BUSINESS INTELLIGENCE USING INFORMATION GAP DECISION THEORY AND DATA MINING APPROACH IN COMPETITIVE BIDDING

Mei-Peng Cheong
Gerald B. Sheblé
Daniel Berleant

ABSTRACT

Since the nineties, many electric utilities and power network companies have undergone and are still experiencing dynamic change in the ways of doing business, from a vertically integrated industry to an open market system. The operational planning activity of a generation company (GenCo) is no longer a cost-minimizing process, but seeks to maximize its net profit subject to physical constraints and market factors. The objective of this research is to develop a strategic bidding decision-making unit that not only considers the technical aspects of unit operation such as capacity limits but also incorporates information about other market participants and the volatility of the market prices. These additional market factors are significant in such an oligopoly market because they influence the amount of electricity sold and purchased, hence affecting the net profit gained. This project proposes an economic model that attempts to data mine the available historical and current market data in a deterministic two-market-participant environment. Further stochastic analysis is performed using the information gap decision theory concept to quantify the uncertainty that arises. The data mining approach can also be justified for information acquisition to reduce uncertainty, hence improving the information gap model.

INTRODUCTION

Traditionally, the electric power industry was dominated by large utilities that manage overall activities in generation, transmission, and distribution of power. The economic incentives to provide cheaper and reliable electricity as well as to encourage efficient capacity expansion and investment planning have opened up the option to introduce competition in the electricity sector. To date, almost half of the states in North America are either fully deregulated or in the transition stages, while others are in the application process[1]. The restructuring process has introduced increased competition and most GenCos no longer employ the conventional unit commitment and economic dispatch techniques to produce the optimal price-quantity bids. Instead, GenCo is now faced with a competitive strategic bidding environment at a higher uncertainty level since every market player has information only about its own production activities and some publicly available information such as market clearing electricity prices and fuel prices.

In this competitive market mechanism, the behavior of each market participant affects but does not control the market, hence leading to an oligopoly market where the market is not perfectly competitive. Market players are faced with uncertainties in which electricity can now be considered a type commodity and its prices are determined by market forces. Every decision made by the market players is dependent on factors that can be described by Porter's five forces. In this framework, Michael Porter illustrates the relationship between competitors within an industry, potential competitors, suppliers, and buyers. The five forces are barriers to entry, rivalry among existing competitors, product substitutes, market power of buyers, and market power of suppliers.

Optimal bidding strategies can be formulated in several ways depending on what type of information is available and accessible. Often times, GenCo models its bidding strategy with the aid of forecasting tools that estimate the market-clearing price (MCP) of the market in the next trading period based on historical fluctuations in prices. A more complicated model will try to include the behavior of the competitors in an attempt to outperform their rivals. The following sections begin with the data mining approach that analyzes the behavior of the competitors' policies or strategies in bidding followed by the information gap decision concept to quantify uncertainties for the dependent input variables in the bidding model.

DATA MINING

Data mining refers to the process of transforming collected raw data into usable information. Although data in its raw form is of limited use, it can be manipulated to realize its potential use. Raw data can be aggregated and analyzed, together with heuristic knowledge about the nature of the problem, providing significant information that can be acted upon in making informed decisions.

This section explains the data mining process with a deterministic model. The organization of the data mining process comprises the forward and backward process. The forward process aims to provide an optimal bidding strategy for GenCos (producers) that includes not only the physical constraints but the

[1] Source: Energy Information Administration (EIA) as of February, 2003.

demand side as well. This model generates a bid-quantity series as the optimal decision with the given information and the inherent structure of the model. The backward process aims to use the output of the forward process to reverse the bidding strategies employed. The basic information on the market structure includes, but is not limited to, the following: have available data on system outages and forecasted loads; loads bid in hourly fashion; forecasted load is available in the day-ahead market; suppliers bid in minimum generation blocks and incremental energy blocks with increasing costs; hourly prices, loads committed, and generation bids are posted.

