Available online at www.sciencedirect.com

ScienceDirect

Energy Conversion and Management 48 (2007) 907–912

ENERGY CONVERSION & MANAGEMENT

www.elsevier.com/locate/enconman

Forecasting electricity market pricing using artificial neural networks

Hsiao-Tien Pao *

Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan, ROC

Received 9 June 2005; received in revised form 8 February 2006; accepted 27 August 2006
Available online 12 October 2006

Abstract

Electricity price forecasting is extremely important for all market players, in particular for generating companies: in the short term, they must set up bids for the spot market; in the medium term, they have to define contract policies; and in the long term, they must define their expansion plans. For forecasting long-term electricity market pricing, in order to avoid excessive round-off and prediction errors, this paper proposes a new artificial neural network (ANN) with single output node structure by using direct forecasting approach. The potentials of ANNs are investigated by employing a rolling cross validation scheme. Out of sample performance evaluated with three criteria across five forecasting horizons shows that the proposed ANNs are a more robust multi-step ahead forecasting method than autoregressive error models. Moreover, ANN predictions are quite accurate even when the length of the forecast horizon is relatively short or long.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Artificial neural network; European energy exchange; Cross validation scheme; Autoregressive error model; Long-term forecasts

1. Introduction

Since the beginning of floating electricity prices, electricity price forecasting has become one of the main endeavors for researchers and practitioners in energy markets. In most commodity markets, the effects of production or supply chain on prices are dampened by surplus storage. By contrast, the electricity market lacks storage for practical purposes, which is an intrinsic source of volatility. The volatility in currency markets makes electricity price forecasting a difficult yet challenging task. To forecast accurately these prices is critical for producers, consumers and retailers. In fact, they must set up bids for the spot market in the short term and define contract policies in the medium term, and in addition, they must define their expansion plans in the long term. For these reasons, all the decisions that each market player must take are strongly affected by price forecasts. Many of these problems can be modeled as mathematical programs. An overview of mathematical programming problems in electricity markets can be found in Conejo and Prieto [1].

Reported techniques to forecast day ahead prices include autoregressive integrated moving average (ARIMA) models [2–4], dynamic regression models [5], other time series techniques [6,7], neural network procedures [8–11] and wavelet transform models [12,13]. Recently, Nogales and Conejo [14] proposed a transfer function model to predict electricity prices based on both past electricity prices and demands. Lu et al. [15] proposed a data mining based electricity price forecast framework, which can predict the normal price as well as the price spikes. Moreover, Conejo et al. [16] present a wavelet transform and ARIMA hybrid model to forecast day ahead electricity prices for the Spanish electricity markets.

Neural network applications for electricity price forecasting have yielded mixed results that may largely by attributed to problems in data selection and sampling variation. Most studies in the literature adopt the practice of arbitrarily splitting available data into a training set for model construction and a test set for model validation.

* Tel.: +886 3 5131578; fax: +886 3 5710102.
E-mail address: htpao@cc.nctu.edu.tw

0196-8904/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enconman.2006.08.016

Cross validation is a re-sampling technique that uses multiple training and test sub-samples. Results from the cross validation analysis will provide valuable insights into the reliability or robustness of neural networks with respect to sampling variation. A rolling validation scheme with increasing length in the training series is used to examine the sampling variation effect.

For long-term forecasting, however, one or more output nodes can be used. If one output node is employed, then the iterative forecasting approach is assumed, and the forecast values are iteratively used as inputs for the next forecasts. On the other hand, if the number of output nodes is equal to the length of the forecasting horizon, then the direct forecasting approach is used in which we forecast the future values directly from the network outputs [17]. The first approach may generate more prediction errors, and the second approach can introduce serious round off errors. This paper proposes an artificial neural network model to predict m daily ahead electricity price on the European Energy Exchange (EEX) market. The characteristic of this model is to employ a single output node structure for m period ahead forecasts using the direct forecasting approach in which we forecast the future values directly from the network outputs. In general, the proposed models can avoid excessive round off and prediction errors. Long-term forecasting is useful for evaluating the robustness of a forecasting technique.

