Statistical Approaches to Electricity Price Forecasting

By J. Stuart McMenamin, Ph.D., Frank A. Monforte, Ph.D.
Christine Fordham, Eric Fox, Fredrick D. Sebold Ph.D., and Mark Quan

1. Introduction

With the advent of competition, hourly electricity prices are being determined by a variety of market mechanisms, rather than cost-based engineering calculations. As a result, electric utilities, generators, and traders face a new set of short-term forecasting problems. These problems are unlike those in other industries, since electricity must be produced at the same time that it is consumed. As a result prices are determined on an hourly basis, 24 hours a day, 7 days a week.

Historically, price forecasting has been performed with least-cost optimization models. These models compute marginal cost based on assumptions about system loads, power plant availability, and fuel prices. These models do not explain price variations related to market strategy and to buyer and seller behaviour in a market system. Statistical models, which by their nature reflect actual market outcomes, are better suited to short-term forecasting in this dynamic environment.

In this paper, a variety of modeling approaches are applied and evaluated for forecasting electricity prices. Methods include time-series models, regression models, and artificial neural network models. The paper discusses the nature of the price-forecasting problem and identifies reasons why flexible approaches, such as neural network models are well suited to this application.

For each approach, model estimates and forecasts are developed using hourly price data for the PJM (Pennsylvania, New Jersey and Maryland) power pool area. The modeling approaches are compared based on accuracy for day-ahead forecasting.

2. Price Modeling Approaches

In competitive electricity markets, the market-clearing price is the hourly bid of the last generation unit to meet system demand. In most systems, all suppliers are paid the hourly market-clearing price. In a perfectly competitive market, the market-clearing price will be equal to the marginal cost of the last supplier. But existing electricity markets are far from meeting the requirements for perfect competition. Reflecting the fact that there are a limited number of suppliers and that customer demand is highly inelastic with respect to the market price, there is significant room for exploitation of market power, especially in periods of high demand.

There are two approaches being used to forecast market prices. The first is a simulation method based on models of production cost. The second involves application of statistical methods to historical market data.

The production cost method involves a simulation of plant dispatch and inter-pool exchanges to meet hourly demands. These methods typically assume that plants are dispatched based on lowest running cost of the next available generation unit, subject to operating and transmission constraints. This approach requires detailed data and assumptions about inventory of generation plant, including operating capabilities, operating costs, and geographic location with respect to transmission facilities. While these models have proven to be very useful for assessing long-term market options, they are not well suited to the modeling of bidding strategies in a market setting.

In contrast, statistical methods relate market prices to observed factors that are believed to impact prices. These factors can include both demand side and supply side variables, and a variety of model specifications and techniques are available. Since bidding strategies are embedded in the observed market outcomes, these methods will work well as long as strategies, constraints, and market rules remain stable or evolve slowly. In what follows, we look at several approaches to short-term statistical modeling using data for the PJM market.

Data

The dependent variable data is the average on-peak price in the PJM market. The on-peak period is defined to be the hours between 8 am and 11 PM, which is a 16-hour block. Data values are available from April 1998 through the present. The PJM market is currently in transition from a power pool using least-cost dispatch to a competitive market based on generator bidding. At this point PJM prices still reflect dispatch costs more than competitor bidding. Still these prices pose a significant modeling challenge, and it is reasonable to believe that methods that work well with these data will also work well in a full bidding context.

Explanatory variables fall into three categories. First, from a time-series perspective, there is the history of the market price itself. In a day-ahead market, lagged price data often have high explanatory power, although the pattern of weekdays, weekends, and seasons introduces some interesting modeling problems for time-series models.

The second set of explanatory factors is demand-side factors. Hourly electricity use reflects the life patterns of people, mechanical systems interacting with weather, cloud cover, timing of sunrise and sunset, water temperatures, and other similar factors. Because system load can typically be modeled and forecasted with high accuracy, we proceed here using the actual demand levels as an explanatory variable, rather than the indirect variables for weather and calendar effects.

The third set of explanatory factors is supply-side factors. These factors include fuel prices, generation unit availability, transmission constraints, and in some markets, hydro flows. Also, in periods of high demand the load levels in surrounding areas can have a significant impact on local prices, reflecting the high price of imported energy and the high opportunity cost of bidding into the local market. The supply-side variables included here are nuclear capacity on-line and natural gas prices. The data are presented below.

Figure 1 shows the PJM average on-peak price. Over the historical period, the mean value is about $25 per MWh, and most observations are between $15 and $30. On several days, however, the price shows a significant and short-lived spike, with hourly values nearing $1000, bringing the average on-peak price for the day above $100 on occasion. (In the numbers shown here, the hourly price has been capped at $200 before computing the daily average).

