ECE 4800 Project Proposal: Estimation of Electric Power Generator's Cost Function Using Historical Market Data

Bo Yuan (by276) Jingwen Ye (jy879)

1 Introduction

In the last two decades, we have witnessed a radical reorganization of the electric utility industry in the United States and other countries. State-owned monopolies have been replaced by regulated and competitive markets where the match between demand and supply takes place in hourly auctions [1]. The auction mechanism is generally based on a uniform price rule, which means all the dispatched producers will receive the same price per MWh of electricity once the market equilibrium is achieved. Besides, the deregulation of the electric utility industry has been taking place in the United States, bringing new benefits and problems [2]. It has shifted the fuel mix of power generation fleetly to a greener form and has made the power plants more efficient and reliable. However, wholesale prices of electricity have increased dramatically in some areas of the United States, market-based investments in transmission have been problematic, and rolling blackouts have been encountered [3]. Hence, the design of the optimal market and the understanding of market participants' behaviors are significant to the ongoing electric industry restructuring process.

Estimating the cost function of electricity generation in a competitive electric market is an important task. The exercise of market power by firms competing in the wholesale market has been a great concern of regulators and governments [4]. Power generating cost functions can be used to measure the extent of market power possessed by a market participant [5]. However, the cost function of a market participant is usually private for protecting the confidential information of the market. Therefore, if we can develop a model to estimate a market participant's cost function by using public market data, like bid quantities and market clearing prices, the extent of market power exercised by that firm can be easily measured. To achieve this, we need to make simplified assumptions about the market participant's bidding behavior, such as it is bidding to maximize the total profit. In previous works, Wolak [5] presented two methods of recovering a firm's cost function estimates in the electricity market using the bids data and market clearing price in Australian National Electricity Market. Maghool et al [6] carried out similar approaches to estimate cost functions under pay-as-bid rules in the Iranian Electric Market. Bosco et al [4] built a model of optimal bidding behavior in the vertically integrated Italian wholesale electricity market. All the aforementioned pieces of literature provide good approaches and useful insights into estimating cost function using publicly available market data. However, as the electric industry has been rapidly restructured over the years, new elements need to be taken into consideration. The growth of power generation capacity, the change of fuel mix, and the increasing impact of transmission loss and congestion may have a significant influence on the market participants' behavior. Therefore, we propose this study of the wholesale electric market in New York State to estimate the cost function of power generating firms, assess their extent of market power exercise ,and provide useful information for understanding the market performance and the market participants' behavior.

2 Materials and Methods

2.1 The model of bidding behavior

In 2002, Wolak [5] devised two models for estimating a power generating firm's cost function under certain assumptions of the firm's optimizing behavior. In the first model, the firm's behavior is not restricted by market rules, and the firm can bid the best-response price that maximizes its profit in each auction, given the market demand and the bids submitted by all other market participants. In the second model, market rules are imposed on the firm's bidding behavior so that the firm may not be able to reach the best-response price that maximizes its profit in each auction. Instead, the first bids to maximize its total profit for each day subject to market constraints, such as limitations on the price segments and bidding quantities. The firm approach is computationally simple and can be a useful diagnosing tool in recovering an estimate of a firm's marginal cost function and its market power. However, the market rules constrain the ability to maximize its profit by setting the best prices all the time. Thus, the deviation of actual price from the best-response price could be large. The second approach respects market rules and should be able to produce more realistic estimations of a firm's cost function, but the computational cost might be larger. In this study, we will carry out both approaches and compare their performance in the wholesale electric market in New York State. A brief description of Wolak's [5] methods is included below.

To begin with, define the following parameters for estimating the cost function of a specific power generating firm A:

Q_{id}	Total market demand in load period i of day d
$SO_{id}(p)$	Total amount of capacity bid by all other firms besides Firm A in load period i of day d at price p
$DR_{id}(p)$	$Q_{id} - SO_{id}(p)$, residual demand faced by Firm A in load period i of day d at price p
QC_{id}	Firm A's contact quantity for load period i of day d
PC_{id}	Firm A's quantity-weighted average contact price for load period i of day d
$\pi_{id}(p)$	Firm A's variable profits at price p for load period i of day d
MC	Firm A's marginal cost of producing 1 MWh of electricity, assumed to be constant
$SA_{id}(p)$	Firm A's bid function for load period i of day d at price p i.e. the amount that Firm A is willing to provide as a function at the price p