The rest of this paper is organized as follows. Section 2 proposes a neural network model and focuses on long-term forecasts using the direct forecasting approach. Section 3 outlines the research design and the data description. The cross validation scheme and three out of sample performance measures are described. Section 4 presents the empirical findings. Conclusions and some further discussions are given in Section 5.

2. Neural networks for time series forecasting

Neural networks are a class of flexible nonlinear models that can discover patterns adaptively from the data. Theoretically, it has been shown that, given an appropriate number of nonlinear processing units, neural networks can learn from experience and estimate any complex functional relationship. Empirically, numerous successful applications have established their role for pattern recognition and forecasting [18]. Otherwise, time series forecasting linear models assume that there is an underlying process from which data are generated and that the future values of a time series are solely determined by the past and current observations. Neural networks are able to capture the autocorrelation structure in a time series even if the underlying law governing the series is unknown or too complex to describe.

The most popular and successful model is the feed forward multilayer network. Fig. 1 shows a three layer feed forward neural network with a single output unit, k hidden units, n input units. w_{ij} is the connection weight from the ith input unit to the jth hidden unit, and T_j is the connecting weight from the jth hidden unit to the output unit. In its applications, the data series is usually divided into a training set (in sample data) and a test set (out of sample). The training set is used for construction of the neural network, whereas the test set is used for measuring the predictive ability of the model. The training process is used essentially to find the connection weights of the networks.

For a univariate time series forecasting problem, suppose we have N observations $y_1, y_2, \ldots, y_N$ in the training set, $y_{N+1}, y_{N+2}, \ldots, y_{N+m}$ in the test set and we need the *m period ahead* forecasts. In order to avoid excessive round off and prediction errors, this research proposes a network with a single output node and p input nodes by using the direct forecasting approach for *m period ahead* forecasting. The $N-m-p+1$ training patterns in the proposed network are

$$y_{p+m} = f(y_p, y_{p-1}, \ldots, y_1) \quad (1)$$

$$y_{p+m+1} = f(y_{p+1}, y_p, \ldots, y_2) \quad (2)$$

$$\vdots$$

$$y_N = f(y_{N-m}, y_{N-m-1}, \ldots, y_{N-m-p+1}) \quad (3)$$

and the m testing patterns are

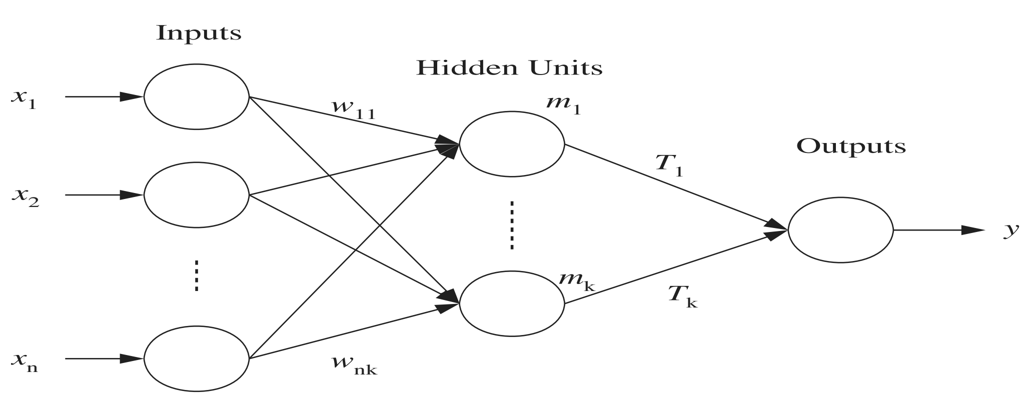

Fig. 1. A three layer feed forward neural network.

$$y_{N+1} = f(y_{N+1-m}, y_{N-m}, \ldots, y_{N-m-p+2}) \quad (4)$$
$$y_{N+2} = f(y_{N+2-m}, y_{N-m+1}, \ldots, y_{N-m-p+3}) \quad (5)$$
$$\vdots$$
$$y_{N+m} = f(y_N, y_{N-1}, \ldots, y_{N-p+1}). \quad (6)$$

