Optimal Combined Forecasts for Electricity Prices: Influence of Clean Energies

Eduardo Caro[1,2], Carolina García-Martos[1], and María Jesús Sánchez[1]
[1] Technical University of Madrid, Madrid, Spain
[2] Corresponding author: Eduardo Caro, e-mail: ecaro@etsii.upm.es

Abstract

In this work, an electricity price forecasting model is developed. The performance of the proposed approach is improved by considering renewable energies (wind power and hydro generation) as explanatory variables. Additionally, the resulting forecasts are obtained as an optimal combination of a set of several univariate and multivariate time series models. The large computational experiment carried out using out-of-sample forecasts for every hour and day allows withdrawing statistically sound conclusions.

Keywords: Optimization, electricity price forecasting, renewable power generation, time series.

1. Introduction

The increasing penetration of renewable energies (particularly, wind power generation) has a significant impact in liberalized power markets. Specifically, the hourly electricity price may be influenced by the generation of clean energies, such as hydro generation and wind power production.

In this work, a one-day-ahead electricity price forecasting method is proposed, based on an optimal combination of several univariate and multivariate approaches. A Design of Experiments (DOE) and an Analysis of Variance (ANOVA) are performed to select those forecasting methods with better forecasting accuracy, and a mathematical programming problem is developed to optimally combine these forecasts.

Technical literature is rich in references pertaining to electricity price forecasting field. These techniques are commonly divided into short-term and medium/long-term forecasting. Medium and long term forecasting are addressed by Vehvilainen and Pyykkonen (2005), Conejo et al. (2010), García-Martos et al. (2011) and Alonso et al. (2011).

With respect to short-term forecasting, several techniques have been proposed in the literature: Conejo et al. (2002), Nogales et al. (2002), Conejo et al. (2004), Conejo et al. (2005), Contreras et al. (2003), Cottet et al. (2003), García-Martos et al (2007A and 2007B), García-Martos et al. (2012).

The contribution of this work is twofold. First, the approaches proposed by García-Martos et al. (2007 and 2012) are extended to include explanatory variables. Second, a mathematical programming problem is proposed to optimally combine the forecast of the previous methods, leading to a forecasting technique with higher accuracy.

The rest of this research is organized as follows. Section 2 briefly describes the Univariate Mixed Model and Dynamic Factor Model, with and without explanatory variables. Section 3 describes the DOE and ANOVA performed. Section 4 details the optimal combined forecasting method. Finally, Section 5 presents some relevant conclusions.

2. Forecasting Models

In this section, the following methods are briefly described: the Univariate Mixed Model (UMM), UMM with explanatory variables, the Dynamic Factor Model (DFM), and DFM with explanatory variables.

2.1. Univariate Mixed Model

The Univariate Mixed Model implements each hourly series by the well-known univariate seasonal ARIMA(p,d,q)x(P,D,Q)$_s$, (also called SARIMA) whose equation is formulated below:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_S^D p_t^h = \theta_q(B)\Theta_Q(B^s)a_t \tag{1}$$

where

$$\begin{aligned}
\phi_p(B) &= (1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\\
\Phi_P(B^s) &= (1-\Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps})\\
\nabla^d &= (1-B)^d\\
\nabla_S^D &= (1-B^s)^D\\
\theta_q(B) &= (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\\
\Theta_P(B^s) &= (1-\Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs})
\end{aligned}$$

and B is the slack operator, and a_t is normally independent and identically distributed random error, and p_t^h is the electricity price at the h-th hour in the t-th day.

From (1), note that the electricity price series is modeled as 24 hourly series of prices $[\boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \boldsymbol{p}_{t+2}, \ldots, \boldsymbol{p}_T]$, where $\boldsymbol{p}_t = [p_t^1, p_t^2, \ldots, p_t^{24}]$. In this model, the historical length used to estimate the parameters is 44 weeks, corresponding to weekdays and/or weekends, depending on the day to be forecasted (García-Martos, 2007).