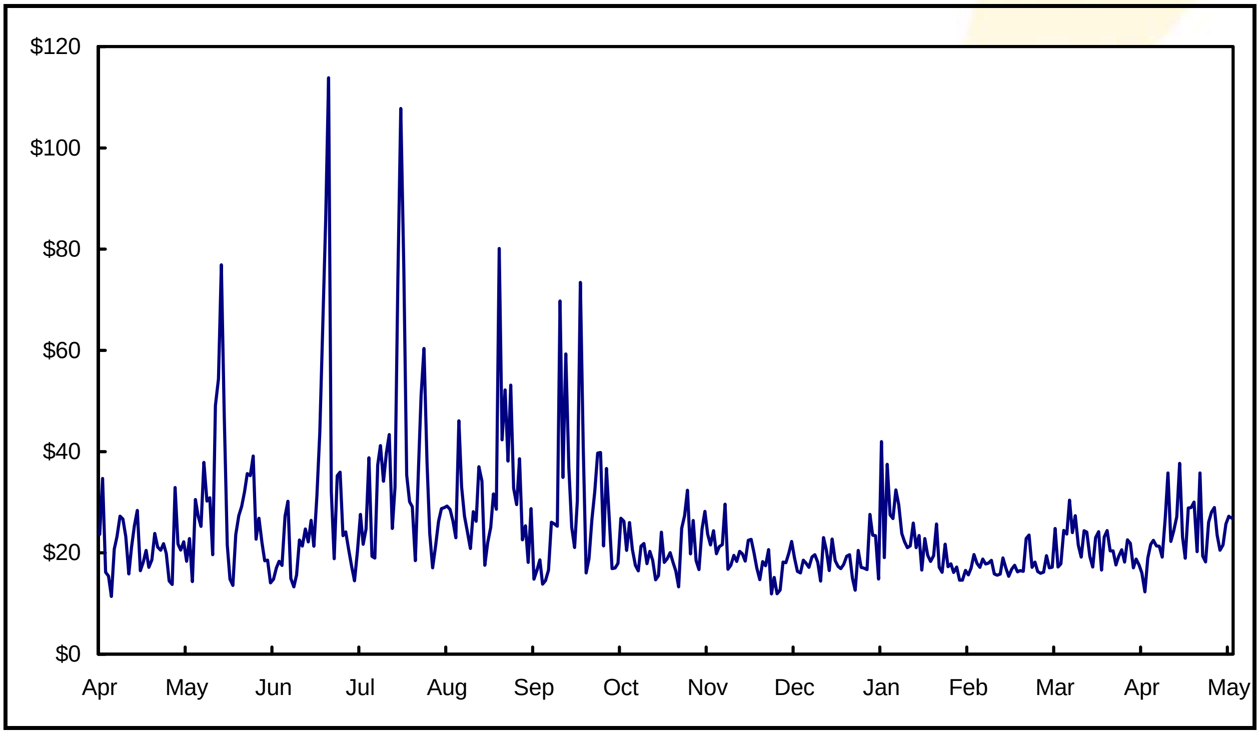

Figure 1. PJM Average On-Peak Price ($/MWh), April 98 to May 99

Figure 2 shows the corresponding values for on-peak energy use. These data have an average value of about 500 billion Watthours (GWh), implying an average load of about 31 GW during on-peak hours. The data show a strong weakly cycle as well as a weather-driven seasonal cycle.

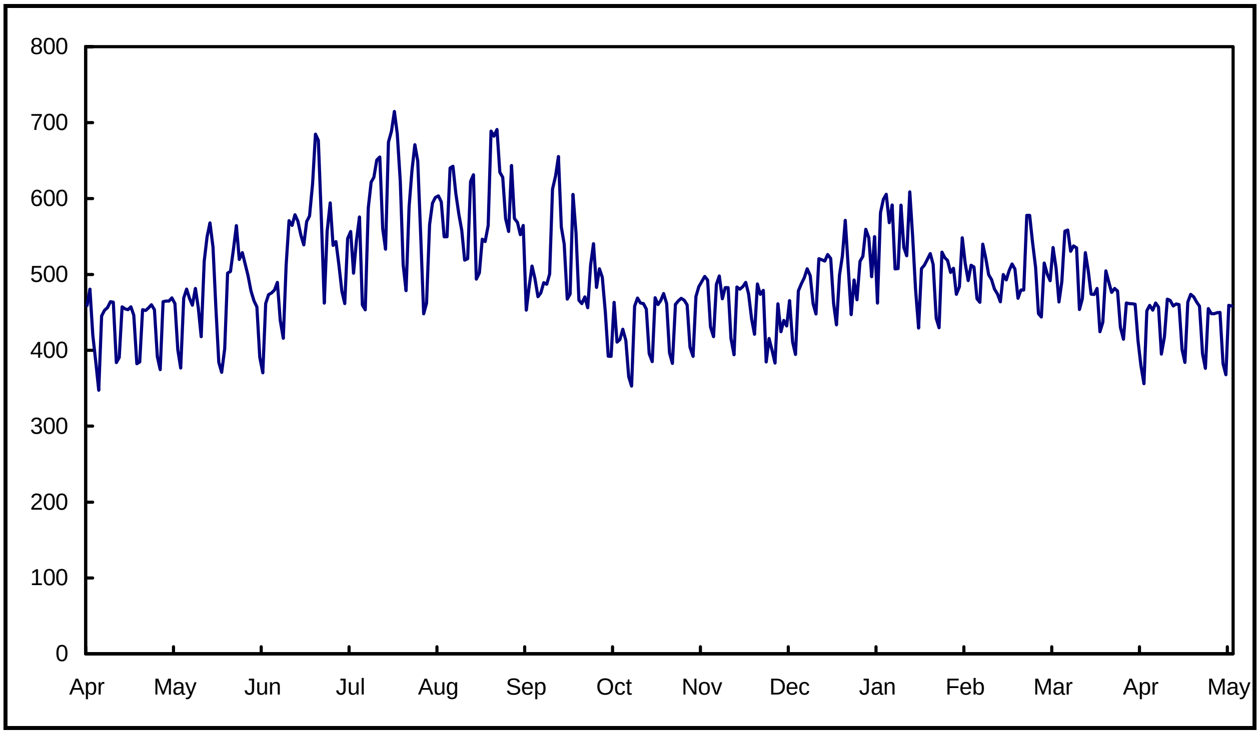

Figure 2. PJM On-Peak Energy Demand (GWh), April 98 to May 99

Figure 3 shows available nuclear capacity measured in million watts (MW). The average value is about 11.5 GW, with a maximum value of about 14 GW. Unit availability reflects both planned and unplanned outages.