The training objective is to find the connection weights such that an overall predictive error means (SSE) is minimized. For this network structure, the SSE can be written as:

$$\text{SSE} = \sum_{i=p+m}^{N} (y_i - \hat{y}_i), \quad (7)$$

where $\hat{y}_i$ is the output from the network. The number of input nodes p corresponds to the number of lagged observations used to discover the underlying pattern in a time series. Too few or too many input nodes can effect either the learning or predictive capability of the network. Experimentation with a pilot sample is often used to find the appropriate numbers of hidden and input nodes.

3. Research methodology

This study focuses on the out of sample performance of the proposed neural networks in electricity prices at the EEX. As discussed earlier, multi-step ahead predictions are modeled. More specifically, the following research questions are addressed: (1) How robust is the neural network out of sample performance with respect to sampling variation and time frame? (2) What is the forecasting ability of neural networks in the long and short forecast horizons? (3) What is the out of sample performance of neural networks relative to the linear time series models such as autoregressive error models (AUTOREG) [19]?

To answer these questions, first, we employ a 15-fold rolling cross validation scheme with increasing length of moving series to deduct the sampling variation effects. Second, five different length forecast horizons with three performance measures are utilized in this study. Finally, AUTOREG models are applied to the data series and the out of sample results are compared to those of neural networks.

3.1. Data analysis

In Germany, the European Energy Exchange based in Leipzig provides day ahead prices for electricity and also forward contracts with varying maturities. Data on prices can be downloaded from their respective websites: www.eex.de. The electricity prices of the Phelix Base at EEX applied here are 1096 daily data recorded in November 2002 through October 2005. Fig. 2 plots this time series. Fig. 3 presents the autocorrelation function (ACF) for this working series. Looking at the ACF, we see that the ACF has spikes at lags 1, 7 and multiples of 7. This implies that the time series observations separated by lags of multiples of 7 time units have a strong positive linear relationship. The information can be used to build linear and nonlinear forecasting models.

3.2. Cross validation

Because the sample autocorrelations for electricity price are statistically large at lags of multiples of 7, we choose

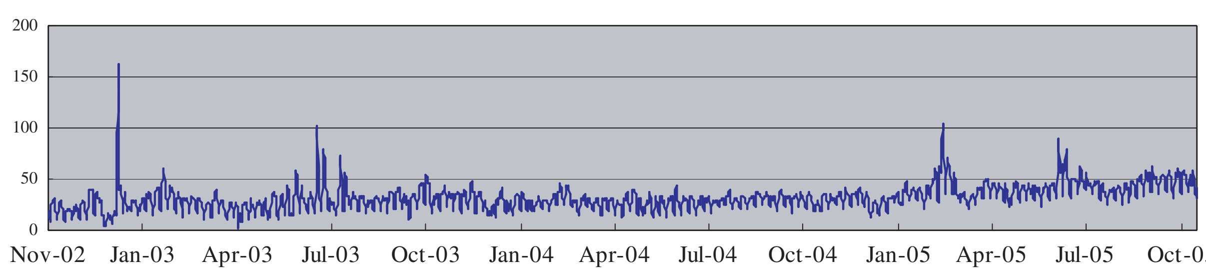

Fig. 2. Day ahead prices for electricity from Dec-2002 to Dec-2004 at EEX.

