Inverse Optimization for the Recovery of Market Structure from Market Outcomes: An Application to the MISO Electricity Market

John R. Birge, Ali Hortaçsu, J. Michael Pavlin

informs
http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Vol. 65, No. 4, July–August 2017, pp. 837–855
ISSN 0030-364X (print), ISSN 1526-5463 (online)

CROSSCUTTING AREAS

Inverse Optimization for the Recovery of Market Structure from Market Outcomes: An Application to the MISO Electricity Market

John R. Birge,[a] Ali Hortaçsu,[b] J. Michael Pavlin[c]

[a] Booth School of Business, University of Chicago, Chicago, Illinois, 60637; [b] Department of Economics, University of Chicago, Chicago, Illinois, 60637; [c] Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada

Contact: jbirge@chicagobooth.edu (JRB); hortacsu@uchicago.edu (AH); mpavlin@wlu.ca, http://orcid.org/0000-0001-8367-8383 (JMP)

Received: October 30, 2015
Revised: October 3, 2016
Accepted: December 15, 2016
Published Online in Articles in Advance: June 19, 2017

Subject Classifications: economics: econometrics; programming: quadratic; natural resources: energy
Area of Review: Computational Economics

https://doi.org/10.1287/opre.2017.1606

Abstract. We propose an inverse optimization-based methodology to determine market structure from commodity and transportation prices. The methods are appropriate for *locational marginal price*-based electricity markets where prices are shadow prices in the centralized optimization used to clear the market. We apply the inverse optimization methodology to outcome data from the Midcontinent ISO electricity market (MISO) and, under noise-free assumptions, recover parameters of transmission and related constraints that are not revealed to market participants but explain the price variation. We demonstrate and evaluate analytical uses of the recovered structure including reconstruction of the pricing mechanism and investigations of locational market power through the transmission constrained residual demand derivative. Prices generated from the reconstructed mechanism are highly correlated to actual MISO prices under a wide variety of market conditions. In a case study, the residual demand derivative is shown to be correlated with coefficients of certain transmission constraints.

Funding: The authors gratefully acknowledge the support of the National Science Foundation [#SES 1426823 and #ICES-1216083]; the U.S. Department of Energy, Office of Science [Contract SC0002587], the Social Sciences and Humanities Research Council of Canada, and the University of Chicago Booth School of Business.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2017.1606.

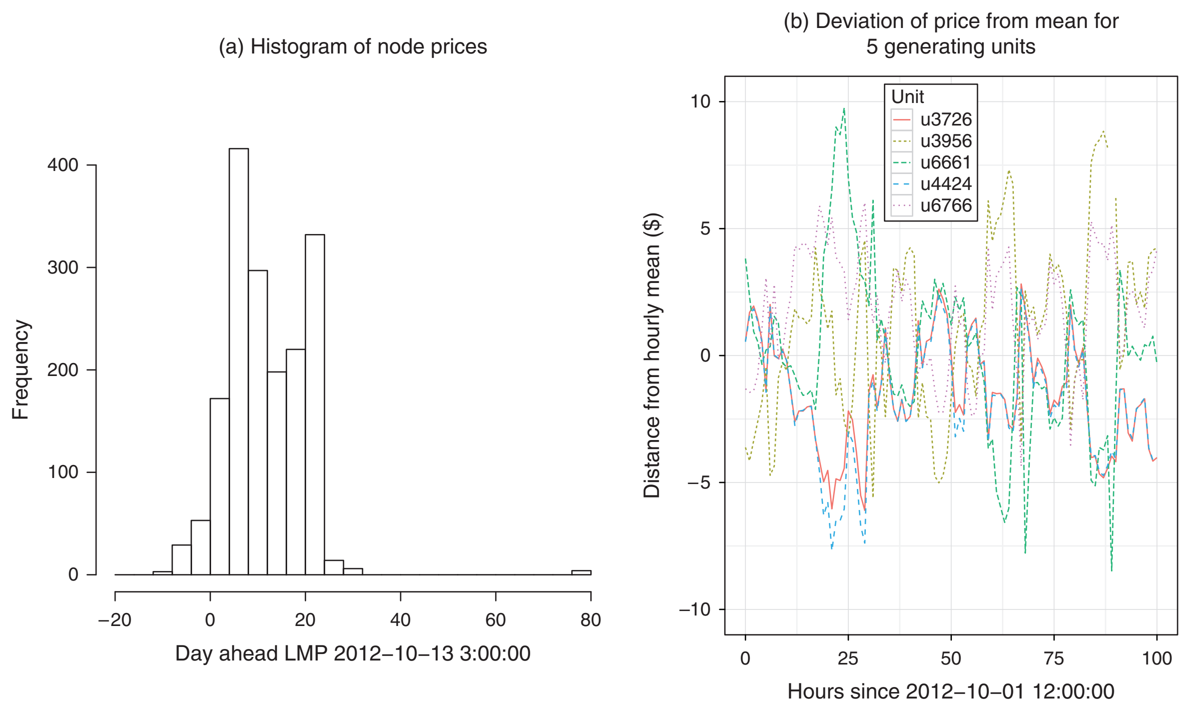
Keywords: inverse optimization • electricity markets • locational pricing • residual demand derivative

1. Introduction

In electricity and many other commodity markets, price varies substantially depending on location. For example, during the month of October 2012 in the Midcontinent Independent System Operator (MISO) day-ahead electricity market, the average hourly price was $25.50 but the average hourly difference between the high and low price was $81.28. The two panels in Figure 1 illustrate both the price variation at a particular time (histogram in Figure 1(a)) and the price volatility over time for individual market participants as different active transmission constraints change the market conditions (time-series plot in Figure 1(b)). Analyses of these markets would benefit from a structural model explaining price formation. However, building this model requires a precise model of the frictions generating these pricing disparities. In particular, to understand how electricity prices are formed, the analyst needs to know how local generation contributes to congestion throughout the market footprint. These operational details are typically private information and difficult to acquire either because the information is dispersed among many private firms or restricted due to legislation such as the Critical Infrastructure Information Act of 2002.[1] However, accurate and detailed pricing data are not subject to these restrictions and are often readily available. In the case of electricity markets that use the dominant locational marginal pricing scheme, the pricing and allocation mechanism centers on the solution to the *economic dispatch problem* (EDP). The EDP is typically a linearly constrained optimization problem whose objective is to optimally allocate production and consumption given market bids. The correspondence between market outcomes and the optimal solution of a partially specified optimization program promotes an inverse optimization-based methodology. We propose a novel inverse optimization-based recovery algorithm where unobservable market structure, which corresponds to parameters of the feasible region of the EDP, is uncovered using observed prices and allocations.

Inverse optimization may be defined as retrieving parameters of an incompletely specified mathematical programming formulation from a partially observable optimal solution. The standard inverse optimization problem, discussed in Zhang and Liu (1996) and elsewhere, is to retrieve unknown parameters of the objective function given an optimal solution. Our problem is novel; we are searching for parameters defining the feasible space of possible allocations. Using knowledge of the EDP, we are able to prespecify a particular form

Figure 1. (Color online) Price Variation in Midcontinent ISO Day-Ahead Market



to the optimization; we also take advantage of a partial solution to the dual problem. Like Zhang and Liu (1996), we leverage necessary conditions for optimality to constrain the missing parameters and motivate an algorithm to uncover them.

We present the methodology in the context of an electricity market using *locational marginal prices* (LMP). An LMP is a price at a particular point in the network and may be defined as the cost of purveying one additional megawatt-hour (MWh) of energy at that particular point. LMPs reflect behavior of neighboring and far flung participants in the network. LMP-based market design has largely come to dominate the distribution of electricity in North America where about 60% of total production is in markets using LMPs (Sahni et al. 2012). The testing of our methodology will focus on the *Midcontinent Independent System Operator* (MISO), which is an LMP-based market spanning 11 states and the Canadian province of Manitoba. Other LMP-based markets include markets facilitated by California ISO, the Electric Reliability Council of Texas, and the Pennsylvania-New-Jersey-Maryland Interconnection (PJM). Together these markets serve over 150 million North American customers.

We apply the inverse optimization methodology to real data from the MISO day-ahead electricity market and evaluate effectiveness for several analytical applications. We take advantage of high granularity data, which includes precise prices and allocations throughout the market footprint as well as prices for transmission lines to derive the market structure. Electricity market data tend to be of high quality, but, when active transmission constraints are taken into account the data become very wide—data sets have few samples but many features. With this type of data, resulting systems are not highly overdetermined, and keeping the scope of the paper in mind, we restrict attention to algorithms based on noise-free assumptions. Given differences in the number of available data samples for different market conditions, this assumption also allows all empirical tests of the algorithm to be placed on a common footing. Noise-free assumptions are standard assumptions for inverse optimization algorithms (e.g., Zhang and Liu 1996).

Using the derived market structure and published bid functions for all market participants, we reconstruct and solve the economic dispatch problem to predict prices and allocations and compare them to the real market outcomes. Results over a wide range of MISO market hours show that the methodology is effective at recovering hidden parameters determining how transmission resources are utilized and the relative losses from different areas in the market. When these parameters are used to reconstruct prices, they are very effective in explaining price variation. In many cases we find near perfect correlation between predicted and actual prices. However, the methodology results in a somewhat larger absolute price error of about 11%. These errors appear to be largely systemic effects due to limitations in the data. Among other factors, covariates of the hours that proxy higher imports for which we have limited data, reduce the accuracy of predicted prices.

While the most obvious audience for these algorithms may be a market participant wishing to understand the optimality of their bids under varying market conditions, we foresee potential use cases from other stakeholders and policy analysts. While the data required to recover the parameters are not available until after the completion of the market, there are a number of potential ex post analyses that are possible due to acquisition of a structural model for these important markets. First, estimation of hidden variables such as local market power may be performed without the endogeneity problems that plague standard econometric techniques. Second, counterfactuals may be evaluated. For instance, an accurate model of the pricing mechanism in an electricity market allows estimation of the value of new production capacity and transmission investment that takes into account the impact of market power on revenues. An ancillary benefit of the inverse optimization methodology as an analytical tool is its efficacy on wide data sets. Under noise-free assumptions, identification of the market structure can be performed with minimal data—identification requires only one more data sample than there are active constraints.

From an operational point of view the most important analytical application of acquiring this structural information is likely to be the ability to evaluate local market power. Electricity generators and demand participants such as large utilities, submit bids daily into the day-ahead and real-time markets. The ability to evaluate market conditions, in particular, how their bids impact the received allocation and price, can have a substantial impact on profitability. The participants' best-response bid is tightly linked to the transmission constrained residual demand derivative that measures how local price changes impact residual demand. An efficient calculation of this quantity for a lossless electricity market is given in Xu and Baldick (2007) and we adapt this calculation to a market with transmission line loss. This measure is a function of bids and the recovered structure, in particular, coefficients describing the contributions of market participants to active congestion constraints and loss parameters.

Following the evaluation of recovery of the pricing mechanism, we present a short study of market power in the MISO market by calculating each participant's transmission constrained residual demand derivative (RDD) over several hours. The RDD identifies the degree to which competitors will increase supply into a particular region as a function of price. The results show substantial variation in the distribution of market power and correlation between market power and contributions to particular constraints. These results reinforce the view that standard market concentration indices such as the Herfindahl-Hirschman Index are of limited value in electricity markets because transmission constraints partially isolate submarkets resulting in frequent changes in market structure (e.g., Karthikeyan et al. 2013, Borenstein and Bushnell 1999). Lee et al. (2011b) and (2011a) argue for the transmission constrained RDD as the basis for a principled measure of market power that reflects actual incentives of individual firms to deviate from competitive market bids into the electricity market.

