter estimates are provided in Appendix Table [X]. We discuss the price spread results here, and the estimated supply functions and gains from trade in Section 7.

Column (1) of Table 6 reports the straight pre- versus post-expansion average (absolute) price spread changes, as a percent of the average (absolute) spread during the pre- period. These are the same percentage changes reported in our interpretation of Table 1, and are provided here as a benchmark for subsequent comparisons. The results here and in column (2) employ the same six months of data pre- and post-expansion as in Table 1. Price spreads for the AEP-Dayton versus PJM delivery point contrasts are omitted from Table 6, because data non-availability prevents us estimating the results in columns (3) and (4) (The AEP-Dayton bilateral market prices were not collected systematically prior to 2004).

To obtain the results in column (2), we fit the supply function model specification (8) separately for each market-specific pricing point pair shown. This was done by nonlinear least squares. Using the fitted models, we can now identify the portion of the total percent change in price spreads (shown in column (1)) that is *not* attributable to changes in fuel prices, emissions permit prices, or variation in retail electricity consumption. This "unexplained" change in the price spread after the market's expansion, evaluated using equation (11), is shown in column (2).

The striking feature of these results is the fact that factor price and retail consumption variation account for little of the observed changes in price spreads. In the peak-period prices shown in Panel A, the estimated percent changes in average spreads in columns (1) and (2) are nearly identical. In the off-peak prices shown in Panel B, the estimated percent changes in average spreads in column (2) are 14 to 26 percentage points lower in magnitude than the percentage changes in column (1). This indicates that perhaps a quarter of the total decline in price spreads off-peak may be attributable to changes in weather, consumption, or factor prices that would have lowered the price spreads in any event.

To help understand why variation in suppliers' costs—which rose significantly in late 2004—had little apparent effect on market prices spreads, we examined to the underlying data. In general, it is asymmetric changes in suppliers' marginal costs that will tend to change the price spread between markets. By examining the raw fuel prices as well as weather data for these delivery areas, one obvious explanation emerges: While there are large changes in these variables over time, these changes rarely asymmetric. The cost of coal delivered to central Pennsylvania producers near the PJM Western Hub pricing point and the cost of coal delivered to southern Ohio producers near the Cinergy Hub pricing point tend to move together. This is sensible: Coal is shipped to both sets of generators (primarily) from the same Appalachian coal mining regions, so the difference in fuel price levels between the two regions largely reflects transportation costs—which are stable over this period.

A similar set of observations applies to retail electricity consumption. The day-to-day variation in electricity consumption is primarily attributable to weather (see, *e.g.*, Reiss and White, 2008). Daily weather data indicate that weather in Pittsburg (near the PJM Allegheny delivery area) and in Columbus (near the AEP-Dayton delivery area) are not appreciably different. In sum, the delivery point areas in Table 1 are not far enough apart to experience asymmetric changes in retail electricity consumption or suppliers' input factor prices.

Overall, there is scant support for the conjecture that price spreads would have converged if the markets' expansion had not occurred, due to changes in producers' input factor prices or due to variation in retail electricity demand.

7.4 Comparison Year Contrasts

To help further rule out possible confounding factors, we have also examined the variation in price spreads during a comparison year. There were no changes in the market's organization (of any kind) during this comparison year.

Making a comparison year contrast serves to net out any possible seasonal effects on price spread changes. Our pre- versus post-expansion comparisons in Table 1, and columns (1) and (2) of Table 6, are based on one year of data centered on the PJM expansion date of October 1, 2004. This means that the six months of pre-expansion data span the summer months (April to September), while the post-expansion data span the winter months (October to March). As a general matter, there is no obvious reason to suspect a purely seasonal effect to price spreads (*i.e.*, why would firms be worse at arbitrage in the summer than in winter?). Nevertheless, we addressed this possibility quantitatively.

To match the seasonal change in columns (1) and (2) exactly, we selected a symmetric comparison period: the 12 months centered on October 1, 2003. We then re-estimated the model using data for the entire two year span from April 1, 2003 to March 31, 2005. As noted above, the price data for the AEP-Dayton bilateral market delivery point do not extend back beyond 2004, so are not included in this analysis. The other two bilateral market delivery points are not available much earlier than 2003, which determined our choice of the comparison year period.