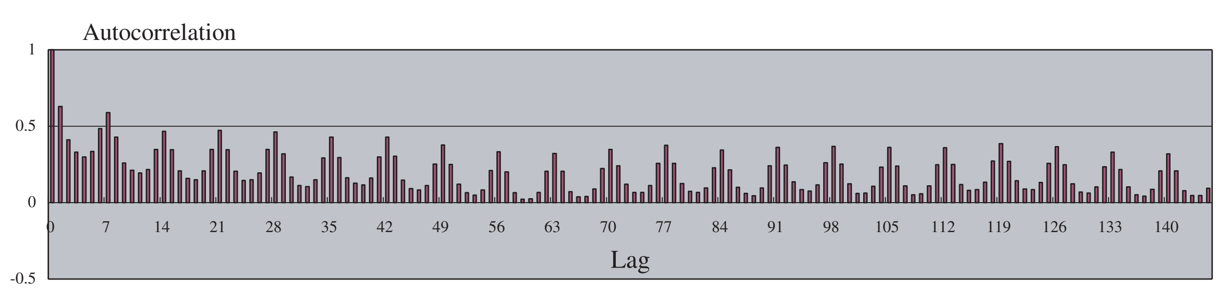

Fig. 3. The autocorrelation function (ACF) for daily electricity prices time series.

y_{t-m}, y_{t-m-7}, y_{t-m-14}, y_{t-m-21} and y_{t-m-28} as the input nodes and y_t as the corresponding output node to build m period ahead forecasting models. The training and testing patterns for the *m period ahead* forecasting neural networks are proposed in the following equations: training pattern:

$$y_t = f(y_{t-m-28}, y_{t-m-21}, y_{t-m-14}, y_{t-m-7}, y_{t-m}), \quad \text{and} \quad t = m + 28 + 1, \cdots, N, \tag{8}$$

testing pattern:

$$y_p = f(y_{p-m-28}, y_{p-m-21}, y_{p-m-14}, y_{p-m-7}, y_{p-m}), \quad \text{and} \quad p = N + 1, \cdots, N + m \tag{9}$$

where $W = N + m$ is the length of the time series in each subset and $N - m - 28$ is the number of training data. Assume the length of the available time series is always longer than the selected time period W. Then, we employ a cross validation scheme to evaluate the performance of the proposed *m period ahead* forecasting structure in Eqs. (3), (6) and (9). A rolling cross validation method with 15 test folds is utilized (see Fig. 4). By rolling, we mean that the in sample periods are expanded sequentially starting from observation one. Each successive fold has 45 observations added. For example, the first in sample period is from y_1 to y_{466-m}, the second from y_1 to y_{511-m} and the last from y_1 to y_{1096-m}. The first out of sample period is from $y_{466-m+1}$ to y_{466}, the second from $y_{511-m+1}$ to y_{511} and the last from $y_{1096-m+1}$ to y_{1096}. Daily observation y_{m+28+1} is the first training output data. For out of sample, we adopt five different time horizons of $m = 7$, 14, 21, 28 and 91 days to examine the effect of forecast horizon. Table 1 shows the training and testing output data of the 15 test folds under different forecasting horizons. To avoid the effects of sampling variation for the out of sample performance, the averages and standard deviations of RMEs, MAEs, and MAPEs of the 15 forecasting periods are compared with autoregressive error models.

3.3. ANN and AUTOREG models

Three layer feed forward neural networks are used to forecast the electricity price. We use node biases except for the input nodes. For m period ahead forecasting, one output node with the training and testing patterns in Eq. (9) is deployed using the direct forecasting approach. We choose y_{t-m}, y_{t-m-7}, y_{t-m-14}, y_{t-m-21}, y_{t-m-28} as the input nodes and y_t as the corresponding output node, where m is the forecasting horizon. Five forecasting horizons, $m = 7$, 14, 21, 28 and 91, are examined in this study. The number of input nodes and the number of hidden nodes are not specified a priori. More than 50 experiments are conducted to determine the best combination of learning rates, momentum, number of input nodes and the number of hidden nodes. Throughout the training, the NeuralWare [20] utility, 'SAVEBEST' is used to monitor and save the lowest root mean square (RMS) error from the training set. The best RMS error results are obtained by using a learning rate of 0.08, a momentum rate of 0.1, 5 input nodes: y_{t-m}, y_{t-m-7}, y_{t-m-14}, y_{t-m-21}, y_{t-m-28} and 7 nodes in a single hidden layer that uses the generalized data learning rule and a sigmoid transfer function ($y = 1/(1 + e^{-x})$). The best architecture of the networks is {5:7:1}. The results are reported in Table 2.