While we develop the structure recovery methodology in the context of electricity markets, there is the potential to apply the inverse optimization-based market structure recovery whenever two structural assumptions hold: the market results in an efficient allocation and pricing corresponds to the value of dual variables of the efficient allocation problem.

Following a discussion of the relationship of this work to the operations research, energy, and economics literatures, we present the electricity market model and efficient allocation problem (Section 2). In Section 3 we present the recovery methodology. We then discuss the application of this methodology to electricity markets and the MISO market in particular in Section 4.

1.1. Literature and Positioning

This paper promotes the analysis of spatial price variation using tools from optimization theory. We provide a specific application to electricity markets that use *locational marginal prices*. Locational marginal prices assign a price at each location in the market that corresponds to the marginal cost of supplying that particular location. Hogan et al. (1996) first proposed the use of mathematical programming to determine electricity prices; since then, locational marginal prices have come to dominate North American electricity markets. While there is some evidence that the locational prices are succeeding at improving market efficiency (e.g., Wolak 2011), there are potential drawbacks such as increased market power, as discussed in Cardell et al. (1997). As identified by Sahni et al. (2012), locational priced markets also present much higher analytical hurdles both for participants and observers that a methodology like ours may contribute to mitigating.

The methodology we introduce to uncover the market structure is a type of *inverse optimization*. Inverse optimization algorithms uncover unknown parameters of a mathematical programming formulation from observations of the optimal solution. Since the important paper of Zhang and Liu (1996), a breadth of research has focused on the following inference problem: given an observable feasible space and an optimal solution to the optimization problem, determine unobservable parameters of the objective function. This problem has been addressed for both linear programming problems (Yang and Zhang 2007; Ahuja and

Orlin 2001; Zhang and Liu 1996, 1999) as well as various nonlinear mathematical programming paradigms (Schaefer 2009, Wang 2009, Iyengar and Kang 2005). Our problem differs in that the unobservable parameters of interest define the feasible region and we assume that relevant parts of both the primal (allocation) and dual (prices) solutions are observable. In the standard inverse optimization problem, the solution is not fully identified. The common practice is to search for norm minimizing values for the missing parameters. In our case, the solution is fully identified and with sufficient data may be determined by solving a system of linear equations.

Optimization paradigms have also been considered in the structural econometric literature. Su and Judd (2012) advocate explicit representation of equilibrium constraints in a mathematical program as a general approach to solving structural estimation problems. Under this approach, the objective function corresponds to the maximum likelihood function. There are clear thematic similarities with inverse optimization when, as in our study, the underlying optimization determines equilibrium outcomes. Our work is distinct from traditional structural econometric models in our not assuming that the equilibrium conditions are smooth functions of a set of defined parameters and in our explicit modeling of constrained resources, boundary solutions, and resulting primal and dual relationships.

In a recent paper, Kekatos et al. (2014a) consider the recovery of electricity grid topology from LMPs in the absence of transmission prices. They develop a maximum likelihood algorithm that takes advantage of structure of the resulting bilinear problem and show results on the 14bus IEEE test set. The Kekatos et al. (2014b) paper extends the Kekatos et al. (2014a) paper. Our paper is differentiated from these papers in its emphasis on extraction of relevant economic market structure rather than physical characteristics of the transmission network. We show that given industry practice, a simple recovery algorithm is sufficient to derive economically relevant information and demonstrate the utility of these recovered parameters in the analysis of the strategic environment of the market participants. To the best of our knowledge, the only other direct investigation of inverse optimization to infer strategic market parameters is Ruiz et al. (2013), who examine algorithms to extract details of bidders cost functions from published allocations and LMPs. Unlike the methodology presented in this paper, the algorithm discussed in Ruiz et al. (2013) is consistent with the standard inverse optimization setting where the unknowns are coefficients in the objective function. It is assumed in Ruiz et al. (2013) that the feasible region as defined by transmission constraints is known. Our paper is particularly unique in this literature as it develops an algorithm in a setting consistent with industry practice and demonstrates the practicality of these algorithms in scaling to real scenarios.

Several clear themes link this work to studies in the economic literature, which examines failure of the "law of one price" in regional trade. Pioneered by the works of Engel and Rogers (1996) and McCallum (1995), these studies have focused on the goal of understanding how distance and border effects contribute to price differences for physically identical goods. Recent treatments of the problem have incorporated sophisticated models of market power (Cosar et al. 2015) and have tested these results with nonaggregated microdata (Broda and Weinstein 2008). These works are characterized by a common underlying model of the contribution of distribution infrastructure to variation in prices: transportation frictions (i.e., borders and distance) are separable, observable, and contribute to price variation in a manner that is consistent across samples. This model may be appropriate for prices of hard goods such as wind turbines, as studied in Cosar et al. (2015). However, electricity differs in a fundamental manner: the preferred and lowest cost distribution method is via a capacitated technology. Adding new transmission lines requires fulfilling time consuming regulatory requirements and providing large capital investments. Capacity issues that make it more difficult to close supply demand gaps are compounded by seasonal demand fluctuations and the inability to relocate production. Natural resources are tied to a specific geography. This applies to natural gas and oil as well as renewable electricity fuels such as hydro and wind.

Which transportation line is at capacity varies sample to sample; as a result, these frictions contribute in different manners to the overall price. In electricity markets, the variation in active constraints is particularly acute. For example, in the MISO market, in 2012, 1,295 unique transmission lines—the relevant resource—were at capacity. Of these, 89% were at capacity in no more than 1% of the hours and no constraint was active in more than 30% of the hours. Efficient use of data is important for deriving estimates from multiple market states. Our results show that identification of market structure sufficient to reconstruct the pricing mechanism requires only one more sample than there are links at capacity.

In natural gas and other energy markets, theoretical models have been proposed that incorporate capacitated technologies into equilibrium models as explanatory of geographic price differences (Mudrageda and Murphy 2008, Cremer et al. 2003). We believe that in addition to applications to electricity markets, this paper may be taken more generally as an extension of this line of theoretical research equating price variation to market-wide capacity constraints.

2. Model and Preliminary Analysis

We focus on a one-shot electricity market facilitated by a centralized market maker. In North American electricity markets, the facilitator would be either an ISO or regional transmission organization (RTO). The model is appropriate to either forward (day-ahead) or spot (real-time) energy markets. In both of these markets, the participants submit bids while the facilitator solves an efficient allocation problem under the assumption that these bids accurately reflect the participants' preferences, i.e., if the bids are honest, the facilitator maximizes social welfare. A bid from a participant consists of a supply or demand function and bounds on the interval where they are willing to supply or consume. The facilitator also takes into account how nodes are connected by transmission resources, the capacity of these transmission resources, and other characteristics determining how electricity flows through the network.

The market features a set of generators and load serving entities who are located at buses connected by transmission lines. Generators and load serving entities correspond to the set of market participants; buses correspond to locations. In practice, it is standard to model line load in electricity markets via a linear approximation of the true AC power flows referred to as the *DC model* (see Stott et al. 2009). Details of the linearization are discussed in Online Appendix A. In the linear model, the net production at a particular bus will have a proportional effect on flow over a particular line. In effect, the transmission lines are shared transmission resources whose utilizations are determined by locational parameters. Other than increasing and decreasing production and consumption at particular locations in the network, the facilitator and participants are not able to influence the routing of electricity as they might be able to in other network settings such as a data network.

The market structure that we are interested in is parameterized by these proportional utilizations of shared transmission resources along with variables that describe how energy is lost throughout the transmission network. These pieces of information are derived from the physical topology of the network, which is discussed in Online Appendix A. Because the pricing mechanism in these markets sets locational prices as the cost of increasing supply at that particular location, the loss and utilization parameters are key pieces of information that explain price variation. This paper considers recovering this economically relevant information and not the physical topology of the network. Despite some data limitations, including the details of the ancillary services market and operational information from generators and missing demand and supply information, the empirical study (Section 4) shows that, at least in the MISO market, these limitations are not critical to understanding price variation. In the large-scale study, the median correlation of predicted prices to real MISO prices is 0.92 and the upper quartile has correlation greater than 0.98.

2.1. Market Model

The electricity market model consists of a set of participants who produce or consume electricity that is distributed through shared transmission resources.

Participants. The market features a set of participants $\mathcal{Q} = \{1..Q\}$, where participant j produces x_j MWh of electricity. With respect to supply and demand the market is symmetric; consumption is interpreted as negative production, i.e., $x_j < 0$.

Transmission Resources. The market is connected via a set of transmission lines, $\mathcal{R} = \{1..R\}$. Each of these lines may be interpreted as a shared resource with a fixed capacity. Line r has capacity d_r. This endowment is utilized heterogenously by participants when they produce or consume electricity.

Each participant produces or consumes electricity at one of a set of possible locations, $\mathcal{L} = \{1..T\}$. Variable b_i denotes the net production at location i. The location determines the rate at which a participant utilizes each resource. Consistent with the DC flow model, for each location i and resource r, there is a locational utilization parameter D_{ri}, such that, if b_i units are produced at location i, then $D_{ri}b_i$ units of resource r are consumed. Allowing b to be the vector of net production for each location, b is feasible if $Db \leqslant d$. We let $l(j)$ be the location of participant j and $z(i)$ be the set of participants operating at location i. So, net production at location i is $b_i = \sum_{j \in z(i)} x_j$. In the context of an electricity market, a location would be equivalent to an electricity bus, although alternative interpretations are possible in other settings.

Transmission lines are imperfect and some proportion of production is lost in the form of heat. We use an affine model to accommodate this line loss, which associates a parameter δ_i equal to the proportion of production lost from each location. This notion of loss is compatible with the MISO electricity market where a similar loss factor is associated with each producer (Ma et al. 2009). The total system production must equal total consumption plus losses: $\sum_{i \in \mathcal{L}} b_i - \sum_{i \in \mathcal{L}} \delta_i b_i = 0$. In other words, the market must clear. Keeping in mind that the DC flow model implies linear line utilizations, the loss model described here is equivalent to the assumption that loss on each line is an affine function of line flow. The affine loss assumption is a significant approximation of AC electricity flow. In the MISO market, this loss factor is determined based on the system's operating point (Ma et al. 2009).[2] Note that participants that are colocated may trade electricity among themselves without loss or impact on transmission capacities; however, positive or negative net production at each location may impact loss from the network and consumes transmission resources.

Market Mechanism. Production and consumption bids are described by a real valued supply function S_j unique to each participant j. The function S_j is defined over an interval $[\ell_j, u_j]$, where $S_j(x_j)$ is the announced total cost to participant j of producing x_j MWh. In the case of consumption, $x_j < 0$, $S_j(x_j)$ is the announced negative of the value to participant j of consuming $-x_j$ MWh. While participants will typically behave as either producer $\ell_j \geqslant 0$ or consumer $u_j \leqslant 0$, there is nothing that precludes participants from behaving as both. For instance, the owner of a partially charged battery operating in an electricity market may wish to either buy or sell electricity depending on price. The total system welfare is $-\sum_{j\in\mathcal{Q}} S_j(x_j)$.

The facilitator allocates and prices the market in a two-stage process. First, the allocation is determined by solving an efficient allocation program referred to as the *unit commitment problem* (UC), which determines the efficient production and consumption from each participant under the assumption that bids are an accurate reflection of true preferences. Then, integer variables are fixed to form the *economic dispatch problem* (EDP) whose dual variables are used to determine prices.