For the results reported in column (3) of Table 6, we first estimated a comparison-year contrast specification without a control function for factor prices and retail consumption:

$$\ln |p_{1t} - p_{2t}| = \alpha_0 + \alpha_1 I_t^{Post} + \alpha_2 I_t^{Year2} + \alpha_3 I_t^{Winter} + \varepsilon_t$$

where the indicator $I_t^{Post} = 1$ for t on or after October 1, 2004 (post-expansion), $I_t^{Year2} = 1$ for t on or after April 1, 2004, and $I_t^{Winter} = 1$ if t falls in the months of October through March (inclusive) in either year. We then transformed the estimated value of α_1 and the variances of ε_t before and after expansion into an estimated percent change using (11). Here this percent change is relative to the change during the comparison year before and after October 1, 2003. These calculations are done separately for all delivery point pairs listed in Table 6, peak and off-peak.

The estimated percent change in spreads, relative to the comparison year change, is shown by delivery point pair in column (3). These results are little changed from those in columns (1) and (2). The reason is straightforward: The comparison year data show little change in the average (absolute) price spread over the six months after October 1, 2003, relative to the six months before October 1, 2003. We infer that without a structural change in the markets' organization, changes in demand between summer and winter do not result in changes in between-market price spreads.

Things are slightly different in column (4), but the same conclusion emerges. Here we first estimated the model in equation (8) by nonlinear least squares, using the full two years of data. We then obtained comparison-year contrasts for the unexplained (log) spread by linearly projecting the residuals from the fitted model (8) onto the time-period indicators used for column (3), or

$$\widehat{\ln \nu_t} = \alpha_0 + \alpha_1 I_t^{Post} + \alpha_2 I_t^{Year\ 2} + \alpha_3 I_t^{Winter} + \xi_t \tag{13}$$

This was done separately for each delivery point pair listed in Table 6, peak and off peak. We then used (11) to obtain point estimates of the percent change in spreads relative to the change during the comparison year, net the effects of any simultaneous changes in factor prices or retail energy consumption levels.

As with the straight pre- versus post-expansion comparisons, controlling for variation in costs and retail consumption results in modest changes in estimated effect of integration on price spreads. Interestingly, the estimated percent changes in spreads (net the effects of the observables) are slightly larger than in columns (1)-(3) for the peak periods in Panel A, and slightly smaller than in the other columns for the off-peak periods in Panel B. In addition, the explanatory power of the control function increases dramatically; while the R^2 values for models in columns (1) and (3) are typically about .1, the R^2 values for the models in column (4) range from .6 to .7. Thus, the control functions are helping to explain a portion of the observed spread changes when contrasted against the relationship between these factors and price spreads during the comparison year.

Summary. In sum, by combining the initial price spread evidence from Section 4 with the analyses performed here, six major observations emerge:

(i) Price spreads between markets fell dramatically after the organized market's expansion, for all contrasting delivery points, in both peak and off-peak periods;

(ii) The magnitude of price spread convergence is substantial, both relative to pre-expansion price spreads and relative to the comparison-year spread changes;

(iii) These changes are apparent within days of the expansion, and price spreads remained smaller persistently thereafter;

(iv) This convergence of price spreads cannot be explained by changes in factor prices, or by demand conditions;

(v) Nor is there any evidence that price spreads change systematically at the same calendar times in prior years, when no structural changes in market organization occurred;

(vi) Nor were there any systematic changes in transmission or productive capacity during the time span studied here.

On this basis, we are led to conclude that the expansion of the organized market caused the substantial change in wholesale electricity market prices we observe in these data. In addition, the quantity evidence in Section 5 suggests that these price changes were accompanied by real changes in which firms and production facilities generated the power consumed by many (indeed, millions) of consumers. Taken together, these results imply that the organized market design was able to identify gains from trade—reallocating production to the lowest-cost producers—that were not being realized by the previous bilateral market arrangements.

8 Efficiency Gains

The model of imperfect trade presented in Section 6 enables us to estimate how large are the newly-realized gains from trade after the organized market's expansion. Here we evaluate its predictions for how trade differed from what would have happened if, counter to fact, the market's expansion had never occurred.