We also apply traditional autoregressive error models to the electricity price data. When time series data are used in regression analysis, often the error term is not independent through time. If the error terms are autocorrelated, the efficiency of ordinary least square parameter estimates is adversely affected and standard error estimates are biased. The autoregressive error model corrects for serial correlation. The AUTOREG model equation is

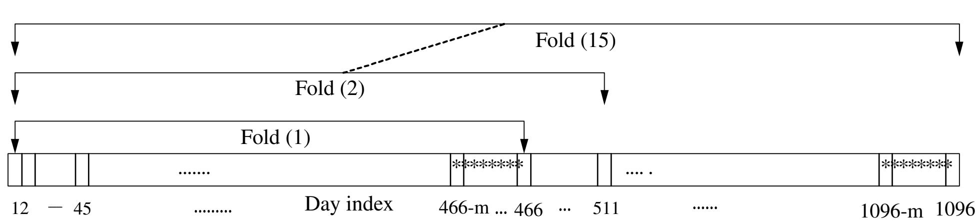

Fig. 4. Rolling cross validation scheme (*: out of sample under different fold).

Table 1
The training and testing output data of 15 forecasting periods for each forecasting horizon

Forecasting horizon	Fold 1		Fold 2			Fold 15	
	Training	Testing	Training	Testing	...	Training	Testing
One week	$y_{36} \cdots y_{459}$	$y_{460} \cdots y_{466}$	$y_{36} \cdots y_{504}$	$y_{505} \cdots y_{511}$	...	$y_{36} \cdots y_{1089}$	$y_{1090} \cdots y_{1096}$
Two weeks	$y_{43} \cdots y_{452}$	$y_{553} \cdots y_{466}$	$y_{43} \cdots y_{497}$	$y_{537} \cdots y_{511}$	...	$y_{43} \cdots y_{1082}$	$y_{1083} \cdots y_{1096}$
Three weeks	$y_{50} \cdots y_{445}$	$y_{446} \cdots y_{466}$	$y_{50} \cdots y_{490}$	$y_{530} \cdots y_{511}$	...	$y_{50} \cdots y_{1075}$	$y_{1076} \cdots y_{1096}$
Four weeks	$y_{57} \cdots y_{438}$	$y_{439} \cdots y_{466}$	$y_{57} \cdots y_{483}$	$y_{523} \cdots y_{511}$	...	$y_{57} \cdots y_{1068}$	$y_{1069} \cdots y_{1096}$
Three months	$y_{120} \cdots y_{375}$	$y_{376} \cdots y_{466}$	$y_{120} \cdots y_{420}$	$y_{421} \cdots y_{511}$	...	$y_{120} \cdots y_{1005}$	$y_{1006} \cdots y_{1096}$

Table 2
Out of sample comparison between neural networks and AUTOREG models

Forecasting horizon	One week		Two weeks		Three weeks		Four weeks		Three months	
	ANN	AUTO	ANN	AUTO	ANN	AUTO	ANN	AUTO	ANN	AUTO
The averages of RMSE, MAE, and MAPE for 15 forecasting periods										
RMSE	3.21	5.70	3.16	6.26	3.25	7.50	3.13	7.73	3.89	9.17
MAE	2.71	4.79	2.47	5.18	2.38	6.15	2.59	6.19	3.06	6.96
MAPE	9.02	15.16	8.24	15.24	8.59	16.39	8.36	16.98	8.85	20.55
The standard deviations of RMSE, MAE, and MAPE for 15 forecasting periods										
RMSE	1.25	2.51	1.13	3.22	1.20	5.45	1.16	5.11	1.12	4.56
MAE	1.09	2.09	1.01	2.78	1.05	4.28	1.02	3.85	0.98	3.17
MAPE	3.18	7.59	2.41	6.04	2.33	6.55	2.30	5.97	2.05	4.42

$$y_t = c + v_t \tag{10}$$

$$v_t = -\psi_1 v_{t-1} - \psi_2 v_{t-2} - \cdots - \psi_m v_{t-m} + \varepsilon_t \tag{11}$$

$$\varepsilon_t \sim \mathrm{IN}(0, \sigma^2). \tag{12}$$

The notation $\varepsilon_t \sim IN(0,\sigma^2)$ indicates that each ε_t is normally and independently distributed with mean 0 and variance σ^2. The cross validation method with 15 forecasting periods is the same as in the neural network model. The averages and standard deviations of RMSEs, MAEs, and MAPEs of the 15 forecasting periods for each forecasting horizon are compared with those of the ANN models.