We focus on the EDP, which may be formulated as follows:

$$\min_x \quad \sum_{j\in\mathcal{Q}} S_j(x_j) \tag{1}$$

$$\text{s.t} \quad b_i = \sum_{j\in z(i)} x_j \quad \forall\, i \in \mathcal{L}, \tag{2}$$

$$\sum_{i\in\mathcal{L}} (1-\delta_i) b_i = 0, \tag{3}$$

$$|Db| \leqslant d, \tag{4}$$

$$x \geqslant \ell, \tag{5}$$

$$x \leqslant u. \tag{6}$$

The market outcome is the allocation x and, prices are assigned to each location using the dual variables of Constraint (2). For the EDP, $[\ell_j, u_j]$ for each participant are set using the integer outcomes of the UC and within these bounds, $S_j(x)$ is a continuous linear or quadratic function.

2.2. Preliminary Analysis

Program (1)–(6) implies a series of optimality conditions gleaned from standard constrained optimization theory (see, e.g., Karush-Kuhn-Tucker conditions in Nocedal and Wright 1999). Consider dual variables $\pi, \lambda, \rho^+, \rho^-, \alpha, \gamma$ corresponding to constraints (2)–(6). Variables ρ^+ and ρ^- correspond to the positive and negative part of the absolute value in Constraint (4) (i.e., both directions of flow over the transmission resource). Then, the necessary conditions for optimality include the following stationarity constraints:

$$\sum_{r\in\mathcal{R}} D_{ri}\rho_r^+ - \sum_{r\in\mathcal{R}} D_{ri}\rho_r^- + \pi_i + (1-\delta_i)\lambda = 0, \quad i \in \mathcal{L}, \tag{7}$$

$$-\pi_{l(j)} + \alpha - \gamma = S'(x_j), \quad j \in \mathcal{P}. \tag{8}$$

Note that these are necessary conditions for optimality under very general continuity conditions (Nocedal and Wright 1999) and will hold even with piecewise and nonconvex supply functions.

With the market clearing constraint in Equation (3), the program is overspecified. To avoid this problem, it is standard in electrical engineering to select a reference location r (Wood and Wollenberg 2012). The rth column of matrix D, which contains the parameters for the utilization of each resource, corresponds to the reference location and is a zero vector; the parameters for other locations will compensate. There is a different matrix D for each reference location, although the optimization problem remains equivalent.

3. Recovery of Market Structure from Outcomes

In this section, we discuss the identification of utilization parameters for resources that are scarce in at least one data sample, i.e., the transmission resource is at full capacity and has a nonzero resource price. Since resource prices are identically equal to zero for any transmission resource with surplus capacity, parameters for these resources cannot be identified; hence, we exclude these resources from the recovery procedure. We refer to resources with a positive resource price as active resources. By convention, each row of D corresponds to a resource in the set $\mathcal{K}$ of resources active in at least one sample. Since we are not able to distinguish between resources that are active due to flow in the positive or negative direction we replace ρ^+ and ρ^- with ρ and use the following constraint in our analysis:

$$\sum_{k\in\mathcal{K}} D_{ki}\rho_k + \pi_i + (1-\delta_i)\lambda = 0, \quad i \in \mathcal{L}, \tag{9}$$

We refer to the vector $-\pi$ as the locational prices or LMPs and the vector $-\rho$ as the resource prices. In the derivation of the algorithm, we assume that samples are noise free. In other words they are free from error and correspond exactly to components of the optimal solution of the EDP. We discuss the noise-free assumption and potential sources of error in the recovery of MISO market prices in Section 4.1 after introducing the empirical methods.

The following observation allows the relevant parameters to be recovered without the unobserved dual variable λ.

Proposition 1. *In an optimal solution of Program (1) λ is equal to the LMP of the reference location corrected by the loss*:

$$\lambda = \frac{-\pi_r}{1-\delta_r}. \tag{10}$$

Proof. Noting that $D_{kr} = 0$ for the reference location r, solving Equation (9) for $i = r$ provides the required result. □

Now, consider a set of M noise-free data samples $\mathcal{M} = \{1..M\}$, where a sample $m = (\pi^m, \rho^m, x^m)$ specifies observed commodity, resource prices, and allocations. Let $\mathcal{K}$ be the set of active resources in the data samples. The identity $\lambda = -\pi_r/(1-\delta_r)$ from Proposition 1 allows λ to be replaced in Equation (9). Rather than trying to identify δ_i, we can consider the relative loss of a location i to the reference location, $q_i = (1-\delta_i)/(1-\delta_r)$. With the relative loss vector q, an equivalent program to the EDP can be formulated by replacing the market clearing constraint $\sum_{i\in\mathcal{L}}(1-\delta_i)b_i = 0$ with the equivalent alternative:

$$\sum_{i\in\mathcal{L}} q_i b_i = 0. \tag{11}$$

The following theorem details how and when the unknown parameters are identified.

Theorem 1. *For an arbitrarily selected reference location r and a set of samples $\mathcal{M}$ of market outcomes $(\pi^m, \rho^m, x^m)_{m\in\mathcal{M}}$, consider equations*

$$\sum_{k\in\mathcal{K}} D_{ki}\rho_k^m = q_i\pi_r^m - \pi_i^m, \quad i\in\mathcal{L},\ m\in\mathcal{M}, \tag{12}$$

then, D and q are identified if the above system of equations has rank $(|\mathcal{K}|+1)|\mathcal{L}|$. The remaining unknown parameters of the primal program, d, may be recovered using equations $\sum_{p\in z(i)} x_p^j = b_i^j$ and $D_k b^j = d_k$ for an arbitrarily selected sample j.

Proof. Equation (12) is simply Equation (7) updated with $(1-\delta_i)\lambda = -q_i\pi_r$ using the identity stated in Proposition 1. The system of equations contains $(|\mathcal{K}|+1)|\mathcal{L}|$ variables that are the elements of matrix D and q, requiring the system to have this rank for a unique solution. Equation $D_k b^j = d_k$ follows directly from the primal formulation. With D_k identified as just stated, and b_j specified from allocation data, d is also identified. □

Equation (12) provides substantial intuition into the identification of loss and locational utilization parameters. The price at a particular location can be expressed as a base price that takes into account losses and a congestion penalty. The base price is equal to the relative loss times the price at the reference location and this proportional difference is constant across samples. The congestion component, on the other hand, depends on line flows, or more precisely if a line flow is at the line limit, and as a result this contribution varies sample to sample. The two components are additively separable allowing full identification.

3.1. Market Structure Recovery Algorithm

The inverse optimization recovery algorithm used in the application section of the text is derived directly from Theorem 1. We will refer subsequently to this algorithm as the IO recovery algorithm. The algorithm takes as input a set of active constraint identities $\mathcal{K}$ and a set $\mathcal{M}$ of $|\mathcal{K}|+1$ samples $\{(\pi^m, \rho^m, b^m)\}$ for which the requirements of Theorem 1 hold and returns recovered model parameters $\bar{D}, \bar{d}$ and $\bar{q}$. The algorithm proceeds as follows:

1. Select an arbitrary reference node r assume all $\bar{D}_{kr} = 0$.
2. Solve the following system of linear equations for $\bar{D}$ and $\bar{q}$:

$$\sum_k \bar{D}_{ki}\rho_k^m = \bar{q}_i\pi_r^m - \pi_i^m \quad i\in\mathcal{L},\ m\in\mathcal{M}.$$

1. Recover transmission capacities from the allocation:

$$\bar{d}_k = \bar{D}_k b^m.$$

This algorithm can fail if the system of equations is not of full rank. This might occur due to lack of variation between samples. In practice however, there appears to be sufficient variation in demand and supply that this is not an issue. The constant loss parameters between samples may also appear problematic because loss over a particular line increases superlinearly and as a result, locational loss parameters are dependent on the system's operating point. As described in Ma et al. (2009) MISO's market design determines linear loss coefficients compatible with the system's final operating point. Our methods assume that the operating point is homogenous for the samples used in the recovery. In Section 4.3 we study the impact of this assumption on parameter recovery.

The algorithm is restricted to the case where there are exactly $|\mathcal{K}|+1$ samples and in practice there may well be additional data samples available. In this case, any noise or error in the data would likely lead to an inconsistent set of equations. A method such as minimizing the least squares error may perform well in this situation, though it is left for future research.

Often, a large set of constraints can be solved for in a more efficient manner by iteratively increasing the size of the set of data samples while ensuring that the requirements of Theorem 1 continue to hold. At each iteration, we can substitute recovered parameters from previous iterations and need only solve a linear system with the newly introduced variables. It is not clear how error will accumulate and depend on the order of samples in this iterative method; so, we focus on the standard IO recovery algorithm in the subsequent sections.

4. Recovery of Market Structure in the MISO Electricity Market

MISO facilitates day-ahead energy, real-time energy, and ancillary services markets for 11 American states and the Canadian province of Manitoba. We perform a study of the application of the IO recovery algorithm in this market for the years 2010 and 2011. MISO solves two sequential optimizations to clear and price the market. The first is the unit commitment problem, which determines binary variables, and the second is the economic dispatch problem, which takes the status of each generator as given and determines the final allocation and prices. The EDP as detailed in Ma et al. (2009) closely fits the model described in Section 2.1. Using the IO recovery algorithm, losses and locational utilizations (D) are recovered, which is privately held information. The algorithm takes as inputs public information including energy and transmission prices (ρ), as well as the energy allocation. Using these recovered parameters, bids and commitment statuses; we reconstruct the EDP in order to predict energy prices.

An important difference between the MISO economic dispatch problem and our model is that MISO clears the energy market alongside the ancillary services market (ASM), which introduces additional constraints to ensure that contingency and regulatory reserves are at required levels. These constraints do not invalidate Theorem 1; so without modification, the recovery method may be used to find locational utilization and loss parameters. The ASM constraints do however have to be incorporated into the reconstructed EDP for prices and allocations to be representative. From data we are able to approximate these constraints as specified in Ma et al. (2009) and incorporate them into the reconstructed EDP. These constraints are discussed in depth along with other institutional details and the precise data provided by MISO in Online Appendix C. We also have limited information regarding ramping constraints that limit deviation from period to period. We are however able to estimate these constraints from data and approximate their effect as changes in the bid upper and lower bounds as detailed in Section 4.1.

In the day-ahead energy market, participants include generators (supply), load-serving entities (demand), and strictly financial players who participate on both sides of the market. We will continue to refer to market submissions on both sides of the market as bids though MISO refers to supply and demand bids, respectively, as offers and bids. Regardless of type, participants are treated similarly and submit marginal cost functions as nondecreasing piecewise linear functions and operational bounds. In the case of generators, these may have up to 10 regions and may be either step functions or continuous. For other participants, only a step function with up to nine components may be submitted. The resulting supply functions, which return the total cost, are either piecewise linear or piecewise quadratic such that the MISO unit commitment problem is a mixed-integer quadratic problem. The EDP fixes integer variables in the MISO unit commitment problem to allow shadow prices to be determined and is a quadratic programming problem.

