Utilities Policy 18 (2010) 227–246

Contents lists available at ScienceDirect

Utilities Policy

journal homepage: www.elsevier.com/locate/jup

Using restructured electricity supply industries to understand oligopoly industry outcomes☆

Frank A. Wolak*

Program on Energy and Sustainable Development, Department of Economics, Stanford University, Stanford, CA 94305-6072, USA

A R T I C L E I N F O

Article history:
Received 7 January 2010
Received in revised form
19 September 2010
Accepted 20 September 2010

Keywords:
Electricity market design
Electricity auction design
Market performance measurement
Diagnosing market power

A B S T R A C T

This paper argues that many determinants of generic oligopoly market outcomes can be studied in bid-based wholesale electricity markets under much weaker assumptions than in other oligopoly industries because of their rich data, regulatory history, and clearly specified market rules. These methods are compared to those used in existing studies of oligopolistic industries where the best data available are market-clearing prices and quantities and demand and cost shifters. The extent to which the methods used in bid-based wholesale electricity markets generalize conventional methods is explained in detail and major applications of these techniques are summarized. Lessons from the study of wholesale electricity markets for the monitoring and design of other oligopolistic markets are also discussed.

1. Introduction

Starting with Chile in the 1980s, a growing number of countries around the world have restructured their electricity supply industries to introduce competition into the wholesale and retail segments. Electricity producers, traders, and retailers are granted open access to the bulk transmission network and local distribution network at regulated prices. All wholesale electricity is traded through a centralized short-term market that sets an hourly or half-hourly price for energy, but market participants are free to enter into forward contracts that clear against this short-term market. Competition among electricity retailers to supply final consumers exists to varying degrees.

Capturing a significant fraction of the potential benefits of industry restructuring has been much more challenging than many observers at first thought. The characteristics of the product and the technology used to produce it make bid-based wholesale electricity markets extremely susceptible to the exercise of unilateral market power. Electricity is very costly to store, its production is subject to severe capacity constraints, and it must be delivered through a transmission network with finite capacity, all of which limit the magnitude of the supply response to a firm's unilateral attempts to raise prices. Finally, because of how electricity has historically been priced to final consumers, the real-time wholesale demand for electricity is close to perfectly price inelastic. For firms that own a substantial fraction of the generation capacity in the industry, the combination of this inelastic aggregate demand with a limited supply response from competitors causes these suppliers to face steep residual demand curves, particularly during the high demand periods of the day. In an oligopolistic industry, the more inelastic the residual demand curve a firm faces, the greater is its unilateral ability to withhold output to raise market prices through higher offer prices or less output made available at the same offer price.

The susceptibility of bid-based wholesale electricity markets to the exercise of unilateral market power makes them an ideal environment in which to study the determinants of oligopolistic firm behavior and industry outcomes. In these markets, highly concentrated industry structures and unexpected high levels of demand are unnecessary for the occurrence of market outcomes that reflect the exercise of substantial unilateral market power. An additional advantage of studying bid-based wholesale electricity markets is the fact that they produce vast amounts of data each day about the strategies firms employ to exercise unilateral market power. Each market participant is required to submit its willingness-to-supply and purchase electricity for each hour or half-hour of the day for all possible market prices. In addition, the regulated or government-owned monopoly history of the industry implies that there is publicly available data on the technological characteristics of the production process—heat rates, operating and maintenance costs and emissions rates of the generation facilities and the physical

☆ I would like to thank the editor, Thomas Tangerås, and two anonymous referees for helpful comments on a previous draft of this paper. Partial financial support for this research was provided by the National Science Foundation.

* Tel.: +1 650 723 3944; fax: +1 650 7241717.
E-mail address: wolak@zia.stanford.edu.

0957-1787/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jup.2010.09.002

characteristics of the transmission network. This is considerably more information about firm strategies and production costs than is available to researchers studying other oligopolistic industries, such as automobiles (Bresnahan, 1981, 1987; Goldberg, 1995; Berry et al., 1995), personal computers (Bresnahan et al., 1997), and breakfast cereals (Nevo, 2001). In these industries, the researcher typically has, at best, data on the market-clearing prices and quantities of each product sold and variables thought to influence production costs and the level of demand for each product.

A final advantage of studying wholesale electricity markets relative to other oligopolistic industries is that the market rules are clearly specified and publicly available. Specifically, the researcher knows precisely how the wholesale prices suppliers receive and demanders pay are determined from the actions taken by these two sets of market participants. This is dramatically different from other oligopolistic industries, where a standard criticism of the empirical model implemented is that the choice of the economic model of firm behavior and the price-setting process is ad hoc because it assumes away important details of the competitive interactions between firms and between firms and consumers. For example, in most industries wholesale prices are set through bilateral negotiations between the seller and each buyer. The usual Cournot quantity-setting or Bertrand price-setting modeling assumptions used to simplify more complex price-setting processes can significantly impact the researcher's conclusions about the ability and incentive of firms to exercise unilateral market power and the competitiveness of observed market outcomes.

In contrast, wholesale electricity markets have detailed rules that must be approved by the regulatory body specifying precisely how market-clearing prices are determined. These rules define the feasible set of market participant actions, what information the market participants know when they take these actions, and how they are compensated for their actions. Consequently, the researcher does not need to make any ad hoc assumptions about how a market participant's actions impact the price it pays or receives; the market rules on file with the relevant regulatory authority specify precisely how each market participant's actions impact its revenues and costs. In this sense, wholesale electricity markets are like experimental economics markets except that experienced and highly compensated market participants play a well-defined game for large sums of money.

Studying wholesale electricity markets allows the empirical researcher to address many of the concerns raised by Fisher (1989, 1991) regarding the usefulness of applying game-theoretical methods to oligopoly contexts. The data on bids and market outcomes, cost data from the former regulated regime, and detailed market rules severely limit the set of possible economic models that can provide institutionally and statistically valid descriptions of observed data. A common problem in modeling market outcomes in oligopolistic industries where only market-clearing price and quantity data are available is the uncertainty about how the actions of market participants impact their profits and those of their competitors. The market rules for bid-based wholesale electricity markets completely eliminate this source of modeling uncertainty. Instead, the major challenge in studying bid-based wholesale electricity markets is formulating a computationally tractable model of market outcomes that is not dramatically inconsistent with the actual market rules.

The goal of this paper is to demonstrate how the features of wholesale electricity markets described above can be used to recover a rich set of economic primitives, direct measures of market performance, and measures quantifying the ability and incentive of a supplier to exercise unilateral market power, while employing fewer untestable assumptions relative to the approaches used to study these issues in other oligopolistic industries. The methods described below cannot be implemented in other oligopoly environments because the necessary data is unavailable and/or there is uncertainty about the form of market institutions. Instead, functional form assumptions for market demand and firm-level cost functions, as well as untestable assumptions about the form of competition (typically, Cournot quantity-setting or Bertrand price-setting behavior), must be made to gain traction on these questions in oligopolistic industries where only market-clearing price and quantity data are available. Consequently, wholesale electricity markets can provide insight about the determinants of market performance and the design of economic institutions that limit the ability and incentive of firms to exercise unilateral market power without many of the untestable assumptions employed in similar studies of other oligopolistic industries.

The four major issues addressed in this paper are: (1) testing the assumption of expected profit-maximizing behavior, (2) quantifying the ability and incentive of a supplier to exercise unilateral market power, (3) measuring overall market performance, and (4) assessing the competitive impacts of mergers. The data available and clearly specified rules for bid-based wholesale electricity markets allow a test of the hypothesis of expected profit-maximizing behavior without any assumptions on the functional form for the demand for electricity, minimal assumptions on the form of the variable cost function for producing electricity, and without assuming a specific oligopoly equilibrium model – the key assumptions required to test this hypothesis with data on market-clearing prices and quantities. Section 3 describes the details of this test and discusses the results of applying it in several electricity markets.

The data availability and precisely specified market rules allow the construction of direct measures of the firm-level ability and incentive to exercise unilateral market power. These measures of the ability and incentive to exercise unilateral market power can be constructed from the data on market participant bids and offers and market-clearing prices and quantities. Section 4 describes the construction of these measures and summarizes applications to wholesale electricity markets in California, Spain, and New Zealand.

The exercise of unilateral market power reduces the economic efficiency of wholesale market outcomes. Information on the technological characteristics of the generation facilities, input fuel prices, and bid and market outcome data can be used to construct a measure of market performance. This measure compares actual market prices to the market prices that would exist if no firms were able to exercise unilateral market power. This procedure accounts for daily input fuel price changes, generation unit outages, and emissions permit price changes in computing the counterfactual no-market-power outcomes. Section 5 describes the details of this procedure and summarizes applications to Spain, California, New England, and the PJM Interconnection region in the United States.

Quantitative merger analysis has recently begun to rely on models of industry equilibrium estimated or calibrated from data on market-clearing prices and quantities (see Werden and Froeb, 1994, 2002; Nevo, 2000). These models rely on functional form assumptions for firm-level demand and cost functions, as well as a model of strategic interaction determining market outcomes. Because of the availability of bid data and information on generation unit-level production costs, these assumptions can be dispensed with when analyzing proposed mergers in bid-based wholesale electricity markets. Section 6 proposes a method for estimating the competitive effects of mergers and provides a hypothetical application of this procedure.

2. Empirical models of oligopoly market outcomes

This section describes the identifying assumptions typically used to estimate models of oligopoly market outcomes that are subsequently used to test the hypothesis of profit-maximizing behavior, measure the ability and incentive of a firm to exercise unilateral market power, quantify overall market performance, and analyze the impact of proposed mergers. The fundamental challenge in modeling oligopoly outcomes is that the payoff of each market participant depends on its own actions as well as the actions of its competitors and consumers. By definition, an oligopolistic industry is one in which there are a small number of firms, so the assumption that each firm is a price-taker is clearly unrealistic. However, imposing the assumption that observed market-clearing prices and quantities are the result of all firms simultaneously playing best-reply strategies to the strategies of their competitors – the observed market outcomes are Nash equilibria – imposes few restrictions on the joint distribution of these variables.

To illustrate this point, consider the following generic N-firm oligopoly. Let $\Pi_i(\theta_1, \theta_2,\ldots,\theta_i, \ldots, \theta_N)$ equal the profit function of firm i and θ_i equal the action taken by firm i and Θ_i is the set of allowable values of θ_i. Consistent with all standard oligopoly models, each firm's profits depend on its own actions as well as those of its competitors. If the only data available to the researcher are the values of θ_i for each firm, then assuming that θ_i^*, the observed value θ_i, is the solution to:

$$\theta_i^* = \mathrm{argmax}_{\theta_i \in \Theta_i} \Pi_i\left(\theta_1^*, \theta_2^*, \ldots, \theta_i, \ldots, \theta_N^*\right), \tag{2.1}$$

is insufficient to test the assumption of profit-maximizing behavior, measure the ability or incentive of a firm to exercise unilateral market power, quantify market performance, or analyze the competitive effects of proposed mergers. In contrast, for bid-based wholesale electricity markets, all of these tasks can be accomplished under minimal assumptions because of data availability and the precision of the market rules. In particular, the market rules specify the elements of θ_i and Θ_i, and the functional form for $\Pi_i(\theta_1^*, \theta_2^*, \ldots, \theta_i, \ldots, \theta_N^*)$. The bid and market outcome data can be used to compute the realized value of $\Pi_i(\theta_1^*, \theta_2^*, \ldots, \theta_i, \ldots, \theta_N^*)$ except for the variable cost of production, which can be estimated with an acceptable degree of precision using data from the former regulated regime or estimated econometrically from the assumption of expected profit-maximizing behavior.

Conventional approaches to measuring the ability and incentive of a firm to exercise unilateral market power, overall market performance, and analyzing mergers require functional form assumptions on the unknown profit functions $\Pi_i(\theta_1^*, \theta_2^*, \ldots, \theta_i, \ldots, \theta_N^*)$, $i = 1, \ldots, N$. In most industries, the best data an empirical researcher can obtain is the price charged and quantity sold for all of the products offered by each firm. The usual assumptions used to specify the functional form for $\Pi_i(\theta_1^*, \theta_2^*, \ldots, \theta_i, \ldots, \theta_N^*)$ given data on observed prices charged and quantities sold are: (1) a functional form for the demand for each product sold by the firm, (2) a functional form for the variable cost of production for the firm, and (3) an equilibrium model of firm behavior.

These techniques can be illustrated using an example based on Rosse (1970). Let $P(q|W, \beta, \varepsilon)$ denote the inverse demand function facing a monopolist and $C(q|Z, \beta, \eta)$ its variable cost function. Both of these functions have known forms that are parameterized by β. The vectors W and Z are demand and cost function shifters respectively, β is the vector of parameters to be estimated, and ε and η are unobserved (to the econometrician) stochastic shocks that are assumed to be at least uncorrelated with both W and Z. For simplicity, assume these shocks are observable to the monopolist. The discussion becomes more complicated if this assumption is relaxed, but the general conclusions remain the same. In terms of the inverse demand and variable cost functions, the profit function of the monopolist is:

$$\pi(q) = P(q|W, \beta, \varepsilon)q - C(q|Z, \beta, \eta), \tag{2.2}$$

where θ in the notation of equation (2.1) is equal to q. The first-order condition for the profit-maximizing choice of q by the monopolist is:

$$\pi'(q) = P'(q|W, \beta, \varepsilon)q + P(q|W, \beta, \varepsilon) - C'(q|Z, \beta, \eta) = 0, \tag{2.3}$$

where $f = (x|y, z)$ denotes the derivative of $f = (x|y, z)$ with respect to x. If the researcher has market-clearing price and quantity data and values of the demand and supply shifters, W and Z, for a cross-section of monopolists selling the same homogenous product, then functional form assumptions on the demand and cost functions and the model of equilibrium outcome determination, combined with the assumption of profit-maximizing behavior, are generally sufficient to estimate β. Rosse (1970) demonstrated this and obtained estimates of β for a cross-section of monopoly newspaper markets.

The logic for this statement is that for certain functional form assumptions on the inverse demand and variable cost functions, $C'(q|Z, \beta, \eta)$, the marginal cost of producing output level q, is the only unobserved variable in equation (2.3). The value of, $P(q|W, \beta, \varepsilon)$, the inverse market demand curve at output level q, is equal to market-clearing price, p, and q is the observed quantity of the product sold. Both of these variables are assumed to be observed. The functional form assumption for the inverse demand curve implies that its slope, $P'(q|W, \beta, \varepsilon)$, is observable given a value for β and ε. If the inverse demand curve is additively separable in ε, then $C'(q|Z, \beta, \eta)$ can be recovered directly from solving (2.3) as:

$$C'\left(q^E|Z, \beta, \eta\right) = P'\left(q^E|W, \beta\right)q + P\left(q^E|W, \beta, \varepsilon\right), \tag{2.4}$$

where I have suppressed the dependence of the slope of the inverse demand curve on ε to reflect the assumed additive separability. Equation (2.4) only allows the researcher to recover the value of the monopolist's marginal cost at the observed level of output, q, using the assumption of profit-maximizing behavior and an assumed inverse demand function. Re-writing equation (2.4) as:

$$\begin{aligned}&\left[P\left(q^E|W, \beta, \varepsilon\right) - C'\left(q^E|Z, \beta, \eta\right)\right] \Big/ P\left(q^E|W, \beta, \varepsilon\right)\\ &\quad = -P'\left(q^E|W, \beta, \varepsilon\right)q / P\left(q^E|W, \beta, \varepsilon\right),\end{aligned} \tag{2.5}$$

demonstrates that the slope of the inverse demand function is the missing ingredient necessary to determine the extent to which the monopolist exercises unilateral market power by setting prices in excess of its marginal cost. Consequently, a researcher's functional form assumptions for $P(q^E|W, \beta, \varepsilon)$ can significantly impact his conclusions about the ability of a firm to exercise unilateral market power through the slope of the inverse demand curve that this functional form assumption implies.