To put our method for calculating efficiency gains on a clear conceptual foundation, return to the simple two-sector model of trade from Section 6. Let Δq_t^a be the actual quantity traded between two market-specific delivery points on day t, for t after October 1, 2004. Similarly, let Δq_t^c be the *counterfactual* quantity that would have been traded after October 2004 'but for' integration. Because price spreads converged post-integration we know $|\Delta q_t^a| > |\Delta q_t^c|$, that is, trade volume increased as a result. Our analysis is based on how suppliers' willingness to sell at different market-specific delivery points varies over the range of quantities between Δq_t^a and Δq_t^c.

At any time t after October 1, 2004, the welfare gain attributable to integration is the difference in sellers' valuations for the incremental quantities traded:

$$W_t = \int_{\Delta q_t^a}^{\Delta q_t^c} S_{1t}(Q_{1t}^d - \theta)\, d\theta - \int_{\Delta q_t^c}^{\Delta q_t^a} S_{2t}(Q_{2t}^d + \theta)\, d\theta \tag{14}$$

The reversal of the two integrals' limits, and the signs on θ, reflect only sign conventions regarding the direction of trade (Δq_t^a and Δq_t^c are positive for net flows into market 1). In the context of Figure 3, the value W_t is the difference between the shaded area depicted in the figure and the smaller shaded area that would apply if, counterfactually, the price spread was at the higher level that would have prevailed without the market's expansion. We will estimate W_t separately for each day and each price block (peak and off peak), then sum these values to obtain total welfare over T-period horizons.

Implementation. To evaluate the gains from trade using (14), we need to estimate the counterfactual price spreads $|\Delta p_t^c|$ and net flows Δq_t^c that would have applied for t after October 1, 2004, 'but for' integration. In addition, we require estimates of the actual between-market flows for the delivery point pairs where these data are not directly observed.

We proceed in two steps. First, we use the fitted spreads model to predict the counterfactual price spread each day t. Second, we then solve the model 'in reverse' for the quantity traded (Δq_t^c) that yields the counterfactual price spread. We replicate this second step with the actual price spreads after the market's expansion, providing the estimate of Δq_t^a. Because our supply function specification (12) is a polynomial in quantity, the integral in (14) is simple to evaluate once we determine the limits Δq_t^a and Δq_t^c.

The first of these steps is straightforward. To estimate the counterfactual price spread, we use the comparison-year contrast model based on (8) and (13). If the market's expansion had not occurred, then the "unexplained" drop in price spreads relative to the comparison year change, or α_1 in (13), would be zero. Setting $\alpha_1 = 0$ implies a counterfactual price spread of

$$|\Delta p_t^c| = |S_{1t}(Q_{1t}^d) - S_{2t}(Q_{2t}^d)| \cdot \exp\left(\alpha_0 + \alpha_2 I_t^{\mathit{Year\ 2}} + \alpha_3 I_t^{\mathit{Winter}} + \xi_t\right) \tag{15}$$

which is equivalent to

$$\Delta p_t^c = \Delta p_t / \exp(\alpha_1). \tag{16}$$

where Δp_t is the observed price spread. The absolute values are unnecessary in (16) as the price difference on each side always has the same sign. We obtain the estimated counterfactual price spread by plugging in the estimate of α_1 from the fitted models in Section 6. The counterfactual price spreads therefore vary (as indicated in Table 6) by delivery point pair and period (peak and off peak). Note that, conceptually, here we are (implicitly) assuming that the same disturbance term ξ_t 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 our model that might affect spreads (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.

To obtain the quantities traded, we use a second implication of the two-sector imperfect trade model from Section 6. Specifically, the price in each market is given by willingness to sell (inverse supply) evaluated at the autarky quantity, less net imports or exports:

$$p_{1t} = S_{1t}(Q^d_{1t} - \Delta q_t) \tag{17}$$

$$p_{2t} = S_{2t}(Q^d_{2t} + \Delta q_t) \tag{18}$$

Here $\Delta q_t > 0$ for exports from market 2 to 1, and negative for imports from 2 to 1. Subtracting gives

$$\Delta p_t = S_{1t}(Q^d_{1t} - \Delta q_t) - S_{2t}(Q^d_{2t} + \Delta q_t) \tag{19}$$

This is a 'markets-clear-in-quantities' condition. It must hold for any set of price spreads; all it says is that imports into market 1 from market 2 must equal exports from market 2 to market 1. Since that must hold for any price spread, it also defines the counterfactual quantity, Δq^c_t, that would yield the counterfactual price spread, Δp^c_t.