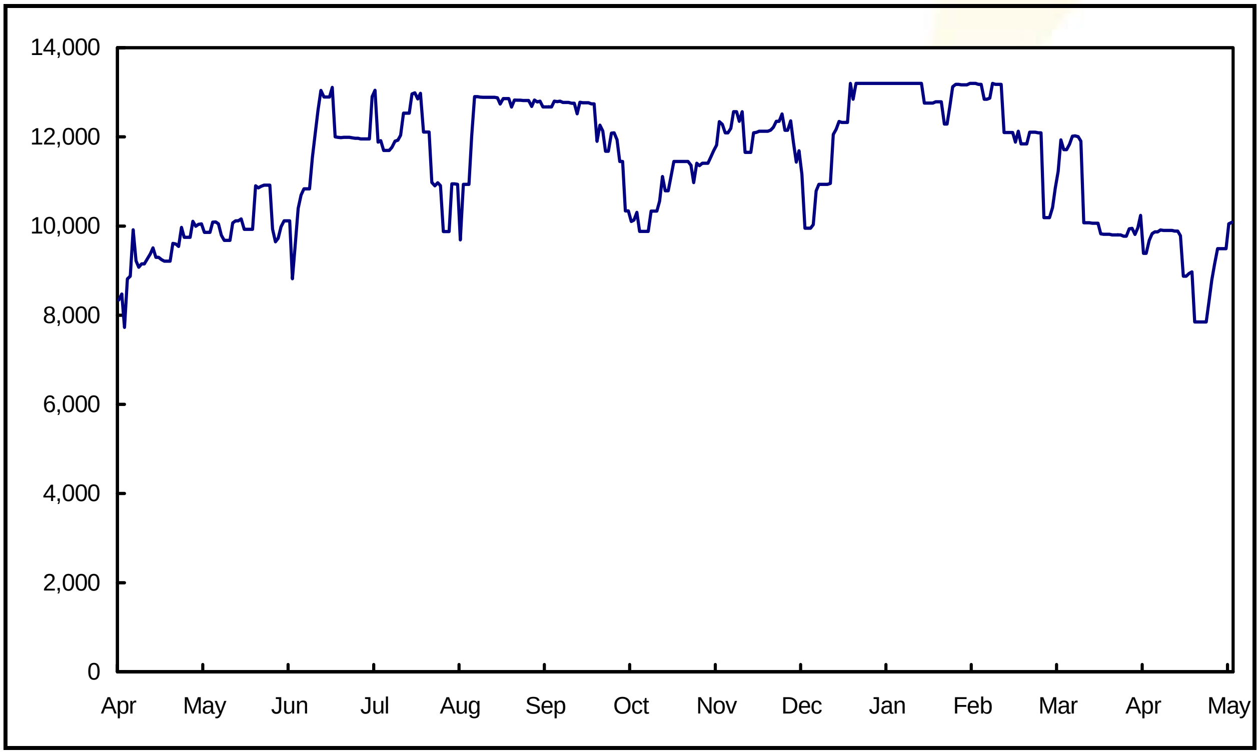

Figure 3. Available Nuclear Capacity (MW), April 98 to May 99

Finally, Figure 4 shows natural gas prices at the Henry Hub. The average price over this period was about $2.00 per mmBtu, which would translate to a marginal fuel cost of about $20 per MWh at a heat rate of 10,000 Btu/KWh.

Figure 4. Gas Prices at Henry Hub ($/mmBtu), April 98 to May 99

Comparing Figures 1 and 2, it is clear that high prices occur in periods of high demand. The relationship is not perfect, however, as shown in Figure 5. This figure provides a scatter plot of on-peak prices versus on-peak loads, coded by type of day. The figure shows that all of the high load and high price days are weekdays. However, not all high-load days have high prices. For example, on days with on-peak energy near 700 GWh, the average on-peak price ranges from $40 to more than $100. Despite this wide dispersion, the chart does suggest that the relationship between loads and price is nonlinear.

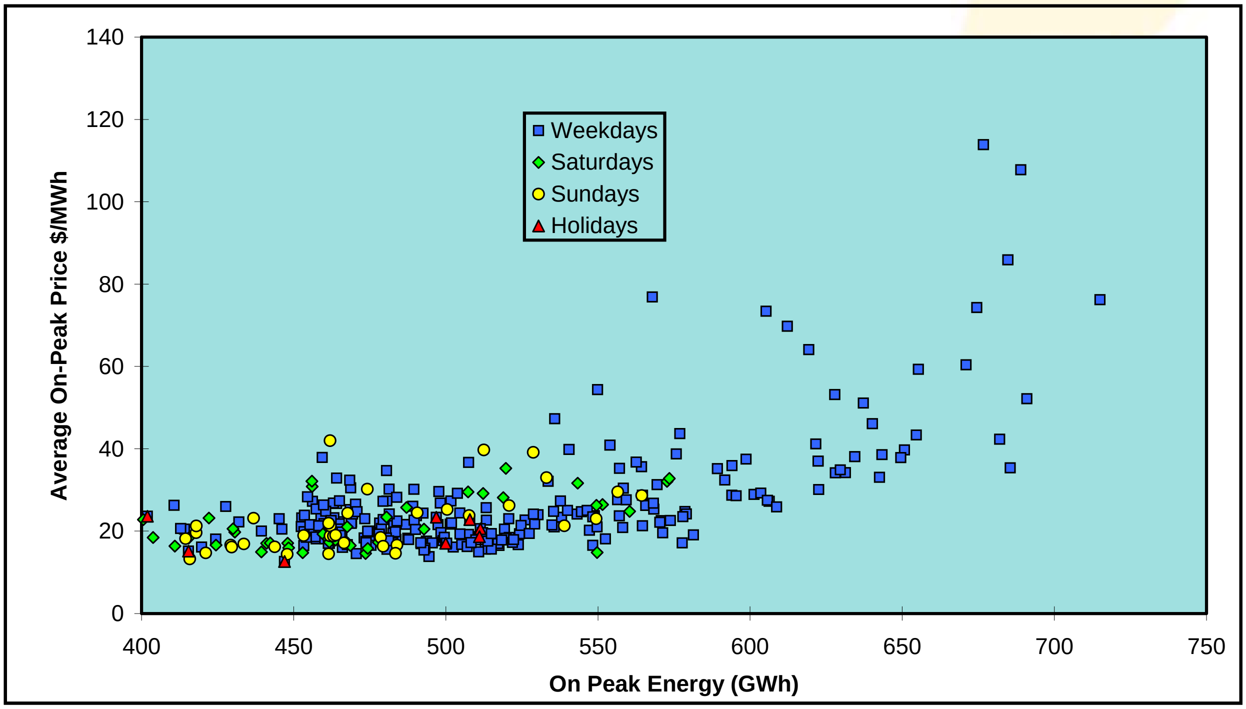

Figure 5: Scatter plot of On-Peak Price vs. Energy

4. Model Specifications

A series of models is estimated using these data. We begin with an exponential smoothing model and an ARIMA model. These data-driven models do not take advantage of the demand and supply-side variables. Next a linear regression model is estimated. Finally, a neural network extension of the linear model is estimated to capture key nonlinearities and interactions.