The Forward Process

The fundamental objective in the forward process is to be able to generate the appropriate price-quantity bidding curve with respect to market movements such as the price of electricity and fuel. In the real market, price and quantity are determined by the market forces, in particular the demand and supply for electricity. MCP depends on fuel price, heat rate curve (to calculate fuel cost), variable cost, operating and maintenance cost, wheeling cost, new equipment installation, future cost evolution, demand variation, and other economic and cost considerations. In our project, we develop a model that consists of the following two components:

- SUPPLY: profit-maximizing competitive producers or utility companies (generation companies)
- DEMAND: price-taking cost-minimizing consumers

In the SUPPLY model, the fuel price is assumed to be given and each supplier is maximizing profit with the following objective function:

MAXIMIZE Profit = Revenue – Cost

Revenue originates from spot market sell and bilateral power sell, and cost is represented as payments for spot market buy, unit operating costs, startup costs, and shut down costs.

$$\underset{\rho,q()}{MAX}\pi = \sum_{k}[\rho_{sk} \cdot q(sell)_k + q(contract)_k \cdot \rho_{ck} - \rho_{sk} \cdot q(buy)_k - \sum_{i=1}^{NG}\{w_{i,k} \cdot c(\min) + qg_{i,k} \cdot gc_i + ust_{i,k} \cdot st_i + usd_{i,k} \cdot sd_i\}]$$

Where:

ρ_{sk}, ρ_{ck} Spot market price and bilateral contract price for time period k

$q(\cdot)_k$ Amount of power sold, contracted, or bought for time period k

$w_{i,k}, ust_{i,k}, usd_{i,k}$ Binary variables for unit status (1=ON/startup/startup, 0=OFF/No startup/No

	shutdown) for time period k for duration i
$c(\min)_i$	Cost of minimum generation for duration i
$qg_{i,k}$	Amount of power generated in addition to minimum generation for time period k for duration i
gc_i	Generation cost beyond minimum generation cost for duration i
st_i, sd_i	Startup and shutdown costs

First of all, we generate a break-even bid curve by excluding the hedging component of the equation.

$$Revenue\ (\rho_{sk}, q(sell)_k) = Cost\ (c(\min)_i, qg_{i,k}, gc_{i,k}, ust_{i,k}, st_i, usd_{i,k}, sd_i)$$

In addition to the quantity sold, we need to consider that there is a fixed payment or minimum payment charged to guarantee future flow of the commodity (electricity) sold. This fixed payment can be treated as a means to recover fixed cost in the long run. This fixed payment is assumed to recapture the minimum generation and startup and shutdown costs.
Deleted the following bracket at the end

$$\sum_k R_{fixed,i} + \rho_{sk} \cdot q(sell)_k = \sum_k \sum_{i=1}^{N} \{w_{i,k} \cdot c(\min)_i + qg_{i,k} \cdot gc_i + ust_{i,k} \cdot st_i + usd_{i,k} \cdot sd_i\}$$

$$\text{where} \quad R_{fixed,i} = w_{i,k} \cdot c(\min)_i + ust_{i,k} \cdot st_i + usd_{i,k} \cdot sd_i$$

The generation cost curve should include the price of fuel (gas) and the average heat rate (AHR) curve of the generating unit. The ρ_{gk} in the following equation represents the price of gas for time period k and the AHR curve is obtained from the input-output (IO) curve averaged over the quantity of power generated.

$$gc_i = \rho_{gk} \cdot AHR \qquad where \qquad TQ = q^2 + q + 18 \qquad \Rightarrow \qquad AHR = \frac{TQ}{q} = q + 1 + \frac{18}{q}$$

The updated revenue equals cost equation:

$$B(q)_k \cdot q(sell)_k = qg_{i,k} \cdot \rho_{gk} \cdot \left(qg_{i,k} + 1 + \frac{18}{qg_{i,k}} \right)$$

In this equation, ρ_{sk} has been replaced with $B(q)_k$ because we want to find the relationship between the bid price and the generation level $qg_{i,k}$ for time period *k*. The spot market price is the price paid, and bids submitted may or may not be accepted. We also know that $q(sell)_k = qg_{i,k}$ (ignoring the hedging component) and hence our break-even bid curve can be represented by:

$$B(q)_k = \rho_{gk} \cdot \left(qg_{i,k} + 1 + \frac{18}{qg_{i,k}} \right)$$

Assuming that the price of fuel is given, we can produce a break-even bidding curve for the supplier. We would then include the maximum profit bid curve by finding the first-order-condition for the profit function. The resulting bidding curve is shown in Figure 1.