In principle, solving (19) for Δq_t using the observed price spreads yields the quantity Δq^a_t needed to evaluate welfare using (14). Repeating the process using the counterfactual price spreads yields the counterfactual traded quantity Δq^c_t that we predict would have transacted had the market's expansion not occurred.

In practice, there is a small wrinkle. Recall that the model in (8) identifies each market's supply function only up to scale. The normalization imposed during estimation implies that the actual and estimated supply function are off by an unknown constant e^c, or

$$\hat{S}_{it}(Q) = e^c\, S_{it}(Q) + \text{error}$$

Estimates of welfare using quantities obtained from (19) using $\hat{S}_{it}(Q)$ would therefore tend to be systematically off by a factor of e^c.

To handle this, we can use the 'markets-clear-in-quantities' condition to estimate c along with Δq_t. Now (17)-(18) become

$$p_{1t} = \lambda\, \hat{S}_{1t}(Q^d_{1t} - \Delta q_t) + \text{error} \tag{20}$$

and

$$p_{2t} = \lambda\, \hat{S}_{2t}(Q^d_{2t} + \Delta q_t) + \text{error} \tag{21}$$

where $\lambda = \exp(-c)$. Since p_1 and p_2 are observed, we know everything except for the value of λ and the T values of Δq_t. That means (20) and (21) comprise $2T$ equations in $T+1$ unknowns. Solving jointly via least squares provides c as well as estimates of the actual quantities for each day, $\Delta \hat{q}^a_t$, that were not observed in our data. Last, we obtain the counterfactual quantities by solving (19) for Δq_t using the counterfactual prices on the left-hand side and the corrected supply function estimates $\hat{\lambda}\hat{S}_{it}$ for S_{it} on the right.

Results. We use the fitted models to evaluate the gains from increased trade between each pair of delivery regions listed in Table 6, for both peak and off peak periods. 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.

These gains from trade represent a substantial increase from the corresponding gains from trade achieved under the preceding bilateral trading system. Under that system, we estimate aggregate gains from trade between regions of approximately \$135 million per annum. After adopting the organized market design, these increased to between \$293 and \$316 million annually. This, it appears that decentralized bilateral trading is able to realize about 45% of the gains from trade achieved with a more efficient market design.

In general, the counterfactual price spreads we obtain from (16) are quite similar to the price spreads that we observed prior to the market's expansion. This is expected, given the results in Table 6; there, we found that little of the total change in price spreads pre- versus post-expansion could be explained (statistically) by the increase in producer's factor prices that occurred in the winter of 2004-05.

Cost of Market Expansion. 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 \$4-5 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 \$162 to \$181 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.

9 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.

10 Conclusion

Our motivation for this paper arose from a vigorous—and, we believe, poorly informed—policy debate about the merits of organized market designs in liberalized electricity markets. This debate reflects two distinct, but related difficulties that confront policy makers. First, the potential for a more efficient market design to reallocate production from high-cost firms to lower-cost competitors creates a political incentive for market participants that stand to lose to oppose it. Second, there is the technocratic challenge that the theoretical appeal of a well-specified market design must be balanced against the cost of implementing it. Given these incentives and challenges, it is not surprising that a consensus among industry participants on this fundamental trade-off has proved elusive.

The central contribution of this paper is to provide a detailed empirical assessment of this problem. The expansion of the organized market design used by the PJM Interconnection in 2004 provides one of the only opportunities to address it empirically. As industry participants rapidly discovered, there were dramatic changes in market outcomes after the expansion: lower-cost facilities increased production, price spreads between Midwestern and Eastern delivery points converged, and the quantities of power transferred between them increased substantially. Indeed, after witnessing the changes that followed PJM's 2004 expansion, several other utilities that previously traded only in bilateral market venues joined PJM's market.

