한국전력시장에서의 단기전력가격 예측

채영진, 김두중, 김은수
한국전력거래소

Forecasting Short-term Electricity Prices in South Korean Electricity Market

Yeoungjin Chae, Doojung Kim, Eunsoo Kim
Korea Power Exchange

Abstract - The authors develop and compare the performance of short-term forecasting models on electricity market prices in Korea. The models are based on time-series methods. The outcome shows that the EGARCH model has the best results in the out-of-sample forecasts.

1. Introduction

The purpose of this study is to estimate the models of real-time hourly wholesale electricity prices of the South Korean CBP market, and to test their performance, which has never been investigated to date. Since the restructuring of the electricity industry around the world during the past 15 years, researchers have used various models to capture the highly volatile characteristics of wholesale electricity prices. These models can be categorized into four groups: (1) time series models including ARIMA, transfer function, dynamic regression, and GARCH variations; (2) artificial intelligence network; (3) wavelets; and (4) jump diffusion/mean reversion.

ARIMA studies includes Fosso et al. (1999), Contreras et al. (2003), Conejo et al. (2005); transfer function papers include Nogales et al. (2002), and Conejo et al. (2005); dynamic regression analysis include Nogales et al. (2002) and Conejo et al. (2005); and GARCH models include Contreras et al. (2003) and Hua et al. (2005) while exponential GARCH models include Knittel et al. (2001) and Bowden et al. (2007). Some studies have used artificial intelligence networks such as neural networks. Examples of these papers are Ramsay and Wang (1997), Szhuta et al. (1999), Gao et al. (2000), Nicolaisen et al. (2000), Zhang et al. (2003), and Conejo at al. (2005). Other studies have used jump diffusion/mean reversion models. These studies include Johnson and Barz (1999), Skantze et al. (2000), and Knittel and Roberts (2005). Bunn and Karakatsani (2003) provide a detailed survey of alternative methods for electricity price forecasting while Alfares and Nazeeruddin (2002) summarize the various methods for electric load forecasting which can be also useful for electricity price forecasting. Among them, Conejo et al. (2005) extensively address time series analysis, neural networks and wavelets using data from the PJM Interconnection. They conclude that dynamic regression and transfer function have better forecasting capability. In addition, Conteras et al (2002), Nogales et al (2003), and Conejo et al (2005) developed a methodology in order to solve the serial correlation problem.

This paper investigates the accuracy of four forecasting models for hourly real-time prices of the South Korean wholesale electricity markets for April 2006 and compares their performance in terms of in-sample and out-of-sample forecasting. The estimated models are ARIMA, transfer function, dynamic regression, and EGARCH. For each model, there are one-week forecasts based on previous four weeks of data. Section 2 discusses the data, models, results and compares the forecasting performance of the four models employed. Section 3 concludes.

2. Data, Models and Results

2.1 Data

This study analyzes hourly real time market clearing prices for the Korean CBP market during April 2006. The data is available on KPX website. April was chosen based on thefact that non-base load generator set the marginal prices during the most of that month. Data from 2nd April to 29th April were used to forecast prices from 30th April to 6th May.

2.2 Models

2.2.1 Procedures

This study follows the methodology proposed by Conteras et al (2002), Nogales et al (2003), and Conejo et al (2005) to estimate forecasting models for wholesale competitive electricity market prices in CBP. In order to solve the serial correlation problems inherent in the price series, they adopt a recursive scheme as follows.

Step 1) Identify a model under certain assumption;
Step 2) Estimate the model parameters;
Step 3) If the assumptions are validated, the procedure continue to step 4,
otherwise go to step 1 to refine the model;
Step 4) Use the model for forecasting.

2.2.2 ARIMA Models

The general ARIMA model has the form $\Phi(B)p_t=c+\Theta(B)\varepsilon_t$ (1), where p_t is the price at hour t, ε_h is the error term, and c is a constant. $\Phi(B)$ and $\Theta(B)$ are polynomial functions of the backshift operator B (note that $B^s p_t=p_{t-s}$).