The assumed functional form for the market demand curve can also impact the researcher's variable cost function estimates. To see this, call the right-hand side of equation (2.4), $C'(q|Z, \beta, \eta)^{\mathrm{obs}}$, because it can be constructed from the observed quantities, p, q, W, and Z, given a functional form assumption for $P(q^E|W, \beta, \varepsilon)$, because are observed by the researcher. If the researcher assumes $C'(q|Z, \beta, \eta)$ is additively separable in η, so that $C'(q|Z, \beta, \eta) = c'(q|Z, \beta) + \eta$, then the potential exists to obtain consistent estimates of the elements of β in the marginal cost function by applying instrumental variables techniques to the equation

$$C'(q|Z, \beta, \eta)^{\mathrm{obs}} = c'(q|Z, \beta) + \eta \tag{2.6}$$

with W and Z as instruments, where $c'(q|Z, \beta)$ is the deterministic portion of the marginal cost function specified by the researcher.

Joint estimation of the inverse demand function and the first-order condition in (2.4) or (2.6) can, but does not always, improve the efficiency of the estimation of β. To the extent that there are over-identifying restrictions (for example, there are more elements of W and Z than there are elements of β) the assumption of profit-maximizing behavior can be tested assuming the validity of functional form assumptions for the inverse demand function and marginal cost function.

The intuition embodied in this monopoly example is employed in all of the papers that estimate models of oligopoly equilibrium in order to assess the validity of the assumption of profit-maximizing behavior, measure the ability and incentive of a firm to exercise unilateral market power, quantify the costs of the exercise of unilateral market power, or analyze the competitive effects of mergers. For example, to compute the no-market-power equilibrium used to measure market performance, the researcher needs to estimate the parametric marginal cost function, $C'(q|Z,\beta,\eta)$. This function can then be used to compute the counterfactual no-market-power equilibrium by solving the equation $P(q|W,\beta,\varepsilon) = C'(q|Z,\beta,\eta)$ for q. If q^{comp} is counterfactual competitive output, then $P(q^{\text{comp}}|W,\beta,\varepsilon)$ is the counterfactual competitive benchmark price.

To understand the role played by the researcher's assumption about the equilibrium model of competition among firms, consider an example patterned after Porter (1983). Suppose that the demand function facing a homogenous product N-firm oligopoly takes the form $D(p|W,\beta,\varepsilon)$ and that $C_i(q|Z,\beta,\eta)$ is now the variable cost function for firm i ($i = 1, \ldots, N$). Let q_i denote the output of firm i and $Q = \sum_{i=1}^{N} q_i$. The solution in p of the equation $D(p|W,\beta,\varepsilon) = Q$, determines the market-clearing price.

In order to derive an estimating equation for each firm similar to (2.4) or (2.6), the researcher must assume a specific equilibrium model of competition among firms. This assumption defines the firm's residual demand curve. For the quantity-setting Cournot equilibrium, competitors to firm i are assumed to sell a fixed quantity of output, regardless of the market price, p. Define $Q_{-i} = \sum_{j=1, j\neq i}^{N} q_j$, as the total output of all other firms besides firm i. The residual demand function facing firm i becomes $\text{DR}_i(p) = D(p,W,\beta,\varepsilon) - Q_{-i}$. In general, the residual demand function faced by firm i gives the difference between the market demand and the amount that firm i's competitors are willing to supply at each price, p.

Firm i's profit function becomes:

$$\pi_i(p) = \text{DR}_i(p)p - C_i(\text{DR}_i(p), Z, \beta, \eta). \tag{2.7}$$

Solving for the value of p that maximizes (2.7) and substituting back into the residual demand curve that firm i faces yields the profit-maximizing quantity choice for firm i as function of Q_{-i}. The usual way of solving for firm i's profit-maximizing quantity choice is to use the inverse demand curve associated with $D(p|W,\beta,\varepsilon)$, say $P(Q|W,\beta,\varepsilon)$, and write firm i's profit function as:

$$\pi_i(q_i) = P(Q_{-i} + q_i|W,\beta,\varepsilon)q_i - C_i(q_i|Z,\beta,\eta). \tag{2.8}$$

Although this is the usual way the profit function for a Cournot competitor is specified, it is not as easy to generalize this form of firm i's profit function to the case of all possible oligopoly models and to the case of wholesale electricity markets, so the remainder of this discussion will use the profit function in equation (2.7).

The first-order condition for maximizing (2.7) with respect to p is equal to:

$$\pi_i'(p) = \text{DR}_i'(p)p + \text{DR}_i(p) - C_i'(\text{DR}_i(p)|Z,\beta,\eta)\text{DR}_i'(p) = 0. \tag{2.9}$$

All of the variables in (2.9) are directly observed except $C_i'(\text{DR}_i(p)|Z,\beta,\eta)$. Firm i's actual output, q_i^0 is equal to $\text{DR}_i(p^0)$, where p^0 is the observed market price. Re-arranging (2.9) yields the implied marginal cost at output level q_i for firm i:

$$C_i'(\text{DR}_i(p)|Z,\beta,\eta) = p + \text{DR}_i(p)/\text{DR}_i'(p), \tag{2.10}$$

which can be computed from observed magnitudes given a parametric functional form for the market demand and an assumed equilibrium model of competition that determines the form of the residual demand curve that firm i faces. Let $C_i'(\text{DR}_i(p)|Z,\beta,\eta)^{\text{obs}}$ denote the value of the right-hand side of (2.10). By the same logic as the monopoly case, assuming certain functional forms for $C_i'(\text{DR}_i(p)|Z,\beta,\eta)$ allows instrumental variables techniques to be applied to estimate the parameters of the cost function using $C_i'(\text{DR}_i(p)|Z,\beta,\eta)^{\text{obs}}$ as the dependent variable. Joint estimation of the market demand function and the first-order conditions in (2.9) for each firm can improve the efficiency of the estimation of β.

To the extent that there are over-identifying restrictions (for example, there are more elements of W and Z than there are elements of β) the assumption of profit-maximizing behavior can be tested assuming the validity of functional form assumptions for the market demand function and the firm-level marginal cost functions and the assumed equilibrium model of competition that specifies each firm's residual demand curve.

Equation (2.9) can be rearranged to obtain a measure of the ability of firm i to impact the market-clearing price. A direct measure of the ability of firm i to raise price by withholding output is:

$$-\text{DR}(p)/(p^*\text{DR}_i'(p)) = -1/e_i(p), \tag{2.11}$$

where $e_i(p) = (p/\text{DR}_i(p))\text{DR}_i'(p)$ is the own-price elasticity of the residual demand curve faced by firm i evaluated at price p. The magnitude, $-1/e_i(p)$, gives the percentage increase in the market-clearing price that results from a one percent reduction in the firm's actual output. In general, computing $-1/e_i(p)$ requires a parametric functional form for the market demand and an equilibrium model of firm behavior to compute firm i's residual demand curve, the same assumptions needed to recover firm i's marginal cost, $C_i'(\text{DR}_i(p)|Z,\beta,\eta)$, from its first-order conditions for profit-maximization.

All homogenous product oligopoly models can be recast as each firm maximizing its unilateral profits against the residual demand curve constructed using the equilibrium actions chosen by its competitors. These include the price-setting Bertrand model, the dominant firm and competitive fringe model, the Stackelberg leader-follower model, and the supply function equilibrium model of Klemperer and Meyer (1989), to name a few. Given functional form assumptions on the market demand curve and the firm-level variable cost curve, the assumption of (expected) profit-maximizing behavior and the assumed equilibrium model can be used to compute implied marginal costs which can then be used to estimate the firm-level marginal cost functions. Firm-level measures of the ability to exercise unilateral market power can be computed in the same manner as shown in equation (2.11) for all of these equilibrium models of competition.

This general approach of specifying functional forms for the market demand and firm-level variable cost functions and an assumed equilibrium model of competition can also be applied to multi-product firm oligopoly industries. Bresnahan (1987) specifies a discrete-choice demand structure where each individual decides whether to purchase an automobile, and if so, which model. He aggregates this discrete-choice demand structure across all consumers to derive a system of aggregate model-level demand equations. Using various assumptions about the nature of strategic interaction – specifically, price-taking competitive behavior, price-setting Nash/Bertrand competition, or collusion among automobile producers – he estimates the parameters of this aggregate demand system along with the parameters of the product-level marginal cost functions implied by the first-order conditions for firm-level profit-maximization implied by each assumed market equilibrium model. Bresnahan (1981) allows for a richer stochastic specification in the aggregate demand system, but follows the same basic procedure to recover estimates of product-level marginal cost functions.

Berry et al. (1995) specify a multinomial logit discrete-choice demand structure at the consumer-level and assume unobservable (to the econometrician) stochastic consumer-level marginal utilities of product attributes. These marginal utilities are assumed to be independent, non-identically normally distributed across product attributes and independent, identically distributed across consumers. Integrating individual-level purchase probabilities with respect to these normal distributions yields product-level aggregate market shares. The authors assume that the conditional indirect utility functions for each consumer contain the same vector of unobservable (to the econometrician) product characteristics, and that these product characteristics are uncorrelated with all observable product characteristics. This stochastic structure and the assumption that market-clearing prices are determined through price-setting Bertrand competition among firm induces correlation between equilibrium prices and the vector of unobserved random product characteristics in the aggregate demand system. Berry et al. (1995) propose and implement an instrumental variables estimation technique that exploits the lack of correlation between observed and unobserved product characteristics to estimate the demand system jointly with the marginal cost function using the first-order conditions for unilateral profit-maximizing product-level pricing analogous to equation (2.10).

Goldberg (1995) uses individual household-level data to estimate a flexible discrete-choice model for automobile purchases at the household level. She then uses weights giving the representativeness of each of these households in the population of US households to produce a system of aggregate demand functions for automobiles based on the choice probabilities implied by her model of household-level automobile demand. Using the assumption of the price-setting Nash/Bertrand behavior among automobile producers, she then computes implied marginal cost estimates similar to those given in (2.10), which she uses to estimate a parametric marginal cost function for each automobile model.

The most important conclusion to draw from this line of research is that the marginal cost function estimates are the direct result of the combination of the assumed functional forms for the aggregate demand and variable cost functions for the products under consideration and the assumed equilibrium model of competition among firms. The first-order conditions for profit-maximization then directly determine the implied marginal cost for that product by an equation analogous to (2.10). Because these parametric demand and cost functions are used to compute an estimate of the ability of a firm to exercise unilateral market power, compute counterfactual competitive benchmark prices, and assess the competitive effects of mergers, these three assumptions – (1) a functional form for aggregate demand and firm-level variable cost functions, (2) an equilibrium model of competition, and (3) profit-maximizing behavior – are required for these analyses as well.

3. Modeling firm behavior and market outcomes in electricity markets

This section describes how to the methods presented in the previous section can be applied to wholesale electricity markets under substantially weaker assumptions. The model of expected profit-maximizing behavior in a bid-based wholesale electricity market from Wolak (2000) forms the basis for deriving these methods and for the tests of the hypothesis of expected profit-maximizing bidding behavior proposed in Wolak (2003a, 2007). I demonstrate how the availability of bids and detailed market rules allow the econometrician to identify the underlying firm-level cost function purely through the assumption of expected profit-maximizing behavior. No functional form assumption for the aggregate demand curve or assumption about an equilibrium model of competition is necessary.

The bid data submitted to the wholesale electricity market is rich enough and the market rules are sufficiently detailed that it is possible to compute a vector composed of the bid and market outcome data that has an expected value equal to the first-order conditions from the firm's expected profit-maximization problem. These equations can be used to recover the firm's variable cost function and to test the assumption of expected profit-maximizing behavior. The functional form and market equilibrium assumptions required to identify a firm's residual demand curve in Section 2 are no longer necessary because the bids and offers submitted by market participants can be used to compute the realized residual demand curve faced by each supplier in a bid-based wholesale electricity market.

3.1. A model of expected profit-maximizing bidding behavior

In bid-based markets, the amount a supplier sells and the price it is paid depends on its bids and the bids of other market participants. Because each supplier's payment depends on the actions of all market participants, bid-based markets are special cases of the oligopoly markets described in Section 2. These markets differ from standard oligopoly markets in the ways that a supplier's actions can influence the amount of energy it produces and the price it is paid. There are detailed market rules approved by the relevant regulatory body, the Federal Energy Regulatory Commission (FERC) in the United States, that precisely specify the manner in which bids must be submitted and how these bids are used to determine the amount each firm produces and the price it is paid for its output.

Unlike standard oligopoly markets, it is possible in all bid-based wholesale electricity markets to take the bids and offers submitted by all market participants and combine them with other inputs to the market operator, such as the physical characteristics of the transmission network, and replicate the observed market prices and quantities produced. This precise relationship between the bids suppliers submit and the amount of output they produce and the prices they are paid can be used to formulate expected profit-maximizing bidding strategies. This section outlines the model of expected profit-maximizing bidding behavior in Wolak (2000, 2003a) that is the foundation for all of the methods described in the remaining sections of the paper.

In order to present this model, I first discuss the market rules that specify how suppliers submit bids and how market prices are determined from these bids for the National Electricity Market (NEM) in Australia, because data from this market are used in the empirical work that I discuss later in this section. Although the details of these rules are specific to the NEM, I will note how this estimation and hypothesis testing procedure can be adapted to any bid-based wholesale electricity market currently operating.

In the NEM, a supplier submits 10 daily price bids and 10 half-hour quantity bids to construct the 48 half-hourly supply curves for the day for each generation unit, or genset. These price bids are required to be greater than or equal to an administratively set minimum price bid and less than or equal to an administratively set maximum price bid. Each quantity bid is required to be greater than or equal to zero, and less than the capacity of the generation unit, and the sum of the 10 half-hourly quantity bids must be less than or equal to the capacity of the genset. These market rules imply that a NEM supplier has a daily strategy set for each genset that is a compact subset of a 490-dimensional Euclidian space – 10 daily price bids and 480 half-hourly quantity bids. Electricity generation plants are typically composed of multiple gensets at the same location. Moreover, suppliers in wholesale electricity markets typically own multiple plants, so the dimension of a supplier's daily strategy set can easily exceed 2000–3000, but it is still a compact subset of a finite-dimensional Euclidian space because of the market rules characterizing the form of feasible price and quantity bids described above.

With bid and market outcome data, the only assumption needed to recover a generation unit-level variable cost functions is that the supplier chooses its daily price bids and half-hourly quantity bids to maximize its expected profits from participating in the wholesale electricity market that day subject to the constraints on the price and quantity bids described above. This assumption yields bid price and bid quantity moment restrictions that can be used to estimate multiple-output, generation unit-level cost functions.

Virtually all wholesale electricity markets set prices on an hourly or half-hourly basis. In the NEM, each day of market operation, d, is divided into half-hour load periods denoted by the subscript i beginning with the 4:00–4:30 am period and ending with the 3:30–4:00 am period the following day. Suppliers are required to submit their price and quantity bids for the entire day by 11:00 am of the day before the energy is to be produced.

Let Firm A denote an expected profit-maximizing supplier that owns multiple gensets. Define the following variables for each day:

Q_{id}: market demand in load period i of day d
$SO_{id}(p)$: amount of capacity bid by all other firms besides Firm A into the market in load period i of day d, at price p
$\mathrm{DR}_{id}(p) = Q_{id} - SO_{id}(p)$: residual demand curve faced by Firm A in load period i of day d, at price p
QC_{id}: fixed-price forward contract quantity for load period i of day d for Firm A
PC_{id}: quantity-weighted average (overall forward contracts signed for that load period and day) fixed-price forward contract price for load period i of day d for Firm A.
$\pi_{id}(p)$: variable profits to Firm A at price p, in load period i of day d
Q_{ijd}: output in load period i from genset j owned by Firm A for day d
Q_{jd}: vector of half-hourly outputs of genset j, $Q_{jd} = (Q_{1jd}, Q_{2jd},\ldots,Q_{48jd})$, for day d
$C_j(Q_{jd},\beta_j)$: daily operating cost of producing the output vector Q_{jd} for genset j
β_j: parameters of daily operating cost function for genset j
$SA_{id}(p,\theta)$: bid function of Firm A for load period i of day d giving the amount it is willing to supply as a function of the price p and bid parameter vector θ defined below.

