

Forecasting Electricity Spot Prices: Assessing the Performance of Linear and Nonlinear Specifications

Fany Nan, Silvano Bordignon, Derek W. Bunn and Francesco Lisi

Abstract In this paper we present an extensive day-ahead forecasting comparison of different classes of models for spot electricity prices, including ARMAX models, constant and time-varying parameter regression models, Markov regime-switching regression models. In particular, forecasting performances are studied for representative trading periods of the day of the UK Power Exchange prompt market. The analyses are conducted considering price series adjusted for the deterministic component with and without spikes and show that relative out-of-sample forecasting performances evaluated through various prediction error statistics are distinct for each trading period, forecasting season and performance indicator, so that it seems that no model consistently outperforms the others. However, we note that ARMAX appears to be the best in most cases. Moreover, the Diebold and Mariano test for equally prediction accuracy indicates that when it is not the best, the ARMAX model is not statistically different from the best. We conclude that simpler linear models may be preferable to more complex nonlinear models in forecasting situations.

Key words: Forecasting Electricity Prices, ARMAX, Regime-Switching, Time-Varying Parameters, Predictive Accuracy Test.

1 Introduction

In recent years electricity price forecasting has become increasingly important in both research and practice [2, 17]. Whilst interest mainly focuses on developing

Fany Nan, Silvano Bordignon and Francesco Lisi
Department of Statistical Sciences, University of Padova, via Cesare Battisti 241/243, Padova
35121 Italy e-mail: [fany.nan,bordigno,lisif]@stat.unipd.it

Derek W. Bunn
Department of Management Science and Operations, London Business School, Regent's Park,
London NW1 4SA UK e-mail: dbunn@london.edu

predictive models that allow producers, retailers and large consumers to optimize their operations, because electricity is not storable, energy traders also need to undertake forecasting to assess forward positions. The peculiar features of electricity prices, such as multi-scale seasonality, high volatility and spiking behaviour, furthermore motivate the active research. Time series and regression models predominate but recently more complex models such as Markov regime switching [14, 11], and time-varying parameter models [15, 11] have worked well, at least in-sample. However, the relative forecasting ability of linear and nonlinear models out of sample is still an open research issue.

The main problem that we have to face, when comparing results from published research, is that the studies proposed are conducted in very different contexts (see [1]). The proposed models are very different from paper to paper: they are estimated on spot or log prices, with or without spikes and there is no convergence on how to treat the deterministic component of price series. Moreover markets are idiosyncratic (eg PJM, EEX, and NordPool have many different characteristics), the forecasting periods are sometimes not long enough to attain reasonable conclusions, recursive re-estimations are often not undertaken and the prediction error statistics also vary. Finally the statistical significance of relative performance is often omitted.

In this paper, we propose an exhaustive day-ahead forecasting comparison that involves four different classes of models, each one characterized by specific features. The key aspect of our study is the fact that the models are considered under equivalent conditions, so that forecasting results can be directly compared. The models are an ARMAX, a constant parameter linear regression model (LR), a time-varying parameter regression model (TVR) and a Markov regime-switching regression model (MS). In the ARMAX model the exogenous factor is the operating margin, that is the excess capacity over demand available in the system. This simple model captures shocks in pricing caused by scarcity. The linear regression model is useful for the analysis of how market fundamentals affect electricity price formation. The time-varying regression model, specified here with random-walk coefficients, allows for price dynamics that continuously adapt to market structure changes. The Markov regime-switching model captures spikes and discontinuities in price series, distinguishing between normal and high-price regimes.

This work considers price data from the UK Power Exchange for the period April 1st, 2005 - September 30th, 2006. Each day consists of 48 half-hourly periods called load periods (load period 1 is defined as 00:00-00:30am, load period 2 as 00:30-01:00am, and so on) for which we have the corresponding prices. We denote by P_{jt} the spot price at time t and load period j ($t = 1, 2, \dots, N$, $j = 1, 2, \dots, 48$). As it is customary, weekend days and national holidays have been removed and the resulting 48 half-hourly time series are considered separately. Each series has length 380. In this work we focus on forecasting prices corresponding to periods that are representative of different moments of the day: trading periods 6 (2:30-3:00am), 18 (8:30-9:00am), 28 (13:30-14:00pm), 38 (18:30-19:00pm) and 44 (21:30-22:00pm). In order to be able to make one day-ahead out-of-sample predictions, the last 189 observations were not considered for model building. Therefore, the sample period can be divided in an in-sample period (April 1st, 2005 - December 31st, 2005) and

