DAY-AHEAD ELECTRICITY PRICE FORECASTING MODELS BASED ON TIME SERIES ANALYSIS

Javier Contreras Sanz [1p], Rosario Espínola Vílchez[2], Francisco J. Nogales Martín[3] and Antonio J. Conejo[2]

[1]Project Engineering

[2]Electrical Engineering

[3]Operations Research

Escuela Técnica Superior de Ingenieros Industriales

University of Castilla – La Mancha

13071 Ciudad Real, Spain

RESUMEN

Este artículo presenta y compara tres herramientas de predicción de precios para el mercado eléctrico diario: regresión dinámica, función de transferencia y modelo ARIMA estacional. Los tres procedimientos se basan en el análisis de series temporales y difieren en el modelado de la relación entre los precios y los errores (errores cuantificados como la diferencia entre el precio real y el predicho). El modelo de regresión lineal relaciona el precio actual con los valores anteriores de los precios. La función de transferencia relaciona el precio actual con los valores anteriores de precios y demandas. Dicha relación puede incluir un error correlado que se puede modelar con un procedimiento ARMA. Finalmente, el tercer modelo relaciona precios actuales con precios pasados y errores actuales con errores pasados. Casos reales del mercado español y californiano ilustran y comparan el comportamiento predictivo de los modelos.

ABSTRACT

This paper presents and compares three price forecasting tools for day-ahead electricity markets: dynamic regression, transfer function and seasonal Auto Regressive Integrated Moving Average models. The three procedures are based on time series analysis and differ when modelling the relationship between prices and error terms (error measured by the difference between the actual price and the one predicted by the model). The dynamic regression model relates the current price to the values of past prices. The transfer function approach relates the current price to the values of present and past prices and demands. This relationship can include a serially correlated error that can be modelled by an Auto Regressive Moving Average process. Finally, the third model relates current prices to the values of past prices, and current error terms to previous errors. Real world case studies from mainland Spain electricity market are presented to illustrate and compare the predictive behaviour of the models.

1. INTRODUCTION

The electric power industry is becoming more sophisticated after a few years of deregulation and restructuring. Electricity markets have emerged from the previous centralized operation in order to supply energy to consumers with the target of attaining high reliability and low cost.

Electricity price forecasting has become an essential tool in competitive electricity markets, both for producers and consumers. The reason is that buying and selling bidding strategies rely on next day price predictions in order to achieve benefit or utility maximization (for buying or selling agents, respectively). In addition, reliable price forecasts have a definitive impact not only in electricity day-ahead markets, but also in monthly schedules and bilateral or financial contracts. For this type of portfolio decisions, it is desirable to have available predictions of price average values over a year horizon.

Currently, there are not many applications of time series forecasting to predict day-ahead electricity prices. The most important ones are Auto Regressive Integrated Moving Average (ARIMA) methods and Artificial Neural Networks (ANN) [Ramsay and Wang 1998, Szkuta et al. 1999]. Ramsay proposes a hybrid fuzzy logic-neural network approach to predict prices in the England-Wales pool, with daily mean errors around 10%. And Szkuta proposes a three-layered ANN using back-propagation in the Victorian (Australia) electricity market, with daily mean errors around 15%.

This paper focuses on short-term decisions associated to the pool and presents three highly effective tools to predict day-ahead prices: dynamic regression, transfer function and seasonal Auto Regressive Integrated Moving Average (ARIMA) models. They are based on time series analysis and applied to forecast actual prices of mainland Spain.

The remainder of the paper is organized as follows. In section 2, a mathematical description of the three models is provided. Section 3 presents numerical results and section 4 states some conclusions.

2. DESCRIPTION OF THE MODELS

In this section, the description of three models based on time series analysis is presented: dynamic regression, transfer function and ARIMA formulations. These three models are a class of stochastic processes used to analyze time series, and have a common methodology. The application of this methodology to the study of time series analysis is due to Box and Jenkins [Box and Jenkins 1976].