I assume that market demand, Q_{id}, is completely price inelastic without loss of generality because any price-responsive demand bid into the wholesale market can be modeled as a "negawatt" supplier, using the following logic. Let $\mathrm{DB}_{id}(p)$ equal the aggregate of all demand bids at price p in period i of day d. Define $\mathrm{SN}_{id}(p) = \mathrm{DB}_{id}(p_{\min}) - \mathrm{DB}_{id}(p)$ (where $p_{\min}$ is minimum allowable price) as the aggregate supply of "negawatts" or demand reductions relative to $\mathrm{DB}_{id}(p_{\min})$, the maximum possible level of demand in period i of day d. Redefining Q_{id} as the sum of all perfectly inelastic demand bids plus $\mathrm{DB}_{id}(p_{\min})$, and treating $\mathrm{SN}_{id}(p)$ as a supply curve, implies that the market demand can be treated as perfectly inelastic for the purposes of constructing residual demand curves.

Wholesale electricity markets differ in the number of bid parameters they allow suppliers to submit, the degree of flexibility suppliers have to alter their price and quantity bids throughout the day, and the mechanism used to determine market prices from the bids submitted, but all existing short-term wholesale electricity markets use supply curves determined by a finite-dimensional parameter vector. These differences in allowable bid functions only impact how $\mathrm{SA}_{id}(p,\theta)$ depends on θ, and the dimension of θ, but do not alter the estimation procedure substantively.

The forward contract variables, QC_{id} and PC_{id}, are set in advance of the day-ahead bidding process. Suppliers sign hedge contracts with large consumers or electricity retailers for a pre-determined pattern of fixed prices (or a single fixed price) throughout the day, week, or month and for a pre-determined quantity each half-hour of the day (or pattern of quantities) throughout the day, week, or month for an entire year or number of years. There is a small amount of short-term activity in the forward contract market for electricity retailers requiring price certainty for a larger or smaller than planned quantity of electricity at some point during the year, but the vast majority of a supplier's contract position is known far in advance of the settlement dates of the contracts. Consequently, from the perspective of formulating its day-ahead expected profit-maximizing bidding strategy, the values of QC_{id} and PC_{id} for all half-hours of the day are known to the supplier at the time it submits its vector of price and quantity bids for the following day. As noted in Wolak (2000) and emphasized in Wolak (2003a, 2007), a supplier's fixed-price forward contract obligations exert an enormous affect on its incentive to take actions to impact prices in the short-term wholesale market. Section 4 discusses this issue with respect to the question of the ability versus the incentive of a firm to exercise unilateral market power in the short-term market.

Recall that a supplier is assumed to set its half-hourly supply curves to maximize expected profits, where this expectation is taken with respect to two sources of uncertainty to the residual demand function the supplier faces each half-hour period of the following trading day. The first is due to the fact that Firm A does not know the precise form of $\mathrm{SO}_{id}(p)$, the aggregate willingness-to-supply curve submitted by all other market participants during load period i of day d, when it submits its bids. The second accounts for the fact that the firm does not know the value of market demand that will set the market price when it submits its bids. As discussed in Wolak (2003a), because I am not solving for equilibrium outcomes in the multi-unit auction, I do not need to be specific about the causes of these two sources of uncertainty in the residual demand function that Firm A faces. Wolak (2003a) notes that the researcher only needs to assume that Firm A knows the joint distribution of the uncertainty in the 48 half-hourly residual demand curves and that it is bidding to maximize expected profits, where the expectation of its realized profit function is taken with respect to this joint distribution.

Let ε_i equal a continuously distributed random vector parameterizing the uncertainty in Firm A's residual demand function in load period i ($i = 1, \ldots, 48$). Re-write Firm A's residual demand function in load period i of day d accounting for this demand shock as $\mathrm{DR}_{id}(p,\varepsilon_i)$. Define $\theta = (p_{11}, \ldots, p_{JK}, q_{1,11}, \ldots, q_{1,JK}, q_{2,11},\ldots,q_{2,JK},\ldots, q_{48,11},\ldots,q_{48,JK})$ as the vector of daily bid prices and quantities submitted by Firm A. For the Australian NEM, there are $K = 10$ increments for each of the J gensets owned by Firm A. As noted earlier, NEM rules require that the price bid for increment k of unit j, p_{jk}, is fixed for each of the $k = 1, \ldots,K$ bid increments for each of the $j = 1, \ldots, J$ gensets owned by Firm A for the entire day. The quantity bid in load period i, for increment k of unit j, q_{ijk} made available to produce electricity in load period i from each of the $k = 1, \ldots, K$ bid increments for the $j = 1, \ldots, J$ gensets owned by Firm A can vary across the $i = 1, \ldots, 48$ load periods throughout the day. In the NEM, the value of K is 10, so the dimension of θ is $10J + 48\mathrm{H}10J$. All wholesale markets require that bid increments are dispatched according to the order of their bid prices, from lowest to highest, which implies that $\mathrm{SA}_{id}(p,\theta)$ is non-decreasing in p.

Deriving the moment conditions necessary to estimate generation unit-level cost functions requires additional notation to represent $SA_{id}(p,\theta)$ in terms of the genset-level bid supply functions. Let

$SA_{ijd}(p,\theta)$ = the amount bid by genset j at price p during load period i of day d,
$SA_{id}(p,\theta) = \sum_{j=1}^{J} SA_{ijd}(p,\theta)$ = total amount supplied by Firm A at price p during load period i of day d.

In terms of this notation, write the realized variable profit for Firm A during day d as:

$$\Pi_d(\theta,\varepsilon) = \sum_{i=1}^{48}[DR_{id}(p_i(\varepsilon_i,\theta))p_i(\varepsilon_i,\theta) - (p_i(\varepsilon_i,\theta) - PC_{id})QC_{id}] - \sum_{j=1}^{J} C_j\left(Q_{jd},\beta_j\right), \tag{3.1}$$

where $\varepsilon = (\varepsilon'_1,\varepsilon'_2,\ldots,\varepsilon'_{48})'$ is the vector of realizations of the ε_i. The function $p_i(\varepsilon_i,\theta)$ is the market-clearing price for load period i for the residual demand shock realization, ε_i, and daily bid vector, θ. It is the solution in p to the equation $DR_{id}(p,\varepsilon_i) = SA_{id}(p,\theta)$. Different mechanisms for setting the prices paid to generation unit owners will result in different price functions that depend on ε_i and θ. For example, a pay-as-bid auction mechanism will give rise to a different price function for each bid quantity segment that is accepted in the auction.

Because $SA_{id}(p,\theta)$ is non-decreasing and $DR_{id}(p,\varepsilon_i)$ is non-increasing in p, the equilibrium price is unique with probability one if the joint distribution function of ε_i is continuous. To economize on notation in the discussion that follows, I abbreviate $p_i(\varepsilon_i,\theta)$ as p_i. The ith element of the vector Q_{jd} is equal to $SA_{ijd}(p_i,\theta)$, the bid supply curve for genset j in half-hour period i of day d evaluated at the market-clearing price in this half-hour period, p_i.

The first term in (3.1) is the daily total revenue received by Firm A for selling the energy it produces in the spot market. The last term is the total daily operating cost to produce the electricity sold. The middle term is the payment made by Firm A if the spot price exceeds the contract price, or received by Firm A if the contract price exceeds the spot price. Forward financial contracts typically settle through these so-called "difference payments" between the buyer and seller of the contract. Under this forward contract settlement scheme, the short-term market operator simply pays for all energy produced and charges for all energy withdrawn from the network at the short-term market price. The market and system operator does not need to know the forward financial contract arrangements between market participants.

Except for the difference payments and the fact that there are 48 half-hour periods in the day, equation (3.1) looks exactly like equation (2.7) for the standard oligopoly model. However, all of the variables in (3.1) except $C_j(Q_{jd},\beta_j)$, the daily variable cost function for generation unit j, are directly observable given the bid data, market outcome data and forward contract data. The residual demand curve, which required a functional form assumption on the market demand function and a quantity-setting Cournot equilibrium model of firm behavior in equation (2.7), can be directly computed from the bids submitted by all market participants for the purposes of computing equation (3.1).

The supplier's expected profit-maximizing bidding strategy maximizes the expected value of $\Pi_d(\theta,\varepsilon)$ with respect to θ, subject to the constraints that all bid quantity increments, q_{ijk}, must be greater than or equal to zero and lesser than the capacity of the unit for all load periods, i, bid increments, k, and gensets, j, and that for each genset the sum of the bid quantity increments during each half-hour period is less than the capacity, CAP_j^{max}, of genset j. All daily price increments must be greater than −\$9999.99/MWh and less than \$10,000/MWh, where all dollar magnitudes are in Australian dollars. All of these constraints can be written as linear combinations of the elements of θ.

In terms of the above notation, the mathematical program characterizing the firm's expected profit-maximizing bidding strategy can be written as:

$$max_\theta E_\varepsilon(\Pi_d(\theta,\varepsilon)) \text{subject to } b_u \geq R\theta \geq b_l \tag{3.2}$$

where $E_\varepsilon(.)$ is the expectation with respect to the joint distribution of ε. The matrix R embodies that the linear inequality constraints on the elements of θ described above. Fig. 1 provides a graphical illustration of the construction of the expected profit-maximizing value of θ. This figure graphs four of the many possible residual demand curve realizations, $DR_a(p), \ldots, DR_d(p)$; the bid curve of Firm A; and the aggregate marginal cost function for Firm A.[1] For each of the four residual demand curves displayed in Fig. 1, the associated market-clearing prices and quantities that result from the intersection of $SA(p,\theta)$ with each of these residual demand curve realizations are shown.

Given Firm A's cost function, the variable profits for each residual demand curve realization can be computed. The solution to (3.2) is the set of values of θ, the vector of bid prices and quantities, that maximizes the expected value of realized profits taken with respect to the joint distribution of ε, the vector of residual demand curve shocks.

The first-order conditions for this optimization problem are:

$$\frac{\partial E_\varepsilon(\Pi_d(\theta,\varepsilon))}{\partial\theta} = R'\lambda - R'\mu, \quad R\theta \geq b_l, b_u \geq R\theta, \tag{3.3}$$

if $(R\theta - b_l)_k > 0$, then $\mu_k = 0$ and $(R\theta - b_u)_k < 0$, then $\lambda_k = 0$, where $(X)_k$ is the kth element of the vector X and μ_k and λ_k are the kth elements of the vectors of Kuhn–Tucker multipliers, μ and λ.

If all of the inequality constraints associated with an element of θ, say p_{jk}, are slack, then the first-order condition reduces to:

$$\frac{\partial E_\varepsilon(\Pi_d(\theta_d,\varepsilon))}{\partial p_{jk}} = 0 \tag{3.4}$$

where θ_d is the value of θ chosen for day d. All of the daily price bids associated with Firm A's gensets over the sample period used in Wolak (2003a, 2007) lie in the interior of the interval of feasible price bids, which implies that all price bids satisfy the first-order conditions given in (3.4) for all days, d, gensets, j, and daily bid increments, k. In Wolak (2003a, 2007), Firm A operates 7 units and each of them has 10 bid increments, which implies 70 daily price moment restrictions. A different number of generation units and bid increments would imply a different number of price moment conditions. These first-order conditions for daily expected profit-maximization with respect to Firm A's choice of the vector of daily price increments can be used to estimate the parameters of the genset-level variable cost functions.

The first-order conditions with respect to the bid quantity increments can also be used to estimate these variable cost functions. There are two conditions that must hold for the quantity bid associated with bid increment k from unit j in load period i, q_{ijk}, to yield a first-order condition of the form:

[1] The assumption of a continuous distribution for ε implies there is an uncountable number of possible residual demand curve realizations.

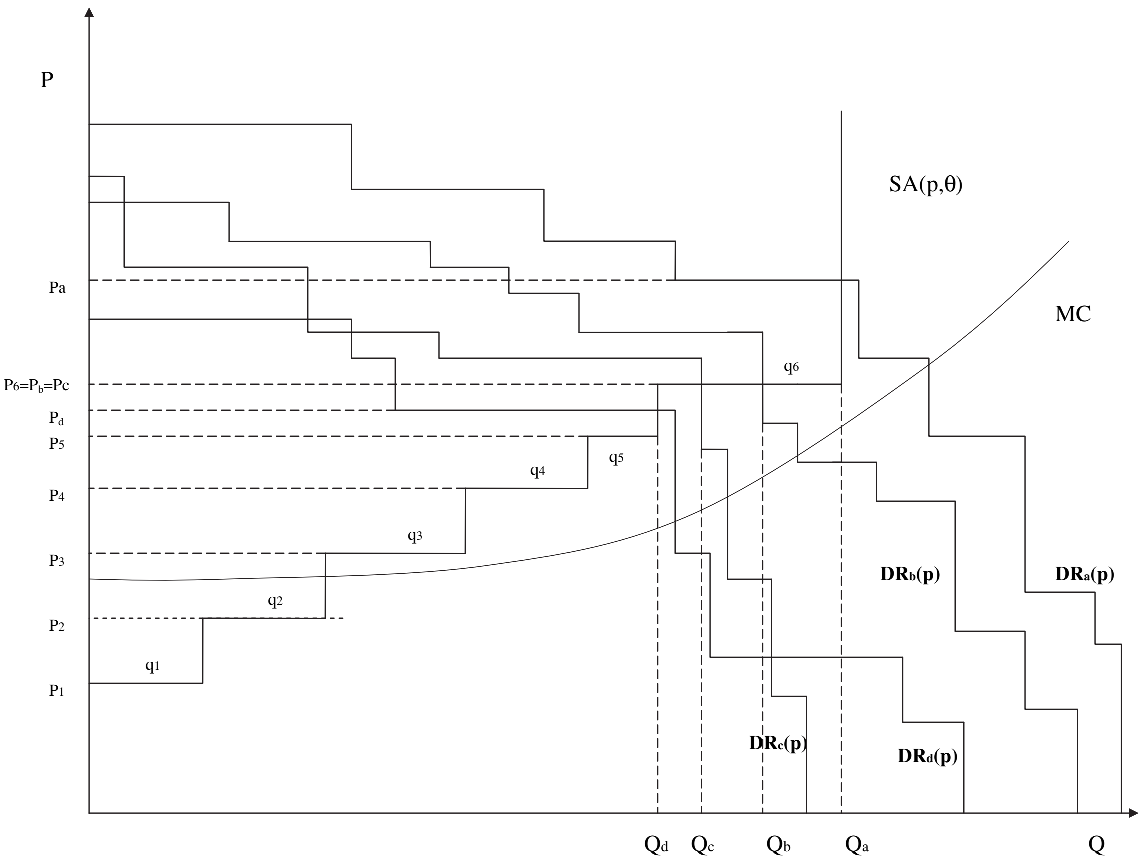


Fig. 1. Determination of expected profit-maximizing bid prices and quantities.

$$\frac{\partial E_\varepsilon(\Pi_d(\theta_d,\varepsilon))}{\partial q_{ijk}} = 0 \tag{3.5}$$

These conditions are: 1) the value of q_{ijk} is strictly greater than zero, and 2) the sum of the q_{ijk} over all bid increments k for unit j in load period i is strictly less than the capacity of the unit.