We are led to the seemingly inexorable conclusion that the organized market design identified new trading opportunities that were simply 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 newfound gains from trade clearly calls into question the assertion that organized market designs are not worth their costs of implementation. We hope these findings prove useful to future decisions regarding electricity market design.

Appendix A.

[TBD]

References

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).

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

Blumsack, S. (2008). get ref.

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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

Schweppe, Fred C., Michael C. Caramanis, Richard D. Tabors, and Roger E. Bohn (1987). *Spot Pricing of Electricity.* Boston, MA: Kluwer Academic Publishers.

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

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

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. Ind. & Ohio)		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 areas of the high-voltage transmission grid, not single points on a map; the points above correspond to geographic regions as follows (approximately): PJM Western Hub is central and western Pa. and northern Maryland; PJM Allegheny is southwestern Pa. and northern W. Virgina; AEP-Dayton is central Ohio and southern W. Virginia; 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, with 1% (***), 5% (**), and 10% (*) significance 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, with 1% (***), 5% (**), and 10% (*) significance levels.

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

Exchange vs. Bilateral Market Delivery Point Pairs[a]		Standard Deviation of Day-Ahead Price Spreads ($ per MWh) Pre-Expansion (Apr.-Sep. 2004)	Post-Expansion (Oct. '04-Mar. '05)	Post / Pre Ratio	F Statistic of Ratio[b]
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. Entries are the change in average between-market daily price differences (basis spreads) between the PJM Allegheny pricing zone and the AEP-Dayton bilateral market hub, for various post- versus pre-expansion time windows on either side of the October 1, 2004 market change date. Percent changes are relative to the average pre-expansion price spread for each window span. Statistical significance of the price changes (in $/MWh) is indicated at 1% (***), 5% (**),and 10%(*) levels for windows exceeding 7 days, using Newey-West standard errors assuming a 5-day lag structure.

TABLE 5
PLANT-LEVEL OUTPUT CHANGES, POST- VS. PRE-EXPANSION

Same calendar-month contrasts: Post-expansion is Oct. 2004-Mar. 2005, pre-expansion is Oct. 2003-Mar. 2004.

Plant Type and Cost, by Decile	Marginal Cost Interval ($/MWh)	Total Capacity (in MW)	*Capacity Utilization* Pre-Expansion	*Capacity Utilization* Post-Expansion	Average Hourly Output Change (In MW)
Nuclear Plants	$ 4.8 to $ 10.0	55,274	84%	83%	-832
Low-Cost Fossil-Fuel Plants					
1st decile	$ 14.6 to $ 19.5	33,407	59%	63%	1079
2nd	$ 19.6 to $ 27.2	33,379	65%	68%	1043
3rd	$ 27.2 to $ 30.1	33,922	66%	69%	1048
High-Cost Fossil-Fuel Plants					
4th decile	$ 30.1 to $ 32.4	33,950	58%	57%	-425
5th	$ 32.4 to $ 33.8	33,338	52%	50%	-767
6th	$ 33.8 to $ 38.5	32,419	49%	47%	-736
7th	$ 38.6 to $ 56.0	34,703	13%	14%	460
8th	$ 56.0 to $ 71.1	33,857	12%	13%	194
9th	$ 71.1 to $ 92.6	32,058	4%	3%	-144
10th	more than $ 92.6	33,351	12%	11%	-196
Low-Cost Plants Overall: Fossil (Deciles 1-3) and Nuclear Plants			70.8%	72.3%	2337
High-Cost Plants Overall: Fossil (Deciles 7-10) Plants			28.6%	27.9%	-1614

Notes. Table includes (essentially) all fossil-fired and nuclear power plants in the contiguous 15-state region east of the Mississippi River and north of Tenn. and N. Carolina (inclusive), excluding New England. Capacity is nameplate. Each plant's marginal cost includes fuel, emissions, and variable O&M costs; for data sources and calculations, see text and Appendix [XX].