The analysis is based on setting up a hypothetical probability model, one for each of the proposed models, to represent the data. The models presented are selected based on a careful inspection of the main characteristics of the hourly price series. In most competitive electricity markets this series presents: high frequency; non-constant mean and variance; multiple seasonality (corresponding to a daily and weekly periodicity, respectively); calendar effect (such as weekends, holidays); high volatility; and high percentage of unusual prices (mainly in periods of high demand).

Moreover, the proposed models can include explanatory variables, for example, the demand of electricity has been included in one model because, a priori, it seems to partly explain the price behavior.

Next, the description of the general statistical methodology to build a final model is presented. The three models have been obtained through the following scheme:

- *Step 0.* A class of models is formulated assuming certain hypotheses.
- *Step 1.* A model is identified for the observed data.
- *Step 2.* The model parameters are estimated.
- *Step 3.* If the hypotheses of the model are validated go to Step 4, otherwise go to Step 1 to refine the model.
- *Step 4.* The model can be used to forecast.

In the following subsections, each step of the above scheme is detailed.

2.1. Step 0

The building of each of the three models presented differs in this step. This step is explained below depending on the selected model.

1) Dynamic regression:

The first proposed method to forecast prices is a dynamic regression model [Nogales et al. 2002]. In this model, the price at hour *t* is related to the values of past prices at hours *t* -1, *t* -2, …, etc. This is done to obtain a model that has uncorrelated errors.

In Step 0, the selected model used to explain the price at hour *t* is the following:

$$p_t = c + w^p(B)p_t + e_t \tag{1}$$

where p_t is the price at time *t* and *c* is a constant. Function $w^p(B) = \sum_{l=1}^{K} w_l^p B^l$ is a polynomial function of the backshift operator $B: B^l p_t = p_{t-l}$, where the total number of terms of the function, *K*, is subject to change in steps 1-3. Function $w^p(B)$ depends on parameters w_l^p, whose values are estimated in Step 1. Finally, e_t is the error term. In Step 0 this term is assumed to be a series drawn randomly from a normal distribution with zero mean and constant variance s^2, that is, a white noise process. The efficiency of this approach depends on the election of the appropriate parameters in $w^p(B)$ to achieve an uncorrelated set of errors. This selection is carried out through Steps 1 to 3 as it is explained below.

2) Transfer function:

A second proposed method that includes a serially correlated error is called transfer function model [Nogales et al. 2002]. Specifically, it is assumed that the price and demand series are both stationary (i.e. with constant mean and variance). The general form proposed to model the (price, demand) transfer function is

$$p_t = c + w^d(B)d_t + N_t \tag{2}$$

where p_t is the price at time *t*, *c* is a constant, d_t is the demand at time *t*, $w^d(B) = \Sigma_{l=0}^{K} w_l^d B^l$ is a polynomial function of the backshift operator, and N_t is a disturbance term that follows an ARMA model of the form

$$N_t = \frac{q(B)}{f(B)} e_t \tag{3}$$

with $q(B) = 1 - \Sigma_{l=1}^{\Theta} q_l B^l$ and $f(B) = 1 - \Sigma_{l=1}^{\Phi} f_l B^l$, both of which being polynomial functions of the backshift operator. The total number of terms of the functions $q(B)$ and $f(B)$, Θ and Φ, respectively, are subject to change in steps 1-3. Finally, e_t is the error term, that is assumed to be a white noise process. The model in (2) relates actual prices to demands through function $w^d(B)$ and actual prices to past prices through function $f(B)$.

3) ARIMA:

The third proposed method is an ARIMA formulation. The proposed general ARIMA formulation in Step 0 is the following:

$$f(B)p_t = q(B)e_t \tag{4}$$

where $f(B)$ and $q(B)$ are functions of the backshift operator such as in (3), and e_t is the error term. But in this case, functions $f(B)$ and $q(B)$ have special forms. They can contain factors of polynomial functions of the form $f(B) = 1 - \Sigma_{l=1}^{\Phi} f_l B^l$ and/or $q(B) = 1 - \Sigma_{l=1}^{\Theta} q_l B^l$, and/or $(1 - B^S)$, where several values of f_t and q_t can be set to 0.

The total number of terms of the functions $q(B)$ and $f(B)$, Θ and Φ, respectively, are subject to change in steps 1-3. It can be observed that the model in (4) relates actual prices to past prices through function $f(B)$, and actual errors to past errors through function $q(B)$.