To implement this approach, define the following two indicator variables:

$$y_{\mathrm{ijk}} = 1 \text{ if } q_{\mathrm{ijk}} > 0 \text{ and zero otherwise} \tag{3.6}$$

$$z_{\mathrm{ij}} = 1 \text{ if } \sum_{k=1}^{10} q_{ijk} < \mathrm{CAP}_{\mathrm{j}}^{\max} \text{ and zero otherwise} \tag{3.7}$$

Because suppliers can and often do produce more than the nameplate capacity of their generation unit, I assume that $\mathrm{CAP}_j^{\max}$ of the generation unit is the sample maximum half-hourly amount of energy produced from that unit over all half-hours during the sample period. This definition of the indicator variables y_{ijk} and z_{ij} implies the following quantity bid increment moment restrictions for all JK quantity bid increments:

$$\frac{1}{48}\sum_{i=1}^{48} y_{ijk} z_{ij} \frac{\partial E_\varepsilon(\Pi_d(\theta_d,\varepsilon))}{\partial q_{ijk}} = 0. \tag{3.8}$$

For cases considered in Wolak (2003a, 2007), there are 70 additional moment restrictions from the first-order conditions for the 70 quantity increments, as long as the population means of the y_{ijk} and z_{ij} are non-zero.

Wolak (2007) shows that there are also moment inequality restrictions implied by expected profit-maximizing bidding behavior and that they form the basis for a specification test of the assumption of expected profit-maximizing bidding behavior. Recall the definitions of y_{ijk} and z_{ij}. If q_{ijk} is such that $y_{ijk} = 0$ and $z_{ij} = 1$, then the following moment inequality restriction holds:

$$\frac{\partial E_\varepsilon(\Pi_d(\theta_d,\varepsilon))}{\partial q_{ijk}} \le 0 \tag{3.9}$$

If q_{ijk} is such that $y_{ijk} = 1$ and $z_{ij} = 0$, then the following moment inequality restriction holds:

$$\frac{\partial E_\varepsilon(\Pi_d(\theta_d,\varepsilon))}{\partial q_{ijk}} \ge 0 \tag{3.10}$$

This implies the following moment inequality restrictions for all 70 quantity bid increments on a daily basis:

$$\frac{1}{48}\sum_{i=1}^{48}\left(y_{ijk} - z_{ij}\right)\frac{\partial E_\varepsilon(\Pi_d(\theta_d,\varepsilon))}{\partial q_{ijk}} \tag{3.11}$$

which should be greater than or equal to zero if the quantity increments are chosen to maximize expected daily profits.

3.2. Implementation of estimation and testing procedures

Deriving a procedure to recover the genset-level daily operating cost function from the first-order conditions for expected profit-maximizing bidding behavior is complicated by the fact that even though $E_\varepsilon(\Pi_d(\theta,\varepsilon))$ is differentiable with respect to θ, the daily realized profit function, $\Pi_d(\theta,\varepsilon)$, is not. The bid functions in the NEM are step functions, which implies that the residual demand curve each supplier faces is non-differentiable, because $\mathrm{DR}_{id}(p) = Q_{id} - \mathrm{SO}_{id}(p)$.

Wolak (2003a, 2007) use a flexible smoothing procedure to construct a differentiable approximation to $\Pi_d(\theta,\varepsilon)$ that is indexed by a smoothing parameter, h. Let $\Pi_d^h(\theta,\varepsilon)$ equal the differentiable version of Firm *A*'s daily variable profit function. When $h = 0$, there is no approximation because $\Pi_d^h(\theta,\varepsilon) = \Pi_d(\theta,\varepsilon)$. Using this smooth, differentiable approximation to $\Pi_d(\theta,\varepsilon)$, the order of integration and differentiation can be switched in the first-order conditions for expected profit-maximizing bidding behavior to produce the equality

$$\frac{\partial E_\varepsilon\left(\Pi_d^h(\theta,\varepsilon)\right)}{\partial\theta}\Big|_{h=0} = E_\varepsilon\left[\frac{\partial\Pi_d^h(\theta,\varepsilon)}{\partial\theta}\right]_{h=0} \tag{3.12}$$

Note that because the distribution of ε is assumed to be continuous, the derivatives on the left-hand side of (3.12) exist.

For the elements of θ which the linear inequality constraints in (3.2) are not binding, the corresponding element of the vector on the right-hand side of (3.12) is equal to zero, so the value of $\frac{\partial\Pi_d^h(\theta,\varepsilon))}{\partial\theta}$ for each day in the sample period can be used to form a sample moment condition. Wolak (2003a, 2007) demonstrates that solving for the cost function parameters (β_j, $j = 1, 2, \ldots, J$) that make these sample moment restrictions as close to zero as possible yields a consistent estimate of these parameters if h tends to zero as the sample size grows arbitrarily large.

This differentiable version of $\Pi_d^h(\theta,\varepsilon)$ is constructed as follows. First, a differentiable residual demand function facing Firm *A* that allows for both sources of residual demand uncertainty is constructed as:

$$DR_{id}^h(p,\varepsilon_i) = Q_{id}(\varepsilon_i) - SO_{id}^h(p,\varepsilon_i) \tag{3.13}$$

where the smoothed aggregate bid supply function of all other market participants besides Firm *A* in load period *i* is equal to

$$SO_i^h(p,\varepsilon_i) = \sum_{n=1}^{N}\sum_{k=1}^{10} qo_{ink}\Phi((p-po_{nk})/h). \tag{3.14}$$

qo_{ink} is the *k*th bid increment of genset *n* in load period *i* and po_{nk} is the bid price for increment *k* of genset *n*, where *N* is the total number of gensets in the market excluding those owned by Firm *A*. Although the qo_{ink} and po_{nk} change on a daily basis, I suppress the subscript *d* from both of these variables to reduce notational clutter. $\Phi(t)$ is the standard normal cumulative distribution function and h is the smoothing parameter. This parameterization of $SO_i^h(p,\varepsilon_i)$ smooths the corners on the step-function bid curves of all other market participants besides Firm *A* to create a supply function that is differentiable in *p*.

The degree of smoothing depends on the value of *h*. Smaller values of h introduce less smoothing. For $h = 0$, the procedure reproduces the original step-function residual demand curve as is required for $\Pi_d^h(\theta,\varepsilon)_{h=0} = \Pi_d(\theta,\varepsilon)$ to hold. This same procedure is used to make $SA_{ij}(p,\theta)$ differentiable with respect to both the market price, *p*, and θ.

In terms of the smooth and realized profit function, the sample moment restriction for bid price increment *t* of genset *s*, is:

$$\begin{aligned}\frac{\partial\Pi_d^h(\theta_d,\varepsilon)}{\partial p_{st}} = \sum_{i=1}^{48}\Bigg[&\left(\partial DR_i^h(p_i(\varepsilon_i,\theta),\varepsilon_i)/\partial p\right)p_i(\varepsilon_i,\theta)\\ &+\left(DR_i^h(p_i(\varepsilon_i,\theta),\varepsilon_i)-QC_i\right)-\sum_{j=1}^{J}\frac{\partial C_j(Q_j,\beta_j)}{\partial Q_{ij}}\left(\frac{\partial SA_{ij}^h}{\partial p_i}\right)\Bigg]\frac{\partial p_i^h}{\partial p_{st}}\\ &-\sum_{j=1}^{J}\frac{\partial C_j(Q_j,\beta_j)}{\partial Q_{ij}}\frac{\partial SA_{ij}^h}{\partial p_{st}}\end{aligned} \tag{3.15}$$

where p_i is shorthand for the market-clearing price in load period *i*. The terms in $(\partial p_i^h)/(\partial p_{st})$ and $\frac{\partial SA_{ij}^h}{\partial p_{st}}$ capture the details of how elements of θ, the bid parameters submitted by Firm *A*, impact market prices. Wolak (2003a, 2007) derive expressions for these terms for a uniform-price wholesale market.

By the assumption of expected profit-maximizing choice of the price bid increments:

$$lim_{h\_>0}E_\varepsilon\left[\frac{\partial\Pi_d^h(\theta_d,\varepsilon)}{\partial p_{st}}\right] = 0. \tag{3.16}$$

For the cases in Wolak (2003a, 2007), let $\ell_d(\beta,h)$ denote the 70-dimensional vector of partial derivatives – ten bid price increments for seven gensets – given in (3.15), where β is the vector composed of β_j for $j = 1, \ldots, J$. Assuming that the functional form for $C_j(Q_j,\beta_j)$ is correct for all gensets, the first-order conditions for expected profit-maximization with respect to the 70 daily bid prices imply that $E(\ell_d(\beta^0,h))_{h=0} = 0$, where β^0 is the true value of β. Consequently, solving for the value of *b* that minimizes:

$$\left[\frac{1}{D}\sum_{d=1}^{D}\ell_d(b,h)\right]'\left[\frac{1}{D}\sum_{d=1}^{D}\ell_d(b,h)\right] \tag{3.17}$$

will yield a consistent estimate of β as both *D* tends to infinity and *h* tends to zero, where *D* is the total number of days in the sample period. Wolak (2001) discusses the technical conditions necessary for the consistency and asymptotic normality of this smoothed estimation problem for the case of multi-unit auctions with step-function bids.

Let $b(I)$ denote this consistent estimate of β, where "*I*" denotes the fact that the identity matrix is used as the generalized method of moments (GMM) weighting matrix. Wolak (2001) describes how to construct $V_D(b(I),h)$, a consistent estimate of the optimal GMM weighting matrix using the $b(I)$ that accounts for potential heteroscedasticity in the covariance matrices of the $\ell_d(\beta^0,h)$ across observations. The optimal GMM estimator finds the value of *b* that minimizes

$$\left[\frac{1}{D}\sum_{d=1}^{D}\ell_d(b,h)\right]'V_D\left(b(I),h\right)^{-1}\left[\frac{1}{D}\sum_{d=1}^{D}\ell_d(b,h)\right], \tag{3.18}$$

as *h* tends to zero. Let $b(O)$ denote this estimator, where "*O*" denotes the fact this estimator is based on a consistent estimate of the optimal weighting matrix.

Wolak (2007) discusses how to incorporate the quantity bid moment restrictions to estimate the elements of β. For the case discussed in Wolak (2003a, 2007), this provides 70 additional moment restrictions that can be used to estimate the parameters of the genset-level variable cost function and test the assumption of expected profit-maximizing bidding behavior. Define $\mathcal{L}_d(\beta,h)$ as the 140-dimensional vector of daily price and quantity moments. Expected profit-maximizing behavior implies that $E(\mathcal{L}_d(\beta^0,h))_{h=0} = 0$. Repeating the logic given in equations (3.17) and (3.18), the optimal GMM estimator of the parameters of the cost functions that incorporates both the price and quantity moment conditions can be derived.

A test of the hypothesis of expected profit-maximizing behavior is equivalent to a test of the over-identifying price and quantity moment restrictions. In particular, if the dimension of β, the vector of parameters of the cost function, is smaller than the number of moment restrictions used to estimate it, then Hansen's (1982) test for the validity of the over-identifying moment restrictions is a test of the null hypothesis of expected profit-maximizing behavior. Wolak (2003a) implements this test and fails to reject the null hypothesis for single period variable cost functions. Wolak (2007) implements this test based on the price moment restrictions only and the combination of the price and quantity moment restrictions for daily variable cost functions and fails to reject the null hypothesis in both instances.

An alternative approach to examining the validity of the moment restrictions implied by expected profit-maximizing behavior is to compute values for the marginal cost of each genset using engineering data on the efficiency of the generation unit and input prices. The efficiency with which heat energy is converted to electrical energy – the heat rate – is typically reported in terms of British Thermal Units (BTU) of heat energy consumed per kilowatt-hour (KWh) of electricity produced. Using the price of the input fuel in $/BTU it is then possible to compute a $/KWh variable cost figure. The regulated or government-owned history of the electricity supply industry in all countries implies there is typically accurate information available on the average heat rates of generation units and other variable operating costs. These can be used to compute marginal cost estimates for each generation unit owned by Firm *A*. Inserting these marginal cost estimates in place of $\frac{\partial C_j}{\partial Q_{ij}}$ in smoothed first-order conditions for expected profit-maximizing bid price and quantity choice yields sample moment restrictions that should have population values equal to zero when evaluated at $h = 0$. This is equivalent to assuming that the true value of β is known so that the test statistic for expected profit-maximizing behavior becomes

$$\left[\frac{1}{D}\sum_{d=1}^{D}\ell_d\left(\beta^0,h\right)\right]' V_D\left(\beta^0,h\right)^{-1}\left[\frac{1}{D}\sum_{d=1}^{D}\ell_d\left(\beta^0,h\right)\right],$$

which is asymptotically distributed as a chi-squared random variable with degrees of freedom equal to the rank of $V_D(\beta^0,h)$ under the null hypothesis of expected profit-maximizing behavior.

To examine the empirical validity of the inequality constraints moment restrictions implied by the expected profit-maximizing choice of the 70 quantity bids in Wolak (2007), compute the following magnitude for each genset *s* and bid increment *t*:

$$MI^h_{dst}(\theta_d,\varepsilon) = \frac{1}{48}\sum_{r=1}^{48}(y_{rst}-z_{rs})\frac{\partial\Pi^h_d(\theta_d,\varepsilon)}{\partial q_{rst}}. \tag{3.19}$$

Define $\Pi_d(\beta,h)$ as 70-dimensional vector composed of the values of $MI^h_{dst}(\theta_d,\varepsilon)$ for genset *s* ($s = 1, \ldots, 7$) and bid increment *t* ($t = 1, \ldots, 10$). The expected profit-maximizing choice of the quantity bids implies that $E(\Pi_d(\beta^0,h))_{h=0} \geq 0$. Evaluating this vector at a consistent estimate of the parameters of the cost functions allows the construction of a test of the null hypothesis that the population value of each of these moment restrictions is greater than or equal to zero. Wolak (2007) implements this test for a supplier in the NEM and finds no statistically significant evidence against the null hypothesis of the validity of these population moment inequalities.

4. Measuring unilateral market power in wholesale electricity markets

This section demonstrates how direct measures of the ability and incentive of suppliers to exercise unilateral market power can be computed using bid and market outcome data from wholesale electricity markets. Because the residual demand curve faced by a firm in a bid-based wholesale electricity market can be computed from the bids and offers submitted by all market participants besides the one under consideration, it is possible to compute an estimate of the inverse of the elasticity of the residual demand curve faced by a supplier without any assumptions on the functional form of the aggregate demand curve or the assumed model of equilibrium outcomes, the two assumptions needed to compute the inverse of the elasticity of the residual demand curve given in equation (2.11).

The only complication with directly computing $-1/e_i(p)$, the inverse of the elasticity of the residual demand curve faced by firm *i* evaluated at price *p*, results from the fact that residual demand curves are non-increasing step functions. This implies that the researcher must choose a value for the percentage change in price from the market-clearing price or the percentage change in quantity from the market-clearing quantity in order to compute $-1/e_i(p)$. Assumptions about the magnitude of the price or quantity change only alter the interpretation given to $-1/e_i(p)$.

Exercising market power requires withholding or increasing output to exploit a downward sloping residual demand curve. Therefore, to compute $-1/e_i(p)$ the researcher should use a quantity change that it is reasonable to believe a supplier would find profitable to make. Quantity changes no larger than 10 percent in the upward or downward direction are reasonable based on this criterion. If the researcher chose a 10 percent quantity reduction relative to the firm's actual output, then the inverse of the elasticity of the residual demand curve for Firm *A* during period *i* of day *d* at price p_{id} is equal to:

$$-1/e_{id}(p) = 0.10(p^*_{id}-p_{id})/p_{id}, \tag{4.1}$$

where p^*_{id} is the lowest price such that residual demand is less than 90% of its value evaluated at the market-clearing price. Mathematically, this is the lowest price, p, such that $DR_{id}(p) \leq DR_{id}(p_{id})\,0.90$. Fig. 2 illustrates the computation of this inverse elasticity.

Bushnell and Wolak (2005) compute hourly values of $-1/e_{id}(p)$ for the five largest suppliers in the Spanish electricity market – Endessa, Iberdrola, Union Fenosa, Viesgo, and Hidroelectrica Del Cantabrico – over the period April 12, 2001 to December 31, 2004 using bids submitted by all market participants. Table 1 presents the 2003 capacity ownership shares for these suppliers. These capacity shares were relatively stable over the sample period, with more than 70 percent of the capacity owned by Endessa and Iberdrola for all years.