The forecasting equation for a one-day-ahead forecast with an exponential smoothing model is as follows.

$$P^{t} = \left(\text{Level}^{t-1} + \text{Trend}^{t-1}\right) \times \text{DayMult}^{t-7} \qquad (1)$$

Where P is the value of on-peak prices, t is the current period, and Level, Trend, and DayMult are variables generated by the smoothing process. The smoothing equations use the multiplicative seasonal form, sometimes referred to as the Holt-Winters method. In this application, the seasonal lag is set to seven, so that the DayMult variables point to the same day in the previous week. With this modification, the smoothing equations are as follows:

$$\text{Level}^{t} = a \times \left(\frac{P^{t}}{\text{DayMult}_{t-7}}\right) + (1-a) \times \left(\text{Level}^{t-1} + \text{Trend}^{t-1}\right) \qquad (2)$$

$$\text{Trend}^{t} = b \times \left(\frac{P^{t}}{\text{DayMult}_{t-7}} - \text{Level}^{t-1}\right) + (1-b) \times \left(\text{Trend}^{t-1}\right) \qquad (3)$$

$$\text{DayMult}^{t} = c \times \left(\frac{P^{t}}{\text{Level}^{t}}\right) + (1-c) \times \text{DayMult}^{t-7} \qquad (4)$$

The forecasting equation for a one-day-ahead forecast with a seasonal ARIMA model is as follows:

$$P^t = \text{SARIMA}(p, d, q)(sp, sd, sq) \qquad (5)$$

Where SARIMA represents a seasonal autoregressive (AR) moving average (MA) process with autoregressive terms of order p and sp, moving average terms of order q and sq, differencing of order d, and seasonal differencing of order sd. In this application, the seasonal periodicity is seven days, implying that seasonal lags point to the same day in the preceding week.

The linear regression model is as follows:

$$P^t = c_0 + \sum_j c_j \times X_j^t + e^t \qquad (6)$$

Where X represents a set of demand and supply-side explanatory variables.

The extension of the regression model to include nonlinear nodes from a neural network model can then be represented as:

$$P^t = c_0 + \sum_j c_j \times X_j^t + \sum_{n=1}^{N} \left(B_n \times \frac{1}{1 + e^{-\left(a_{n,0} + \sum_{k=1}^{K} a_{n,k} X_k^t\right)}} \right) + u^t \qquad (7)$$

The first summation in (7) repeats the linear regression from equation (6). The second summation includes a set of N nonlinear nodes from a simple feedforward artificial neural network. The specific form uses logistic transformation functions, which can capture a variety of nonlinear responses. By construction, the variables included in a node are multiplicatively interactive if they have nonzero parameters in the sum that appears in the logistic exponent.

The neural network approach and this specific functional form are widely used in day-ahead forecasting of system loads. The approach is well suited to this problem because the response of system load to weather is nonlinear and because there are significant interactions among explanatory variables. (For a complete discussion of the neural network functional form, see McMenamin and Monforte, 1998). As seen in Figure 5, the relationship between system load and prices also appears to be nonlinear. Further, there may be important interactions between demand and supply variables that help explain daily variations in price. The neural network equation provides a simple way to allow nonlinearities and interactions in the model without imposing restrictive assumptions about the structure of the relationship.

5. Estimation Results

The first model estimated is an exponential smoothing model using a Holt-Winters multiplicative form with trend and seasonal elements. As discussed above, the periodicity of the seasonal term is set to 7 days for this application with daily data. The results are summarized below. This naïve model will serve as a reference point. The mean absolute percent error (MAPE) is 22.6% and the mean absolute deviation is about \$6 per MWh. These statistics reflect the accuracy of the model for purposes of day-ahead forecasting.

Exponential Smoothing Model Summary	
Simple Smoothing Parameter	.619
Trend Parameter	-.006
Seasonal Parameter	.321
Adjusted Observations	397
Deg. of Freedom for Error	394
Adjusted R-Squared	0.379
Std. Error of Regression	9.76
Mean Abs. Dev. (MAD)	6.02
Mean Abs. % Err. (MAPE)	22.57%
Durbin-Watson Statistic	1.816

Table 1: Exponential Smoothing Summary

The second model estimated is an ARIMA model. After examination of time-series diagnostics and experimentation with various forms, the final model is a (0,1,4) (1,0,0), implying that the data are differenced, and a model is fit with an MA4 and a seasonal AR1. As with the smoothing model, the seasonal periodicity is set to 7 days. The coefficients and summary statistics for this model are presented in Table 2. As is evident in Table 2, the ARIMA model provides only a modest improvement with respect to day-ahead accuracy, with a MAPE of 22% and a MAD of $5.8 per MWh.

Variable	Coefficient	StdErr	T-Stat
CONST	-0.007	0.025	-0.265
SAR(1)	0.219	0.052	4.225
MA(1)	-0.380	0.051	-7.460
MA(2)	-0.247	0.052	-4.717
MA(3)	-0.296	0.053	-5.627
MA(4)	-0.031	0.051	-0.595
Summary Statistics			
Adjusted Observations			391
Adjusted R-Squared			0.415
AIC			4.528
BIC			4.589
Std. Error of Regression			9.55
Mean Abs. Dev. (MAD)			5.78
Mean Abs. % Err (MAPE)			22.03%
Durbin-Watson Statistic			1.992

Table 2: ARIMA (0, 1, 4) (1, 0, 0) Summary

The regression model uses a combination of lagged dependent variables and explanatory variables. The three explanatory variables are on-peak energy use (OnPeakEnergy), nuclear availability (NukeAvail), and the price of natural gas (HHPrice). As shown in Table 3, this specification improves the MAPE value to 19.5% and reduces the MAD to about $5 per MWh.