The market clearing price p is the smallest price that makes the equation $SA_{id}(p) = DR_{id}(p)$ hold. The variable profits earned by Firm A for load period i of day d can be written as:

$$\pi_{id}(p) = DR_{id}(p)(p - MC) - (p - PC_{id})QC_{id} \tag{1}$$

where the first term is the variable profits by selling electricity in the spot market. The second term represents the payoffs to the generator from buying and selling hedge contracts. The wholesale electric market in NYS clears in each hour, so there are 24 load periods for each day. The residual demand function can be constructed using historical bid data and smoothed by Gaussian kernel smoother for differentiability. Suppose that there are stochastic demand shocks to the price-setting process in each period, Firm A will have to maximize its expected profits given an assumed distribution for this uncertainty. Let ε_i be the stochastic shock and $DR_i(p, \varepsilon_i)$ denote the re-written residual demand function in load period i, where $i = 1, ..., 24$. We define $\Theta = (p_{11}, ..., p_{JK}, q_{1,11}, ..., q_{1,JK}, q_{2,11}, ..., q_{2,JK}, , q_{24,11}, ..., q_{24,JK})$ as the vector of daily bid prices and quantities submitted by firm A. We assume Firm A has J gensets, and the market rule requires Firm A to submit K price increments for each genset, which are constant for the entire day. The parameters in Θ are defined as follows:

p_{kj}	The price set for each of the $k = 1, \ldots, K$ bid increments for genset $j = 1, \ldots, J$
q_{ikj}	The quantity made available to produce electricity in load period i for each of the $k = 1, \ldots, K$ bid increments for genset $j = 1, \ldots, J$
$SA_i(p, \Theta)$	Bid function of firm A in load period i given bid vector Θ
$p_i(\varepsilon_i, \Theta)$	Market clearing price for load period i, given the residual demand shock realization ε_i and the daily bid vector Θ
$f(\varepsilon)$	Probability density function of $\varepsilon = (\varepsilon_1, ..., \varepsilon_{24})$

Firm A's expected profits given the daily bid vector Θ is:

$$E(\Pi(\Theta)) = \int_0^\infty ... \int_0^\infty \sum_{k=1}^{24} \left[DR_i\left(P_i(\varepsilon_i, \Theta)\right)\left(P_i(\varepsilon_i, \Theta) - MC\right) - (P_i(\varepsilon_i, \Theta) - PC_i)QC_i\right] f(\varepsilon) d\varepsilon_1, \ldots, d\varepsilon_{24} \tag{2}$$

The optimal-reply bidding strategy of firm A should be the solution to the following optimization problem.

$$\begin{aligned} \max_{\Theta} \quad & E(\Pi(\Theta)) \\ s.t. \quad & b_U \geq R\Theta \geq b_L \end{aligned} \tag{3}$$

The constraints represented by the matrix R and vectors of upper and lower bounds b_U and b_L imply that the q_{ik} cannot be negative and the sum of the q_{ik} relevant to a given genset should be smaller than or equal to the capacity of the genset.

In the first method, we assume that Firm A can choose the market clearing price to maximum its profit based on its competitors' bid in each load period. We define $C(q)$ as the total variable cost with output level q. The profit of Firm A in each load period is

$$\Pi(p) = DR(p, \varepsilon)p - C(DR(p, \epsilon)) - (p - PC)QC \tag{4}$$

The optimal-reply price is conducted by differentiating Equation 4 and set it to 0. we have:

$$\Pi'(p) = DR'(p, \varepsilon)(p - C'(DR(p, \epsilon))) + (DR(p, \epsilon) - QC) = 0 \tag{5}$$

Using the first-order condition, the estimation of the marginal cost at the observed market-clearing price p^E is

$$C'(DR(p^E, \varepsilon)) = p^E - \frac{QC - DR(p^E, \varepsilon)}{DR'(p^E, \varepsilon)} \tag{6}$$

The residual demand function $DR(p^E, \varepsilon)$ can be approximated by:

$$DR(p, \epsilon) = Q_d(\varepsilon) - SO_h(p, \varepsilon) \tag{7}$$

where the total bid supply function of all other market participants besides Firm A is equal to:

$$SO_h(p, \varepsilon) = \sum_{n=1}^{N} \sum_{k=1}^{10} qo_{nk} \Phi(\frac{p - po_{nk}}{h}) \tag{8}$$

qo_{nk} is kth bid increment of genset n and po_{nk} is 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. $\Phi(t)$ is the standard normal cumulative distribution function, h is the user selected smoothing parameter, and $\varphi(t)$ is the standard normal density function. As a result, we have

$$DR'_h(p, \varepsilon) = -\frac{1}{h} \sum_{n=1}^{N} \sum_{k=1}^{10} qo_{nk} \varphi(\frac{p - po_{nk}}{h}) \tag{9}$$

This model succeeds only if the best-reply prices can be gotten for all realization of ε_i, which is impossible since market rules introduce restrictions on generators to set best reply prices.