2.2. UMM with Explanatory Variables

The UMM model described above is extended in order to consider explanatory variables, leading to a SARIMAX model. The exogenous variables considered in this paper are: the forecasted wind power production for each day and hour, and the hydro reservoirs for each day. These two variables can be represented with $X_{h,t}$ and H_t, respectively, where subindexes h and t represents the hour and day considered.

Using the renewable information {$X_{h,t}$ and H_t}, a SARIMAX(p,d,q)x(P,D,Q)$_s$ model is developed, leading to parameters $\{\beta_{Wind,h}, \beta_{Hydro}\}$, which are the regression coefficients for the wind and hydro explanatory variables, respectively.

The SARIMAX above can be rewritten using daily information regarding the wind power production ($Y_t = \sum_{h=1}^{24} X_{h,t}$). Then, an alternative SARIMAX model is proposed.

From models above, the three series (prices, wind power production, and hydro reservoirs) have to be differenced. This is so because these three series have a common regular unit root and, then, differentiating them we avoid estimating spurious regression coefficients.

2.3. Dynamic Factor Model

Models above do not consider multivariate cross-correlation of the panel of 24 hourly prices. These correlations can be considered by using a VARIMA (Vector-ARIMA) model, with the 24-dimensional vector of hourly series. However, this model would imply that the dimension of each parameter would by 24x24, leading to the well-known problem called "curse of dimensionality".

In order to reduce the dimensionality, the DFM model proposed by García-Martos et al. (2012) is used in this paper. The model is presented below:

$$\widetilde{\mathbf{p}}_t = \mathbf{f}_t\mathbf{P}^T + \mathbf{e}_t \tag{2a}$$

$$\phi_p(B)\Phi_P(B^s)\nabla_s^D \widetilde{f'_{k,t}} = \theta_q(B)\Theta_Q(B^s)a_t \quad (2b)$$

where $\tilde{\boldsymbol{p}}_t$ are the centered prices, $\boldsymbol{f}_t$ is the vector of unobserved *r* common factors, $\tilde{f'}_{k,t}$ refers to each centered differenced common factor, and $\mathbf{P}^T$ is the transpose of the loading matrix whose columns are the weights of the original series used to build the unobserved common factors. Vector **e** is a stationary process which explains a small portion of the variability of hourly prices, and it is model as an ARMA process.

Numerical simulations denote that the optimal number of common factors is 2 or 3. In this paper, we DFM with r = 2 and r = 3, denoting them as DFM-2 and DFM-3, respectively.

2.4. DFM with Explanatory Variables

The model detailed in Section 2.3 above is extended in order to consider explanatory variables, leading to parameters $\{\beta_{Wind,h}, \beta_{Hydro}\}$ which are the regression coefficients for the wind and hydro explanatory variables, respectively.

Again, the three series (prices, wind power production, and hydro reservoirs) have to be differenced. This is so because these three series have a common regular unit root and, then, differentiating them we avoid estimating spurious regression coefficients.

3. Design of Experiments and ANOVA

Section 2 above presents three different models for forecasting electricity prices considering exogenous variables: UMM, DFM-2, DFM-3. For each model, depending on the explanatory variables used, will result a different "forecasting method". Each forecasting method is determined by: (i) a model (UMM, DFM with 2 common factors, and DFM with 3 common factors), (ii) considering no wind power generation, hourly generation, or daily generation, (ii) considering no hydro reservoirs or daily reservoirs (See Table 1).

Table 1. ANOVA Model

	Name	Levels
Factors	Model	UMM DFM-2 DFM-3
	Explanatory Variable 1: Wind	No Wind Hourly Wind Daily Wind
	Explanatory Variable 2: Hydro	No Hydro Hydro
Block	Day	$1, \dots, T$

From Table 1, note that there are 3x3x2 = 18 different forecasting methods. These methods are compared using the ANOVA technique (see Montgomery (1984)). The main objective is to determine: (i) which model produces the most accurate forecasts, and (ii) whether consider or not the explanatory variables concerning wind power generation and/or hydro reservoirs.