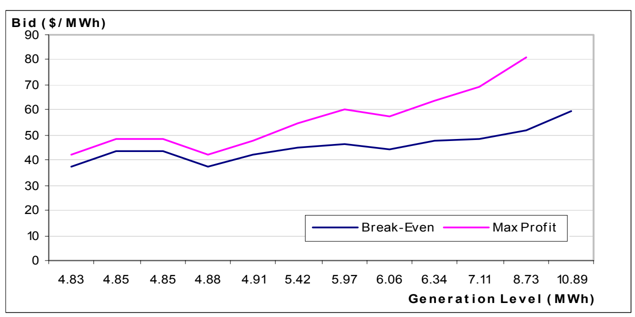


Figure 1. The maximum profit and break-even bid at each generation level.

The maximum profit is obtained by using the forecasted electricity price that maximizes our profit. The break-even curve is found by the equation derived above and is dependent on the natural gas prices. Therefore, we have a range of bids at each cross section of the generation level. The amount to bid depends on the risk profile of the GenCo, and later sections will explain how to quantify uncertainty in determining the price-quantity bid.

The Induction Procedure

As mentioned in the earlier section, data mining seeks to discover hidden patterns that can inform the analyst about the strategy or types of generating unit used. The first data mining approach used in this research is the rule or decision tree induction to deduce the pattern discovered from the data. It produces patterns that relate to the bidding decision made or other data fields (attributes). The resulting patterns are typically generated as a tree with splits on data fields and terminal points (leafs). In order to study the bidding behavior of a supplier, we first use this rule induction to separate the *active* (when bid is submitted) bid-quantity-price group from the *passive* (when no bids are submitted) group.

The algorithm used for the decision tree induction is the standard C4.5 algorithm [1,2]. Similar to its predecessor, the ID3 algorithm, C4.5 is a top down induction of decision tree that employs the divide and conquer concept. However, to select the best splitting attribute, it uses the highest *information gain ratio*[2] based on the entropy formula to generate decision tree. The optimal splitting attribute chosen is the attribute that results in the smallest tree. The general steps for C4.5 are as follows: (1) Attribute is selected for root node and branch is created for each possible attribute; (2) The instances are split into subsets (one for each branch extending from the node); (3) Procedure is repeated recursively for each branch, using only instances that reach the branch; (4) Procedure stops when all instances have the same class. It infers decision tree from the training set to convert the learned tree into an equivalent set of rules.

We have developed a basic forward process and used the Java program (Weka3.2) to run C4.5 on the data generated by the forward process. The first rule induction process yields the following simple result shown in Figure 2.

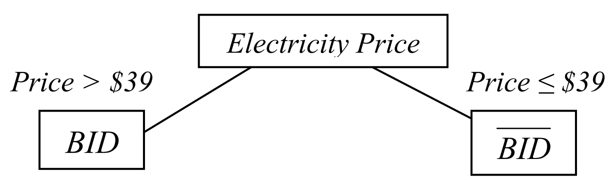


Figure 2. The induced decision tree based on the data generated from the forward process.

Simple as it seems to be, this is a strong result in that this rule matches all the data generated from the forward process. From the competitor's viewpoint, it may only rely on the forecasted electricity price to determine whether we, as GenCo 1, will bid or not and adjust its bid accordingly. However, this set of data may over-fit the algorithm and thus more data should be used to verify this decision tree. Moreover,