Roadmap to the empirical tests. Empirical tests of the algorithm and analyses of the MISO market are presented in the following three subsections. First, in Section 4.1 we describe the approximate version of the MISO EDP used to reconstruct prices with recovered parameters and market data. In Section 4.2 we present a large-scale test of the algorithm performance by reconstructing prices from MISO data for the years 2010 and 2011. As detailed below, in this study we exclude hours with large numbers of active constraints to maintain tractability. Otherwise, on a monthly basis we recover all identifiable parameters and determine market outcomes by reconstructing the EDP for each hour and compare generated prices and allocations to the actual MISO market results. Section 4.3 presents a study of the consistency of parameters recovered using different sets of market hours. In Section 4.4 we present a detailed case study of three hours with differing levels of congestion. In this case study, we examine the distribution of error, consistency of recovered constraint coefficients, and compare solution quality to benchmark price prediction algorithms. Finally, in Section 4.5 we specify the transmission constrained RDD appropriate for the MISO market. We evaluate the RDD for all locations for the market hours in the case study. We examine the distribution of the RDD and its correlation to particular constraint coefficients.

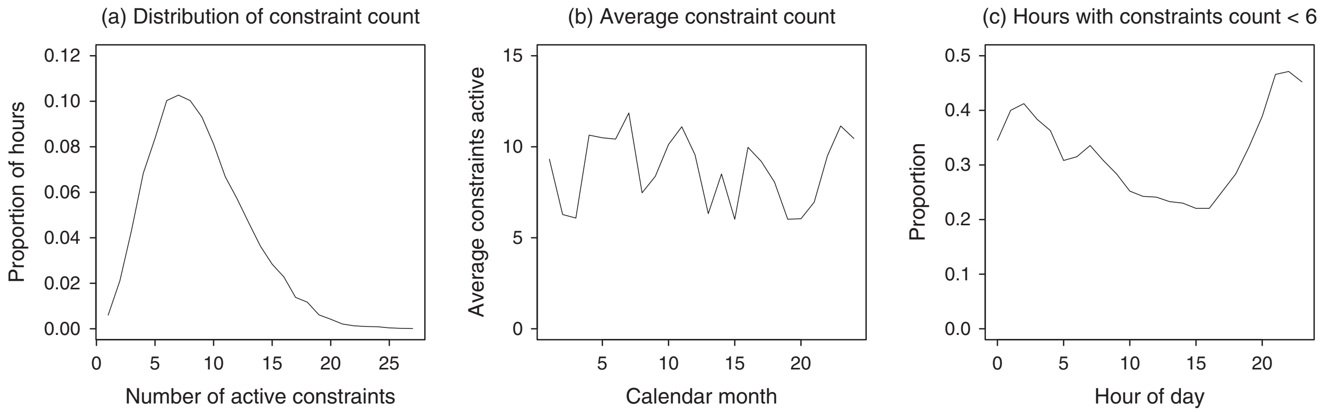
Details of the selection of test cases and hours is discussed below with respect to each study. We restrict our attention to test cases with at most six active constraints. Figure 2 provides a picture of joint constraint activity in the MISO market for years 2010 and 2011. Panels (a), (b), and (c), respectively, show the distribution of the number of jointly active constraints, the average number of active constraints and the proportion of hours with at most six active constraints. Thirty-two percent of all day-ahead market hours have fewer than six jointly active constraints. This number of constraints is computationally tractable for parameter recovery.[3]

Before discussing these empirical studies and results we discuss the details of determining market outcomes for the MISO market.

4.1. Reconstructing the MISO Pricing and Allocation Mechanism

The intent of these experiments is to gauge the accuracy and utility of the reconstructed MISO EDP on

Figure 2. Constraint Activity in the MISO Market



Notes. Constraint count is the number of jointly active constraints in a day-ahead market hour. Calendar months continue from January 2010 (1) until December 2011 (24).

a wide range of instances. The key inputs to this problem are market bids and commitment statuses, which are directly evident from published market data and the transmission constraints that are recovered. Contingency and regulatory reserves are both allocated through the ASM. Approximate versions of these constraints are incorporated using ASM bids. The reconstructed EDP has a requirement that each of the ancillary services are at the observed volumes for all sets of nodes with common ASM prices. This is discussed in greater detail in Online Appendix C.

The reconstructed EDP is summarized in the following program, which is quadratic due to the supply functions S_j^m:

$$\min_{x,y}\left\{\sum_{j\in\mathcal{E}\cup\mathcal{R}} S_j^m(x_j) + y_j^{rg}p_j^{rg,m} + y_j^{sp}p_j^{sp,m} + y_j^{su}p_j^{su,m}\right\} \quad (13)$$

$$\text{s.t } b_i = \sum_{j\in z(i)} x_j, \quad \forall\, i\in\mathcal{L}, \quad (14)$$

$$\sum_{i\in\mathcal{L}} \bar{q}_i b_i = \sum_{i\in\mathcal{L}} \bar{q}_i b_i^m, \quad (15)$$

$$\bar{D}b \leqslant \bar{d}^m, \quad (16)$$

$$\ell_j^m \leqslant x_j \leqslant u_j^m, \quad \forall\, j\in\mathcal{E}, \quad (17)$$

$$\ell_j^{rg,m} \leqslant x_j + y_j^{rg} + y_j^{sp} + y_j^{su} \leqslant u^{rg,m}, \quad \forall\, j\in\mathcal{A}, \quad (18)$$

$$x_{j,t+1} - r_j \leqslant x_j \leqslant x_{j,t+1} + r_j, \quad \forall\, j\in\mathcal{Q}, \quad (19)$$

$$x_{j,t-1} - r_j \leqslant x_j \leqslant x_{j,t-1} + r_j, \quad \forall\, j\in\mathcal{Q}, \quad (20)$$

$$\sum_{j\in z} y_j^{rg} \geqslant c_z^{rg,m}, \quad \forall\, z\in Z^m, \quad (21)$$

$$\sum_{j\in z} y_j^{rg} + \sum_{j\in z} y_j^{sp} \geqslant c_z^{sp,m}, \quad \forall\, z\in Z^m, \quad (22)$$

$$\sum_{j\in z} y_j^{rg} + \sum_{j\in z} y_j^{sp} + \sum_{j\in z} y_j^{su} \geqslant c_z^{su,m}, \quad \forall\, z\in Z^m. \quad (23)$$

In Equations (13)—(23), $\mathcal{A}$ and $\mathcal{E}$, respectively, denote participants committed to the ASM market and participants committed to only the energy market. Sets $\mathcal{Q}$ and $\mathcal{L}$, respectively, denote the full sets of market locations and participants. The objective is to minimize the sum of the bid costs S_j^m, which are linear and quadratic functions determined by the unit commitment, and the ASM costs. Variables y_j^a and parameters $p_j^{a,m}$ denote the ASM allocation and ASM bid cost in sample m from participant j for ancillary service a. The ancillary service may be regulatory (r), spinning (sp), or supplemental (su). Constraints (17) and (18) denote bid limits that depend on commitment status. Constraints (21)–(23) denote the regulatory requirements for each zone and service. These constraints are discussed in greater detail in Online Appendix C. The predicted outcome includes the allocation x and the dual prices of Constraint (14).

Ramping constraints shown in Constraints (19) and (20) are generator dependent bounds on the difference between generation in the current and adjacent hours. Data regarding physical limitations of generators including ramping constraints are not public data. However, generator specific ramping quantities r_j may be estimated from allocation data by finding the largest difference between adjacent hours in a large data sample. These quantities were estimated using the full data set from 2010 and 2011. The actual allocation received by generator j in the periods immediately preceding and following sample m are denoted by $x_{j,t-1}$ and $x_{j,t+1}$. This approximates the true model, which is cooptimized across hours. The constraints that we are incorporating are more strict than those that are used in the actual MISO EDP. In the multiperiod formulation, neighbouring hours may be adjusted together to coordinate supply. Note that while this may introduce error into the EDP, the IO recovery algorithm is not impacted since the multiperiod constraints affect the KKT conditions for the participants (Equation (8)) but not the locations (Equation (7)). It should be noted that ramping constraints are taken into account by the unit commitment so that when integer constraints are fixed in the EDP the ramping constraints are unlikely to play a large role in the prices and allocation. In preliminary experiments, ramping constraints were not found to play a major role. Additional concessions due to limited data are discussed in the following paragraphs.

Missing demand and supply data. Demand and supply resulting from imports, exports, and grandfathered bilateral transactions are substantial and not publicly available. On average, however, imports accounted for 7% of total load in 2011[4] and their influence must be considered.

Heterogeneity in unobservable demand and supply across samples, may impact market clearing and resource utilization constraints in the reconstructed EDP. For instance, if in a sample imports at a particular node i increase, the market will appear to be undersupplied and resources will appear underutilized if node i has positive coefficients and overutilized if node i has negative coefficients. To correct for these effects, we recalculate the right-hand side of both market clearing and resource utilization constraints for each sample using the observed allocation. For sample m the market clearing constraint will be adjusted such that $\sum_{i\in\mathscr{L}} \bar{q}_i b_i = \sum_{i\in\mathscr{L}} \bar{q}_i b_i^m$. The resource constraints are adjusted by including only binding constraints and adjusting capacities such that for sample m and resource k, $\bar{d}_k^m = D_k b^m$. These new coefficients should appropriately reflect the change in contribution from unobservable supply.

Changes in transmission network. Short- and long-term modifications to the electricity network may also result in error. These changes are the results of scheduled maintenance, unscheduled faults, and new investment. By minimizing the time span of each experiment the chance of these changes causing significant error is reduced. This has been taken into account when producing the map from participant to location. For each month, we repeatedly cluster participants and locations with persistent identical prices.

4.2. Aggregate Analysis of MISO Market Years 2010 and 2011

For each month of 2010 and 2011, all identifiable sets of six or fewer simultaneously active constraints were found. Each nonempty set of constraints was used as a *test case*. For each test case, the set of *potential hours* was found consisting of all hours from the month of the test case where the hour's active constraints were a subset of the test-case constraints. From these potential hours, a minimal set of in-sample hours were found by iteratively adding the potential hour with the largest number of active constraints until a set of hours sufficient to identify all constraints was found. This resulted in $k+1$ hours for the in-sample, where k is the number of active constraints in the test case. The remaining potential hours were placed into the collection of out-of-sample hours.

Characteristics of these hours are shown in Table 1. In total, parameters were assessed for 1,411 test cases from these years. In-sample and out-of-sample sets for these test cases were made up of 2,846 day-ahead market hours, which, because some hours are in multiple test cases, resulted in runs being tabulated for 16,431 individual test-case/hour combinations[5]. Details of the hours and test cases are reported in Table 1.

Reported results compare the LMP and MWh allocations from the recovered mechanism with the corresponding actual MISO LMP and allocations. Table 2 reports the correlation for the LMP and error for both LMP and allocated MWh over all samples as well as a number of cuts of the data. Table 3, reports the results of the regression of predicted LMP on the true LMP ($lmp_{pred} = slope \cdot lmp_{true} + int$). The regression results provide a more precise picture of the aggregate structure and biases of the error. Across all samples, the predicted LMP has a median correlation of 0.92 with the true LMP. The mean correlation (*mn corr* column) is much lower due to the long left tail, which is visible in the histogram of the correlation over all samples shown in panel (a) of Figure 3. The *avg-err* and *mw avg-err* are the medians of the average of the absolute error for the LMP and allocation. For all samples they are, respectively, \$2.74/MWh and 1.92 MWh. The median regressed slope of the predicted LMP is 0.93 and is shifted by the intercept \$3.52 (*int* column). A slope less than one indicates that the relative price difference between high and low cost resources is underestimated. The results are much stronger for more highly

Table 1. Summary of Hours Studied in Aggregate Analysis

	Active constraints							
	0	1	2	3	4	5	6	Total
Test cases:		43	128	247	342	359	292	1,411
Unique hours:	14	100	350	613	717	612	435	2,846
All test-case hours:	758	4,757	4,949	2,994	1,652	886	435	16,431
In-sample hours:	113	700	1,631	1,994	1,542	871	433	7,284
Out-of-sample hours:	645	4,057	3,318	1,000	110	15	2	9,147

Note. Entries are the number of test cases/hours with the corresponding number of active constraints.