In this paper, the metric used to compare the forecasting accuracy are the percentage errors $e_{h,t}$, which are used to compute the daily MAPE as follows:

$$MAPE_t = \frac{1}{24}\left(e_{1,t} + e_{2,t} + \cdots + e_{24,t}\right)$$

where

$$e_{h,t} = \left|\hat{p}_{h,t} - p_{h,t}\right| / \; p_{h,t}$$

and $p_{h,t}$ corresponds to the actual value for the electricity price in the *h*-th hour at the *t*-th day.

The hourly forecasting error in the *h*-th hour is linearly modeled as:

$$e^h_{ijkd} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_d + (\alpha\beta)_{ij} + u_{ikjd}$$

where parameter μ represents the global mean, parameter α_i is the main effect related to the factor "Model", parameter β_j is the main effect related to the factor "Wind", parameter γ_k is the main effect related to the factor "Hydro", parameter δ_d is the main effect related to the factor "Day", and parameter $(\alpha\beta)_{ij}$ corresponds to the interaction between factors "Model" and "Wind". Term u_{ikjd} stands for normally-distributed random error.

The values of main effects $\{\alpha_i, \beta_j, \gamma_k\}$ quantify the average increment or decrement of the response variable (hourly percentage error) for the corresponding forecasting method with respect to the average response.

For each hour, the eighteen forecast series have been compared using the ANOVA technique, and those forecasting methods which produce statistically-significant smaller errors are selected as "candidates". Table 2 indicates which are those methods selected as candidates.

Table 2. Candidate methods for each hour

	Model	Hydro	Wind	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Method 1	UMM	None	None								☑	☑							☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 2		None	Hourly	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 3		Hydro	None								☑	☑							☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 4		Hydro	Hourly	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 5		None	Daily			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 6		Hydro	Daily			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 7	DFM-2	None	None			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑				☑	☑		☑	☑	☑	☑	☑	☑
Method 8		None	Hourly			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 9		Hydro	None			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑				☑	☑		☑	☑	☑	☑	☑	☑
Method 10		Hydro	Hourly			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 11		None	Daily			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 12		Hydro	Daily			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 13	DFM-3	None	None			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑				☑	☑		☑	☑	☑	☑	☑	☑
Method 14		None	Hourly			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 15		Hydro	None			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑				☑	☑		☑	☑	☑	☑	☑	☑
Method 16		Hydro	Hourly			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 17		None	Daily			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Method 18		Hydro	Daily			☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑

4. Combined Forecast

The main idea of the "combined forecast" is to make use of the set of candidate methods (see Table 2) and compute a forecasting series as the weighted average of the selected methods. Then, the forecasted price value for the *h*-th hour on the *t*-day ($\hat{p}^{comb}_{h,t}$) is computed as:

$$\widehat{p}^{comb}_{h,t} = \omega_{h,1}\widehat{p}^{m_1}_{h,t} + \ldots + \omega_{h,18}\widehat{p}^{m_{18}}_{h,t}$$

where $\hat{p}^{m}_{h,t}$ corresponds to the forecast value obtained with the *m*-th forecasting method for the *h*-th hour on the *t*-day, and weights $\{\omega_{h,1}, \ldots, \omega_{h,18}\}$ range from 0 to 1, subject to:

$$\omega_{h,1} + \ldots + \omega_{h,18} = 1$$

The weight values are optimally computed by means of an optimization problem. To compute these values, the approach developed in Lam et al. (2001) is extended, proposing a mathematical programming problem which minimizes the MAPE of the combined forecast, i.e.,:

$$e^{comb}_{h,t} = \frac{\left| \sum_{M=1}^{18} \omega_{h,M}\widehat{p}^{m_M}_{h,t} - p_{h,t} \right|}{p_{h,t}}$$

Resulting the following non-linear optimization problem:

$$\underset{\omega_{h,1},\ldots,\omega_{h,18}}{\text{minimize}} \quad \frac{\left|\sum_{M=1}^{18} \omega_{h,M}\widehat{p}_{h,t}^{m_M} - p_{h,t}\right|}{p_{h,t}} \tag{3}$$

Problem (3) has to be solved for each hour, considering the candidates of Table 2.