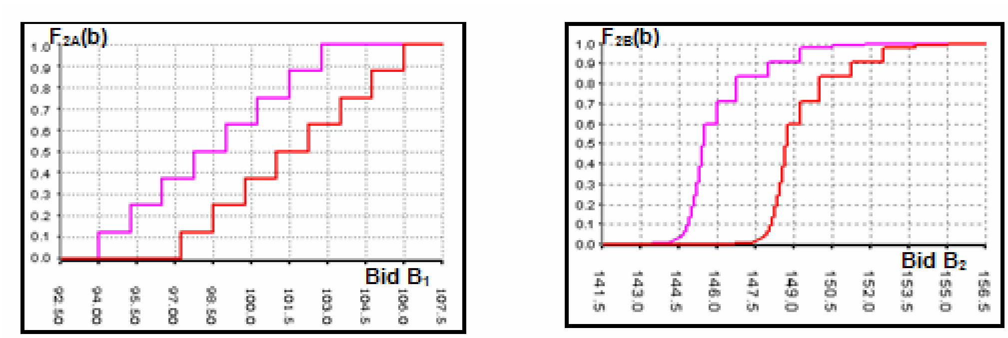
---

[2] *Information gain ratio* is used to compensate for the number of attributes by normalizing the information encoded in the split itself since the *information gain* formula ignores the number of regions used, hence may lead to bias and over-fitting problem.

we cannot conclude that the fuel price is insignificant when it comes to bidding decision. This induction process will be extended to include other market factors, such as the fuel prices and market demand, to discover the price signals that are used to infer the strategies employed by other market participants. Since electricity prices, fuel prices, and other market variables are fuzzy or uncertain, we will next describe a model that helps to quantify uncertainty in decision making.

USING INFORMATION GAP THEORY TO QUANTIFY UNCERTAINTY

Information gap theory [3] is useful for making decisions in cases where uncertainty is present and severe. We have developed an info-gap model based on the bidding model described in Reference [4]. Information gap theory handles distributions that may not be fully specified, such as in Figures 3a–3b.



(a) GenCo 2 cost function F_{2A} for G_{2A}. *(b) GenCo 2 cost function F_{2B} for G_{2B}.*

Figure 3. The cost functions with respect to bids for the two generators of GenCo 2.

This problem is formulated with two generation companies—GenCo 1 and a competitor, GenCo 2. Both GenCos are competing to sell X_D megawatt-hours (MWh) of electric energy. GenCo 1 attempts to model GenCo 2 to determine a bid for an amount and a price that will serve its profit-making interests.

Suppose that we wish to ensure that the expected monetary value (*EMV)* of a bid (corresponding to the expected profit) meets or exceeds a given minimum value. An information gap model helps to identify bids that meet that requirement and the uncertainty-reducing information needed to ensure that other, possibly more desirable, bids meet that requirement. An example of such a potentially more desirable bid would be one that corresponds to a wide range of possible *EMV* values, some quite high and desirable and others below a minimum tolerable *EMV*. For example, in Figure 4, the bidder may enjoy a high *EMV* of 74200, at a bid of \$145/MWh, but that bid may also result in an *EMV* of 29680 if the true curve happens to be the lowest *EMV* curve shown.

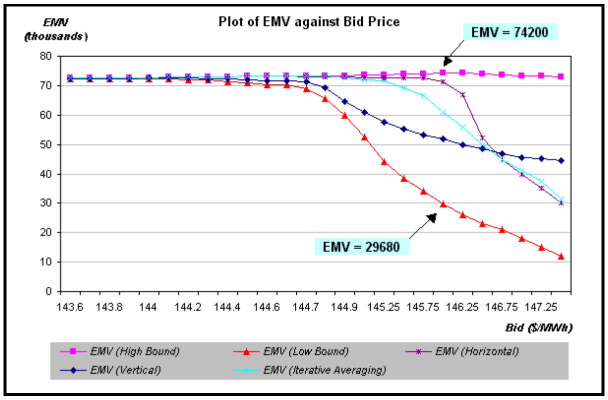


Figure 4. A wide range of possible *EMV* values for a given bid.

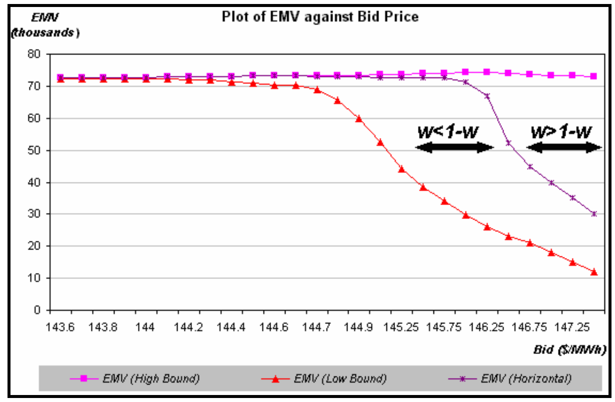


Figure 5. *EMV* curves corresponding to the left envelope (lowest curve), the right envelope (highest curve), and the horizontal average envelope of Figure 6.