Table 2. Summarized Results for Recovery of Prices and Allocation for MISO Market Years 2010 and 2011

	All hours						In-sample		Out-sample	
Hours	count	corr	mn corr	sd corr	avg-err	mw avg-err	count	corr	count	corr
All	16,431	0.92	0.82	0.24	2.74	1.92	7,284	0.95	9,147	0.88
High correlation (⩾ 0.92)	8,063	0.98	0.97	0.02	2.12	1.99	4,404	0.98	3659	0.98
Low correlation (< 0.92)	8,368	0.75	0.67	0.26	3.33	1.86	2,880	0.79	5,488	0.72
Peak	4,641	0.88	0.79	0.23	3.07	1.57	2,285	0.91	2,356	0.83
Off-peak	11,790	0.93	0.83	0.24	2.59	2.05	4,999	0.97	6,791	0.90
High abs change (⩾ 4,000 MWh)	647	0.83	0.76	0.24	3.71	2.06	344	0.91	303	0.74
Low abs change (⩽ 500 MWh)	424	0.92	0.83	0.21	2.52	1.76	176	0.92	248	0.91
Low timespan (⩽ 24 hrs)	4,227	0.92	0.80	0.26	2.89	1.86	1,917	0.95	2,310	0.87
High timespan (⩾ 336 hrs)	4,024	0.90	0.81	0.22	2.76	2.00	1,746	0.94	2,278	0.85
High fraction (= 1)	2,342	0.94	0.87	0.18	2.47	1.83	2,209	0.94	133	0.93
Low fraction (⩽ 0.2)	2,967	0.83	0.73	0.31	3.40	1.91	330	1.00	2,637	0.81

Note. All quantities are averages across test-case hours.

correlated results as shown in the two rows that follow. As reported in row 2 of Table 3, when correlation between recovered and MISO prices is above the median 0.92, there is no median pivot (*slope* = 1.01) and the shift is \$1.61. The standard deviation of both these quantities and the standard error is also significantly reduced. Basic investigation shows that poorer performance is tightly linked to several problem characteristics. We have divided the hours based on these characteristics and report the performance metrics for these subsets in the later rows of Tables 2 and 3. These cuts provide strong indications about which problem characteristics are associated with better performance; however, there is substantial overlap between the cuts, which leads to confounding and makes it difficult to reach firm conclusions about which parts of the model are responsible for the error. Discussion of these results as well as in-sample and out-sample performance are below. It should be noted that the reconstruction of the pricing mechanism performs well, achieving median correlation of at least 0.8 even on these collections of hours designed to be challenging.

Peak hours. Hours are divided according to whether they are peak or off-peak hours. Peak hours are defined as weekday hours between 6 A.M. and 9 P.M. CST. Loads are higher during peak hours and these hours are associated with higher imports, congestion, and higher ASM requirements. Off-peak hours show higher performance across all measures. In particular off-peak correlation is 0.93 versus 0.88 for peak hours. Peak versus off-peak is also associated with a large change in the slope of the regression line, 0.86 versus 0.95.

Changing demand. Results for hours with high and low absolute change from the previous hour are reported in the following pair of rows in both tables. These cuts are an attempt to isolate the impact of ramping constraints on the predicted LMP. When change is higher, correlation is substantially lower (0.88 versus 0.93). This is largely driven by poor out-sample performance, which we discuss below. The slope of the regression lines are very close in this case.

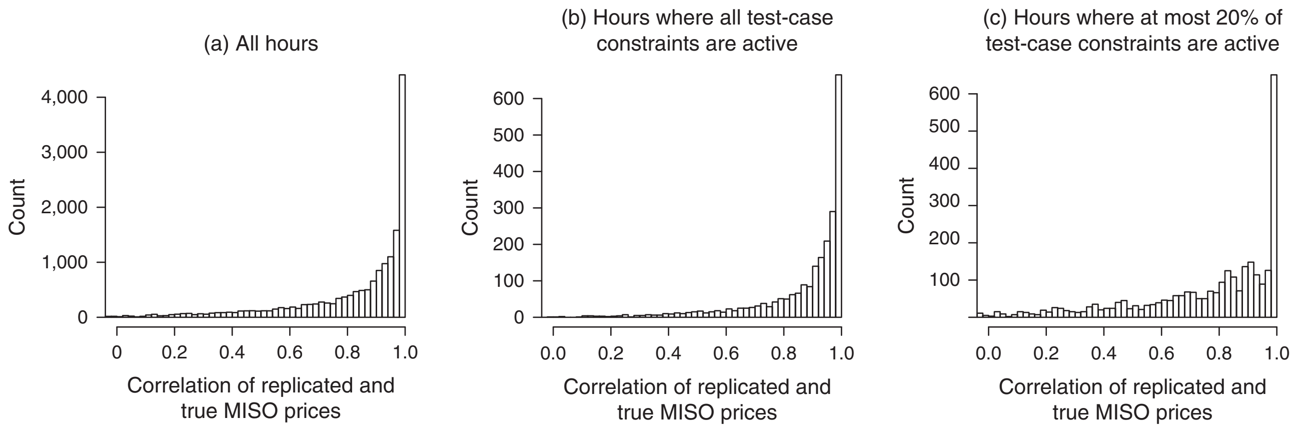
Time span. The time span over which in-sample hours varied was used to cut the data to gain an indication as to whether unobservable changes in the transmission network (e.g., transmission line outages) were having a substantial impact on the results. The results, shown in rows 8 and 9, show only a very modest improvement

Table 3. Summary of Regression of Recovered Prices to True Prices

	slope	mn slope	sd slope	int	mn int	sd int	std err	sd std err
All	0.93	0.89	0.43	3.52	6.29	15.22	1.01	93.73
High correlation (⩾ 0.92)	1.01	1.01	0.18	1.61	1.72	4.74	0.33	0.96
Low correlation (< 0.92)	0.76	0.77	0.55	10.14	10.71	19.83	2.56	131.00
Peak	0.86	0.84	0.40	6.36	8.25	17.00	1.59	82.07
Off-peak	0.95	0.91	0.44	2.94	5.52	14.38	0.85	97.93
High abs change (⩾ 4,000 MWh)	0.90	0.91	0.65	5.18	6.21	29.00	1.52	427.52
Low abs change (⩽ 500 MWh)	0.93	0.87	0.49	2.63	4.94	15.46	0.65	139.47
Low timespan (⩽ 24 hrs)	0.92	0.86	0.36	3.98	7.47	12.98	1.02	26.25
High timespan (⩾ 336 hrs)	0.93	0.92	0.40	3.73	5.82	14.51	1.18	28.41
High fraction (= 1)	0.93	0.91	0.31	3.40	5.73	12.16	1.15	18.01
Low fraction (⩽ 0.2)	0.92	0.87	0.62	3.46	6.32	20.08	1.36	174.30

Note. All quantities are averages across test-case hours.

Figure 3. Histograms of Correlation Levels Between Predicted and MISO Prices in Aggregate Experiments



when all in-sample hours are within a single day. The regression lines are also almost identical. At least over a one-month period, it does not appear that longitudinal changes in the transmission network are a significant cause of error.

Fraction of active constraints. This cut takes into account the fraction of the test-case constraints active in a particular hour. We consider this fraction to be a proxy for differences in the system operating point and note that the linear loss coefficients used by MISO depend on the operating point (i.e., the specific loads and transmission conditions). When all test-case constraints are active (high fraction rows) the reconstruction has substantially higher correlation and LMP error. When all test-set constraints are active, the median correlation is 0.94 as opposed to 0.83 when fewer than 20% of active constraints are active. These results are consistent with incompatible loss coefficients resulting from heterogeneity in the system operating point. This is investigated in more depth in Section 4.3.

In-sample vs. out-sample hours. In-sample versus out-sample performance is important for determining the degree to which derived constraints are generalizable to other market conditions. As expected, correlation across all samples and cuts is higher for in-sample than out-sample hours but correlation is relatively high in both cases (0.95 versus 0.88). We find that cuts associated with increased heterogeneity have much poorer out-sample performance. In particular, there is a large gap in median correlation between in-sample and out-sample performance for high change hours and low fraction hours (respectively, 0.17 and 0.19) and a very small gap of 0.1 for low change and high fraction hours.

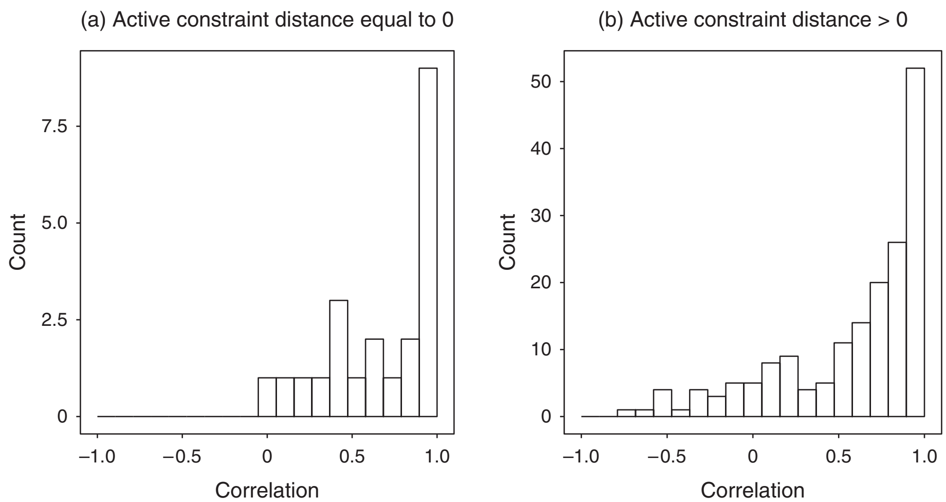
Additional sources of error include the possibility of multiple optimal solutions and changes in the system operating point (i.e., the net production at each bus and the structure of the transmission network). Since the objective function is affine with respect to certain parameters, multiple optimal solutions cannot be excluded. Loss parameters are dependent on the system operating point. Implicit in our estimation methodology is the assumption that loss parameters are constant across the samples used in the estimation. Substantial error may be introduced if actual loss parameters vary significantly between samples used in the recovery of parameters and prices. This is investigated further in the following Section 4.3.

4.3. Consistency of Parameter Recovery

To study the impact of the range of system operating points in the data samples on the parameter recovery we use the set of active constraints in a particular hour as a proxy for the operating point. This proxy invites a simple distance metric, termed the *active constraint distance*, to assess variation in the operating point for a particular parameter recovery: for a test case containing active constraint set $\mathcal{K}$, the *active constraint distance* is the number of constraints from $\mathcal{K}$ that are not active over all market samples used to recover the parameters.