Table 2 computes the annual average of the hourly values of the inverse of the elasticity of the residual demand curve for each of the five suppliers. Table 2 also lists the annual standard deviation of these hourly values. Several conclusions emerge from this table. First, the annual average values of the inverse elasticities are significantly higher for Endessa and Iberdrola than for the other three suppliers, which imply that these two firms have a significantly higher ability to exercise unilateral market power than the other firms. Because these two firms have substantial fixed-price

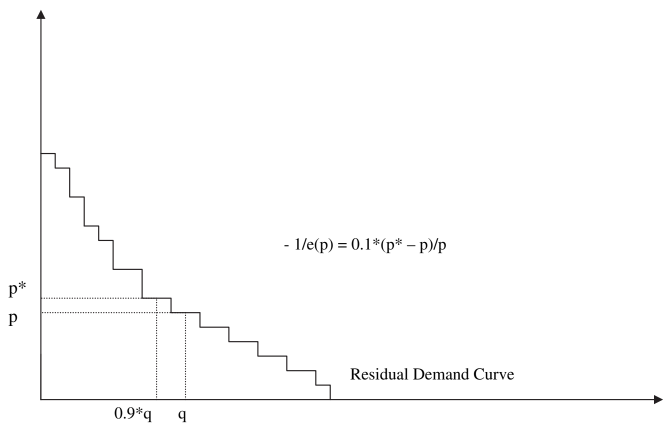


Fig. 2. Inverse of residual demand elasticity at market price ($-1/e(p)$).

Table 1
Capacity ownership shares in Spanish market for 2003.

Firm	Coal	Gas	Hydro	Nuclear	Oil	Total	Share
Endessa	5521	1406	5293	3511	1760	17490	38%
Hidro Cantabrico	1523	387	542	156		2607	6%
Iberdrola	1189	2370	8676	3213	1741	17190	37%
Union Fenosa	1946	0	1778	702	747	5173	11%
Other	864	1607	664		718	3853	8%
Total	11043	5770	16954	7581	4966	46314	

retail load and forward contract obligations relative to the amount of electricity they produce during this time period, these two firms have little incentive to exploit their ability to raise market prices.

A second conclusion is the fact that Iberdrola appears to have experienced a significant decline in its ability to exercise unilateral market power from 2001 to 2004. In particular, the average value of the inverse elasticity in 2001 was 2.19, whereas the average value in 2004 was 1.18, an almost 50 percent reduction in its ability impact the market price through its unilateral actions. On the other hand, Endessa experienced virtually no decline in its ability to influence the market price from 2001 to 2004. The average value of its inverse elasticity in 2001 was 1.98 and the average value in 2004 was 1.92. The remaining three firms all experienced little, if any, reduction in their ability to exercise unilateral market power from 2001 to 2004.

Figs. 3 and 4 plot the annual hourly means of the inverse elasticities for Endessa and Iberdrola for each of the 24 h of the day. These figures yield a third conclusion: The reductions in the ability of Iberdrola to exercise unilateral market power from 2001 to 2004 are constant across hours of the day and for Endessa the annual hourly values are roughly constant across the four years.

It is important to note that values of $-1/e_{id}(p)$ greater than 1 are not inconsistent with expected profit-maximizing behavior. First, this magnitude only measures the ability of the supplier to raise or lower market prices by withholding or increasing output, respectively, not the supplier's incentive to do so. These two concepts differ because the fixed-price forward market obligations of a supplier, either in the form of fixed-price forward contracts or fixed-price retail load obligations, significantly reduce the incentive to exercise unilateral market power. Second, even if a supplier has no fixed-price forward contract obligations, values of $-1/e_{id}(p)$ greater than 1 are consistent with expected profit-maximizing bidding behavior because all residual demand curve realizations that a supplier faces are step functions and firms choose the parameters of their willingness-to-supply curve to maximize their expected profits, where this expectation is taken with respect to the distribution of these step-function residual demand realizations. Further, as shown in Wolak (2003a, 2007), the assumption of expected profit-maximizing bidding behavior does not imply that the equation $[p_{id} - C'(\mathrm{DR}_{id}(p_{id}))]/p_{id} = -1/e_{id}(p_{id})$ holds for any residual demand curve realization, $\mathrm{DR}_{id}(p_{id})$, and associated market-clearing price, p_{id}. This result implies that the Hortacsu and Puller (2008) rejection of expected profit-maximizing bidding behavior by some suppliers in the Electricity Reliability Council of Texas (ERCOT) wholesale electricity market is not because these firms fail to choose their willingness-to-supply curves to maximize

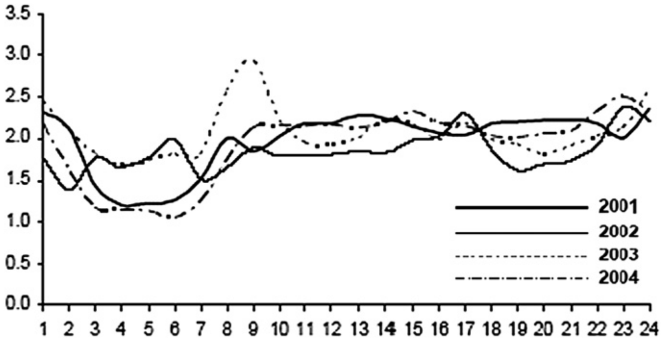


Fig. 3. Annual average hourly values of $-1/e(p)$ for Endessa.

expected profits. Hortacsu and Puller (2008) test and reject a much stronger restriction that is not implied by expected profit-maximizing bidding behavior: that the equation $[p_{id} - C'(\mathrm{DR}_{id}(p_{id}))]/p_{id} = -1/e_{id}(p_{id})$ holds for all residual demand curve realizations and all price and quantity pairs along the firm's willingness-to-supply function. Consequently, the correct conclusion to draw from the Hortascu and Puller empirical results is that their simplified model of expected profit-maximizing bidding behavior is inconsistent with the ERCOT market rules, not that suppliers do not maximize expected profits.

As discussed in Wolak (2000, 2003a, 2007), a supplier's incentive to exercise unilateral market power can differ dramatically from its ability to exercise unilateral market power due to fixed-price forward market obligations. For example, a supplier that has sold 500 MWh of energy to be delivered during a pre-specified hour in the future for \$35/MWh has an incentive to minimize its cost of supplying that quantity of energy because the revenues it earns from producing 500 MWh during this hour are fixed through the difference payment mechanism associated with the supplier's fixed-price forward contract obligations. If the short-term price, p, is above \$35/MWh, then the supplier must make the payment $(p - 35) \times 500$ to the buyer of the forward contract, which implies that any revenues on the 500 MWh of sales at a short-term price above the forward contract price are made in difference payments to the buyer of the contract. Conversely, if the short-term price is below the forward contract price, then the buyer of the contract is obligated to make the payment $(35 - p) \times 500$ to the supplier, so that any revenues below the forward contract price for the 500 MWh of sales at a short-term price below the forward contract price are made up for by difference payments by the buyer of the contract.

This logic implies that a supplier with a significant ability to exercise unilateral market power that has a fixed-price forward market obligation of 500 MWh and only expects to sell 400 MWh of energy in the short-term market will use this ability to exercise unilateral market power to reduce the short-term price and thereby reduce the cost of purchasing the 100 MWh difference between its fixed-price forward market obligation and its sales in the short-term market. If this supplier expects to sell 500 MWh in the short-term market, it has no incentive to exercise unilateral market

Table 2
Annual mean and standard deviation (SD) of hourly values of $-1/e(p)$.

	Endessa		Iberdrola		Union Fenosa		Hidro D.C.		Viesgo	
Year	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2001	1.975	1.305	2.191	2.794	0.520	0.518	0.310	0.362	0.228	0.277
2002	1.842	3.224	2.408	4.419	0.530	2.648	0.355	2.501	0.261	2.479
2003	2.079	2.987	1.859	3.091	0.612	2.236	0.391	1.937	0.231	1.583
2004	1.918	1.271	1.186	0.892	0.465	0.469	0.316	0.379	0.197	0.230

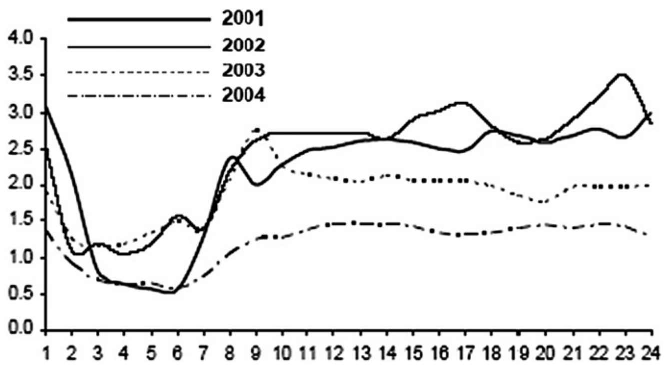


Fig. 4. Annual average hourly values of $-1/e(p)$ for Iberdrola.

power, regardless of its ability to do so, because its short-term market sales equal its fixed-price forward market obligation. Only a supplier who expects to sell more than its fixed-price forward market obligation in the short-term market has an incentive to use its ability to exercise unilateral market power to raise the short-term price. Wolak (2000) provides a detailed discussion of the impact of a supplier's fixed-price forward market obligations on its incentive to exercise unilateral market power.

Consistent with the above logic, Wolak (2000) introduces the following measure of the incentive of a supplier to exercise unilateral market power. Define the residual demand curve net of the supplier's fixed-price forward market obligations as $DR^C_{id}(p) = (DR_{id}(p) - QC_{id})$. Note that depending on the value of the short-term price, $DR^C_{id}(p)$ can be either positive or negative because the supplier's fixed-price forward market obligations, QC_{id}, do not depend on the short-term market price. The inverse elasticity of the residual demand curve net of the Firm A's fixed-price forward market obligations measures the incentive it has to withhold output in order to raise the short-term market price (a large positive value) and increase output to in order to lower the short-term market price (a large, in absolute value, negative value).

In terms of the above notation, the net-of-forward-contracts inverse elasticity of the residual demand curve for a ten-percent quantity change is defined as

$$-1/e^C_{id}(p) = 0.10\left(p^{C*}_{id} - p_{id}\right)\Big/p_{id}, \tag{4.2}$$

where p^{C*}_{id} is the lowest price such that $DR^C_{id}(p)$, the net-of-forward-contract residual demand of Firm A, is less than 90 percent of its value at the market-clearing price. Mathematically, p^{C*}_{id} is the lowest price p such that $DR^C_{id}(p) \le 0.9 \times DR^C_{id}(p_{id})$. If QC_{id} is a substantial fraction of $\mathrm{DR}_{id}(p_{id})$ then a 10 percent reduction in $DR^C_{id}(p)$ gives rise to a very small percentage price increase, implying a very small value for $-1/e^C_{id}(p)$.

These measures of the ability and incentive to exercise unilateral market power are related by the following equation:

$$-1/e^C_{id}(p) = -1/e_{id}(p)[(DR_{id}(p_{id}) - QC_{id})/DR_{id}(p_{id})]. \tag{4.3}$$

The numerator of $(DR_{id}(p_{id}) - QC_{id})/DR_{id}(p_{id})$ is the firm's exposure to spot market prices and the denominator is the firm's total production and sales in the short-term market during load period i of day d. In most wholesale electricity markets, the ratio of the firm's exposure to spot market prices divided by its total production during the hour is typically less than 0.30 and can often be less than 0.20. This implies that $-1/e_{id}(p)$ can be 5 times the value of $-1/e^C_{id}(p)$, implying that the incentive to withhold output in order to raise the market price can be significantly less than the ability to withhold output in order to raise prices, because the supplier has substantial fixed-price forward contract obligations.

In most markets, a supplier's fixed-price forward market obligations are treated as confidential, so it is only possible to compute $-1/e_{id}(p)$ given bid and market outcome data. Fortunately for researchers, but unfortunately for California consumers, the levels of fixed-price forward contract obligations of the five largest suppliers in California were known for the period April 1998 to December 2000. As discussed in detail in Wolak (2003c), the three largest investor-owned load-serving entities in California – Southern California Edison, San Diego Gas and Electric, and Pacific Gas and Electric – together signed virtually no fixed-price forward contracts with the five largest fossil-fuel suppliers in California. Consequently, $-1/e_{id}(p)$ provides an accurate measure of both the ability and incentive of these five suppliers to exercise unilateral market power in the California market during this time period.

Wolak (2003b) computes hourly values of the inverse of the elasticity of the residual demand curve of the five largest suppliers in the California real-time energy market during the four-month period from June 1 to September 30 for 1998, 1999 and 2000. This calculation is done only for hours during this four-month period in which there is no transmission congestion that could shrink the size of the market in which any supplier might compete in and therefore increase $-1/e_{id}(p)$ substantially. Because the incidence of congestion in 2000 was substantially higher than in 1998 or 1999 this sample selection procedure biases against a finding of higher values of $-1/e_{id}(p)$ for the four months of 2000. This analysis also excludes hours when market prices are below \$20/MWh, because this price is estimated to be less than the lowest variable cost of the units owned by these five suppliers. Because input fuel costs are significantly higher in 2000 relative to 1998 and 1999, this assumption also biases against a finding of substantially higher values of $-1/e_{id}(p)$ in 2000 relative to 1998 and 1999.

Wolak (2003b) reports the average hourly value of the inverse of the firm-level residual demand elasticity over the period June 1 to September 30 of each year as a summary measure of the extent of unilateral market power possessed by each supplier. Table 3 reproduces the results of this analysis. Despite the two assumptions that bias against a finding of higher values of $-1/e_{id}(p)$ in 2000 relative to 1998 and 1999, for each firm the average value of this measure of unilateral market power is an order of magnitude higher in 2000 than the average values in 1998 and 1999. Including hours with congestion in the analysis and excluding only hours with market prices below the lowest variable cost unit owned by these five firms in each year makes the differences between the firm-level average values of $-1/e_{id}(p)$ for 2000 relative to 1998 and 1999 substantially larger.

These firm-level results are consistent with the view that the enormous increase in prices in the California market beginning in June of 2000 documented in Borenstein et al. (2002) was due to a substantial increase in the ability of the five large suppliers in California to exercise unilateral market power during the summer of 2000. Wolak (2003c) documents the factors that led to this increased ability to exercise unilateral market power. The many investigations of the causes of the California Electricity Crisis have failed to uncover evidence that these five suppliers coordinated their actions to raise prices in the California market. The changes in these firm-level measures of market power suggest that coordinated actions by suppliers were unnecessary to bring about the substantial price increases that occurred during the period June 1, 2000 to September 30, 2000. The results in Table 3 are consistent with the enormous price increases in the summer of 2000 relative to the summers of 1998 and 1999 being the result of the unilateral expected profit-maximizing response of each of the five suppliers to the bidding behavior of all other market participants in the California market.

McRae and Wolak (2009) compute measures of the unilateral ability and incentive to exercise unilateral market power for the four

Table 3
Average hourly value of $-1/e_i(p)$ by year for June 1 to September 30 for hours with no congestion and prices above \$20/MWh.

Firm	1998 Mean (standard error)	1999 Mean (standard error)	2000 Mean (standard error)
1	0.045 (0.007)	0.035 (0.008)	0.164 (0.029)
2	0.039 (0.007)	0.028 (0.007)	0.164 (0.029)
3	0.054 (0.009)	0.032 (0.008)	0.095 (0.028)
4	0.065 (0.013)	0.032 (0.007)	0.189 (0.032)
5	0.055 (0.009)	0.035 (0.008)	0.161 (0.029)

largest suppliers in the New Zealand wholesale electricity market. Because a hydroelectric dominated system like New Zealand's can sometimes have extremely low or even negative prices caused by sustained periods with large water inflows, a large value of the inverse elasticity of a supplier's residual demand curve can occur not because a supplier can increase the market price by changing its output level, but simply because the market price is extremely small and even a small \$/MWh price increase is a large percentage of a small market price. For this reason, McRae and Wolak define the inverse semi-elasticity of Firm *A*'s residual demand curve as $\eta_{id} = -(1/100) \times (\mathrm{DR}_{id}(p_{id})/\mathrm{DR}'_{id}(p_{id}))$, which gives the \$/MWh price increase associated with a one percent reduction in Firm *A*'s output, and the inverse semi-elasticity of Firm *A*'s net-of-forward contract residual demand curve as $\eta^C_{id} = -(1/100) \times (DR^c_{id}(p_{id})/DR^{c\prime}_{id}(p_{id}))$. Similar to the case of the two inverse elasticities, these two measures are related by the equation: $\eta^C_{id} = \eta^*_{id} \times [(DR^c_{id}(p_{id}) - QC_{id})/DR_{id}(p_{id})]$.