An information gap model for this example problem may be specified as follows.

1. ***Decision variable.*** This is our bid B_2 in *\$/MWh*.
2. ***Uncertain variable.*** Define a cumulative distribution function (CDF) for the competitor's bid that serves in the role of nominal best guess. Any CDF judged to fill this role could be used. For purposes of illustration we use "horizontal averaging" of the left and right CDF envelopes of Figure 3b, giving the intermediate curve of Figure 6. In horizontal averaging, for each vertical axis value y_i, the corresponding horizontal axis values of the left envelope, B_l, and of the right envelope, B_r, are averaged, giving a value $B_i=(B_l+B_r)/2$. The point (B_i, y_i) is on the average CDF curve, which may be plotted as precisely as desired by using an appropriate set of values for i. The average CDF serves as a nominal best guess CDF. Accounting for weights generalizes the averaging formula to $B_i=(wB_l+(1-w)B_r)/2$. Let the uncertain variable in the info-gap model be the weight w of the left envelope, with the weight of the right envelope then being $1-w$. Then Figure 5 describes the *EMV* values calculated from the CDF envelopes of Figures 3b and 6.
3. ***Nominal value of uncertain variable.*** There seems to be no particular reason to prefer weighting one envelope more than the other when doing horizontal averaging, so the default nominal value of weight w is $\tilde{w}=0.5$.
4. ***Uncertainty parameter.*** The amount of uncertainty in the model, α, is the amount of deviation from the nominal value of the uncertain variable that is to be considered. In this model, that is the amount of deviation from $\tilde{w}=0.5$. In the worst case, this might be ±0.5, giving a range of weights from 0 to 1. Determining this is the goal of the information gap analysis.

5. ***Uncertainty model.*** This is the function $\mathbf{u}(\alpha, \widetilde{w})$ that describes the amount of uncertainty in the uncertain variable w in terms of its nominal value $\widetilde{w}$ and uncertainty parameter α. Consistent with points 2–4 above, we have $\mathbf{u}\ (\alpha, \widetilde{w})=\{w\!: w \leq |\,\widetilde{w}+\alpha|\}$.

6. ***Reward function.*** The reward is the *EMV* of a bid. It is determined by the bid value and the *EMV* curve that applies. In this problem, the *EMV* curve is uncertain, so the worst possible *EMV* curve is used to allow the reward function to provide the minimum *EMV* that the bid could be associated with, as required by the info-gap analysis. The worst possible *EMV* curve is in turn determined by the leftmost possible CDF curve for the competitor's bid. This curve is found by horizontal averaging with an averaging weight of $\widetilde{w}+\alpha$. Using this curve is consistent with the ultimate goal of designing the bid and any necessary information-seeking activities to ensure at least a minimum *EMV*. Thus reward function *R* is defined by

$$R(B_j, w) = EMV\,(B_j, w \cdot \overline{F_{2B}}(B_j) + (1-w) \cdot \underline{F_{2B}}(B_j))$$

   Where $\overline{F_{2B}}(B_2)$ is the highest possible envelope around the function F_{2B} (i.e., the left envelope in Figure 3b), and $\underline{F_{2B}}(B_2)$ is the corresponding lowest possible envelope (i.e., the right envelope).

7. ***Critical reward.*** This is the minimum acceptable value of the reward function, call it r_c. The results of an information gap analysis differ depending on the value assigned to r_c,.

8. ***Robustness function.*** This function, $\hat{\alpha}(b, r_c)$, returns the greatest value of uncertainty parameter α for which falling below the critical reward r_c is not possible in the model. It therefore measures the ability of the model to deliver an acceptable reward in the presence of uncertainty, hence the term *robustness*. Its value is therefore dependent on acceptable reward r_c. It is also dependent on the bid B_2, because the reward is dependent on B_2.