In the second method, Firm A's bid need to respect market constraints, so that it may not able to reach the market clearing price in each load period. The firm choose to maximize its total profit through out the day instead. Generalized method moments (GMM) estimation is introduced to recover the cost functions. Define the following parameters:

$SA_{ij}(p, \Theta)$	The amount bid of genset j at price p in load period i
$C_j(q, \beta_j)$	The variable cost of producing output q from genset j
β_j	The vector of parameters of the cost functions for genset j
$SA_i(p, \Theta) = \sum_{j=1}^{J} SA_{ij}(p, \Theta)$	The sum bid of firm A at price p in load period i.

Variable profit for firm A during day d is:

$$\Pi_d(\Theta, \varepsilon) = \sum_{i=1}^{24} \left[DR_i(p_i(\varepsilon_i, \Theta), \varepsilon_i) p_i(\varepsilon_i, \Theta) - \sum_{j=1}^{J} C_j(SA_{ij}(p_i(\varepsilon_i, \Theta), \Theta), \beta_j) - (p_i(\varepsilon_i, \Theta) - PC_i) QC_i \right] \tag{10}$$

where ε is the vector of realization of $\varepsilon_i, i = 1, \ldots, 24$. $p_i(\varepsilon_i, \Theta)$ is the market clearing price for load period based on the residual demand shock realization ε_i, and the daily big vector Θ. The optimal bidding strategy builds an optimization problem:

$$\begin{aligned} \max_{\Theta} \quad & E(\Pi(\Theta, \varepsilon)) \\ s.t. \quad & q_{ikj} \geq 0, \forall i, j, k \\ & b_U \geq R\Theta \geq b_L \end{aligned} \tag{11}$$

The first order condition is applied to the moment restrictions, and we get:

$$E_\varepsilon \left(\frac{\partial \prod_d(\Theta_d, \varepsilon)}{\partial p_{kj}} \right) = 0 \tag{12}$$

for all gensets, j, and bid increments, k. Equation 12 defines the $J \times K$ moment restrictions used to match parameters of cost functions. The sample analog of this moment restriction is:

$$\begin{aligned} \frac{\partial \Pi_d(\Theta_d, \varepsilon)}{\partial p_{kj}} = & \sum_{i=1}^{24} [(DR'_i(p_i(\varepsilon_i, \Theta), \varepsilon_i) p_i(\varepsilon_i, \Theta) + (DR_i(p_i(\varepsilon_i, \Theta), \varepsilon_i) - QC_i) - \\ & \sum_{j=1}^{J} C'_j(SA_{ij}(p_i(\varepsilon_i, \Theta)), \beta_j) \frac{\partial SA_{ij}}{\partial p_i} \frac{\partial p_i}{\partial p_{kj}} - \sum_{j=1}^{J} C'_j(SA_{ij}(p_i(\varepsilon_i, \Theta)), \beta_j) \frac{\partial SA_{ij}}{\partial p_{kj}}] \end{aligned} \tag{13}$$

where p_i is the shorthand for the market-clearing price in load period i. The values of partial derivatives in Equation 13 are computed as follows:

$$SA^h_{ij}(p, \Theta) = \sum_{k=1}^{10} q_{ikj} \Phi(\frac{p - p_{kj}}{h}) \tag{14}$$

$$SA_i^h(p,\Theta) = \frac{1}{h}\sum_{j=1}^{J}\sum_{k=1}^{10} q_{ikj}\Phi(\frac{p-p_{kj}}{h}) \tag{15}$$

$$\frac{\partial SA_{ij}}{\partial p} = \frac{1}{h}\sum_{k=1}^{10} q_{ikj}\phi(\frac{p-p_{kj}}{h}) \tag{16}$$