Finally, certain hypotheses on the three models must be assumed. These hypotheses are imposed on the error term, e_t. In Step 0, this term is assumed to be a randomly drawn series from a normal distribution with zero mean and constant variance s^2. In Step 3, a diagnostic checking is used to validate these model assumptions, as explained in subsection 2.4.

2.2 Step 1

Depending on the selected approach, a trial model must be identified for the price data. First, in order to make the underlying process stationary (more homogeneous mean and variance), a transformation of the original price data may be necessary. In this step, if a logarithmic transformation is applied to the price data, a more stable variance is attained for the three models. For the dynamic regression and the transfer function approaches, all the initial parameters are set to zero. Particularly, in the ARIMA formulation, the inclusion of seasonal factors of the form $(1-B^S)$ may be necessary to make the process more stationary. And, to attain a more stable mean, factors of the form $(1-B)$, $(1-B^{24})$, $(1-B^{168})$ may be necessary, depending on the particular type of electricity market. The initial selected parameters for the ARIMA formulation are based on the observation of the autocorrelation and partial autocorrelation plots. In successive trials, the same observation of the residuals obtained in Step 3 (observed values minus predicted values) can refine the structure of the functions in the model.

2.3 Step 2

After the functions of the models have been specified, the parameters of these functions must be estimated. Good estimators of the parameters can be found by assuming that the data are observations of a stationary time series (Step 1), and by maximizing the likelihood with respect to the parameters [Box and Jenkins 1976].

The SCA System [Liu and Hudak 1994] is used to estimate the parameters of the corresponding model in the previous step. The parameter estimation is based on maximizing a likelihood function for the available data. A conditional likelihood function is selected in order to get a good starting point to obtain an exact likelihood function, as described in [Box and Jenkins 1976].

2.4 Step 3

In this step, a diagnosis check is used to validate the model assumptions of Step 0. This diagnosis checks if the hypotheses made on the residuals (actual prices minus predicted prices by estimated model in Step 1 are true. Residuals must satisfy the requirements of a white noise process: zero mean, constant variance, uncorrelated process and normal distribution. These requirements can be checked by taking tests for randomness, such as the one based on the Ljung-Box statistic, and observing plots, such the autocorrelation and partial autocorrelations plots.

If the hypotheses on the residuals are validated by tests and plots, then, the corresponding model can be used to forecast prices. Otherwise, the residuals contain a certain structure that should be studied to refine the model in Step 1. This study is based on a careful inspection of the autocorrelation and partial autocorrelation plots of the residuals.

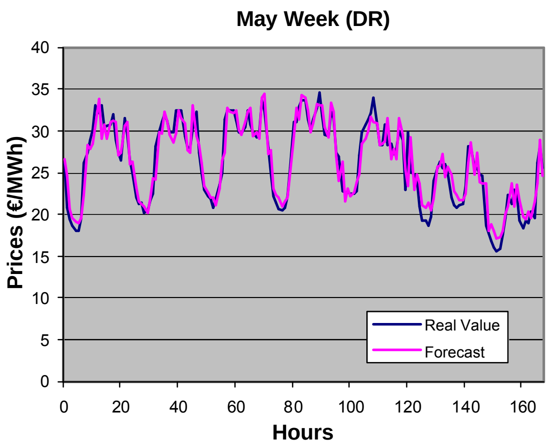
2.5 Step 4

In Step 4, the corresponding model from Step 2 can be used to predict future values of prices (typically 24 hours ahead). Due to this requirement, difficulties may arise because predictions can be less certain as the forecast lead time becomes larger.

The SCA System is again used to compute the 24-hour forecast. The exact likelihood function is selected to obtain a very accurate prediction.