TABLE 6

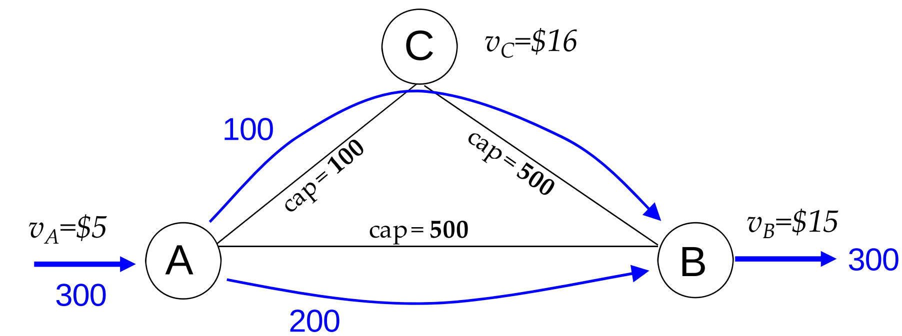
ESTIMATED EFFECTS OF MARKET EXPANSION ON PRICE SPREADS

Percent change in average daily absolute price spreads between markets,
for eight delivery point pairs by delivery period. Standard errors in parentheses.[a]

	Pre v. Post		*Difference-in-Difference*	
Exchange v. Bilateral Market Contrast[b]	(1)	(2)	(3)	(4)
Controls for Factor Price and Demand Variation?	NO	YES	NO	YES
Panel A: Peak-Period Delivery [c]				
1. PJM Western Hub – MichFE	-0.34	-0.28	-0.36	-0.52
	(0.11) ***	(0.13) **	(0.17) **	(0.13) ***
2. PJM Western Hub – Cinergy	-0.41	-0.35	-0.48	-0.68
	(0.10) ***	(0.11) ***	(0.13) ***	(0.08) ***
3. PJM Allegheny – MichFE	-0.37	-0.34	-0.43	-0.62
	(0.11) ***	(0.13) **	(0.16) ***	(0.11) ***
4. PJM Allegheny – Cinergy	-0.39	-0.41	-0.48	-0.65
	(0.11) ***	(0.10) ***	(0.13) ***	(0.09) ***
Panel B: Off-Peak Period Delivery [c]				
5. PJM Western Hub - MichFE	-0.72	-0.58	-0.67	-0.60
	(0.04) ***	(0.09) ***	(0.09) ***	(0.12) ***
6. PJM Western Hub - Cinergy	-0.74	-0.48	-0.63	-0.51
	(0.04) ***	(0.11) ***	(0.09) ***	(0.13) ***
7. PJM Allegheny - MichFE	-0.80	-0.69	-0.78	-0.75
	(0.02) ***	(0.06) ***	(0.06) ***	(0.07) ***
8. PJM Allegheny - Cinergy	-0.80	-0.61	-0.72	-0.62
	(0.03) ***	(0.08) ***	(0.07) ***	(0.10) ***
N. Observations	261	261	522	522

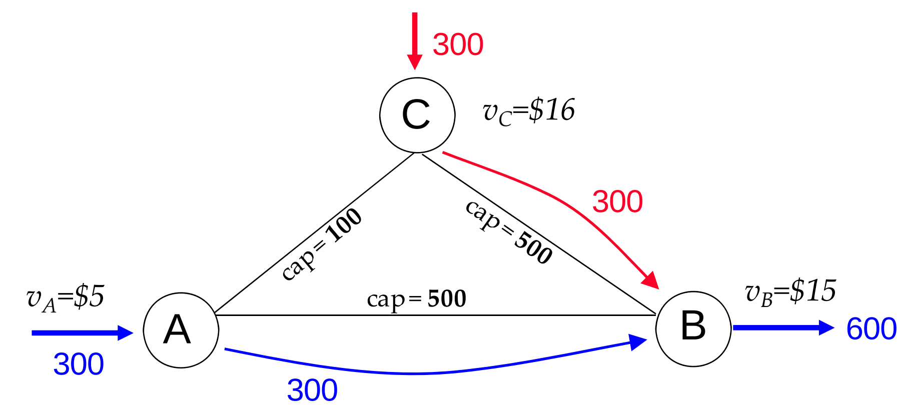
Notes. Summary results from separate regression estimates for each of eight delivery point price-spread pairs shown. Complete parameter estimates reported in Appendix Table A[XX]. Point estimates shown are (approximate) post-expansion percent changes in average absolute price spreads for day-ahead forward contracts, by delivery period, with indicated controls/contrasts (see text). (a) Newey-West standard errors, assuming a five-day lag structure; significance indicated for 1% (***), 5% (**), and 10% (*) levels. (b) See Table 1 notes for delivery point geographic areas. (c) Separate contracts are traded for peak (7am to 11pm) and off-peak (11pm to 7am) delivery.