The final specification introduces two nonlinear nodes to the linear model presented above. The first node includes only on-peak energy as an input variable. This node will allow representation of nonlinear effects, to the extent these effects are present. The second node includes the two supply-side variables, gas prices and nuclear availability, allowing the modeling of nonlinearities and interactions with respect to these variables. The extended model is estimated using nonlinear least squares applied to normalized data.

Variable	Coefficient	StdErr	T-Stat
CONST	-28.576	5.896	-4.847
Saturday	-1.060	1.574	-0.673
WkDay	-2.017	1.287	-1.567
Lag1OnPk	0.306	0.050	6.159
Lag2OnPk	-0.021	0.050	-0.419
Lag3OnPk	-0.099	0.050	-1.975
Lag4OnPk	0.095	0.050	1.877
Lag5OnPk	-0.012	0.054	-0.232
Lag6OnPk	-0.032	0.056	-0.582
Lag7OnPk	0.046	0.051	0.890
HHPrice	6.595	1.822	3.620
NukeAvail	-0.002	0.000	-4.974
OnPeakEnergy	0.116	0.011	10.920
Summary Statistics			
Adjusted Observations		392	
Adjusted R-Squared		0.569	
Durbin-Watson Statistic		1.695	
AIC		4.267	
BIC		4.531	
Std. Error of Regression		8.18	
Mean Abs. Dev. (MAD)		5.06	
Mean Abs. % Err. (MAPE)		19.52%	

Table 3: Regression Model Results

As shown in Table 4, this specification provides a further improvement in model accuracy, with day-ahead MAPE values dropping to 17% and MAD values dropping to $4.5 per MWh. The actual and predicted values are presented in Figure 6. As is evident, the model does not fully predict the price spike values that occurred in the summer of 1998. This is not surprising given the price dispersion that is evident for high load levels in the scatter plot in Figure 5. Otherwise, however, the day-ahead model tracks actual outcomes and changes in price fairly well.

Comparing the autoregressive terms, the coefficient on the one-day lag of price (Lag1OnPk) is about half the level in the neural network model as it is in the regression model. This indicates that the neural network model places higher reliance on the explanatory variables and lessor reliance on the time-series properties of the residuals.

Coefficient	Value	StdErr	T-Stat
Linear: Intercept	2.417	1.169	2.068
Linear: Saturday	-0.008	0.039	-0.205
Linear: WkDay	-0.029	0.045	-0.651
Linear: Lag1OnPk	0.162	0.047	3.469
Linear: Lag2OnPk	-0.094	0.045	-2.065
Linear: Lag3OnPk	-0.065	0.045	-1.442
Linear: Lag4OnPk	0.090	0.045	2.021
Linear: Lag5OnPk	0.007	0.048	0.137
Linear: Lag6OnPk	-0.003	0.050	-0.053
Linear: Lag7OnPk	0.013	0.047	0.283
Linear: HHPrice	-0.263	0.230	-1.141
Linear: NukeAvail	0.003	0.165	0.018
Linear: OnPeakEnergy	0.479	0.068	7.086
Node1: Slope	-3.243	0.953	-3.404
Node1: Bias	8.425	2.472	3.408
Node1: OnPeakEnergy	-3.470	1.226	-2.830
Node2: Slope	1.659	1.427	1.163
Node2: Bias	-0.348	0.326	-1.066
Node2: NukeAvail	-0.695	0.431	-1.615
Node2: HHPrice	1.612	1.093	1.475
Summary Statistics			
Adjusted Observations			392
Adjusted R-Squared			0.663
AIC			4.037
BIC			4.372
Std. Error of Regression			7.23
Mean Abs. Dev. (MAD)			4.49
Mean Abs. % Err. (MAPE)			17.15%
Durbin-Watson Statistic			1.619