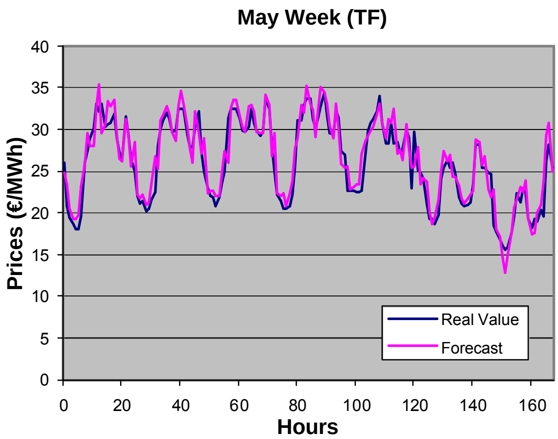
3. NUMERICAL RESULTS

The three models described in the previous section have been applied to predict the electricity prices of mainland Spain. The week selected to validate the model corresponds to the second week of May 2001 (from days May 11th to 17th), which is typically a high demand week. The hourly data used to forecast are from January 1st to May 10th, 2001. Numerical results for the three proposed models are presented below. Fig. 1 to 3 show the forecasted prices for each of the three models together with the actual prices and daily mean errors. All the study cases have been run on a DELL Precision 620 Workstation with two processors Pentium III, 1 Gb of RAM, and 800 MHz. Running time, including estimation and forecasting, has been under three minutes in all cases.



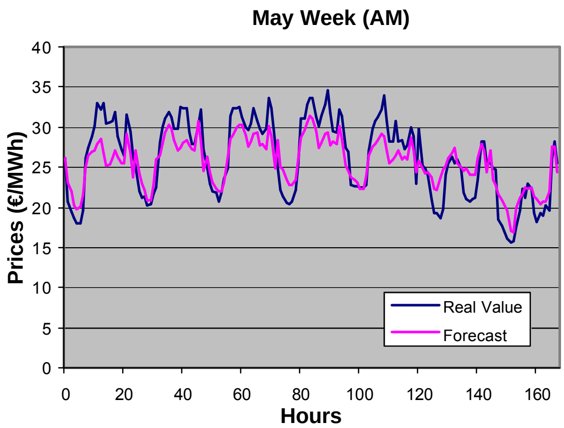
Days	1	2	3	4	5	6	7
Error (%)	5.24	3.94	3.06	3.34	5.19	7.23	7.55

Figure 1. Forecast of May week using a dynamic regression model.



Days	1	2	3	4	5	6	7
Error (%)	5.5	4.82	4.07	4.85	5.09	4.81	7.25

Figure 2. Forecast of May week using a transfer function model.



Days	1	2	3	4	5	6	7
Error (%)	10.2	7.35	6.47	8.96	6.78	9.53	8.94

Figure 3. Forecast of May week using an ARIMA model.

4. CONCLUSIONS

This paper has proposed three forecasting models: dynamic regression, transfer function and ARIMA, to predict hourly electricity prices in the Spanish electricity day-ahead market. These models are based on time series analysis. The dynamic regression and transfer function models have performed better than the ARIMA model, though the three techniques provide reasonable predictions. The difference between the ARIMA model and the other two may be due to the lack of flexibility of the ARIMA formulation when including multiple seasonality terms.

5. REFERENCES

G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, 2nd ed., Holden-Day, San Francisco, USA, 1976.

L. Liu and G. P. Hudak, *Forecasting and Time Series Analysis using the SCA Statistical System*, Scientific Computing Associated, 1994.

F. J. Nogales, J. Contreras, A. J. Conejo and R. Espínola, *Forecasting Next-Day Electricity Prices by Time Series Models*, IEEE Transactions on Power Systems, Vol. 17, no. 2, pp. 342-348, May 2002.

B. Ramsay and A. J. Wang, *A Neural Network Based Estimator for Electricity Spot-Pricing with Particular Reference to Weekend and Public Holidays*, Neurocomputing, vol. 23, pp. 47-57, 1998.

B. R. Szkuta, L. A. Sanabria and T. S. Dillon, *Electricity Price Short-Term Forecasting using Artificial Neural Networks*, IEEE Transactions on Power Systems, vol. 14, no. 3, pp. 851-857, August 1999.

CORRESPONDENCE

Javier Contreras Sanz
E.T.S. de Ingenieros Industriales
University of Castilla – La Mancha
13071 Ciudad Real, Spain
Phone: +34 926 295464
FAX: +34 926 295361
e-mail: Javier.Contreras@uclm.es