A total of 147 test cases out of the original 1,411 were identified where the parameter recovery could be duplicated at least once with distinct sets of data samples. Because some parameter sets could be duplicated more than twice, this resulted in 342 unique parameter recoveries. Correlation and absolute error between parameters recovered were obtained for test cases by comparing parameters recovered with distinct sets of data samples. We find that loss parameters for cases where all constraints are active in all samples (active constraint distance equal to zero) have on average lower mean absolute error (0.068 versus 0.34) and higher correlation (0.66 versus 0.58). These results are driven largely by the chance of a very bad estimate that increases as differences in the operating point increase. The probability of a correlation coefficient less than 0.2 increases from 0.09 to 0.21 when not all constraints are active in all samples. These results are illustrated in the histograms shown in Figure 4 where panel (a)

Figure 4. Histograms of Correlation Levels Between Recovered Loss Parameters



shows the distribution of correlation coefficients when the active constraint distance is zero and panel (b) when the active constraint distance is greater than zero. Results are similar for the utilization parameters as shown in Figure 5.

These results indicate the importance of controlling for the system operating point when selecting the data samples for the IO recovery algorithm. While the operating point is likely an important source of error, it also appears that by selecting these samples carefully, the error can be minimized and that the methodology and applications are still sound. Even when there are at least six missing active constraints in the set of samples used in the recovery, in 25% of cases correlation between loss parameters is over 0.9. The role of the system operating point adds variation to the data, which is not easy to control for statistically and reinforces the need to use data efficiently.

4.4. Case Study of March 2011

To obtain a more nuanced understanding of the algorithm performance we focus attention on a small number of test cases taken from March of 2011. These test cases have the characteristics of hours where the algorithm performs well. In particular, they are from primarily off-peak hours from a shoulder month with a high fraction of active constraints. In this section, we examine the structure of pricing error and compare how predicted prices compare to prices estimated from simple benchmark algorithms.

Three cases were selected (Test 1, Test 2, and Test 3), where, respectively, 1, 2, and 3 unique constraints are active and identified. For each case, $|\mathcal{K}|+1$ in-sample hours are selected that yield a full rank system of linear equations allowing parameter identification. For each test case a pair of out-sample hours were also selected from March of 2011 by selecting hours with

Figure 5. Histograms of Correlation Levels Between Recovered *D* Parameters

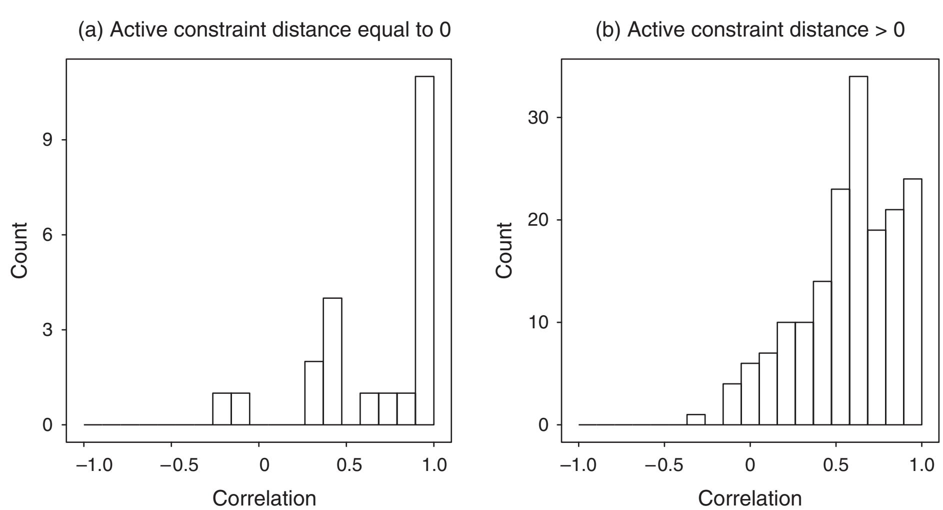


Table 4. Price and MW Predictions from Using IO Recovery Algorithm

ID	slope	int	corr-coef	std-reg-err	avg-abs-err	mw avg-abs-err
IS-1-0	1.29	−3.30	1.00	0.02	4.22	2.32
IS-1-1	0.99	1.97	1.00	0.01	1.73	2.00
OS-1-0	1.04	1.00	1.00	0.00	1.91	1.80
OS-1-1	0.67	13.34	0.95	0.60	2.45	1.73
IS-2-0	1.29	−3.30	1.00	0.02	4.22	2.32
IS-2-1	0.95	4.00	1.00	0.08	2.65	2.72
IS-2-2	0.99	1.97	1.00	0.01	1.73	2.00
OS-2-0	1.04	1.00	1.00	0.00	1.91	1.80
OS-2-1	1.04	2.41	1.00	0.03	3.34	2.54
IS-3-0	1.12	−1.78	1.00	0.05	2.68	1.05
IS-3-1	1.03	2.74	1.00	0.07	3.63	1.27
IS-3-2	0.70	13.33	0.98	0.36	5.08	1.85
IS-3-3	0.95	5.42	0.99	0.20	3.85	1.85
OS-3-0	1.02	0.89	1.00	0.05	1.54	1.15
OS-3-1	0.63	15.98	0.94	0.56	4.47	1.01

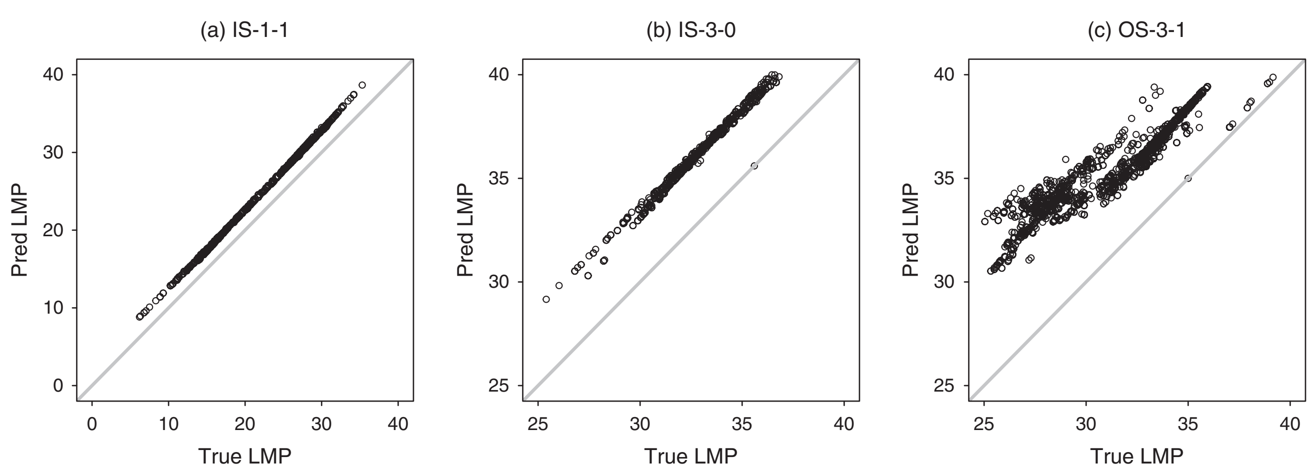
common active constraints closest to the in-sample hours. The precise active constraints are reported in Online Appendix C.2. In Test 1, in-sample hours are IS-1-0 and IS-1-1, out-sample hours are OS-1-0 and OS-1-1, and all constraints are active in all hours. In Test 2, in-sample hours are IS-2-0 to IS-2-2, out-sample hours are OS-2-0 and OS-2-1, and only one of the two identified constraints is active in IS-2-0, IS-2-2, and OS-2-1. In Test 3, in-sample hours are IS-3-0 to IS-3-3, out-sample hours are OS-3-0 and OS-3-1, and only two of the three identified constraints are active in OS-3-1.

Pricing results: Results of applying the IO pricing mechanism for the three test cases are summarized in Table 4. The same quality measures are considered as in Section 4.2 including correlation to true prices and regression results.

These test cases were found to have prices nearly perfectly correlated for most of the hours with correlation equal to 1.00 in 11 of 15 hours and at minimum equal to 0.94. The slope is more variable being between 0.63 and 1.29 with a mean of 0.98. The intercept is positive in almost all cases, consistent with a small positive mean bias—locations tend to be overpriced and most severely at low priced locations. This bias can be present even in highly correlated instances with slope close to 1.0, as in OS-2-1 and IS-3-1. This can be the cause of large absolute errors even in highly correlated instances, as in hours IS-1-0 and IS-2-0.

The scatter plots shown in Figure 6 show the relationship between predicted and true prices for samples with a range of correlation values. Figure 6(a) shows IS-1-1, which has correlation of 1.0 and slope close to 1. The positive intercept indicates a constant shift where the predicted prices are overpriced by about $1.97. In this case all locations are overpriced, however, the error is greater for lower priced locations. Figure 6(b) shows IS-3-0, which has positive slope and negative intercept. The plot shows that locations are overpriced and that higher priced locations suffer from larger pricing error.

Figure 6. Predicted vs. True LMPs



Note. The gray line is the reference line where Pred LMP = True LMP.

Figure 6(c) shows OS-3-1, which yielded the lowest correlation (0.63). Examining this figure, it is apparent that there are several superimposed subsets of locations that have different linear relationships with the true prices. For instance, there are two subsets both of which in isolation have predicted prices highly correlated with true prices. The two lines are parallel but shifted one from the other by about \$4/MW. On inspection, the subsets of shifted nodes have similar coefficients for particular active constraints. The shift likely results from inaccurate demand or supply within that subset of nodes that is isolated from other nodes by the active constraint. For instance, the approximate ASM constraints may not be capturing the full regulatory requirements of the ASM in the smaller correlated subset resulting in a surfeit of supply, which results in the relative underpricing of this set of nodes.

The consistent overpricing indicates that we are either overconstraining the problem globally or neglecting lower cost sources. While specification errors in ASM and ramping constraints may play some role, lack of cooptimization across hours and the handling of imports are likely to be significant factors. Cooptimization across hours allows potential adjustments in other hours to relax ramping constraints and thus increase available supply in the current hour. Since we are considering only the current hour, these adjustments are not available, raising the marginal price of energy. With respect to imports, by adjusting market clearing and resource constraints to observed allocations, we are effectively treating unobserved supply as fixed at current levels. If there is elasticity in these sources of supply, the marginal price may decrease. Nonetheless, the relationship between prices at different nodes is dominated by the resource utilization constraints, or in MISO's language congestion costs. The high correlation observed between real and predicted prices supports accuracy of recovered parameters in the resource utilization matrix D.

Comparison to benchmarks. To test the model quality, we compare the accuracy to prices generated with two benchmark algorithms. The first benchmark (NTC) is a model where we assume that there is no congestion and perform an unconstrained uniform price auction with the observed bids. The second benchmark (DC) produces prices using a naive data-centric approach: assume prices remain unchanged from the first sample.

Table 5 shows the outcome of the NTC benchmark algorithm on the case-study examples. This algorithm is identical to the IO recovery algorithm except that transmission constraints are removed from the efficient allocation problem. Price differences produced in this model are then exclusively a result of the recovered loss coefficients. The IO recovery results are consistently better than the NTC benchmark results on all of our quality measures. The highest correlation produced by the NTC benchmark is 0.90 on these test cases. This comparison confirms that the recovered transmission constraints contain useful information for explaining price variation.