Table 4: Neural Network Model Results

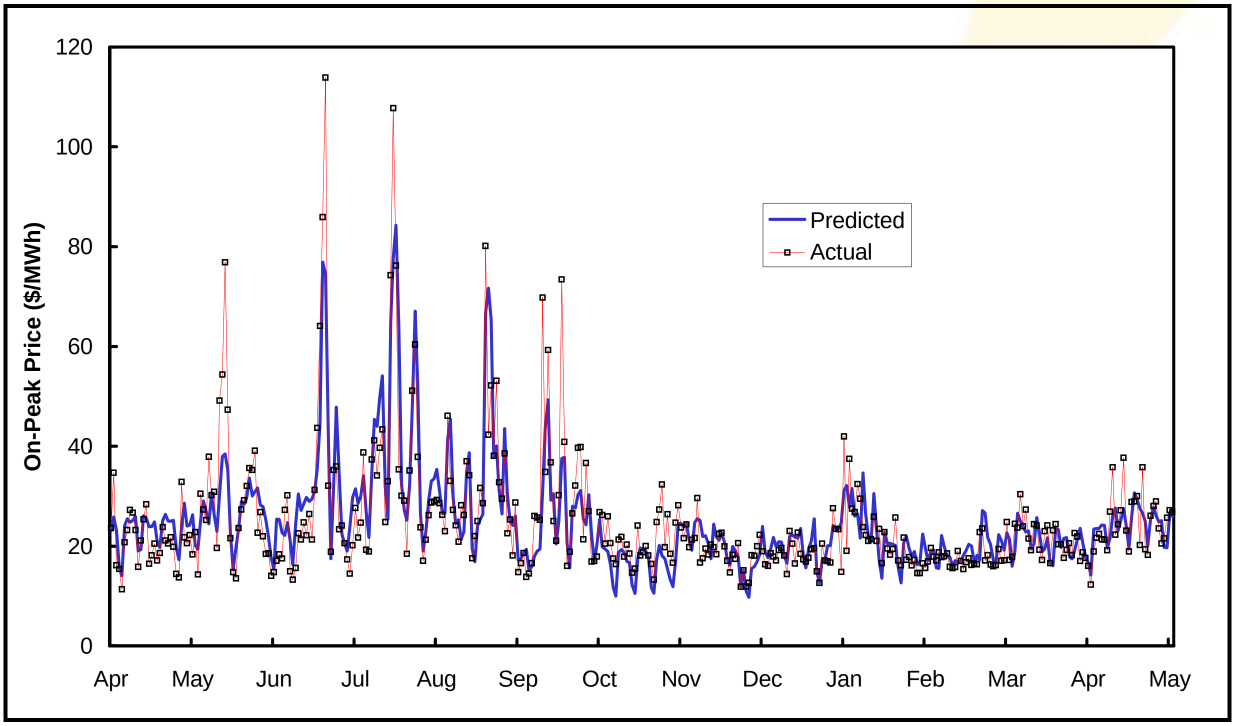

Figure 6: Actual and Predicted Values – Neural Network Model

To understand the role of the parts of the neural network model, the contribution of each component (the linear component and the two nonlinear nodes) to the total predicted value is presented in Figure 7, Figure 8, and Figure 9. The linear node appears to account for most of the seasonal variation and also captures weekly cycles. The first nonlinear node fires only under price-spike conditions. The second nonlinear node is driven mostly by natural gas prices. This node has the greatest contribution when gas prices are high and nuclear capacity is low, as was the case in both the spring of 98 and the spring of 99.