Figure 1



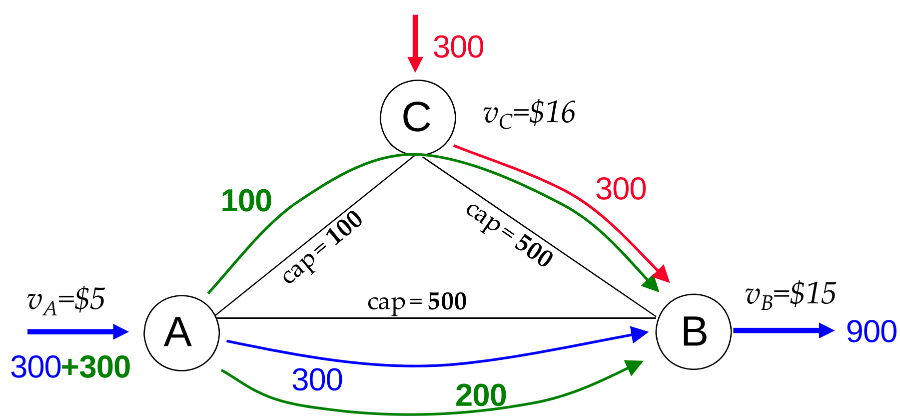
C
v_C=$16
100
cap = 100
cap = 500
v_A=$5
cap = 500
v_B=$15
A
B
300
300
200

(a)



300
C
v_C=$16
300
cap = 100
cap = 500
v_A=$5
cap = 500
v_B=$15
A
B
600
300
300

(b)



300
C
v_C=$16
100
300
cap = 100
cap = 500
v_A=$5
cap = 500
v_B=$15
A
B
900
300+300
300
200

(c)

Figure 2. Quantities Traded:
Day−Ahead Net Exports from Midwest to East (PJM)