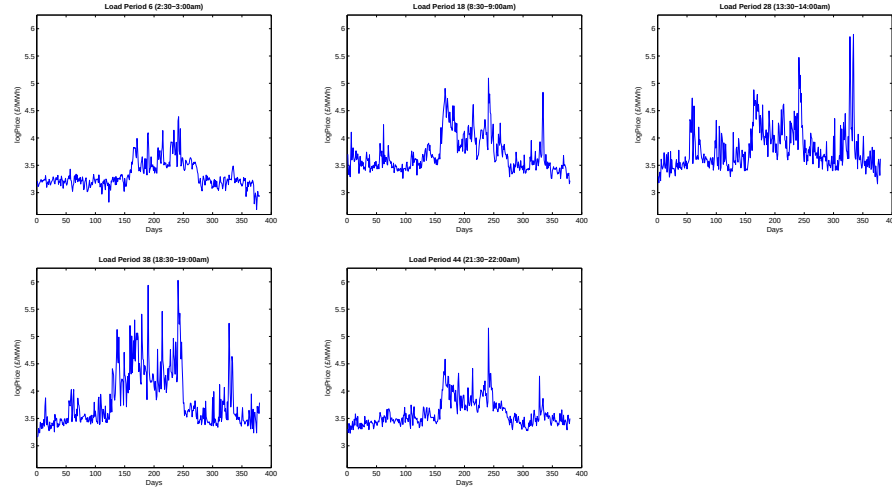


Fig. 1 Electricity log prices for the considered load periods from April 1st, 2005 to September 30th, 2006.

an out-of-sample period (January 1st, 2006 - September 30th, 2006). Predictions are evaluated considering not only the whole forecasting period (January-September 2006, 189 data), but also the three sub-samples which reflect winter, spring and summer. The scope is to understand how the forecasting accuracy changes by season and if there is a particular class of models that outperforms the others in terms of predictive performance. Predictive performance of the models, estimated on the adjusted log price series with and without spikes, are evaluated according to both prediction error statistics and equal forecasting accuracy test.

2 Data Analysis of the UKPX electricity market

The prices are characterized by very high peaks in winter and summer, especially during high demand load periods. This spiky behaviour is mainly connected with supply shocks. To attain a more stable variance, logarithmic transformations of the price series are used for the calibration of the models.

As figure 1 shows, a feature of our data is the presence of a deterministic component that changes according to the considered load period. The problem is how to treat it. Some authors prefer to consider it as a regressor in the models [11], others prefer to remove it and to work on the adjusted price series [8, 14]. We followed this second approach. Thus we removed the deterministic component from the log price series and we estimated the models on the adjusted series. Common tools for modeling deterministic components include functions with dummies [9],

seasonal functions with trend [18, 3] derived from the sinusoidal approach applied by Pilipovic [16] or a combination of both [13, 14], but these methods depend on the considered market and the type of data. Instead of using a parametric method we pursued a nonparametric one. The Friedman's Supersmoother [7] is a very flexible method to estimate a deterministic component of the price series in our context. It is a nonparametric technique based on the nearest neighbor method characterized by specific procedures for the selection of the smoothing parameter.

Let D_{jt} be the estimate of the deterministic component obtained applying the Friedman's Supersmoother to the log price series $\log P_{jt}$. Then, the adjusted series of log prices is given by $p_{jt} = \log P_{jt} - D_{jt}$ ($t = 1, 2, \dots, N$ and $j = 6, 18, 28, 38, 44$). After adjusting for the deterministic component, the Augmented Dickey-Fuller and Phillips-Perron unit root tests for p_{jt} across the five load periods rejected the null hypothesis of unit root at the 5% significance level. In the same way, the KPSS test did not reject the null hypothesis of stationarity at the 5% significance level.