4.1. Results

In this subsection, the electricity price series is forecasted using the following methods: DFM, as developed in García-Martos et al. (2012); UMM as proposed in García-Martos et al. (2007); and COMB, which corresponds to the approach proposed in this paper; for the period between January 1st, 2007 to December 31th, 2009, for the Iberian Market. Figure 1 provides the actual price and forecasts for August 25th, 2009.

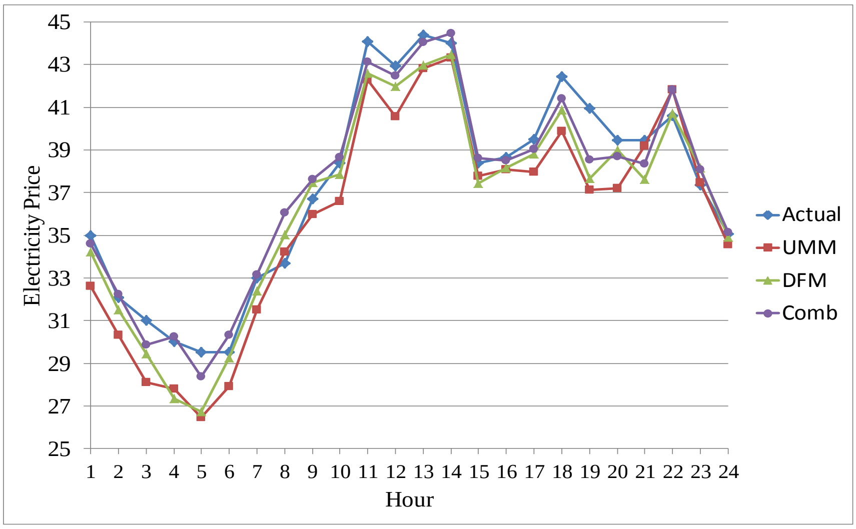

Figure 1. Real price and forecasts for August 25th, 2009.

Figure 2 depicts the monthly average forecasting errors. From Figs. 1 and 2 it can be observed that the forecasting method proposed in this paper outperforms other traditional approaches in terms of forecasting accuracy.

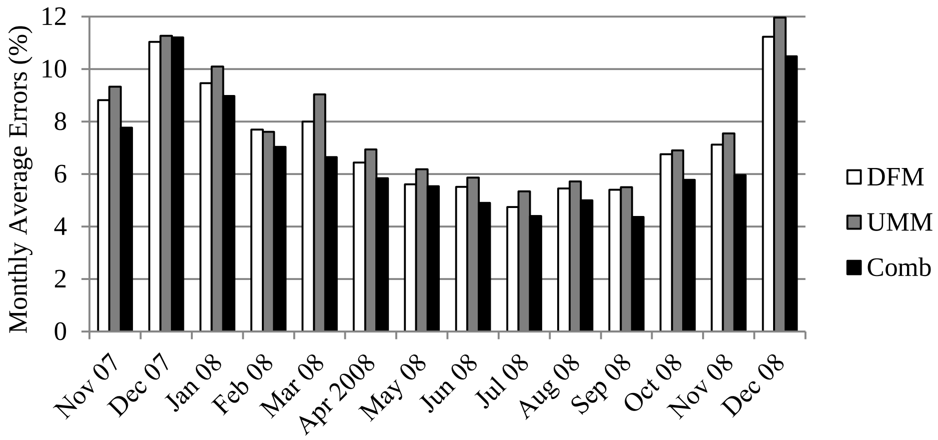

Figure 2. Average monthly errors for DFM, UMM, and Combined models

5. Conclusions

A one-day-ahead combined forecasting method is developed in this paper, based on optimization techniques and considering renewable energies.