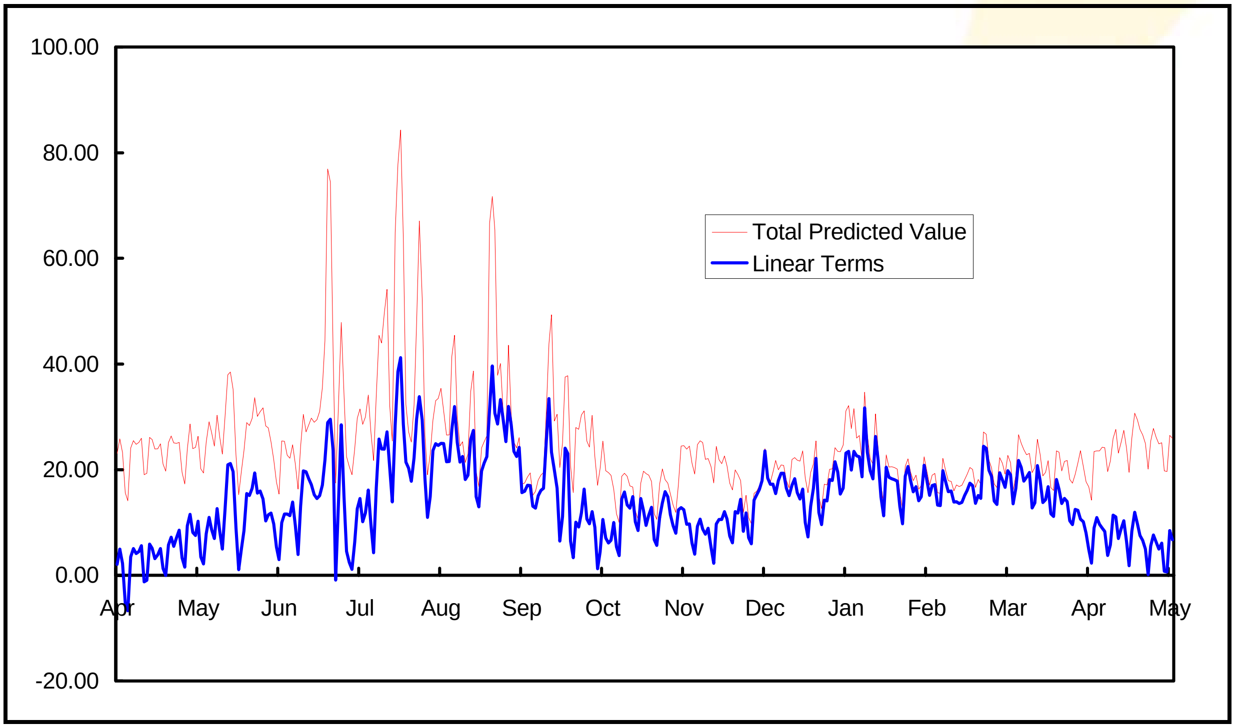

Figure 7. Contribution of Linear Terms to Predicted Value

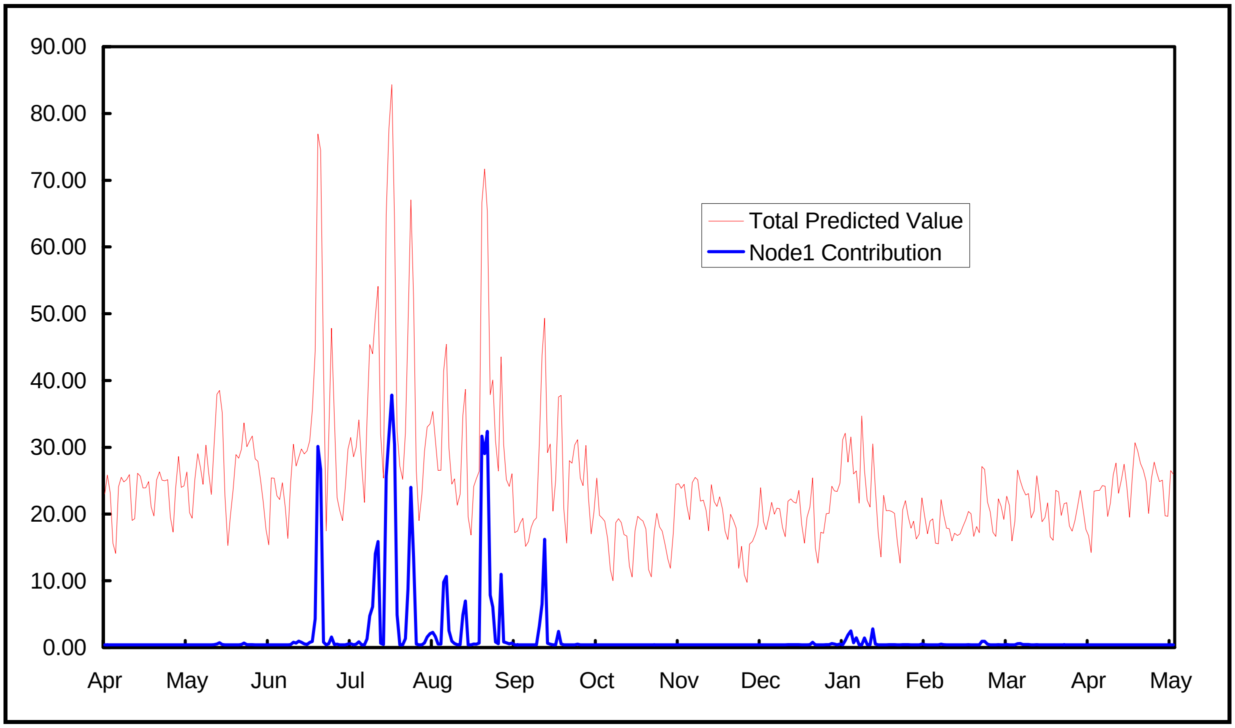

Figure 8. Contribution of Node 1 (OnPeakEnergy) to Predicted Value

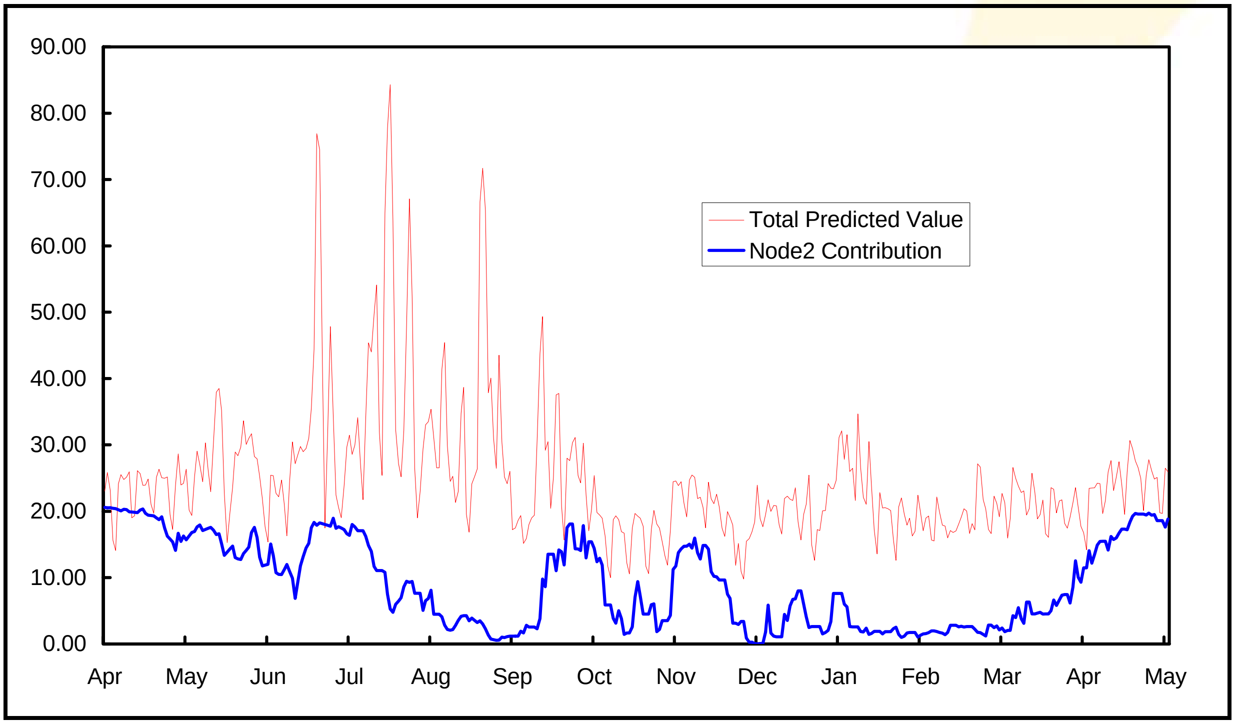

Figure 9. Contribution of Node 2 (Supply Terms) to Predicted Value

6. Conclusion

As competitive electricity markets evolve, interest in price forecasting will increase significantly. Whereas under the regulatory compact, utilities are responsible for prudent planning and are guaranteed a reasonable rate of return on investment decisions, in a competitive generation market, profits will be determined by the relationship between price and cost. Owners of generation assets will want the best possible price forecasts to make the best decisions about contracting and bidding strategies. Retailers will want the best possible price forecasts to develop strategies for covering the loads of their customers. And utilities, which may be both owners of generation and retailers, will be interested in price forecasting for purposes of trading and risk management.

As the analysis above suggests, electricity price forecasting is a significant challenge. Price variation is significant on a day-to-day basis, and prices are even more volatile on an hourly basis. The analysis suggests that more advanced modeling methods will produce better forecasts. Moving from naïve methods to advanced methods, such as neural networks reduces the day-ahead forecasting error from about $6 per MWh to $4.5 per MWh. By developing better data about supply side factors, it is reasonable to expect that further improvements can be made. Because of the nonlinear and interactive nature of the price responses, this is a good application for neural network approaches, which provide a flexible nonlinear method.