As figure 1 shows, our series present very high sudden spikes (further evident if we consider spot prices), which could affect model estimation and, consequently, the forecasting experiment. A possible solution of the problem is to treat these spikes as outliers and use some procedure to preprocess the data. With respect to this issue we decided to conduct analysis on the forecasting accuracy of the models estimated on the series with and without peaks. According to [17], we did not cut the spikes at a specified threshold, but we dampened them differentiating between jumps and extreme jumps. The method was applied iteratively (until a stop criterion was satisfied) to the spot price series P_{jt} . The threshold is calculated from the adjusted series: if the adjusted price is higher than the threshold, it is replaced by a logarithmic function depending on both the price and the threshold. Then the series is transformed again into spot prices and the deterministic component is recalculated. Section 4 contains the forecasting results for the models estimated on the adjusted logarithmic price series with and without spikes.

3 Predictive models

As Karakatsani and Bunn [11] pointed out studying the UK market, there is a strong linkage between prices and market fundamentals. In our research we take into account this characteristic: to improve forecasting accuracy all the proposed models are based on important market outcomes. We considered the following variables (in the log scale):

Demand Forecast ($demF_t$). This is the national day-ahead demand forecast published by the system operator for each load period. The term 'day ahead' means that $demF_t$ is available at time $t - 1$.

Indicated Margin ($margin_t$). It is the available capacity margin and it is defined as the difference between the sum of the maximum export limits nominated by each generator prior to each trading period, as its maximum available output capacity, and the demand forecast.

To face possible non linear relations between price and demand, and price and margin, we introduce a quadratic polynomial of $demF$ and $margin$.

Gas Price. It is the daily UK natural gas one-day forward price, from the main National Balancing Point (NBP) hub. The strong relation with price series, especially during winter spikes, suggests to include it into the models. In particular, our models include the series of deviations ($gasF.res_t$) of gas prices from their deterministic components calculated with the Friedman's Supersmoother.

Past Prices (p_{t-j}). They are the lagged spot prices. In particular, lags 1 and 5, corresponding to the previous day price and to the previous week price, were considered.

Volatilities (Vol_t). It is an indicator of instability and risk for electricity price series ($priceVol_t$) and for demand forecast series ($demVol_t$). Volatility is computed as the coefficient of variation calculated on a rolling windows of the last 5 days.

The first class of models we consider is the ARMAX(p,q,s). For our dataset the identified model is the ARMAX(1,1,1)

$$p_{jt} = \phi_j p_{j(t-1)} + \varepsilon_{jt} + \theta_j \varepsilon_{j(t-1)} + \beta_j z_{jt}, \quad \varepsilon_{jt} \sim WN(0, \sigma_j^2), \quad (1)$$

where ε_{jt} is the error term and z_{jt} is the exogenous variable (margin). $\phi_j, \theta_j, \beta_j$ are constant coefficients. This model, estimated through maximum likelihood methods, is the simplest among our models.

The second model is a linear regression (LR), which accounts for relation between prices and market fundamentals. The model is specified as:

$$p_{jt} = \beta_j' \mathbf{X}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim \text{i.i.d}(0, \sigma_j^2) \quad (2)$$

where β_j is a $k \times 1$ vector of constant coefficients and ε_{jt} is an i.i.d. error term. $\mathbf{X}_{jt}$ is a $k \times 1$ vector of regressors selected with stepwise backward techniques (AIC criterion). Parameters β_j are estimated with standard least square method.

To face non-linearity and heteroskedasticity problems, we consider two further classes of models. The first one is a time-varying parameter regression model (TVR):

$$p_{jt} = \beta_{jt}' \mathbf{X}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim WN(0, \sigma_{\varepsilon_j}^2), \quad (3)$$

$$\beta_{j(t+1)} = \beta_{jt} + \mathbf{v}_{jt}, \quad \mathbf{v}_{jt} \sim WN_k(0, \mathbf{H}_j), \quad (4)$$

where $\mathbf{v}_{jt}$ is independent from ε_{jt} and $\mathbf{H}_j = \text{diag}\{\sigma_{v_{jk}}^2\}$. In the above formulation, the regression coefficients are not unknown constants but latent stochastic variables following the random walk dynamics (4). The estimation of this model was performed using state space methods and the Kalman filter [10, 6].