Table 6 shows the same outcome using the DC benchmark. For each test case, this "data-centric" approach uses the true LMPs and allocation for one hour, empirically observed from MISO data, as the predicted LMPs and allocation for all other hours. The hours used for the prediction are, respectively, IS-1-0, IS-2-0, and IS-3-0. This may be interpreted as a very naive statistical approach where we are conditioning on the set of active constraints. In comparison to the DC benchmark, the IO recovery algorithm is competitive at predicting prices. In particular, of the 12 hours in the case study, the IO algorithm correlation is higher in three instances, tied in eight instances, and lower in only one instance. The absolute error is lower in nine of twelve instances. It is worth noting that the IO recovery does far better than the DC benchmark at predicting the allocation. This is not surprising given that the IO recovery algorithm incorporates bidding information

Table 5. Benchmark 1 NTC: No Transmission Constraints

ID	slope	int	corr-coef	std-reg-err	avg-abs-err	mw avg-abs-err
IS-1-0	0.33	21.41	0.89	0.84	4.13	3.25
IS-1-1	0.24	20.30	0.87	0.74	4.35	3.41
OS-1-0	0.24	20.40	0.87	0.75	4.39	3.17
OS-1-1	0.72	14.41	0.95	0.70	5.27	1.58
IS-2-0	0.33	21.41	0.89	0.84	4.13	3.25
IS-2-1	0.36	20.54	0.90	0.78	3.76	2.50
IS-2-2	0.24	20.30	0.87	0.74	4.35	3.41
OS-2-0	0.24	20.40	0.87	0.75	4.39	3.17
OS-2-1	0.36	20.18	0.90	0.78	3.80	2.44
IS-3-0	0.32	28.22	0.75	1.73	2.94	1.05
IS-3-1	0.36	24.75	0.76	1.33	4.32	1.58
IS-3-2	0.31	25.73	0.72	1.39	6.64	1.90
IS-3-3	0.33	25.82	0.75	1.33	6.41	1.74
OS-3-0	0.25	29.85	0.71	1.82	3.61	1.15
OS-3-1	0.48	22.71	0.81	1.16	6.37	1.06

Table 6. Benchmark 2 DC: Uses First IS Sample Prices and Allocations for Each Test

ID	slope	int	corr-coef	std-reg-err	avg-abs-err	mw avg-abs-err
IS-1-0						
IS-1-1	0.85	5.75	1.00	0.06	2.27	10.81
OS-1-0	0.86	6.07	1.00	0.07	2.76	11.41
OS-1-1	1.57	−25.84	0.97	1.96	7.04	16.22
IS-2-0						
IS-2-1	1.07	−2.25	1.00	0.08	0.42	8.07
IS-2-2	0.85	5.75	1.00	0.06	2.27	10.81
OS-2-0	0.86	6.07	1.00	0.07	2.76	11.41
OS-2-1	1.10	−2.33	0.99	0.37	0.62	11.07
IS-3-0						
IS-3-1	1.22	−1.25	1.00	0.04	5.84	2.02
IS-3-2	1.19	4.62	1.00	0.09	9.85	4.26
IS-3-3	1.12	5.11	0.97	1.46	8.65	8.68
OS-3-0	0.84	6.57	1.00	0.14	0.79	7.11
OS-3-1	1.45	−7.58	0.99	0.41	6.50	7.25

and unit commitment outcomes ignored by the DC benchmark. This is nonetheless important since it is not trivial to describe a principled statistical approach that efficiently incorporates this information.

The encouraging performance of the IO recovery algorithm relative to the benchmark algorithms promotes the application of the algorithm for predicting prices when we can assume that the constraint structure is stable and consistent. However, in the case of MISO, due to a policy of delaying the publication of bidding data for three months, potential changes in the transmission topology, regulations, and market behavior it is likely that the constraint structure will be disrupted. It should be kept in mind that the IO recovery algorithm is not targeted only to predicting prices. The market structure recovered by the IO recovery algorithm allows ex post analyses including counterfactuals and estimation of secondary market characteristics; studies that cannot be directly arrived at from data-centric approaches such as the DC benchmark or more sophisticated statistical studies.

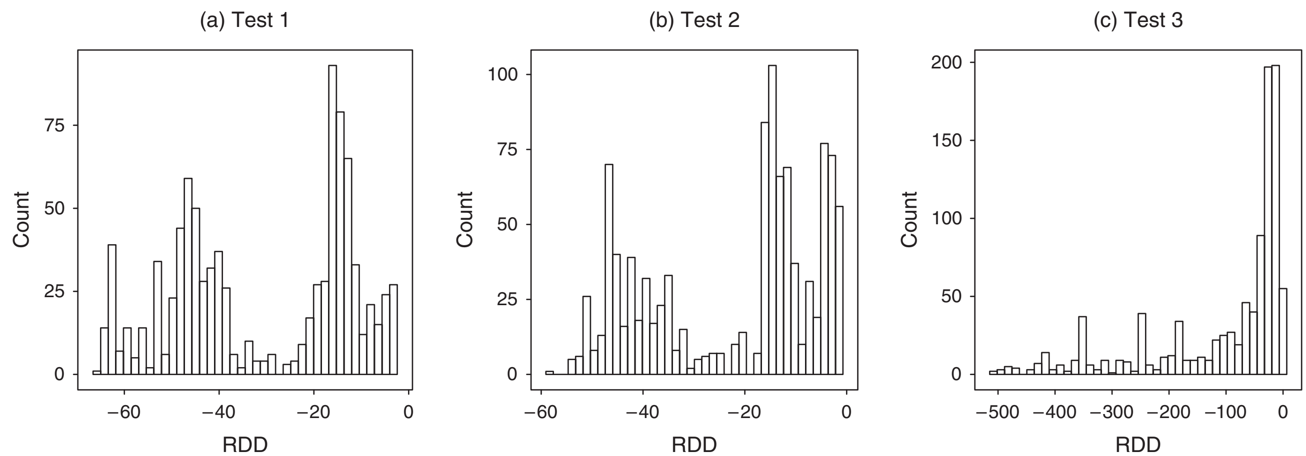
4.5. Evaluating Market Power Variation

The transmission constraint parameters are useful for measuring the distribution of market power through the market footprint. As noted in (Borenstein and Bushnell 1999), (Karthikeyan et al. 2013) and others, market power can vary substantially in an electricity market and this variation can be both over the footprint and over time. This spatial and temporal variation depends on the impact of a particular participant on active constraints and the ability of other participants to make up for production shortfalls that depend on participant bids and loss parameters. To capture this variation, Lee et al. (2011b) and (2011a) argue that useful market power measures can be constructed by using the RDD for each participant in the market. The RDD defines the demand surplus that a particular participant would observe for a marginal increase in price. In the transmission constrained electricity market this depends both on the transmission constraint contribution factors D and the loss parameters δ as well as the bid functions of other participants. The RDD is a precise reflection of the realized market state and does not necessarily reflect the expectations upon which a participant would condition their bid. For instance, only the realized set, and no highly likely alternative sets, of active constraints are taken into account. Xu and Baldick (2007) show that in a lossless locationally priced electricity market, this realized RDD can be efficiently calculated for each location given bids and the constraint factors of active constraints. To adapt this analysis to our model requires only a minor modification. In short, replacing the lossless energy balance equation with Constraint (3) and carrying through the same derivation as in Xu and Baldick (2007), the RDD for a particular participant i located at the reference bus becomes

$$\mathrm{RDD}(i) = -\bar{q}_i^T \Lambda_i \bar{q}_i + \bar{q}_i^T \Lambda_i \bar{D}_i (\bar{D}_i^T \Lambda_i \bar{D}_i)^{-1} \bar{D}_i^T \Lambda_i \bar{q}_i,$$

where, $\bar{q}_i$ and $\bar{D}_i$ are, respectively, the participant relative loss vector and the participant active constraint contribution matrix with entries, rows, and columns corresponding to i removed. Vector $\bar{q}_i$ has $Q-1$ entries and $\bar{D}_i$ is a $(Q-1)\times(Q-1)$ dimension matrix. With $b(j)$ denoting the location of participant j and for convenience allowing $i = Q$, $(\bar{q}_i)_j = q_{b(j)}$ and $(\bar{D}_i)_{rj} = D_{rb(j)}$. Matrix Λ_i is diagonal with entries corresponding to $S_j''(x_j)$ of each participant except i (i.e., Λ_i is a $(Q-1)\times(Q-1)$ dimension matrix). To calculate the RDD for a participant i not at the reference bus, δ and D can easily be refactored so that $b(i)$ is the reference bus by dividing q by $q_{b(i)}$ and zeroing the $b(i)$th row of D using the market clearing constraint. This calculation does not require resource capacities d for which allocation data are required for precise identification. As a

Figure 7. Distribution of Residual Demand Derivative for Samples with Differing Active Constraints

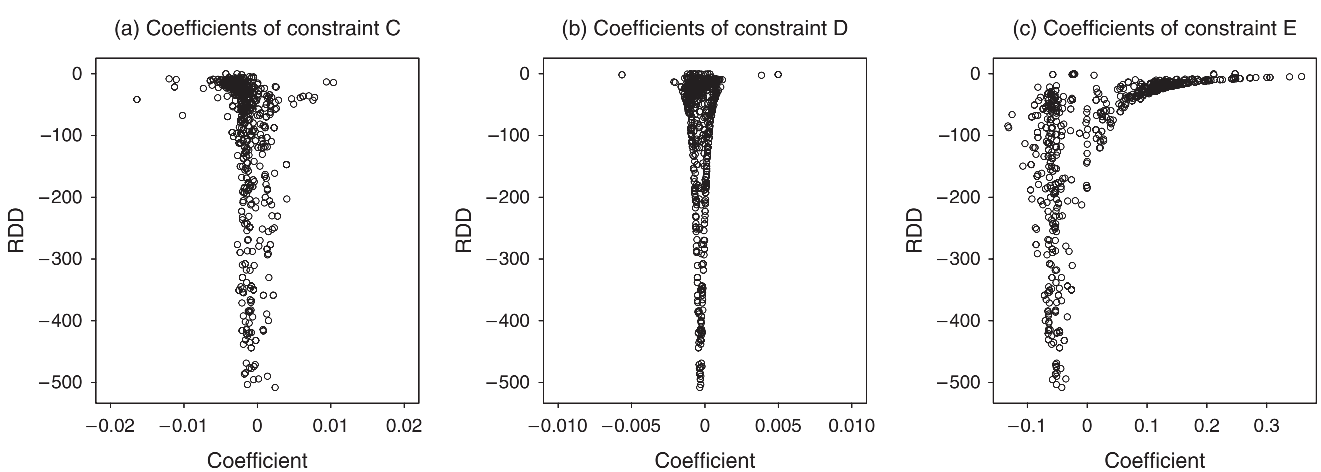


result, the residual demand derivatives shown in this section are resilient to systemic error from missing supply and demand data unlike the reconstruction of the pricing mechanism.

We have performed this calculation for hours from each of the test cases from the case study of Section 4.4. This study does not provide an exhaustive study of market power variation. However, we are able to observe the RDD under higher and lower market congestion conditions as proxied by the number of active constraints. Histograms showing the distribution of the RDD for each participant are shown in Figure 7. This figure shows an outcome for each of the three test sets with one, two, or three active constraints. Panels (a) and (b) of Figure 7 show very similar ranges of market power though the less constrained example in panel (a) appears to show a slightly more uniform distribution of the RDD. Panel (c) on the other hand is quite distinct. In this case three constraints are active, indicating that it is a higher congestion market hour. While the bulk of the participants face similar RDDs in the −70 to 0 range, there is a long left tail where participants face substantially higher magnitude RDDs. This is consistent with higher congestion being associated with greater magnitude and variation of market power.