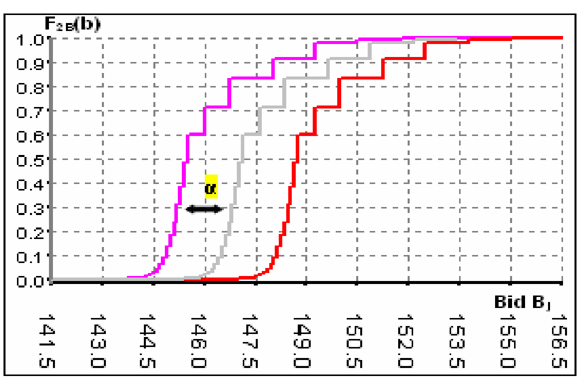


Figure 6. Best-guess curve between the left and right envelopes computed by horizontal averaging and the maximum uncertainty α, showing that the space of plausible curves is within the envelopes.

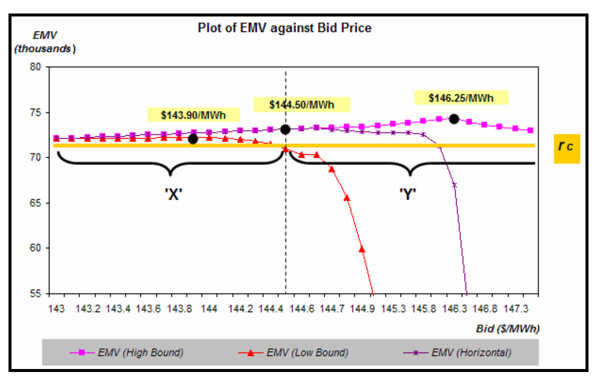


Figure 7. Critical reward separates the *EMV* curves into regions.

Figure 7 gives information about $\hat{\alpha}(b,r_c)$ for a range of bid values, and a specific value of r_c, which would be a business decision provided as a model input. For bid values toward the left of Figure 7 (denoted as region 'X'), the *EMV* is above r_c regardless of which *EMV* curve is considered, so r_c will be safely met for any of those bids. However it may be desirable to consider bidding higher in order to reap the potential opportunity for greater gain due to potentially higher *EMV* values such as, for example, the peak feasible *EMV* of \$146.25/MWh for region 'Y'. This is simply an analytical elaboration of the intuition that the higher the bid the greater the profit, if the bid is successful, but the higher the chance that a competitor will undercut the bid resulting in no profit. Thus, bids in the range designated by 'Y' in Figure 7 are not guaranteed to have an *EMV* of at least r_c unless new information is obtained that rules out values of w that are too close to 0 (thereby moving the worst-case *EMV* curve upward). The result of an information gap analysis is a validation (or invalidation) of a bid value based on whether its *EMV* meets minimal requirements. However, the analysis does not tell us what the best bid is.

In Reference [4], we presented a discussion on the different decision criteria used to determine the bid such as maximizing worst-case *EMV*s, maximizing expected *EMV*s, and converting *EMV*s to utilities using risk profiles. Another approach is to seek information by using the data mining techniques to reduce the uncertainty in the model. However acquisition of information requires an expenditure of resources. Unless the new information leads us to a more informative and robust info-gap model $\mathbf{u}_{new}$, which is a subset of $\mathbf{u}(\alpha, \widetilde{w})=\{w\!: w \leq |\widetilde{w} + \alpha|\}$ mentioned earlier, the acquisition of information is of no value added to our model.

CONCLUSIONS

This research shows that data mining using an evolutionary technique can be used to infer rivals' strategy. Based on available market information, one can develop algorithms to deduce the policy employed by competitors. More work will be implemented to include the forecast of market variables, such as prices. However, not all information is free and the acquisition cost of information varies. A support tool such as the information gap theory can assist in quantifying severe uncertainty when information is scarce and expensive. It helps decision makers to develop preferences, assess risks and opportunities, and seek information, given a minimum required level of reward. This minimum level of reward could be determined by incorporating risk management methodologies such as value at risk or profit at risk. Understanding how to balance the cost of new information with its benefits is an important next step.

ACKNOWLEDGEMENT

This work was supported in part by the Electric Power Research Center (EPRC) and the Power Affiliate Research Program at Iowa State University.