The second class is a Markov regime-switching model (MS) defined as:

$$p_{jt} = \beta_{jS_t}' \mathbf{X}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim N(0, \sigma_{jS_t}^2), \quad (5)$$

$$\Pr(S_t = i | S_{t-1} = h) = \pi_{ih}, \quad \forall i, h \in S \quad (6)$$

where S_t is the latent regime at time t , $S = \{1, 2\}$ is the set of possible states (say, base and peak), β_{jS_t} is a $k \times 1$ vector of coefficients in regime S_t , $\mathbf{X}_{jt}$ is a $k \times 1$ vector

of regressors, $\sigma_{js_t}^2$ is the error variance in regime S_t and π_{ih} is the transition probability between states i and h . This class of models assumes that the market at each time point is in one of 2 possible states, indexed by the unobservable discrete variable S_t , which evolves according to a first-order irreducible homogeneous ergodic Markov process. Each market regime is characterized by a distinct regression price model, i.e. the model parameters are a function of the prevailing state S_t at each time point. Maximum likelihood estimates of β_{js_t} and $\sigma_{js_t}^2$ are performed using the EM algorithm while for smoothed inferences of regimes, Kim's algorithm was used [10, 12]. Once a MS model has been estimated, price forecasts are the linear combination of predicted prices across regimes weighted by predicted regime probabilities:

$$\hat{p}_{t+1} = \sum_{i=1}^2 \hat{p}_{t+1}^i \cdot \hat{\Pr}(S_{t+1} = i | I_t)$$

In our work, the parameters of the Markov regime-switching models were also estimated on rolling windows with fixed length of 6 months (MS6).

4 Forecasting results

The predictive accuracy of the four classes of models was assessed, overall and for each season, by means of some commonly used measures: the Mean Squared Error (MSE), the Mean Squared Percentage Error (MSPE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). These prediction error statistics are evaluated on the original spot prices. The analysis was performed for two study cases: one in which models are estimated on the adjusted log prices and another one in which models are estimated on the same series but without spikes.

Prediction accuracy results, with and without spikes, indicate that no model outperforms all the others in all the situations. Differences in forecasting accuracy depend on several factors, such as the considered load period, the forecasting period and the performance indicator. Moreover, generally, different seasons show different models that prevail for best predictive performance and often the results depend on the particular statistics that we are using for evaluate the predictive accuracy, so that no clear conclusion is possible. However, we note that the ARMAX model gives the best results in 44% of cases, followed by TVR which is the best in 26% of cases. These percentages go up to 60% and 28% respectively if we consider the estimation without spikes. It follows the Markov regime-switching model estimated on a rolling window of 6 months which is the best forecasting model in 13% of cases, but only in 1% if we exclude spikes. This behaviour can be explained considering that the MS6 is a specific model for capturing spikes, while ARMAX and TVR not, so they benefit from the exclusion of high jumps. Thus, ARMAX model produces better forecasts in most cases, even if it is the simplest among the considered predictive models. So it is interesting to conduct the analysis on the predictive differences with respect to the ARMAX model. As an example, table 1 contains forecasting results

Table 1 *Forecasting results for load period 38, expressed in terms of ratios between the prediction error statistics values of the best models and the statistics values of ARMAX.*

Forecasting Period		R _{MSE}	R _{MSPE}	R _{MAE}	R _{MAPE}
Estimation with spikes					
January-March	Statistics Values	0.946	0.849	0.934	0.918
	Best Model	MS	MS6	MS6	MS6
April-June	Statistics Values	1.000	1.000	1.000	1.000
	Best Model	ARMAX	ARMAX	ARMAX	ARMAX
July-September	Statistics Values	0.543	0.781	0.867	0.920
	Best Model	TVR	TVR	TVR	TVR
Whole Period	Statistics Values	0.948	1.000	0.992	1.000
	Best Model	MS	ARMAX	MS6	ARMAX
Estimation without spikes					
January-March	Statistics Values	1.000	1.000	1.000	1.000
	Best Model	ARMAX	ARMAX	ARMAX	ARMAX
April-June	Statistics Values	1.000	0.942	0.976	0.965
	Best Model	ARMAX	MS	TVR	MS
July-September	Statistics Values	0.732	0.992	0.967	1.000
	Best Model	TVR	TVR	TVR	ARMAX
Whole Period	Statistics Values	1.000	1.000	1.000	1.000
	Best Model	ARMAX	ARMAX	ARMAX	ARMAX

expressed in terms of ratios between the prediction error statistics value of the best models and the statistics value of ARMAX. We show results for period 38, that is a load period with very high volatility and peaks. As pointed out above, there is no agreement about which predictive model is the best, neither among the seasons nor among the prediction error statistics. MS6 model shows good performances during the winter season: it can reach spikes that the other models cannot. But if we remove spikes from the series during the calibration of the model, it turns out to be penalized in terms of forecasting accuracy. Models ARMAX and TVR are mostly the bests.