Figure 8 shows the relationship between the RDD (y-axis) and constraint coefficients (x-axis). Details of IS-3-0 and the active constraints in this hour are reported in Online Appendix C.2. If the constraint is associated with higher levels of congestion, we expect to see a skewed distribution of the RDDs. In panels (a) and (b) the RDD is symmetrically distributed with respect to the constraint coefficients for Constraints C and D. On the other hand Figure 8(c) shows that the RDD is highly skewed with respect to coefficients for Constraint E. Locations with larger magnitude negative parameters for Constraint E are associated with higher market power for this hour. This demonstrates how different "sides" of the transmission constraint, likely associated with particular regions of the market, may face very different market conditions in LMP based markets.

Figure 8. Relationship Between Residual Demand Derivative and Active Constraint Coefficients for Example IS-3-0



5. Conclusion

This paper develops an inverse optimization-based method for recovering market structure from allocation and pricing data. It is justified when the allocation and locational prices correspond to the primal and dual of the optimal assignment problem. The IO recovery algorithm implements the methodology in a context specialized for electricity markets where prices are determined by competition for transmission resources and line loss. This institutional context features high-quality and wide data sets, with few samples and many attributes, which supports the noiseless assumptions made in our algorithms. In a broader context, incorporation of data errors should be considered carefully on a case-by-case basis. For more narrow data sets our algorithm admits a least-squares approach to estimate the parameters of the linear system with little modification but the appropriateness of implied assumptions on the source of error should be carefully evaluated.

The primary attributes identified with the inverse optimization algorithm are utilizations of transmission resources and loss parameters that can be identified from locational and transmission prices alone. Additional data including the observed allocation and bids allow for reconstruction of the pricing and allocation mechanism as well as derivation of the transmission constrained RDD. The inferences and analyses possible with different data sets are summarized in Table 7.

The methodology was applied to the MISO electricity market where we have evaluated the quality of recovered market structure by contrasting real market prices with predicted market prices. Correlation between predicted and true prices is very high in most cases indicating that the recovered market structure explains price variation very well. Nonetheless, there are hours where predicted prices are poor and even when correlation is high, prices are often shifted. We suspect these errors are due to incomplete market specification and unobservable market volumes that restrict supply and demand over the market footprint. In off-peak hours, when import demand is lower, correlation is much higher and absolute error is much lower. The simple model of losses that implicitly assumes a similar operating point for market samples used in the recovery is another source of error and we have shown the consistency of parameter estimations is contingent on the similarity of the operating point. In addition, the requirement that network topology is constant over the set of market samples may also contribute to error when line outages occur. This limitation may become more relevant if transmission switching (e.g., Hedman et al. 2010) becomes more prevalent.

The recovered market structure is of managerial value for both day-to-day operational decisions and longer-term investment decisions. The reconstruction of the pricing mechanism would allow a participant to evaluate ex post optimality of their bidding decisions by rerunning the mechanism with alternative bids. We show an efficient means to calculate the RDD from the recovered market structure, including losses, and publicly available data, which can help guide optimal bidding strategies. Empirical results on a very small set of MISO market hours show substantial variation in market power both within and between market hours. Various counterfactual analyses such as the estimation of the value of a generation capacity investment are possible with the reconstructed pricing mechanism.

Markets where transportation links between regions are capacitated are challenging for managers because market conditions can shift dramatically depending on which constraints are active. The methodology described in this paper can help managers determine operational strategies by using structural models to make market inferences and perform counterfactuals. Similarly, this may provide a mechanism for regulators to make inferences about markets where they cannot observe all transactions but have some knowledge of prices and allocation volumes. In the applications and extensions we have discussed estimation of missing market data, estimation of market power, and analysis of investment decisions as particularly worthy of further study. In addition, future work should consider robust handling of data error into the methodology to allow for the system of equations solved when the inverse optimization is overspecified and improve resilience to mispecification. The simplest method is to add an additive residual for each of the equations. Then, under the assumption that residuals are independent and have zero mean and common variance, the least-squares solution is an appropriate estimator. However, the independence assumption may be problematic.

With further refinement of these and related techniques, structure recovery methodologies based on inverse optimization may be a useful analytical tool for participants in institutionalized markets like electricity markets as well as decentralized but efficient markets exhibiting locational price variation. For participants,

Table 7. Required Data for Identification and Analyses

Identified attribute	D, λ	d	Mechanism	Reconstruct π, x	Residual demand derivative
Required data	π, ρ	π, ρ, x	π, ρ, x	π, ρ, x, bids	π, ρ, bids

Note. All identification is with respect to active constraints.

market structure inferences may allow managers to improve operational strategies. For market regulators, when they cannot observe all transactions but have some partial observations, the inferences allowed via these tools may be useful for setting policy.

Acknowledgments
The authors would like to thank seminar participants at the Becker-Friedman Institute, the University of Michigan, the University of Toronto, the University of Waterloo, and the University of Wisconsin-Madison.

Endnotes

[1] FERC Orders 630, 643, 683, and 715 detail restrictions to Critical Energy Infrastructure Information (CEII) related to transmission. In these orders, FERC sets boundaries for who it will divulge CEII that it holds and removes the requirement for companies to divulge transmission information including power flow data. Commenters cited in FERC Order 643 express concern that restrictions of this information will restrict market opportunities.

[2] Further details of the calculation of loss factors and their role in determining energy prices were discussed during the 2014 MISO Marginal Loss Workshop and are reported at https://www.misoenergy.org/_layouts/MISO/ECM/Redirect.aspx?ID=180736.

[3] The linear system that must be solved for the IO recovery is sparse and can be solved very quickly, though with larger systems there are substantial memory requirements. For the experiments in this section, the sparse linear solver from the Python SciPy package (http://www.SciPy.org) was used.

[4] 2011 State of The Market Report for the MISO Electricity Markets (https://www.potomaceconomics.com/wp-content/uploads/2017/02/2011-State-of-the-Market-Report.pdf).

[5] The procedure was run on an additional 267 hours, which resulted in an infeasible EDP because of errors in the reported bids. Problematic bids included ASM bids with declared minimum and allocated volumes greater than declared maximum volumes. Because it is not clear how MISO handles these bids and they were only present in a small number of cases, these hours were excluded from the analysis.

References

Ahuja RK, Orlin JB (2001) Inverse optimization. *Oper. Res.* 49(5):771–783.

Borenstein S, Bushnell J (1999) Market power in electricity markets: Beyond concentration measures. *Energy J.* 20(4):65–88.

Broda C, Weinstein DE (2008) Understanding international price differences using barcode data. NBER Working Paper No. 14017.

Cardell JB, Hitt CC, Hogan WW (1997) Market power and strategic interaction in electricity networks. *Resource Energy Econom.* 19(1):109–137.

Coşar AK, Grieco PLE, Tintelnot F (2015) Borders, geography, and oligopoly: Evidence from the wind turbine industry. *Rev. Econom. Statist.* 97(3):623–637.

Cremer H, Gasmi F, Laffont J (2003) Access to pipelines in competitive gas markets. *J. Regulatory Econom.* 24(1):5–33.

Engel C, Rogers JH (1996) How wide is the border? *Amer. Econom. Rev.* 86(5):1112–1125.

Hedman KW, Ferris MC, O'Neill RP, Fisher EB, Oren SS (2010) Co-optimization of generation unit commitment and transmission switching with $n-1$ reliability. *IEEE Trans. Power Systems* 25(2):1052–1063.

Hogan WW, Read EG, Ring BJ (1996) Using mathematical programming for electricity spot pricing. *Internat. Trans. Oper. Res.* 3(4):209–221.

Iyengar G, Kang W (2005) Inverse conic programming with applications. *Oper. Res. Lett.* 33(3):319–330.

Karthikeyan SP, Raglend IJ, Kothari DP (2013) A review on market power in deregulated electricity market. *Internat. J. Electrical Power Energy Systems* 48:139–147.

Kekatos V, Giannakis G, Baldick R (2014a) Grid topology identification using electricity prices. *2014 IEEE Power Energy Soc. General Meeting, PES '14* (IEEE, Piscataway, NJ), 1–5.

Kekatos V, Giannakis GB, Baldick R (2014b) Online energy price matrix factorization for power grid topology tracking. Preprint arXiv: 1410.6095.

Lee Y, Baldick R, Hur J (2011a) Firm-based measurements of market power in transmission-constrained electricity markets. *Power Systems, IEEE Trans.* 26(4):1962–1970.

Lee YY, Hur J, Baldick R, Pineda S (2011b) New indices of market power in transmission-constrained electricity markets. *IEEE Trans. Power Systems* 26(2):681–689.

Ma X, Song H, Hong M, Wan J, Chen Y, Zak E (2009) The security-constrained commitment and dispatch for midwest ISO day-ahead co-optimized energy and ancillary service market. *2009 IEEE Power Energy Soc. General Meeting, PES '09* (IEEE, Piscataway, NJ), 1–8.

McCallum J (1995) National borders matter: Canada-U.S. regional trade patterns. *Amer. Econom. Rev.* 85(3):615–623.

Mudrageda M, Murphy FH (2008) An economic equilibrium model of the market for marine transportation services in petroleum products. *Oper. Res.* 56(2):278–285.

Nocedal J, Wright SJ (1999) *Numerical Optimization*, 2nd ed. (Springer, New York).

Ruiz C, Conejo AJ, Bertsimas DJ (2013) Revealing rival marginal offer prices via inverse optimization. *Power Systems, IEEE Trans.* 28(3):3056–3064.

Sahni M, Jones R, Cheng Y (2012) Beyond the crystal ball. *IEEE Power and Energy Magazine* 10(4):35–42.

Schaefer AJ (2009) Inverse integer programming. *Optim. Lett.* 3(4):483–489.

Stott B, Jardim J, Alsac O (2009) DC power flow revisited. *IEEE Trans. Power Systems* 24(3):1290–1300.

Su CL, Judd KL (2012) Constrained optimization approaches to estimation of structural models. *Econometrica* 80(5):2213–2230.

Wang L (2009) Cutting plane algorithms for the inverse mixed integer linear programming problem. *Oper. Res. Lett.* 37(2):114–116.

Wolak FA (2011) Measuring the benefits of greater spatial granularity in short-term pricing in wholesale electricity markets. *Amer. Econom. Rev.* 101(3):247–252.

Wood AJ, Wollenberg BF (2012) *Power Generation, Operation, and Control* (John Wiley & Sons, New York).

Xu L, Baldick R (2007) Transmission-constrained residual demand derivative in electricity markets. *Power Systems, IEEE Trans.* 22(4):1563–1573.

Yang X, Zhang J (2007) Partial inverse assignment problems under norm. *Oper. Res. Lett.* 35(1):23–28.

Zhang J, Liu Z (1996) Calculating some inverse linear programming problems. *J. Computat. Appl. Math.* 72(2):261–273.

Zhang J, Liu Z (1999) A further study on inverse linear programming problems. *J. Computat. Appl. Math.* 106(2):345–359.

John R. Birge is the Jerry W. and Carol Lee Levin Distinguished Service Professor of Operations Management at the University of Chicago Booth School of Business. His work focuses on theory, computation, and, as in this paper, model development for decision making under uncertainty.

Ali Hortaçsu is the Ralph and Mary Otis Isham Professor of Economics at the University of Chicago. He works on modelling market mechanisms, with applications in energy, finance, and e-commerce.

Michael Pavlin is an assistant professor of operations and decision sciences at the Lazaridis School of Business and Economics, Wilfrid Laurier University. His research interests include data analytics and the economics of services with applications in finance, healthcare and energy.