References

1. Azoff, E. M. *"Neural Network Time Series Forecasting of Financial Markets,"* John Wiley & Sons, 1994.
2. Kuan, C and H White, *"Artificial Neural Networks, An Econometric Perspective,"* UCSD Discussion Paper, June 1992
3. McMenamin, J. S. and F. A. Monforte, *"Short-term Energy Forecasting With Neural Networks,"* Energy Journal, Volume 19, Number 4, 1998.
4. White, H, *"Neural-Network Learning and Statistics,"* AI Expert, December, 1989.

Author Biographies

J. Stuart McMenamin, Ph.D. Vice President at Itron, where he specializes in the fields of energy economics, statistical modeling, and software development. Over the last 20 years, he has managed numerous projects in the areas of system load forecasting, price forecasting, retail load forecasting, end-use modeling, regional modeling, load shape development, and utility data analysis. In addition to directing large analysis projects, Dr. McMenamin directs the development of Itron's forecasting software products. He has also directed the development of software packages for external clients, including the EPRI end-use models, regional economic forecasting packages, and home energy rating software. Dr. McMenamin received his B.A. in Mathematics and Economics from Occidental College and his Ph.D. in Economics from UCSD.

Frank A. Monforte, Ph.D. is Vice President of Forecasting at Itron, where he is a leading authority in the areas of short-term load forecasting, load profiling, retail scheduling, end-use forecasting, and statistical and mathematical modeling. Dr. Monforte directs the development, support, and implementation of Itron's forecasting and load profiling tools. In addition to his forecasting responsibilities, he is a nationally recognized authority in the area of industrial end-use analysis. Dr. Monforte has co-authored award-winning publications on a range of problems including the use of neural networks for short-term load forecasting, long-term end-use forecasting, and the use of nonlinear programming techniques for development of a least cost gas supply planning tool. Dr. Monforte received his B.A. in Economics from the University of California, Berkeley and his Ph.D. in Economics from the University of California, San Diego.

Christine Fordham is a Senior Consultant in the Forecasting Division, where she specializes in design and implementation of leading-edge forecasting systems. Ms. Fordham has extensive experience with implementing forecasting systems using enterprise relational databases including Oracle 8i and 9i and Microsoft SQL Server. Ms. Fordham has implemented systems to forecast system loads, transmission zone loads, market prices, retail profiles, and utility or provider of last resort loads for utilities in the United States and Canada. She has developed short-term gas forecasting systems for gas companies in Europe and the United States. She has also implemented long-term forecasting systems using end-use models for numerous utilities. Ms. Fordham received her B.S. in Economics from the Massachusetts Institute of Technology in 1988.

Eric Fox is a Vice President of Forecasting at Itron, where he directs forecasting and energy analysis projects and manages Itron's Northeast office. Mr. Fox has over 15 years of forecasting experience with extensive expertise in electric and gas demand forecasting. Mr. Fox is one of Itron's primary instructors. He provides forecast training through workshops sponsored by Itron, utility onsite training programs and workshops held by other organizations including EPRI and the Institute of Business Forecasting. Mr. Fox has also provided testimony and directed regulatory workshops related to forecasting and rate design issues. Mr. Fox received his B.A. and M.A. in Economics from San Diego State University.

Frederick D. Sebold , Ph.D. Vice President at Itron. Dr. Sebold has been actively engaged in the energy industry for over twenty years. His chief areas of expertise are energy demand analysis, energy efficiency program planning and analysis, econometrics, and cost-benefit analysis. His research background includes a wide range of consulting activities for state and local government as well as private industry. Dr. Sebold has directed numerous end-use analysis projects in the residential and commercial sectors. These projects have encompassed both data collection and analysis, and have entailed the use of a wide range of analytical techniques including building simulations and econometric approaches like SAE and conditional demand analysis. Dr. Sebold has also directed a wide range of studies in the area of energy efficiency. These projects have included impact and process evaluation of resource acquisition programs, market assessments and market effects analyses of market transformation programs, and assessments of demand-side management potential. Dr. Sebold has also played a central role in the design of multi-utility energy efficiency programs. Dr. Sebold received his B.A. in Economics from Saint Vincent College, and his M.A. and Ph.D. in Economics from Boston College.

Mark Quan is a Senior Forecast Analyst with Itron's Forecasting Division. Since joining Itron in 1997, Mr. Quan has specialized in energy forecasting and mathematical modeling. Mr. Quan has developed and implemented several automated forecasting systems to predict next day system demand, load profiles, and retail consumption for companies throughout the United States and Canada. Mr. Quan conducts several of Itron's forecasting workshops and customized training sessions. His breadth of forecasting experience in both regulated and deregulated markets makes Mr. Quan ideally suited to teaching practical solutions to difficult forecasting problems. Mr. Quan received an M.S. in Operations Research from Stanford University and a B.S. in Applied Mathematics from the University of California at Los Angeles.

Itron Inc.
Itron is a leading technology provider and critical source of knowledge to the global energy and water industries. Nearly 3,000 utilities worldwide rely on Itron technology to deliver the knowledge they require to optimize the delivery and use of energy and water. Itron delivers value to its clients by providing industry-leading solutions for meter data collection, energy information management, demand response, load forecasting, analysis and consulting services, transmission and distribution system design and optimization, web-based workforce automation, C&I customer care, as well as enterprise and residential energy management.

To know more, start here: www.itron.com

Itron Inc.
Corporate Headquarters
2111 North Molter Road
Liberty Lake, Washington, 99019
U.S.A.
Phone: 1.800.635.5461
Fax: 1.509.891.3355

Itron Inc.
Energy Forecasting - West
11236 El Camino Real
San Diego, California, U.S.A. 92130-2660
Phone: 1.800.755.9585
Fax: 1.858.481.7550

Itron Inc.
Energy Forecasting - East
20 Park Plaza, Suite 910
Boston, Massachusetts 02116-4399
Phone: 1.617.423.7660
Fax: 1.617.423.7664

100670W-02
12/06