3.4. Forecasting evaluation methods

For the purpose of evaluating out of sample forecasting capability, we examine the forecasting accuracy by calculating three different evaluation statistics: the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). They are expressed in the following:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} (P_i - A_i)^2 / n} \tag{13}$$

$$\mathrm{MAE} = \sum_{i=1}^{n} |P_i - A_i| \tag{14}$$

$$\mathrm{MAPE} = \sum_{i=1}^{n} |(P_i - A_i)/A_i| * 100 \tag{15}$$

where P_i and A_i are the ith predicted and actual values, respectively, and n is the total number of predictions.

4. Empirical results

The electricity prices at the EEX applied here are 1096 daily data recorded in 2002 through 2005. For long-term forecasting, in order to avoid excessive round off and prediction errors, this research proposes a neural network with a single output node and p input nodes by using the direct forecasting approach. The cross validation method with 15 forecasting periods is used to deduce the sampling variation effect. Three performance measures are utilized to investigate the forecasting ability in long and short forecast horizons. Additionally, the SAS-AUTOREG procedures [19] are used to select available autoregressive error models to the 15 test folds under a different forecasting horizon. It is not surprising to find that lags 1, 7, 12, 14, 21 and 28 are included in all of the models. The averages and standard deviations of RMSEs, MAEs and MAPEs for the 15 forecasting periods under different forecasting horizons are compared with those of the ANN models. Table 2 shows the results. Several observations can be made from it. First, ANNs perform better than autoregressive error models for out of sample forecasting capability. For each forecast horizon, the average values of the three evaluation statistics of the 15 forecasting periods for AUTOREG are larger than those of the ANN models. Forecasting accuracy decreased as forecasting horizon extended for the AUTOREG model. For the three month forecast, neural networks still have smaller average values of the three statistics, but the autoregressive model does not. These results may be expected since the linear model is not available for longer prediction periods.

Second, neural networks have smaller sampling variation effects than the autoregressive model under different forecasting horizons. Sampling variations become smaller for long forecast horizons. Finally, for m period ahead forecasting, the proposed ANNs use the direct forecasting approach with a single output node structure to avoid excessive round off and prediction errors. The model input nodes are y_{t-m}, y_{t-m-i}, y_{t-m-j}, ..., where the autocorrelation for time series $\{y_t\}$ are statistically large at lags $m+i$, $m+j$. Overall, depending on the rolling cross validation scheme, the proposed long-term forecasting ANNs are better than the autoregressive error models.

5. Conclusions

This paper proposes an ANN with a single output node structure to forecast m daily ahead electricity price using the direct forecasting approach. The model input nodes are y_{t-m}, y_{t-m-7}, y_{t-m-14}, y_{t-m-21}, y_{t-m-28} and the corresponding output node is y_t, where the autocorrelation for time series $\{y_t\}$ are statistically large at lags $m+7$,

$m + 14$, $m + 21$, $m + 28$. For m period ahead forecasting, other techniques use an iterative approach with one output node structure or the direct approach with m output nodes structure. However, both approaches may entail more prediction and round off errors. In general, the proposed models can avoid excessive round off and prediction errors. We investigated the potential of neural network models utilizing a rolling cross validation scheme. Out of sample performance with five forecast horizons is evaluated along three criteria, RMSE, MAE and MAPE.