The final forecasting model for CBP during April 2006 is explained below. The final ARIMA model is: $(1 - \Phi_1 B^1)(1-\Phi_{24}B^{24}-\Phi_{168}B^{168})p_h=c+(1-\Theta_1 B^1 - \Theta_2 B^2)(1-\Theta_{24}B^{24})\varepsilon_h$ (2). The ARIMA estimated model (2) for April

2006 depends on previous values of prices as a product of 3 terms: 1 hour ago, 1 day ago and 1 week ago. Differentiation was not needed. It also depends on the previous values of errors as a product of 3 terms: 1 hour ago to 2 hour ago, and 1 day ago.

2.2.3 Dynamic Regression Model

The general dynamic regression model has the form: $p_t=c+w^d(B)d_t+w^p(B)p_t+\varepsilon_t$ (3), where p_t is the price at hour t, d_t is demand at hourt, ε_t is the error term that follows a white noise process, and c is a constant. $w^d(B)$ and $w^p(B)$ are the polynomial functions of the backshift operators (note that $B^s d_h=d_{h-s}$).

The final dynamic regression model estimated is: $p_t=c+w^0 d_t+w^1 d_{t-1}+w^{167} d_{t-167}+w^{192} d_{t-192}+w^1 p_{t-1}+w^{24} p_{t-24}+w^{168} p_{t-168}+\varepsilon_t$ (4). The dynamic regression model (4) for April 2006 depends on previous values of prices at 3 terms: 1 hour ago, 1 day ago and 1 week ago. It also depends on the values of demand at 4 terms: current demand, 1 hour ago, 167 hour ago, and 192 hour ago.

2.2.4 Transfer Function Model

The transfer model has the form: $p_t=c+w^d(B)d_t+N_t$ (5), $N = (\Theta(B)/\Phi(B))\varepsilon_t$ (6), where p_t is the price at hour t, d_t is demand at hour t, N_t is the error term that follows ARMA(p,q) process of the form(6), and c is a constant. $w^d(B)$ is the polynomial functions of the backshift operators (note that $B^s d_h=d_{h-s}$). In this model, actual price is related to the values of demand through function $w^d(B)$, actual prices to past prices through function $\Phi(B)$, and actual error to past errors through $\Theta(B)$.

The final transfer function model estimated is: $d(p_t)=w^0 d(d)_t+w^{167} d(d)_{t-167}+(1-\Theta_1 B^1)(1-\Theta_1 B^1-\Theta_{24} B^{24})\varepsilon_t$ (7). Like dynamic regression model, all of the coefficients of the selected model (7) are statistically significant at 1% level. The transfer function model (7) for 4 weeks of April 2006 shows that the difference in prices at t and at t-1 depends on values of demand at 2 terms: current demand and 167 hour ago. The error term, N_t follows pure MA process. N_t depends on the previous values of the error term at 1 hour ago, seasonal MA at 1 hour and 1day.

2.2.5 ARIMA-EGARCH Model

The EGARCH model has the form:

$\Phi(B)p_t=c+\Theta(B)\varepsilon_t$ (8),

$$\log(h_t^2)=\omega+\alpha\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}^2}}\right|+\gamma\left(\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}^2}}\right)+\beta\log(h_{t-1}^2) \quad (9),$$

where p_t is the price at hour t and ε_t is the error term that has the normal distribution of zero mean and variance, h^2. Equation (9) shows the structure of the conditional variance. The basic idea of this model is to capture the autoregressive conditional heteroskedasticity in the residuals, which suggests the possible existence of the inverse leverage effect(Bunn and Karakatsani (2003), Knittel and Roberts(2005), and Bowden and Payne(2007)). The intuition behind inverse leverage effect is that since marginal costs become much higher when demand spikes, positive demand shocks have a larger impact on price movements relative to negative shocks. ω is the mean of the volatility equation (6) while α reflects the size effect of a shock irrespective of the sign of the shock. γ shows the sign effect of a shock while β indicates the degree of volatility persistence (Bowden and Payne (2007)).