McRae and Wolak find that half-hourly market prices are highly positively correlated with the value of the average of the four firm-level half-hourly values of these measures of the ability and incentive to exercise unilateral market power over the period January 2001 through May 2008. Moreover, McRae and Wolak (2009) find a strong positive predictive relationship between η_{id}, the half-hourly value of the firm-level inverse semi-elasticity of the residual demand curve, and the half-hourly offer price (the highest accepted price step on the supplier's half-hourly willingness-to-supply curve) submitted by that firm during that same half-hour period. McRae and Wolak find an even stronger positive predictive relationship between η^C_{id}, the half-hourly value of the firm-level inverse semi-elasticity of the net-of-fixed-price-forward-market-obligations residual demand curve, and that supplier's half-hourly offer price. These results are consistent with the logic that during half-hours when a supplier has a greater ability or incentive to exercise unilateral market power by raising the market-clearing price, that supplier submits a higher offer price. Because the half-hourly values of η^C_{id} for some suppliers are often negative, the second set of empirical results imply that lower values of the supplier's offer price are associated with larger (in absolute value) negative values of η^C_{id}. Running firm-level regressions with separate coefficients for negative and positive values of η^C_{id} yields precisely estimated positive coefficient estimates for $\max(0,\eta^C_{id})$ and $\min(0, \eta^C_{id})$ that are not statistically different from each other.

5. Measuring market performance in wholesale electricity markets

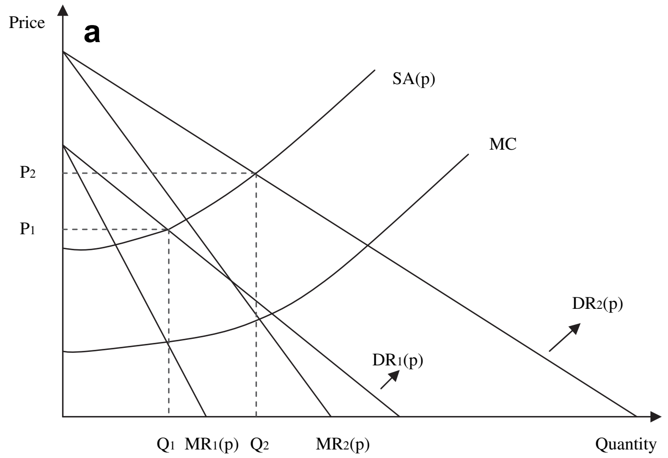
This section describes a procedure for measuring market inefficiencies in wholesale electricity markets that does not rely on parametric functional form assumptions for demand or variable cost functions or an assumed equilibrium model of competition. I describe an application of this procedure to the California electricity market over the period June 1998 to December 2000 and an application comparing the performance of two wholesale markets in the eastern US – New England and PJM – to the performance of the California market in order to assess the impact of the level of fixed-price forward contract obligations held by large suppliers on wholesale market outcomes.

To clarify the differences between this approach and usual approaches used to assess market performance, consider the Cournot example in Section 2. To compute the overall cost of market inefficiencies caused by the market power exercised by the *N* firms, the researcher would need to compute a counterfactual no-market-power or competitive benchmark price using the estimated market demand function, $D(p|W,\beta,\varepsilon)$, and variable cost functions, $C_i(q|Z,\beta,\eta)$ $i = 1, \ldots, N$. Computing the intersection of this aggregate demand curve with the aggregate marginal cost curve for the industry yields the perfectly competitive price-taking market equilibrium. Note that this counterfactual no-market-power equilibrium cannot be computed without assumptions on the form of the market demand function and the industry-wide marginal cost curve. If all the researcher has is market-clearing price and quantity data and information on *W* and *Z*, the demand and supply shifters, then the demand function cannot be estimated from this information without a functional form assumption. Likewise, the firm-level variable cost curves cannot be estimated without functional form assumptions and an assumed equilibrium model of competition, in this case quantity-setting Cournot behavior. I will now show how the availability of bid and market outcome data and information on the technical characteristics of generation facilities can be used to compute this counterfactual no-market-power price without any of these assumptions.

To determine the competitive benchmark equilibrium price, I must first determine a supplier's expected profit-maximizing bid if it had no ability to influence the market price through its unilateral actions. Fig. 5a and b provide a graphical answer to this question. For simplicity, I assume that there are only two possible residual demand curve realizations and that these residual demand curves are continuous. This makes it possible to compute the supplier's optimal bid curve graphically rather than by maximizing equation (3.1), but does not compromise the generality of the result that I would like to demonstrate.

Fig. 5a shows that for a pair of downward sloping residual demand curves, Firm *A*'s expected profit-maximizing bid curve, SA(p), is any curve that passes through the ex post profit-maximizing price and quantity pairs for the two residual demand curve realizations. Specifically, if the residual demand curve realization $\mathrm{DR}_1(p)$ occurs, the price and quantity pair (P_1,Q_1) shown in Fig. 5a is the ex post profit-maximizing price and output level for Firm *A* given this residual demand curve realization. For residual demand curve realization $\mathrm{DR}_2(p)$, the price and quantity pair (P_2,Q_2) is the ex post profit-maximizing price and output level. Consequently, any bid curve that passes through (P_1,Q_1) and (P_2,Q_2) is an expected profit-maximizing bid curve.

Fig. 5b considers the case in which Firm *A* is unable to raise market price through its own unilateral actions for the case of two residual demand curve realizations. The inability of Firm *A* to impact market prices unilaterally is reflected in the fact that both residual demand curves are infinitely price elastic. Following precisely the same logic as given above for the case of Fig. 5a, the ex post profit-maximizing price and output pairs for these two residual demand curve realizations are (P_1,Q_1) and (P_2,Q_2), which implies that the expected profit-maximizing bid curve passes through these two points. For simplicity, I have drawn it as being equal to Firm *A*'s marginal cost curve, because (P_1,Q_1) and (P_2,Q_2) are the points where the marginal cost curve crosses the relevant residual demand curve realization. This logic implies that if Firm *A* has no ability to influence the market price for all possible residual demand curve realizations, then it will find it unilaterally expected



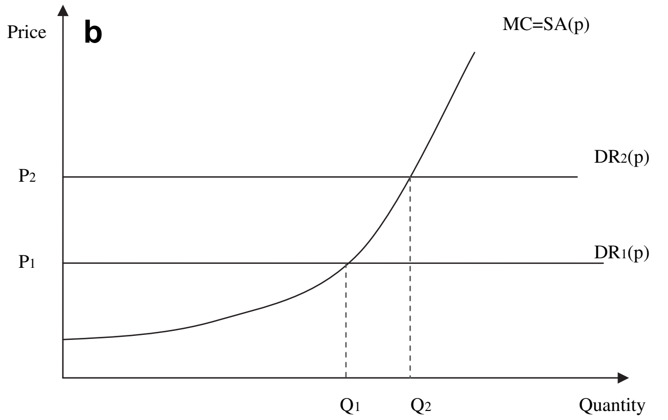


Fig. 5. (a) Expected profit-maximizing bidding with ability to influence market price. (b) Expected profit-maximizing bidding with no ability to influence market price.

profit-maximizing to submit an SA(p) equal to its marginal cost curve. Applying this logic to all firms implies that the counterfactual no-market-power aggregate bid curve is the industry marginal cost curve.

As discussed in Wolak (2003a), step-function residual demand curves and the fact that θ, the vector of bid parameters, is finite-dimensional, considerably complicates the process of solving for Firm A's expected profit-maximizing bid curve. However, the basic insight of Fig. 5b remains: a supplier facing a distribution of residual demand curve realizations that are close to perfectly price elastic over the relevant range of output would submit a bid curve as close as possible to its marginal cost curve.

Because the aggregate demand is observed and is completely price inelastic for the reasons discussed in Section 3, the only remaining complication associated with computing the no-market-power market equilibrium is the construction of the industry marginal cost curve. I discuss the details of this procedure for the case of California because a number of complications that arise there are also relevant to other wholesale electricity markets.

5.1. Computing the no-market-power aggregate bid curve

Information collected during the former regulated regime and information that is still collected to monitor system reliability provides valuable input to the construction of the no-market-power aggregate bid curve. For all US markets and virtually all other markets around the world, there is detailed information on generation unit-level heat rates, the rate at which a MWh of electricity can be produced using British Thermal Units (BTUs) of heat energy. In all US electricity markets, there are short-term markets for most of the input fuels that can be used to compute an estimated variable fuel cost for each generation unit given its heat rate. The second component of the estimated marginal cost of producing electricity is the variable operating and maintenance (O&M) cost. Operating and maintenance costs are primarily the labor costs associated with operating and maintaining the generation unit in working order. During the former regulated regime, detailed information was collected on these costs, so there is fairly accurate information on this component of variable costs. Fortunately, these variable costs are also a small fraction of variable fuel costs, and a declining one at that, as fossil-fuel prices continue to increase, so that errors in measuring these variable costs are not likely to exert a significant impact on the resulting competitive benchmark prices. The final component of the variable cost of producing electricity in California during this time period is due to the existence of a nitrogen oxides (NO_X) emissions permit market in Southern California. Consequently, suppliers faced an opportunity cost of producing electricity in the form of the emissions permits that they needed to own in order to produce electricity. Similar to the generation unit's heat rate, all generation units located in Southern California have NO_X emissions rates – the pounds of NO_X emissions emitted per MWh of energy produced.

Putting all of these facts together, the marginal cost of producing electricity in day d for each fossil-fuel generation facility j is equal to:

$$MC(j,d) = \text{O\&M}(j) + \text{Fuel\_Price}(j,d) \times \text{Heat\_Rate}(j) + \text{NO}_\text{X}\text{\_Price}(d) \times \text{NO}_\text{X}\text{\_Rate}(j) \tag{5.1}$$

where O&M(j) is the variable operating and maintenance cost for unit j, Fuel_Price(j,d) is the price of the input fuel for unit j on day d, Heat_Rate(j) is the average heat rate of unit j, NOx_Price(d) is the price of NO_X emission permits on day d, NOx_Rate(j) is the rate at which NO_X emissions are produced per MWh for unit j. For all units located outside of the South Coast Air Quality Management District (SCAQMD) market, the emission permit component of equation (5.1) is not relevant. Borenstein et al. (2002) [hereafter, BBW] discuss the sources of this information for all generation units in California. Kolstad and Wolak (2003) demonstrate how the existence of the SCAQMD market facilitated the actions of suppliers in Southern California to raise wholesale electricity prices during the period June 2000 to March 2001. Fig. 6, taken from BBW, plots the instate fossil-fuel generation marginal cost curve constructed using this algorithm for September of 1998, 1999, and 2000.

There are two other types of suppliers to the California market that complicate the construction of the no-market power aggregate supply curve. The first is importers and the second is hydroelectric suppliers. California depends on imports from neighboring states for 20–25 percent of its annual electricity consumption. Because these importers submit bids to supply energy at various points of interconnection to the California market and because electricity flows according to the laws of physics rather than where a seller would like its energy to flow, it is impossible to identify what generation unit located outside of California is supplying imported energy. BBW address this complication by assuming that importers are unable to exercise unilateral market power, so that the bids they submit in fact reflect their true marginal cost of supplying energy to California during that hour. This assumption biases upward the value of the resulting competitive benchmark price, which biases downward the estimated cost of unilateral market power.

O&M costs are the only incurred variable cost of supplying hydroelectric or geothermal energy. These O&M costs are significantly lower than those for fossil-fuel generation units, so these units are always the cheapest variable cost units. However, because some

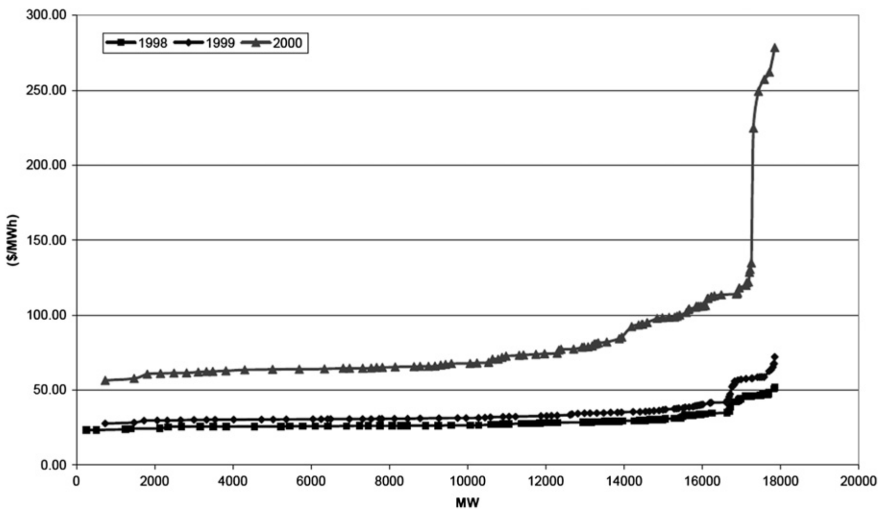


Fig. 6. Instate fossil-fuel generation marginal cost curve for September (by year).

hydroelectric and geothermal energy can be stored, there is an opportunity cost to using this energy because it could be sold at another price in a future period. BBW avoid the very complex problem of determining the hourly counterfactual no-market-power opportunity cost of reservoir resources (hydro and geothermal) by assuming that the amount of reservoir resource energy produced each hour during their sample period would not change in the no-market-power solution. BBW demonstrate that this assumption biases the average competitive benchmark price upwards, which biases the estimated cost of market power downward. The appendix of BBW provides a detailed analysis of the claim that their assumption leads to a downward bias in the estimated cost of market power.

Fig. 7, taken from BBW, puts all of these components of the aggregate no-market-power supply curve together in a manner that makes explaining the computation of the counterfactual competitive benchmark price more straightforward. Let q_{tot} equal the total demand in California. The variable q_{mt} is the generation capacity that is under a regulatory restriction to bid into the market at price of zero, q_{rsv} is supply of reservoir resources (hydroelectric and geothermal) that are assumed to supply their actual production

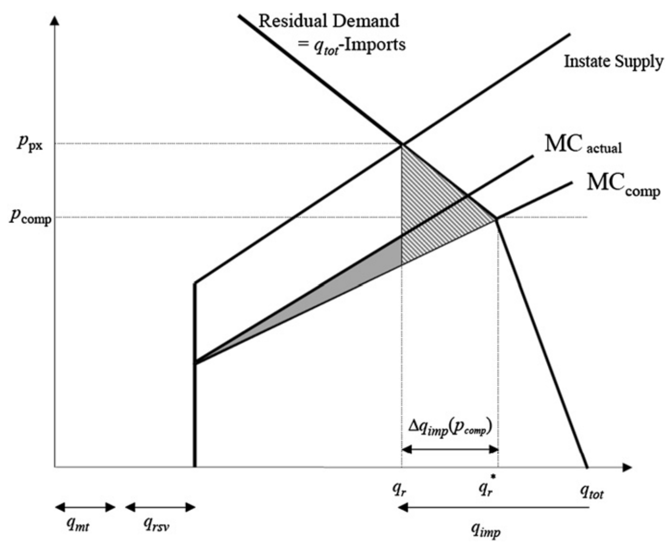


Fig. 7. Computing no-market-power competitive benchmark price.

in the counterfactual pricing process. The import supply curve is plotted from right to left so that where it intersects the actual aggregate bid curve of all instate resources yields the actual market-clearing price, what is labelled p_{px} in Fig. 7. The distance from q_{tot} to q_r, the amount supplied by instate generation units is equal to q_{imp}, the actual imports into California.