BIBILIOGRAPHY

1. Quinlan, J. R., "Induction of decision trees." *Machine Learning*, vol. 5, 1986, pp. 239–266.
2. Quinlan, J. R., *C4.5: Programs for Machine Learning*. San Mateo, California, 1993.
3. Y. Ben-Haim, *Information Gap Decision Theory*. California: Academic Press, 2001.
4. M.-P. Cheong, D. Berleant, and G. Sheblé, "Information gap decision theory as a tool for strategic bidding in competitive electricity markets," submitted to *Probabilistic Methods Applied to Power Systems (PMAPS) International Conference*, 2004.
5. M.-P. Cheong, D. Berleant, and G. Sheblé, "On the dependency relationship between bids," in *Proceedings of the 35th North American Power Symposium*, October 20–21, 2003, pp. 277–281.
6. D. Berleant, L. Xie, and J. Zhang, "Statool: A tool for distribution envelope determination (DEnv), an interval-based algorithm for arithmetic on random variables," *Reliable Computing*, vol. 9, 2003, pp. 91–108.
7. S. Ferson and L. R. Ginzburg, "Different methods are needed to propagate ignorance and variability," *Reliability Engineering and Systems Safety*, vol. 54, 1996, pp. 133–144.
8. S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, "Constructing probability boxes and Dempster-Shafer structures." *Rep. SAND2002-4015*, January 2003.
9. C. Mattson, D. Lucarella, and C. C. Liu, "Modelling a competitor's bidding behavior using fuzzy inference networks," in *Proc. 2001 Intelligent System Applications in Power Systems (ISAP) International Conf.*, 2001.
10. C. W. Richter and G. B. Sheblé, "Genetic algorithm evolution of utility bidding strategies for the competitive marketplace," *IEEE Transactions on Power Systems*, vol. 13, no. 1, 1998, pp. 256–261.
11. C. W. Richter, G. B. Sheblé and D. Ashlock, "Comprehensive bidding strategies with genetic programming/finite state automata," *IEEE Transactions on Power Systems*, vol. 14, no. 4, 1999, pp. 1207–1212.
12. G. B. Sheblé, *Computational Auction Mechanisms for Restructured Power Industry Operation*. Massachusetts: Kluwer Academic Publishers, 1999.
13. D. C. Skinner, *Introduction to Decision Analysis, 2nd edition*. Florida: Probabilistic Publishing, 2001.

14. H. L. Song, C. C. Liu, J. Lawarree, and R. W. Dahlgren, “Optimal electricity supply bidding by Markov decision process,” *IEEE Trans on Power Systems*, vol. 15, May 2000, pp. 618–624.
15. S. Stoft, *Power System Economics: Designing Markets for Electricity*. New Jersey: Wiley, 2002.
16. L. Yang, F. S. Wen, F. F. Wu, Y. Ni, and J. Qiu, “Development of bidding strategies in electricity markets using possibility theory,” Int. Conf. on Power System Technology, Kunming, China, October 13–17, 2002.
17. F. S. Wen and A. K. David, ”Optimal bidding strategies and modeling of imperfect information among competitive generators,” *IEEE Transactions on Power Systems*, vol. 16, February 2001, pp. 15–21.
18. F. Allen and R. Karjalainen, “Using genetic algorithms to find technical trading rules,” *Journal of Financial Economics*, vol. 51, 1999, pp. 245–271.
19. L. A. Berker and M. Seshadri, *GP-evolved Technical Trading Rules Can Outperform Buy and Hold*. Worcester Polytechnic Institute, Computer Science Technical Report.
20. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Pennsylvania: Addison-Wesley, Reading, 1989.
21. J. H. Holland *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: University of Michigan Press, 1975.
22. R. C. Holte, “Very simple classification rules perform well on most commonly used datasets,” *Machine Learning*, vol. 11, 1993, pp. 63–91.
23. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 2000.
24. J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
25. J. R. Koza, *Genetic Programming II: Automatic Discovery of Reusable Programs*. MIT Press, 1994.
26. J. Koza, D. Goldberg, D. Fogel, and R. Riolo, (ed.), *Proceedings of the First Annual Conference on Genetic Programming*. MIT Press, 1996.
27. J. Li and E. P. K. Tsang, “Improving technical analysis predictions: An application of genetic programming,” in *Proceedings of the 12th International Florida AI Research Society Conference*, pp. 108–112, Orlando, Florida, May 1–5, 1999.