The final ARIMA-EGARCH model estimated is: $(1 - \Phi_1 B^1)(1-\Phi_{24}B^{24}-\Phi_{168}B^{168})d(p_t)=c+(1-\Theta_1 B^1)(1-\Theta_{168}B^{168})\varepsilon_t$ (10),

$$\log(h_t^2)=0.88+0.37\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}^2}}\right|+0.12\left(\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}^2}}\right)+0.57\log(h_{t-1}^2) \quad (11).$$

The ARIMA-EGARCH estimated model (10) for April 2006 shows that the difference in prices depends on the past prices in 3 terms: 1 hour ago, 1 day ago and 1 week ago. It also depends on the previous values of errors as a product of 2 terms: 1 hour ago and 1 week ago. Equation (11) shows the coefficients of the volatility equation. The coefficient of a shock to hourly SMP irrespective of its sign is 0.37 and it is positive and significant at 1% level. Likewise the coefficient of sign effect is 0.12 and it is also is positive and significant at 5% level, which implies the existence of inverse leverage effects. In other words, there is a chance that an unexpected demand spike may result in positive price shocks in the wholesale prices. The volatility persistence coefficient is 0.57 which is also positive and significant at 1% level.

2.3 Results and Performances

2.3.1 Error Measures

To compare the prediction power of all proposed models, two types of measures are used for daily and weekly comparison. One is Mean Absolute Percentage Error (MAPE), which is compute as follows:

$$MAPE_{day}=\left[\sum_{t=T+1}^{T+h}\left|\frac{\hat{p}_t^j-p_t^j}{p_t^j}\right|\Big/h\right]\times 100$$

, where $MAPE_{day}$ is the daily mean absolute percentage error, p^j_t is the actual price at hour t, $\hat{p}^j_t$ is the forecasted price for that hour.

The other is Root Mean Squared Error (RMSE).

$$RMSE_{day}=\sqrt{\sum_{t=T+1}^{T+h}\frac{(\hat{p}_t^j-p_t^j)^2}{h}}$$

,where $RMSE_{day}$ is the daily root mean squared error, p^j_t is the actual price at hourt, $\hat{p}^j_t$ is the forecasted price for that hour.

2.3.2 In-sample Results

The daily and weekly results of in-sample forecasting for the four models are reported in Table 1. Among the model estimated, dynamic regression shows the best result in daily and weekly forecasts while ARIMA model show consistently lower performance. The next best model is transfer function. The objects of in-sample comparison are the last 7 days of 4 weeks. It is noted that EGARCH has better performance in the later days of the week. Appendix IV includes the table for actual and in-sample forecasted prices for day 1.

Table 1 In-sample forecasting error comparison (%)

	ARIMA		Dynamic Regression		Transfer function		EGARCH	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
day1	4.89	4.47	2.68	2.53	3.08	2.92	3.02	2.90
day2	6.47	7.70	3.68	3.97	3.71	3.90	5.88	7.08
day3	6.95	7.70	3.14	3.78	3.46	3.83	5.35	6.22
day4	4.77	6.13	4.27	4.31	5.21	5.13	3.85	5.26
day5	4.39	5.47	3.32	3.34	3.88	3.91	3.90	4.92
day6	3.29	4.34	3.31	3.57	3.90	4.10	3.65	4.51
day7	4.97	4.86	4.36	4.01	4.75	4.45	3.82	3.92
week	5.10	5.81	3.54	3.64	4.00	4.03	4.21	4.98

2.3.3 Out-of-sample Results

Table 2 describes the results of out-of-ample forecast from four models. Unlike in-sample forecast results, EGARCH is the best estimated model in terms of MAPE and RMSE while the transfer function model shows the lowest performances in both MAPE and RMSE. It is noted that EGARCH achieved the average 2.1% forecasting MAPE except for the day 3 and day 6, which shows high MAPE between 10% - 12%. The forecast results from dynamic regression are very close to ARIMA results by .15% differences in weekly forecasting.