$$\frac{\partial SA_{ij}}{\partial p_{kj}} = -\frac{1}{h}q_{ikj}\phi(\frac{p-p_{kj}}{h}) \tag{17}$$

$$\frac{\partial p_i(\varepsilon,\Phi)}{\partial p_{kj}} = \frac{\frac{\partial SA_i(p_i(\varepsilon_i,\Theta),\Theta))}{\partial p_{kj}}}{DR_i'(p_i(\varepsilon_i,\Theta),\varepsilon_i) - SA_i'(P_i(\varepsilon_i,\Theta),\Theta)} \tag{18}$$

$$DR(p,\varepsilon) = Q_d(\varepsilon) - SO_h(p,\varepsilon) \tag{19}$$

$$SO_h(p,\varepsilon) = \sum_{n=1}^{N}\sum_{k=1}^{10} qo_{nk}\Phi(\frac{p-po_{nk}}{h}) \tag{20}$$

Then, a functional form for $C_j(q,\beta_j)$ is required for estimating cost function. Let $l_d(\beta)$ denotes the $J\times K$ dimensional vector of partial derivatives of $\Pi_d(\Theta_d,\varepsilon)$ given in Equation 13, Assuming the functional form of $C_j(q,\beta_j)$ is correct, we have $E(l_d(\beta^0)) = 0$. Solving the value of b that minimizes:

$$[\frac{1}{D}\sum_{d=1}^{D} l_d(b)]'[\frac{1}{D}\sum_{d=1}^{D} l_d(b)] \tag{21}$$

will yield a consistent estimate of β. Let $b(I)$ denotes this consistent estimate of β, where I denotes the fact that the identity matrix is used as the GMM weighting matrix. Construct a consistent estimate of the optimal GMM weighting matrix using the consistent estimate of β as shown below:

$$V_D(b(I)) = \frac{1}{D}\sum_{d=1}^{D} l_d(b(I))l_d(b(I))' \tag{22}$$

The optimal GMM estimator finds the value of b that minimizes:

$$[\frac{1}{D}\sum_{d=1}^{D} l_d(b)]'V_D(b(I))^{-1}[\frac{1}{D}\sum_{d=1}^{D} l_d(b)] \tag{23}$$

2.2 The New York State wholesale electric market

The New York Independent System Operator (NYISO) directs the operation of the New York State (NYS) power system to supply power to loads while maintaining safety and reliability [7]. The NYISO provides open access to the NYS transmission system for the market participants and facilitates a Day-Ahead Market (DAM), and a Real-Time Market (RTM) using market participants' bid data. The energy market provides a mechanism for market participants to buy and sell Locational Based Marginal Price (LBMP) energy and to bid various kinds of bilateral transactions [8]. Suppliers may sell energy directly into the market at LBMP or be party to a bilateral contract selling directly to purchasers. Load Serving Entities (LSEs) and others may purchase energy at LBMP by submitting bids and/or they may be party to a bilateral contract purchasing directly from a supplier.

The NYISO energy market uses a two-settlement process [7]. The first settlement is based upon the day-ahead bids and the corresponding schedule and prices determined by the day-ahead Security-Constrained Unit Commitment (SCUC). The second settlement is based upon the real-time bids, the corresponding real-time commitment (RTC), and real-time dispatch (RTD). Market participants may participate in the DAM and/or the real-time market. As a result of the day-ahead commitment process, a set of generators are scheduled to be available for dispatch in each hour of the day and a set of LSEs are scheduled to buy a certain amount of load at the day-ahead price. The generators designated by SCUC to be available for the next day are dispatched against the LSE bid-in load and losses. From the dispatch, LBMPs are computed and forward contracts are established for generation and load accordingly. Subsequently, during real-time operation, changes in operating conditions, the influence of additional real-time-supply bids, and variations in actual load will cause the real-time schedules and prices to be different from the day-ahead schedules and prices. The difference in generation levels and load consumption as compared to the first settlement values are settled at the second settlement or real-time price. For our study, we are only focusing on the day-ahead market and using day-ahead bids and prices as our model inputs.