From the case studies analyzed, we noticed that the forecasting accuracy of the neural networks is not very sensitive to the length of forecast horizon, but the autoregressive error models are. For the three month forecast, neural networks still have smaller average values of the three statistics, but the autoregressive models do not. These results may be expected since the linear model is not available for longer prediction periods. Additionally, neural networks have smaller sampling variation effects than autoregressive models under different forecasting horizons. Statistic MAPE notices that sampling variations become smaller as the forecast horizon extends. Furthermore, the proposed neural networks have a better ability to learn from data patterns in the training time period and successfully predict m period ahead outcomes for electricity prices on the EEX. Overall, depending on the rolling cross validation scheme, ANNs are better than autoregressive error models for long-term forecasting.

One fruitful area for further research is a combined methodology of multivariate linear models such as dynamic regression and neural networks. It is suspected that electricity prices time series contain a linear and nonlinear component. Since it has been well established in the literature that Box–Jenkins types of models are particularly effective for linear patterns, whereas neural networks are the preferred models for nonlinear patterns, a combination approach should produce even better results than either linear or nonlinear approach used singly.

References

[1] Conejo AJ, Prieto FJ. Mathematical programming and electricity markets. Top 9 2001;1:1–54.

[2] Box GEP, Jenkins GM. Time series analysis: forecasting, and control. San Francisco, CA: Holden-Day; 1976.

[3] Contreras J, Espinola R, Nogales FJ, Conejo AJ. ARIMA models to predict next-day electricity prices. IEEE Trans Power Syst 2003;18:1014–20.

[4] Fosso OB, Gjelsvik A, Haugstad A, Birger M, Wangensteen I. Generation scheduling in a deregulated system. IEEE Trans Power Syst 1999;14:75–81.

[5] Nogales FJ, Contreras J, Conejo AJ, Espinola R. Forecasting next-day electricity prices by time series models. IEEE Trans Power Syst 2002;17:342–8.

[6] Obradovic Z, Tomsovic K. Time series methods for forecasting electricity market pricing. In: Proceedings of the IEEE power Eng society summer meeting, vol. 2; 1999. p. 1264–5.

[7] Crespo J, Hlouskova J, Kossmeier S, Obersteiner M. Forecasting electricity spot prices using linear univariate time series models. Appl Energy 2002;77:87–106.

[8] Szkuta BR, Sanabria LA, Dillon TS. Electricity price short-term forecastin using artificial neural networks. IEEE Trans Power Syst 1999;14:851–7.

[9] Ramsay B, Wang AJ. A neural network based estimator for electricity spot-pricing with particular reference to weekend and public holidays. Neurocomputing 1998;23:47–57.

[10] Zhang L, Luh PB, Kasiviswanathan K. Energy clearing price prediction and confidence interval estimation with cascaded neural networks. IEEE Trans Power Syst 2003;18:99–105.

[11] Rodriguez CP, Anders GJ. Energy price forecasting in the Ontario competitive power system market. IEEE Trans Power Syst 2004;19:366–74.

[12] Yao SJ, Song YH. Prediction of system marginal prices by wavelet transform and neural network. Elect Mach Power Syst 2004;19:983–93.

[13] Kim CI, Yu IK, Song YH. Prediction of system marginal price of electricity using wavelet transform analysis. Energ Convers Manage 2002;43:1839–51.

[14] Nogales FJ, Conejo AJ. Electricity price forecasting through transfer function models. J Oper Res Soc 2005:1–7.

[15] Lu X, Dong ZY, Li X. Electricity market price spike forecast with data mining techniques. Elect Power Syst Res 2005;73:19–29.

[16] Conejo AJ, Plazas MA, Espinola R, Molina AB. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans Power Syst 2005;20:1035–42.

[17] Hu MY, Zhang G, Jiang CX, Patuwo BE. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. Decision Sci 1999;30:197–216.

[18] Zhang G, Qi M. Neural network forecasting for seasonal and trend time series. Eur J Oper Res 2005;160:501–14.

[19] SAS/ETS User's Guide, Version 8. Cary, NC: SAS Institute Inc.; 1999.

[20] Neural Ware. Neural computing: neural works professional II/PLUS and neuralworks explorer. NeuralWare; 1993.