Table 2 Out-of-sample forecasting error comparison (%)

	ARIMA		Dynamic Regression		Transfer function		EGARCH	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
day1	4.96	4.60	3.41	3.22	5.54	4.43	1.42	2.38
day2	3.84	3.95	5.80	5.13	11.84	9.72	1.45	3.90
day3	4.73	6.83	6.27	7.40	6.80	7.66	12.08	8.00
day4	4.18	4.25	4.36	4.48	4.13	4.65	2.81	5.53
day5	3.90	4.54	5.97	5.76	6.58	6.62	3.40	6.39
day6	8.10	7.87	4.35	3.71	5.11	4.46	10.41	5.02
day7	5.27	5.17	5.92	5.65	6.00	5.92	1.39	3.57
week	5.00	5.32	5.15	5.05	6.57	6.21	4.66	4.97

3. Conclusion

Short term future price information is important for buyers and sellers in the wholesale electricity market for their optimal operation and risk management. This paper investigated forecasting accuracy of four different time-series methods to predict the 24 hourly system marginal prices of the real time wholesale electricity market in the Korean CBP. Model estimation has been carried out using hourly data from the Korean CBP in April 2006.

Among the four models, EGARCH and ARIMA are more effective than transfer function for out-of-sample forecasting. For in-sample forecasting, dynamic regression shows the best prediction power. In addition all four models present relatively modest forecasting error about 5~6% in terms of MAPE, which is good enough considering the highly volatile characteristics of electricity prices and power demand. Market participants in CBP can use the forecasting results from these models for their short-term business decision regarding bidding and operation strategies.

However, there are some limitations in this study. First, this paper deals with only 4 weeks of hourly data in April 2006, which is clearly not enough to capture the true volatility in the longer horizon. Second, it should be also noted that we may need different models for different period of time. Third, the study does not consider the influence of market structure and regulatory aspects of the Korean CBP forecasting the wholesale prices. These points suggest future research directions: to develop more comprehensive models to accommodate various aspects of volatile electricity prices.

[References]

[1] Borenstein, S. (2002), "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster", *Journal of Economic Perspectives*, 16,191-211.

[2] Bowden, N. and J. Payne (2007), "Short Term Forecasting of Electricity Prices for MISO Hubs: Evidence from ARIMA-GARCH Models", a study paper, Illinois State University.

[3] Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis, Forecasting ,and Control*, 2nd edition, San Francisco, CA: Holden-Day.

[4] Bunn, D.W. and N. Karakatsani (2003), "Forecasting Electricity Prices", Working Paper, London Business School.

[5] Conejo, A. J., J. Contreras, R. Espinola, and M. A. Plazas (2005), "Forecasting Electricity Prices for a Day-Ahead Pool-Based Electric Energy Market", *International Journal of Forecasting*,21,435-462.

[6] Contreras, J., R. Espinola, F. J. Nogales, and A. J. Conejo (2003), "ARIMA Models to Predict Next-Day Electricity Prices", *IEEE Transactions on Power Systems*, 18, 1014-1020.

[7] Nogales, F. J., J. Contreras, A. J. Conejo, R. Espinola (2002) "Forecasting Next-Day Electricity Prices by Time Series Models", *IEEE Transactions on Power Systems*, 17, 342-348.

[8] Escribano, A. and J.I. Pena, and P. Villaplana (2002), "Modeling Electricity Prices: International Evidence" Universidad Carlos III De Madrid Working Paper 02-27.

[9] Federal Energy Regulatory Commission (1999). Order No. 2000 Regional Transmission Organizations.

[10] Fosso, O.B., A. Gjelsvik, A. Haugstad, M. Birger, and I. Wangensteen (1999), "Generation Scheduling in a Deregulated System: The Norwegian Case", *IEEE Transactions on Power Systems*, 14, 75-81.

[11] Husiman, R. and R. Mahieu (2003), "Regime Jumps in Electricity Prices", *Energy Economics*, 25,425-434.

[12] Johnson, B. and G. Barz (1999), "Selecting Stochastic Process for Modeling Electricity Prices" *Energy Modelling and the Management of Uncertainty*, Risk Publications.

[13] Kim, C.I., I. Yu, Y.H. Song (2002), "Prediction of System Marginal Price of Electricity Using Wavelet Transform Analysis", *Energy Conversion and Management*, 43, 1839-1851.

[14] Korea Power Exchange (2006), "Cost-Based Pool: Market Operation", .