Because the identity and amount of energy supplied from each generation unit in California is known, it is possible to construct the actual instate marginal cost curve using equation (5.1). The step of this aggregate marginal cost curve corresponding to generation unit j has height equal to $MC(j,d)$ and quantity equal to amount that unit j actually produced during hour h of day d.

The computation of MC_{comp} and the associated competitive benchmark price, p_{comp} in Fig. 7, involves one additional complication that BBW argue should increase the realism of the competitive benchmark price. Generation units can fail when they are called upon to operate or can simply be unavailable because of planned outages. One of the reasons MC_{actual} may be higher or lower across days, even if input fuel prices do not change across days, is because of such planned and unplanned generation outages. The North American Electricity Reliability Council (NERC) collects detailed information on generation unit outages on a historical basis by generation unit characteristics as part of its mandate to monitor the reliability of the US electricity supply.

For each of the instate fossil-fuel generation units, BBW compute a forced outage factor based on a representative sample of similar generation units. For each hour during the sample period, a draw is taken from the joint distribution of these outages (assuming that the probability of an outage is independent across generation units) and only those generation units that are not forced out according to this draw are included in the realization of MC_{comp}. This realization of MC_{comp} is intersected with the import supply curve and q_r^*, the no-market-power output of the instate suppliers, and p_{comp} given in the figure are determined. For each hour during the sample period 100 draws of MC_{comp} are taken and q_r^* and p_{comp} are computed for each draw. BBW define the no-market-power price for hour h as the mean of these 100 values of p_{comp} for that hour. The authors note that failure to account for this outage risk or derating of capacity in the aggregate bid curve by the unit's outage rate, as several authors have done, biases downward the competitive benchmark price, which in turn biases upward the estimated cost of the exercise unilateral market power.

Fig. 8, taken from BBW, decomposes the total cost of meeting q_{tot} into a number of components, most notably total payments under competitive benchmark outcomes, actual production costs, competitive benchmark production costs, competitive rents, and oligopoly rents (the difference between actual price and the no-market-power price times q_{tot} minus q_{mt}. Table 4, also taken from BBW, shows an enormous increase in total payments between June to October of 1998 and 1999 and that same period in 2000, with the change in the oligopoly rents explaining the vast majority of that difference. In particular, oligopoly rents were $425 million in 1998, $382 million in 1999, but $4.448 billion in 2000. BBW provide a detailed diagnosis of the factors that led to this enormous increase in oligopoly rents between 1998 and 1999 and 2000. The results of Wolak (2003b), discussed in the previous section, provide an important part of this explanation. These results demonstrate that the unilateral ability and incentive of the five large California fossil-fuel suppliers to exercise unilateral market power by withholding output or bidding a higher price for the same amount of output increased substantially between 1998 and 1999 versus 2000.

Bushnell et al. (2008) [hereafter, BMS] use this general procedure to compare the performance of the California, PJM (major parts of Pennsylvania, New Jersey, Maryland, Delaware, Virginia and the District of Columbia), and New England electricity markets in order to assess the relationship between market outcomes and the extent of vertical arrangements, either through fixed-price forward contract obligations or fixed-price retail load obligations, on the performance of wholesale electricity markets. The authors compute no-market-power counterfactual prices for the three markets from June 1999 to September 1999 to compare actual market performance to the no-market-power counterfactual for each of the three markets.

In order to assess the impact of vertical arrangements on wholesale market outcomes, BMS also construct a non-cooperative equilibrium Cournot model for each market that involves the large suppliers competing against a fringe of price-taking small suppliers, including importers. As an upper bound on actual market outcomes, they compute this Cournot equilibrium assuming no vertical arrangements between any of the large suppliers and final demand in each of the three markets.

Table 4
Yearly production costs and rent distribution in millions of dollars (June to October).

	1998	1999	2000
Total actual payments	1672	2041	8977
Total competitive payments	1247	1659	4529
Production costs – actual	759	1006	2774
Production costs – competitive	715	950	2428
Competitive rents	532	708	2101
Oligopoly rents	425	382	4448
Oligopoly inefficiency – instate	31	31	126
Oligopoly inefficiency – imports	13	24	221

BMS find that the no-forward-contracts Cournot equilibrium prices are slightly lower but come close to replicating actual market outcomes in the California market over their sample period. However, for the PJM and New England markets the no-forward-contracts Cournot equilibrium prices are substantially higher than actual market outcomes, meaning that actual market outcomes are much closer to the no-market-power outcomes than would be predicted by the no-forward-contracts Cournot equilibrium. To assess the impact of vertical arrangements, the authors use publicly available information on the fixed-price retail load obligations of the major suppliers in each of the markets and compute the Cournot equilibrium outcomes over their sample period incorporating these forward contract obligations into each firm's objective function.

Fixed-price retail load obligations are defined as the amount of energy a supplier's retailing affiliate sold to final customers at a retail price that does not vary with the hourly wholesale price. The major suppliers in New England and PJM had substantial fixed-price retail load obligations, but as noted above, the five major suppliers in California did not. BMS find that once these retail load obligations are included in the objective functions of the large suppliers, Cournot equilibrium prices in New England and PJM fall substantially and are much closer to actual market outcomes.

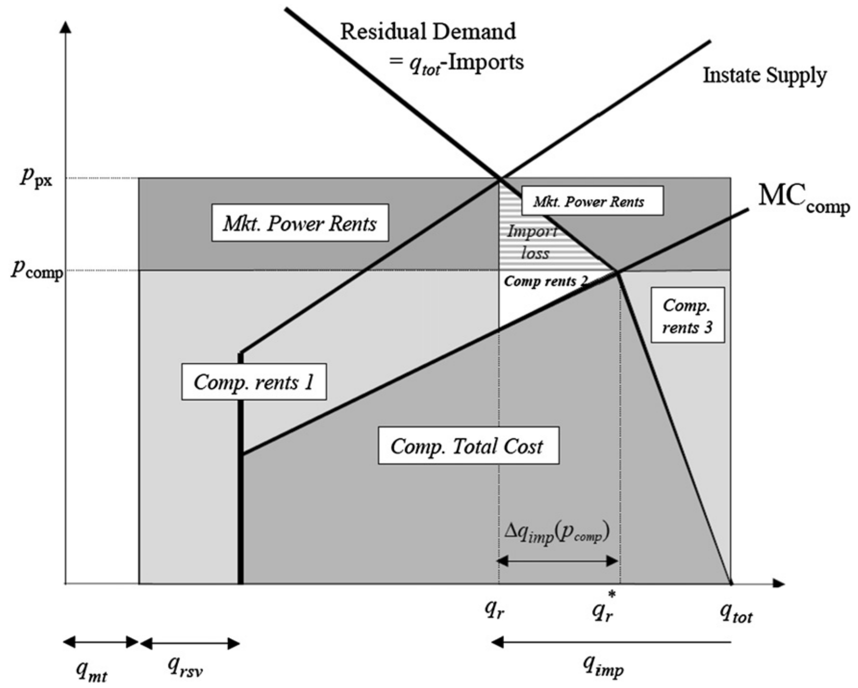


Fig. 8. Calculation of division of rents.

These results reinforce the analysis in Wolak (2000) that fixed-price forward contracts limit the incentive of suppliers to exercise unilateral market power in short-term wholesale electricity markets. BMS make the very important point that horizontal market structure – specifically, concentration of generation capacity ownership – does help to explain differences in market performance, but only after accounting for these vertical arrangements. The authors conclude that their results support the hypothesis that fixed-price long-term forward contracts and other vertical arrangements are a major source of the differences in the performance of wholesale electricity markets.

There have been a number of applications of this approach to computing the no-market-power prices and assessing the costs of the exercise of unilateral market power. Joskow and Kahn (2002) consider the case of the California electricity market during the summer of 2000. They also devote considerable effort in their analysis to quantifying the magnitude and frequency that specific generation units were withheld from the market in order to raise market prices. Mansur (2007) applies the competitive benchmark pricing procedure to the PJM Interconnection during its first summer of operation to assess the impact of vertical integration between generation and retailing on the behavior of specific market participants. Bushnell and Wolak (2005) apply this procedure to the Spanish electricity market for the years 2000–2004 to assess changes in competitiveness of market outcomes in Spain over that time period.

6. Merger analysis in wholesale electricity markets

This section presents an approach to merger analysis that makes use of the bid, market outcome, and variable cost data to compute counterfactual post-merger prices and to assess the impact of potential divestiture scenarios on post-merger market outcomes. I present an application to a hypothetical merger where data on forward contract positions for market participants were made available on a confidential basis. The results of this analysis emphasize the crucial role that post-merger forward market positions play in drawing conclusions about the adverse competitive effects of a proposed merger.

To contrast this approach with the usual approach to merger analysis, recall the Cournot example of Section 2. To assess the impact of firms j and k merging, the profit function of the merged firm becomes:

$$\pi_i\left(q_j, q_k\right) = P\left(Q_{-(j+k)} + q_j + q_k | W, \beta, \varepsilon\right)\left(q_j + q_k\right) - C_j\left(q_j | Z, \beta, \eta\right) - C_k(q_k | Z, \beta, \eta). \quad (6.1)$$

The profit functions of the remaining firms are the same as the one given in equation (2.8). The first-order conditions for firms j and firm k in the pre-merger Cournot equilibrium solution are replaced by the partial derivatives of (6.1) with respect to q_j and q_k, but the first-order conditions for the remaining firms are unchanged from the pre-merger model solution. Solving these N-equations in the N unknown quantities yields the post-merger Cournot quantities for all firms, including the new firm. Substituting the post-merger value of $Q = \sum_{i=1}^{N} q_i$ into the inverse demand function yields the post-merger price. Therefore, if the analyst only has market-clearing prices and quantities pre-merger for a time series of observations for the industry or for a cross-section of similar industries, the researcher must assume parametric functional forms for the inverse demand and variable cost functions and an equilibrium model of competition to obtain estimates of the inverse demand function and variable cost functions necessary to compute the post-merger price and quantities.

For the case of wholesale electricity markets, it is possible to assess the competitive effects of mergers using the bids of all other suppliers besides the ones merging to compute the residual demand curve faced by the merged entity. In addition, because virtually all wholesale electricity markets require bids to be submitted at the generation unit level, it is possible to evaluate potential divestiture packages by allocating the bids of the divested units to the residual demand curve faced by the merged entity. For this same reason, it is also possible to perform merger analyses for different size geographic markets based on the location of the generation units, by excluding bids of specific generation units from the residual demand curve faced by the merged entity. Choosing a smaller geographic area (than the entire market) over which to compute the residual demand curve the merged entity faces may make more sense if transmission network constraints may often prevent suppliers outside of the region from competing with the merged entity.

As the discussion in Section 5 emphasizes, the levels of fixed-price forward contract obligations of the merging suppliers are crucial ingredients for any merger analysis, because the incentive a supplier has to take actions to raise market price depends on its level of fixed-price forward contract obligations. Another important concern in any merger analysis is whether the parties will retain their pre-merger forward contract levels after the merger takes place. Consequently, any remedy for a proposed merger must also address the incentives the merged entity has to take on less fixed-price forward contract obligations after the merger in order to exploit its greater ability to raise wholesale prices post merger.

Often wholesale market operators have access to QC_{gid}, the total quantity of energy Firm A has sold in fixed-price forward contract g during hour i of day d because of arrangements these market operators have to settle forward contract commitments between market participants. However, PC_{gid}, the price at which forward contract g makes deliveries during hour i of day d, is rarely available to the wholesale market operator. Suppose that information on the values of QC_{gid} have been collected for the merging parties for two years before a proposed merger, although information on PC_{gid} is not available. For the reasons discussed below, this fact does not adversely impact our ability to undertake a merger analysis.

Suppose that Firms A and B are the merging parties. Let $\mathrm{DR}^G(p)$ equal Firm G's ($G = A,B$) residual demand curve evaluated at price p. The following decomposition of these residual demand curves is useful:

$$\mathrm{DR}^A(p) = Q^{\mathrm{tot}} - \mathrm{SO}^{-(A+B)}(p) - S^B(p) \quad (6.2a)$$

$$\mathrm{DR}^B(p) = Q^{\mathrm{tot}} - \mathrm{SO}^{-(A+B)}(p) - S^A(p), \quad (6.2b)$$

where Q^{tot} is the market demand, $\mathrm{SO}^{-(A+B)}(p)$ is the supply of all firms besides Firms A and B at price p, and $S^G(p)$ is supply of firm G ($G = A,B$) at price p. The residual demand curve of the merged entity (M) is

$$\mathrm{DR}^M(p) = Q^{\mathrm{tot}} - \mathrm{SO}^{-(A+B)}(p). \quad (6.3)$$

It is important to note that an implicit assumption in this analysis is that the aggregate willingness-to-supply curve for all other firms besides the merging parties, $\mathrm{SO}^{-(A+B)}(p)$, remains unchanged after the merger. Comparing (6.2a) and (6.2b) to (6.3) reveals that the merged entity must face a residual demand curve with a smaller price response than either of the merging parties individually. This result and the fact that the combined output of the merged entity at the pre-merger price is more than the output of either party implies that the elasticity of the residual demand curve faced by merged entity at the actual market-clearing price must be smaller in absolute value than the pre-merger elasticity of the residual demand curve at the actual market-clearing price of either merging party.

In terms of this notation, the realized variable profit for Firm *A* during hour *i* at price p_i is:

$$\pi_i^A(p_i) = DR_i^A(p_i)p_i - C^A\left(DR_i^A(p_i)\right) - \left(p_i - PC_i^A\right)QC_i^A, \tag{6.4}$$

where a superscript "*A*" now denotes a value for Firm *A* to distinguish these variables from those for the other merging party, Firm *B*. The realized variable profit for Firm *B* during hour i at price p_i is:

$$\pi_i^B(p_i) = DR_i^B(p_i)p_i - C^B\left(DR_i^B(p_i)\right) - \left(p_i - PC_i^B\right)QC_i^B. \tag{6.5}$$

By the logic discussed in Section 5 for computing the expected profit-maximizing bid curves given in Fig. 5a and b, if the market rules do not restrict the ability of market participants to extract all available variable profits, the following results should hold:

$$p_i \approx p_i^{A,pre} = argmax_p \pi_i^A(p) \text{ and } p_i \approx p_i^{B,pre} = argmax_p \pi_i^B(p) \tag{6.6}$$

implying that the observed market-clearing price is close to the value of p that maximizes the ex post variable profits of Firms *A* and *B*. Consequently, the first step in the merger analysis is to determine the extent to which this result holds. Although *PC* is not available, this does not affect the values of p that maximize (6.4) and (6.5), because *PC* only appears in each firm's profit function multiplied by *QC*. It is straightforward to maximize (6.4) and (6.5) with respect to *p* using a univariate optimization routine that does not require derivatives, such as the golden search routine described in Press et al. (1992).

To compute the counterfactual price for the merged firm, construct the variable profit function for hour *i* at price *p* for the merged firm as:

$$\pi_i^M(p) = DR_i^M(p)p - C^M\left(DR_i^M(p)\right) - \left(p - PC_i^M\right)QC_i^M, \tag{6.7}$$

where $DR_i^M(p)$ is the residual demand curve faced by the merged entity. $C^M(q)$ is the variable cost of the merged entity, which can be written as:

$$C^M(q) = \int_0^q MC^A(s)ds + \int_0^q MC^B(s)ds, \tag{6.8}$$

where $MC^J(q)$ is the marginal cost curve of Firm *J*, $J = A,B$. The marginal cost curve of each of the merged firms is computed from the generation unit-level marginal costs computed in Section 5. Each step of the marginal cost curve has length equal to the capacity of the generation unit and height equal to the $MC(j,d)$ computed in equation (5.1). The remaining variable in (6.7) is defined as follows:

$$QC_i^M = QC_i^A + QC_i^B.$$

Recall that the value of PC_i^M is not necessary to compute the value of p that maximizes (6.7). The counterfactual merged entity profit-maximizing market-clearing price is defined as p_i^M.