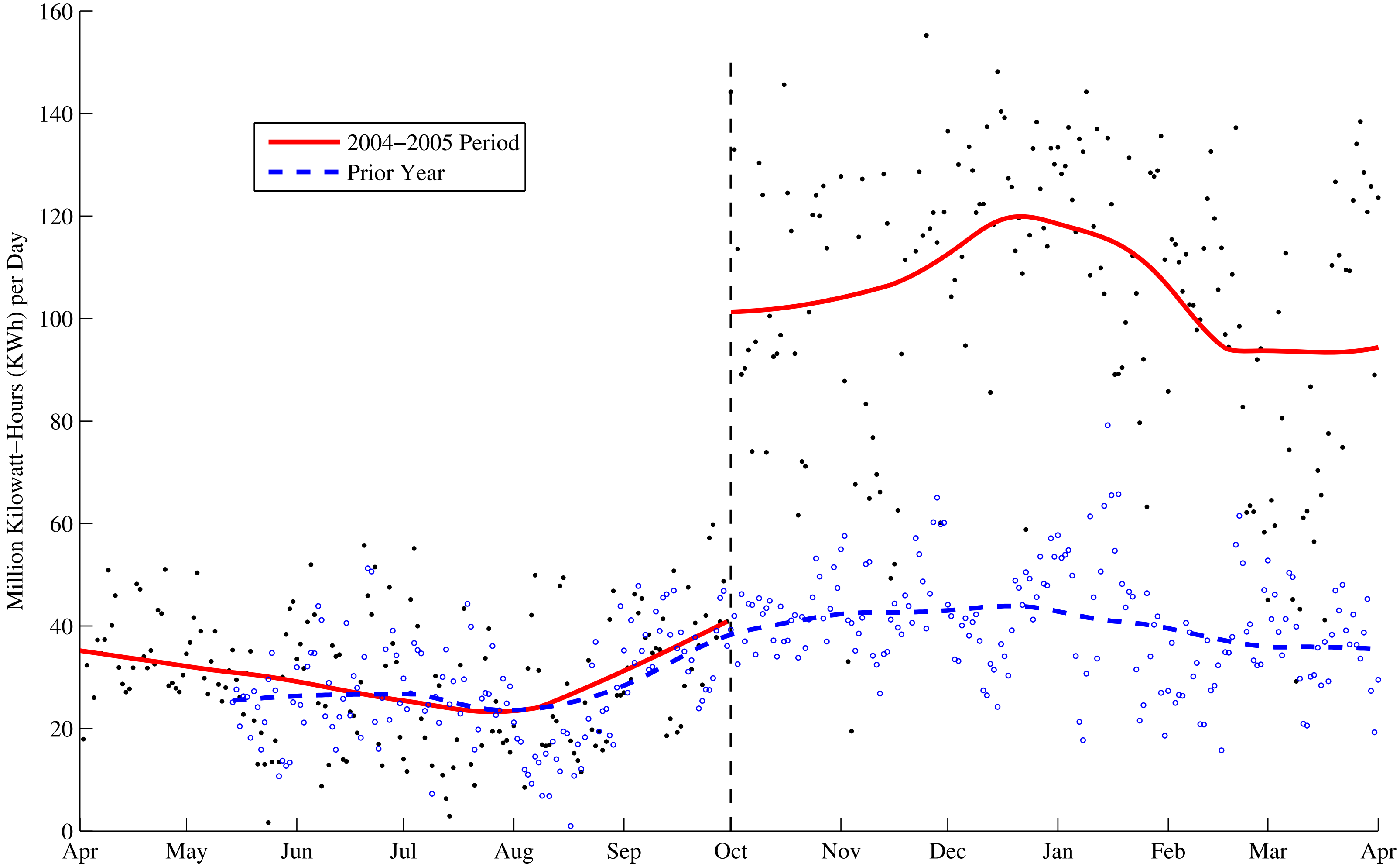
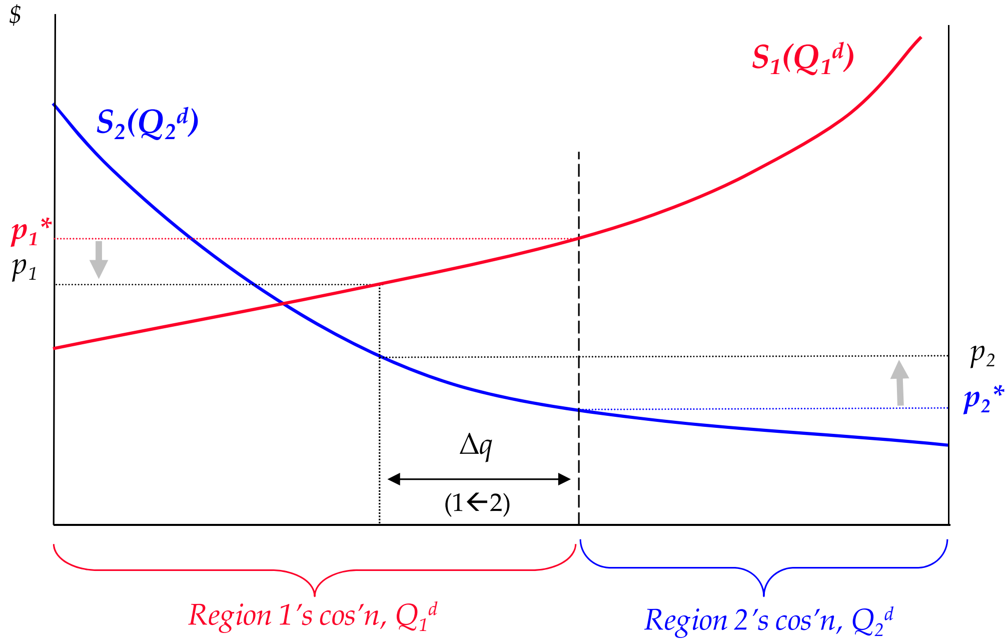


Figure 3



$
$S_1(Q_1^d)$
$S_2(Q_2^d)$
p_1*
p_1
p_2
p_2*
Δq
(1←2)
Region 1's cos'n, Q_1^d
Region 2's cos'n, Q_2^d

Figure 4