The combined method is based on the solution of a linear mathematical programming problem which computes the optimal combination of forecast series pursuing a minimal MAPE. The forecasting is based on the Univariate Mixed Model (developed by García-Martos et al. (2007)), and on the Dynamic Factor Model (proposed in García-Martos et al. (2012)). These two methods are extended by including information concerning clean energies. Specifically, daily hydro reservoirs and hourly/daily wind power production are considered.

The proposed method is applied to the Iberian Market, employing data from the 1st of January 2007 until the end of December of 2009. Numerical results denote that the proposed model exhibits higher forecasting accuracy than other traditional methods.

Here we show additional results to those that can be encountered in the full paper García-Martos et al (2013).

6. References

Alonso, A.M., García-Martos, C., Rodríguez, J. and Sánchez, M.J. (2011). "Seasonal Dynamic Factor Analysis and Bootstrap Inference: Application to Electricity Market Forecasting", Technometrics, 53 (2), 137-151.

Conejo, A.J., Nogales, F.J., and Arroyo, J.M. (2002). ”Price-taker bidding strategy under price uncertainty,” IEEE Trans. Power Systems, 17, 1081-1088.

Conejo, A.J., Nogales, F.J., Arroyo, J.M. and Garcia-Bertrand, R. (2004). ”Risk constrained self scheduling of a termal power producer,” IEEE Transactions on Power Systems, 19, 1569-1574.

Conejo A. J., Contreras J., Espínola R., Plazas M. A. (2005), ”Forecasting electricity prices for a day-ahead pool-based electric energy market,” Int. Journal of Forecasting, 21, 3, 435-462.

Conejo, A.J., Carrion, M., Nogales, F.J. and Morales, J.M. (2010), ”Electricity Pool Prices: Long-Term Uncertainty Characterization for Futures-Market Trading and Risk Management,” Journal of the Operational Research Society, 61, 235-245.

Contreras, J., Espínola, R. Nogales, F.J., Conejo, A.J. (2003), ”ARIMA Models to Predict Next-Day Electricity Prices,” IEEE Transactions on Power Systems, 18, 3, 1014-1020.

Cottet R., and Smith M. (2003), ”Bayesian Modeling and Forecasting of Intraday Electricity Load”, Journal of the American Statistical Association, 98, 464, 839-849.

García-Martos, C., Rodríguez, J. and Sánchez, M.J. (2007), ”Mixed models for short-run forecasting of electricity prices: application for the Spanish market,” IEEE Transactions on Power Systems, 2, 2, 544-552.

García-Martos, C., Rodriguez, J. and Sánchez, M.J. (2011), ”Forecasting electricity prices and their volatilities using Unobserved Components,” Energy Economics, 33, 6, 1227-1239.

García-Martos, C., Rodriguez, J. and Sánchez, M.J. (2012), ”Forecasting electricity prices by extracting dynamic common factors: application to the Iberian Market,” IET Generation, Transmission and Distribution, 6, 1, 11-20.

García-Martos, C., Caro, E. and Sánchez, M.J. (2013). "Electricity Price Forecasting Accounting for Renewable Energies: Optimal Combined Forecasts”. Working Paper, ETSII-UPM.

Lam, K.F. Mui, H.W. and Yuen, H.K. (2001). ”A note on minimizing absolute percentage error in combined forecasts,” Computers & Operations Research, 28, 1141-1147

Montgomery, D.C. (1984), ”Design and Analysis of Experiments,” New York: Wiley.

Nogales, F.J., Contreras, J., Conejo, A.J., Espínola, R. (2002), ”Forecasting Next-Day Electricity Prices by Time Series Models,”. IEEE Transactions on Power Systems, 17, 2, 342-348.

Vehvilainen, I. and Pyykkonen, T. (2005), ”Stochastic factor model for electricity spot price: the case of the Nordic market,” Energy Economics, 27, 351-367.