Fig. 9 graphs the average hourly actual price and the average hourly values of $p_i^{A,pre}$ and $p_i^{B,pre}$, the prices that maximize (6.4) and (6.5) for Firms *A* and *B* for all 24 h of the day for all days during the sample period. All of these prices are normalized by the sample mean of the actual price. The hourly value of *QC* in (6.4) and (6.5) equals the best estimate of that supplier's forward contract position (including its fixed-price retail load obligations). This figure also graphs the average hourly post-merger price that maximizes (6.7) and a post-divestiture (of certain generation units owned by both parties) average hourly price that maximizes the ex post profits of the merged entity. Call this post-merger and post-divestiture price $p_i^{M,D}$. Comparing the average daily pattern of actual prices to the average daily pattern of $p_i^{A,pre}$ and $p_i^{B,pre}$ suggests that the model does a respectable job of finding the pre-merger actual prices. Wolak (2000) performed similar analysis for a supplier in the Australian NEM and found that the half-hourly ex post profit-maximizing price obtained from solving (6.4) yielded prices very close to actual prices. Comparing the average daily pattern of actual prices and $p_i^{A,pre}$ and $p_i^{B,pre}$ to the average daily pattern of the post-merger prices, p_i^M, shows limited average predicted price increases associated with the merger at the current level of fixed-price forward contract obligations of the two merged firms.

Fig. 10 graphs the average hourly actual price and the average hourly values of $p_i^{A,pre}$ and $p_i^{B,pre}$, for the case that the hourly values of *QC* for each firm are such that their sum is at the low end of the level of forward contract obligations that one might expect the merged entity to hold after the merger. These prices are also normalized by the sample average of actual prices. As expected, the average pattern of daily actual prices is significantly lower than the average pattern of $p_i^{A,pre}$ and $p_i^{B,pre}$ because of the greater incentive to exercise unilateral market power by both Firms *A* and *B* pre-

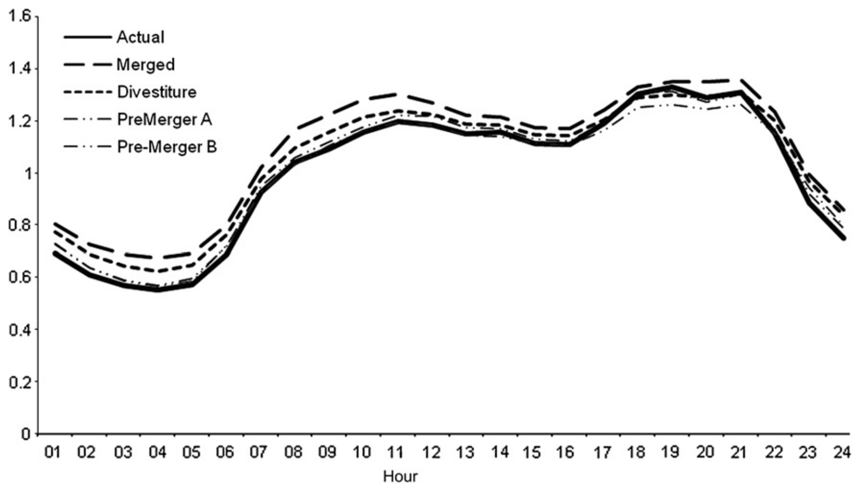


Fig. 9. Average prices by hour normalized by sample mean of actual price (actual forward contract levels).

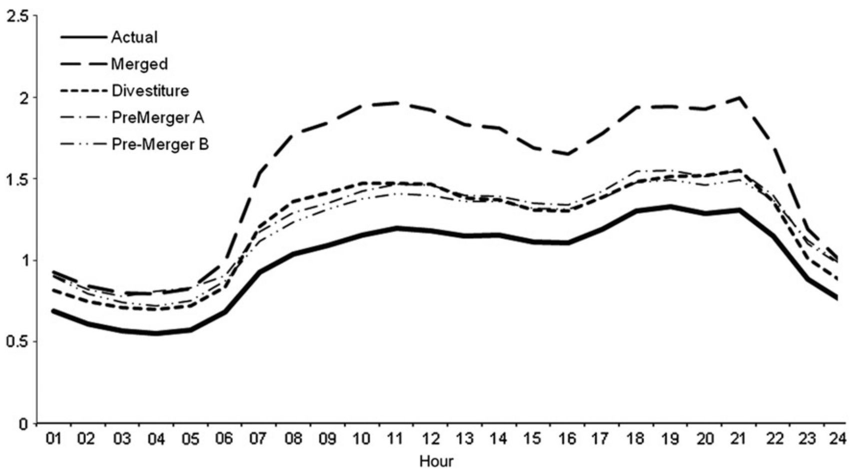


Fig. 10. Average prices by hour normalized by sample mean of actual price (minimal forward contract levels).

merger due to a lower level of fixed-price forward contract obligations. The merger results in much higher prices relative to actual prices and even relative to $p_i^{A,pre}$ and $p_i^{B,pre}$. The proposed divestiture to remedy the merger essentially returns the average pattern of daily prices to the pre-merger levels for both parties. This result suggests that the divestiture scenario would limit the incentive of the merged entity to exercise unilateral market power beyond what either firm could do pre-merger, even if the merged entity decided to significantly reduce its forward contract obligations before the merger.

This hypothetical analysis demonstrates the usefulness of this approach to merger analysis for wholesale electricity markets. The availability of generation unit-level bid data allows the construction of the residual demand curve faced by any set of firms over any geographic market. In addition, information on the hourly values of QC for each of the merged parties can be used to assess the sensitivity of any assessment of the competitive effects of a merger to the post-merger forward contract levels of the merging parties. Finally, it is important to emphasize that the analyst is not required to use actual bids to construct residual demand curves. The unit's marginal cost or other credible offer curves could be used instead.

7. Lessons for market monitoring and design in oligopoly industries

This section draws some lessons from wholesale electricity for market monitoring and design in other oligopoly industries. The richer data sets and precisely specified market rules allow a number of conclusions about potential determinants of firm behavior and market outcomes in oligopoly industries. There are also lessons for antitrust and other market monitoring entities.

The first lesson is that there is little evidence against the hypothesis of expected profit-maximizing behavior in wholesale electricity markets. This result suggests that evidence against profit-maximizing behavior in other oligopoly industries may be the result of incorrect specifications for the demand or variable cost functions or an invalid model of competition among firms, rather than any valid evidence against (expected) profit-maximizing behavior. For the case of wholesale electricity markets, the results of Hortacsu and Puller (2008) discussed in Section 3 are instructive on this point. Their apparent rejection of the assumption of expected profit-maximizing bidding behavior by some suppliers in the ERCOT market is the result of the assumptions of their simplified model of expected profit-maximizing bidding behavior being inconsistent with the ERCOT market rules and not that the suppliers do not choose the parameters of their willingness-to-supply curves in a manner consistent with the ERCOT market rules to maximize expected profits.

A second lesson is the importance of fixed-price forward market commitments in determining the incentive firms have to exercise unilateral market power in short-term markets. Substantial forward contract obligations or fixed-price retail load obligations by suppliers to the major load-serving entities significantly limits the incentive to raise these prices. Assessment of the extent of unilateral market power suppliers possess can be badly distorted by failing to account for the fixed-price forward market commitments of the suppliers. For example, a supplier that owns a substantial amount of productive capacity may in fact have very little incentive to exercise unilateral market power because of its forward market commitments. Particularly, in industries with high fixed costs and substantial delays between the decision to produce and ability to produce due to siting and constructing the necessary productive capacity, forward markets that clear far enough in advance to allow new entrants to compete are significantly more competitive than shorter-term forward markets. This result further emphasizes the need to develop active forward markets in these types of industries to discipline the ability of suppliers to exercise unilateral market power.

A third lesson is the fact that prices substantially in excess of no-market-power levels can occur without coordinated actions to raise prices among the major suppliers. The combination of capacity constraints in production, inelastic final demand, the non-storability of the final product and the fact that it must be delivered through a specialized network all combine to enhance the ability of suppliers to exercise unilateral market power, thereby imposing substantial economic harm on consumers.

A fourth lesson concerns the definition of significant unilateral market power. One feature of wholesale electricity markets is that the ability to exercise harmful levels of unilateral market power requires certain preconditions, and even though these preconditions can occur for a significant fraction of the year, this can take the form of a number of episodes of very short duration. For example, a supplier may be able to exercise substantial unilateral market power or a merger may enable the merged entity to exercise substantial unilateral market power during 4 h of the day on days where certain system conditions occur.

The typical view of market power analysis is that there is an antitrust concern if the price increase is sustained and significant. Consequently, it may be necessary to revise the meaning of the word sustained to mean a significant fraction of the year, rather than a single sustained period of time, because it is rarely the case that unilateral market power in wholesale electricity markets is exercised during all or even a substantial fraction of the hours of the day for a number of consecutive days. Instead, market power is exercised for a significant fraction of the hours of the year when certain system conditions arise. Similar logic applies to other industries that face large demand fluctuations throughout the year, capacity constraints in production and non-storability of the product they sell.

The final lesson concerns the important role future forward market positions play in determining appropriate remedies for proposed mergers. Although the existing level of forward commitments may limit the incentive of the merged parties to exercise unilateral market power, it is important to put incentives in place for the merged party to engage in sufficient fixed-price forward contracting in the future. As the California Electricity Crisis convincingly demonstrated, the exercise of unilateral market power is not a long-term problem. However, as the experience of California from June 2000 to June 2001 also demonstrated, enormous amounts of economic harm can occur from the exercise of substantial unilateral market power for short period of time. Load-serving entities will sign fixed-price forward contracts to commit suppliers to the short-term market and to encourage new entry. However, the difficulty described in detail in Wolak (2003c) is that final demand will be forced to buy out the future market power that market participants expect to be exercised in the short-term market by these suppliers because the current lower levels of fixed-price forward contract obligations carryover into the future. This implies significantly higher prices for the forward contracts that will be passed on to final consumers, as Wolak (2003c) shows occurred in California.

8. Concluding comments

This paper has demonstrated that many determinants of oligopoly market outcomes can be studied in wholesale electricity markets under weaker assumptions than in other oligopoly industries because of their rich data, regulatory history and clearly specified market rules. This paper has contrasted these methods with those used in existing studies of oligopoly industries where the best data available are market-clearing prices and quantities and demand and cost shifters. The extent to which the methods used in electricity supply industries generalize these conventional methods is explained in detail and major applications of these techniques are summarized.

Although it is difficult to argue that restructuring of electricity supply industries to create bid-based electricity markets has benefitted US consumers, the research summarized in this paper suggests there is reason for optimism that the promised benefits will eventually come. The very rich data available to researchers from these markets and the precisely specified market rules make it possible to undertake detailed analyses of the incentives created by specific market rules and the costs of these incentives in terms of adverse market outcomes.

References

Berry, Steven T., Levinsohn, James L., Pakes, Ariel, 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Borenstein, Severin, Bushnell, James, Wolak, Frank A., 2002. Measuring inefficiencies in California's re-structured electricity market. American Economic Review (December), 1376–1405.

Bresnahan, Timothy, 1981. Departures from marginal-cost pricing in the American automobile industry. Journal of Econometrics 17, 201–227.

Bresnahan, Timothy, 1987. Competition and collusion in the American automobile market: the 1955 price war. Journal of Industrial Economics 45 (4, June), 457–482 (special issue).

Bresnahan, Timothy, Stern, Scott, Trajtenberg, Manuel, 1997. Market segmentation and the sources of rents from innovation: personal computers. RAND Journal of Economics 28, S17–S44.

Bushnell, James, Wolak, Frank A., November 2005. Report on the Current and Future State of Competition in the Spanish Electricity Supply Industry.

Bushnell, James, Mansur, Erin, Saravia, Celeste, 2008. Vertical arrangements, market structure, and competition: an analysis of restructured U.S. electricity markets. American Economic Review (March), 237–266.

Fisher, Franklin M., 1989. Games economists play: a noncooperative view. RAND Journal of Economics 20 (1, Spring), 113–124.

Fisher, Franklin M., 1991. Organizing industrial organization: reflections on the handbook of industry organization. Brookings Papers on Economic Activity – Microeconomics, 201–240.

Goldberg, Penny K., 1995. Product differentiation & oligopoly in international markets: the case of the U.S. automobile industry. Econometrica 63 (4), 891–952.

Hansen, Lars P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hortacsu, Ali, Puller, Steven, 2008. Understanding strategic bidding in multi-unit auctions: a case study of the Texas spot market. The RAND Journal of Economics (Spring), 86–114.

Joskow, P.L., Kahn, E., December 2002. A quantitative analysis of pricing behavior in California's wholesale electricity market during summer 2000: the final word. The Energy Journal 23 (3), 1–25.

Klemperer, Paul D., Meyer, Margaret A., 1989. Supply function equilibria in oligopoly under uncertainty. Econometrica 57 (6), 1243–1277.

Kolstad, Jonathan, Wolak, Frank, 2003. Using Environmental Emissions Permit Prices to Raise Electricity Prices: Evidence from the California Electricity Market. http://www.stanford.edu/~wolak Available from.

Mansur, Erin T., 2007. Upstream competition and vertical integration in electricity markets. Journal of Law and Economics 50, 125–156.

McRae, Shaun D., Wolak, Frank A., 2009. How Do Firms Exercise Unilateral Market Power? Evidence from a Bid-Based Wholesale Electricity Market. http://www.stanford.edu/~wolak Available at.

Nevo, Aviv, 2000. Mergers with differentiated products: the case of the ready-to-eat breakfast cereal industry. RAND Journal of Economics 31 (3), 395–421.

Nevo, Aviv, 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica 69 (2), 307–342.

Porter, Robert H., 1983. A study of cartel stability: the joint executive committee, 1880–1886. Bell Journal of Economics 14 (2, Autumn), 301–314.

Press, William H., Teukolsky, Saul A., Vetterling, William T., Flannery, Brian P., 1992. Numerical Recipes in Fortran 7: The Art of Scientific Computing. In: Fortran Numerical Recipes, second ed.., vol. 1. Cambridge University Press.

Rosse, James N., 1970. Estimating cost function parameters without using cost data: illustrated methodology. Econometrica 38 (2), 256–275.

Werden, Gregory, Froeb, Luke, 1994. The effects of mergers in differentiated products industries: structural merger policy and the logit model. Journal of Law, Economics, & Organization 10, 407–426.

Werden, Gregory, Froeb, Luke, 2002. The antitrust logit model for predicting unilateral competitive effects. Antitrust Law Journal 70 (1).

Wolak, Frank A., 2000. An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market. International Economic Journal 14 (2), 1–40.

Wolak, Frank A., February 2001. Identification, estimation and inference for cost functions in multi-unit auction markets. http://www.stanford.edu/~wolak Available from.

Wolak, Frank A., 2003a. Identification and estimation of cost functions using observed bid data: an application to electricity markets. In: Dewatripont, M., Hansen, L.P., Turnovsky, S.J. (Eds.), Advances in Economic and Econometrics: Theory and Applications, Eight World Congress, vol. II. Cambridge University Press, New York, pp. 133–169.

Wolak, Frank A., May 2003b. Measuring unilateral market power in wholesale electricity markets: the California market 1998 to 2000. American Economic Review, 425–430.

Wolak, Frank A., August/September 2003c. Diagnosing the California electricity crisis. The Electricity Journal, 11–37.

Wolak, Frank A., 2007. Quantifying the supply - side benefits from forward contracting in wholesale electricity markets. Journal of Applied Econometrics 22, 1179–1209.