2.3 Historical market data

Historical market data are downloaded from NYISO's Energy Market and Operational Database [9]. Day-ahead locational based marginal price (LBMP) at each bus is available, as well as the aggregated zonal LBMPs. The LBMPs represent the incremental value of an additional MW of energy injected at a particular location. The LBMP at each bus is calculated using the price of energy at a reference bus, the cost of losses, and the cost of congestion relative to the reference bus. The LBMP for the reference bus is defined as the least cost to supply a 1-MW load attached to the reference bus. The loss and congestion costs at the reference bus are zero [7]. At other buses, the transmission grid determines the losses, and transmission constraints determine congestion values. To simplify this model, we are not considering the effect of transmission losses and congestion on LBMP, since the behavior of the system operator under congestion may differ from the one described above. We will use only the data of those auctions in which the cost of losses and congestion are small, so that we can model the entire New York State market as a whole. In short, we will use the hourly LBMP data at the reference bus when losses and congestion are low for our study.

Zonal load data are also available from the NYISO's Energy Market and Operational Database [9]. However, we can only get the total amount of energy bid load and bilateral load in each zone in each hour, while the amount of load served by individual generators are not available for protecting market information. Fortunately, we can get unit-level power generation data from the United States Environmental Protection Agency's Air Market Program Database [10]. This database contains hourly power generation, heat input, and air pollution emission data of most thermal power generators in New York State. Using this database, we can construct the bid function of an arbitrary power generating firm and the residual demand function that the firm had to face in each load period.

The marginal cost of power generation can also be estimated by calculating its fuel cost and variable operation and maintenance (OM) cost. To validate the results, we will estimate the marginal cost of the firm from its fuel cost and OM cost, and compare it with the cost function we got using the market bid data. The reference level of fuel cost and OM cost of typical thermal power generators are given in the National Renewable Energy Laboratory's Annual Technology Baseline [11]. We can also use NYISO's Reference Level Manual in 2019 [12] as a complementary data source.

3 Anticipated Results

We anticipate getting a good estimation of cost functions of various power generating firm in the New York State wholesale electric market. We anticipate observing different features of the cost functions of different firms due to the differences in their generating capacity, unit type, fuel type, geographic location, and so on. We also anticipate observing monthly or seasonal trends of variation of the cost functions due to the change in fuel price, and electricity demand. We also anticipate getting measures of market power exercise of the power generators.

4 Intellectual Merit

The proposed work will advance knowledge of the relation between power generation cost and the market clearing price, and the extent of exercise of market power and its impact on market performance. It is potentially transformative because it could be widely used by power system operators and electric utility owners to improve the efficiency of the electricity market. The researchers are qualified based on previous research experience and knowledge background. Adequate resources are provided by the Department of Electrical and Computer Engineering at Cornell University.

5 Broader Impacts

The results of the proposed research could affect many aspects of the wholesale power market. It has the potential to help power system operators get good estimations of power generation cost of the market participants, improve the efficiency of the market system, and preventing the exercise of market power by competing firms. The research will also benefit the power generator owners themselves by understanding their competitor's market behavior and maximize their profits. A more efficient electric market will eventually benefit the general public by providing greener and cheaper electricity for everyone to use.

References

[1] Matteo Pelagatti. *Supply Function Prediction in Electricity Auctions*, pages 203–213. Springer Milan, Milano, 2013.

[2] S. Blumsack. How the free market rocked the grid. *IEEE Spectrum*, 47(12):44–59, 2010.

[3] J. Duncan Duncan Glover and Mulukutla S. Sarma. *Power System Analysis and Design*. Brooks/Cole Publishing Co., USA, 3rd edition, 2001.

[4] Bruno Bosco, Lucia Parisio, and Matteo Pelagatti. Strategic bidding in vertically integrated power markets with an application to the Italian electricity auctions. *Energy Economics*, 34(6):2046–2057, 2012.

[5] Frank a Wolak. Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets. Technical Report January, Stanford University, 2002.

[6] E. Maghool, A. Naghavi, and S. F. Ghaderi. Estimating cost function in power markets under pay-as-bid pricing rules using observed bid data. *2008 IEEE Electrical Power and Energy Conference - Energy Innovation*, pages 1–7, 2008.

[7] New York Independent System Operator. Market Participants User's Guide, 2020.

[8] New York Independent System Operator. Day-Ahead Scheduling Manual, 2020.

[9] New York Independent System Operator. Energy Market and Operational Data, 2020.

[10] United States Environmental Protection Agency. Air Markets Program Data, 2020.

[11] National Renewable Energy Laboratory. 2020 Annual Technology Baseline, 2020.

[12] New York Independent System Operator. Reference Level Manual, 2019.