In order to assess the statistical significance of the predictive differences among the models, we performed the Diebold and Mariano test [5] for equal prediction accuracy (5% significance level). Results indicate that ARMAX model produces significantly better forecasts with respect to the other models during, for example, spring in load period 18, 28 (estimation with spikes), and spring in load periods 18, 28, whole in load period 44 if estimated without spikes. Furthermore, no model is significantly better than the ARMAX (detailed results of the D-M test are available at http://homes.stat.unipd.it/lisif/?file=NBBL_wp.pdf).

We considered also a direct comparison between forecasting accuracy of each model estimated on the detrended log price series with spikes and the same model estimated on the series without spikes. The scope is to analyze if removing spikes helps to produce better forecasts. The exclusion of the peaks from the data produces

better predictive results in the 66% for the ARMAX model, 70% for the linear regression, 56% for the MS model, 36% for the MS6 model and 65% of the cases for the TVR model. The Markov regime-switching model appears to be disadvantaged because of despiking. The other models behave better even if during the peak load periods (18, 28, 38) predictive performance are often negatively affected by the exclusion of spikes. The Diebold-Mariano test highlights that these improvements are statistically significant only in very few cases clustered in load periods 6 and 44.

Thus, it seems that simpler linear models can perform better in terms of forecasting accuracy than more complex nonlinear methods. This is in line with findings in [4]: while nonlinear models are superior to linear ones in capturing certain features of the data series, from a forecasting perspective linear models reveal to be robust for prediction, even when the data are generated by a nonlinear model.

References

1. Aggarwal, S.K., Saini, L.M., Kumar, A.: Electricity price forecasting in deregulated markets: A review and evaluation. *Electrical Power and Energy Systems* **31**, 13–22 (2009)
2. Bunn, D.W.: *Modelling prices in competitive electricity markets*. Wiley (2004)
3. Cartea, A., Figueroa, M.G.: Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Appl. Math. Finance* **12**(4), 313–335 (2005)
4. Clements, M.P., Krolzig, H.M.: A Comparison of the Forecast Performance of Markov-switching and Threshold Autoregressive Models of US GNP. *Econometrics Journal*, **1**(1), C47–C75 (1998).
5. Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. *J. Bus. Econ. Statist.* **13**, 253–263 (1995)
6. Durbin, J., Koopman, S.J.: *Time Series Analysis by State Space Methods*. Oxford University Press (2001)
7. Friedman, J.H.: A variable span scatterplot smoother. Technical Report No.5, Laboratory for Computational Statistics, Stanford University (1984)
8. Geman, H., Roncoroni, A.: Understanding the fine structure of electricity prices. *J. Bus.* **79** (2006)
9. Haldrup, N., Nielsen, M.Ø.: A regime switching long memory model for electricity prices. *Journal of Econometrics* **135**, 349–376 (2006)
10. Hamilton, J.D.: *Time Series Analysis*. Princeton University Press, Princeton, NJ (1994)
11. Karakatsani, N., Bunn, D.W.: Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *Int. J. Forecasting* **24**, 764–785 (2008)
12. Kim, C.J.: Dynamic linear models with Markov-Switching. *J. Econometrics*, **60**, 1–22 (1994)
13. Kosater, P., Mosler, K.: Can markov regime-switching models improve power-price forecasts? evidence from german daily power prices. *Appl. Energy* **83**, 943–958 (2006)
14. Misiorek, A., Trueck, S., Weron, R.: Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. *Stud. Nonlinear Dynam. Econometrics* **10**(3), Article 2 (2006)
15. Pedregal, D.J., Trapero, J.R.: Electricity prices forecasting by automatic dynamic harmonic regression models. *Energy Convers. Manage.* **48**, 1710–1719 (2007)
16. Pilipovic, D.: *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill, New York (1998)
17. Weron, R.: *Modeling and Forecasting Electricity Loads and Prices*. Wiley, New York (2006)
18. Weron, R., Bierbrauer, M., Trück, S.: Modeling electricity prices: jump diffusion and regime switching. *Physica A* **336**, 39–48 (2004)