Working Paper Series, N. 4, April 2010

Department of Statistical Sciences

*University of Padua*

*Italy*

# The Forecasting Accuracy of Electricity Price Formation Models

**Fany Nan**[1], **Silvano Bordignon**[1], **Derek W. Bunn**[2], **Francesco Lisi**[1]

[1] Department of Statistical Sciences, University of Padua,
Via Cesare Battisti 241, 35121 Padova, Italy.

[2] Department of Management Science and Operations,
London Business School, Regent's Park, London NW1 4SA UK

**Abstract:** In this paper we present an extensive comparison of four different classes of models for daily forecasting of spot electricity prices, including ARMAX, constant and time-varying parameter regression models as well as non linear Markov regime-switching regressions. They are selected for particular reasons related to the emerging body of research on the price formation processes observed in electricity markets. The analyses are conducted for representative trading periods of the day in the UK Power Exchange prompt market, with the price series adjusted for their deterministic components and spikes. They show that relative out-of-sample forecasting performances are distinctly different for each trading period, season and across the actual performance metrics. No model consistently outperforms the others, but the ARMAX approach performs well in most cases and the Diebold and Mariano test indicates that, when it is not the best, the ARMAX model is not statistically different from the best. Nevertheless, we suggest that subtle differences in performance between different methods under different conditions are consistent with the apparent variations in the price formation processes by time of day and by season. We conclude with some observations on the disparities between the model specifications appropriate for understanding in-sample price formation and those for accurate out-of-sample predictions.

**Keywords:** Forecasting, Electricity, Prices, ARMAX, Regime-Switching, Time-Varying Parameters, Accuracy.

UNIVERSITÀ DEGLI STUDI DI PADOVA DIPARTIMENTO DI SCIENZE STATISTICHE

**Final version (2010-04-27)**

# Contents



**Department of Statistical Sciences**
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

**Corresponding author:**
Fany Nan
tel: +39 049 827 4124
fany.nan@stat.unipd.it
http://homes.stat.unipd.it/fany

# The Forecasting Accuracy of Electricity Price Formation Models

**Fany Nan**[1], **Silvano Bordignon**[1], **Derek W. Bunn**[2], **Francesco Lisi**[1]
[1] Department of Statistical Sciences, University of Padua,
Via Cesare Battisti 241, 35121 Padova, Italy.

[2] Department of Management Science and Operations,
London Business School, Regent's Park, London NW1 4SA UK

**Abstract:** In this paper we present an extensive comparison of four different classes of models for daily forecasting of spot electricity prices, including ARMAX, constant and time-varying parameter regression models as well as non linear Markov regime-switching regressions. They are selected for particular reasons related to the emerging body of research on the price formation processes observed in electricity markets. The analyses are conducted for representative trading periods of the day in the UK Power Exchange prompt market, with the price series adjusted for their deterministic components and spikes. They show that relative out-of-sample forecasting performances are distinctly different for each trading period, season and across the actual performance metrics. No model consistently outperforms the others, but the ARMAX approach performs well in most cases and the Diebold and Mariano test indicates that, when it is not the best, the ARMAX model is not statistically different from the best. Nevertheless, we suggest that subtle differences in performance between different methods under different conditions are consistent with the apparent variations in the price formation processes by time of day and by season. We conclude with some observations on the disparities between the model specifications appropriate for understanding in-sample price formation and those for accurate out-of-sample predictions.

**Keywords:** Forecasting, Electricity, Prices, ARMAX, Regime-Switching, Time-Varying Parameters, Accuracy.

## 1 Introduction

Whilst price forecasting is clearly an important activity for managing operational, financial and trading risks in the liberalised electricity sector, the substantive body of research which has emerged to model power price formation does not generally have a predominantly predictive orientation. The research challenges of specifying adequate econometric models to describe the delicate, nonlinear and evolutionary interactions of fundamental and market conduct variables known to influence power price formation have sustained an increasing amount of work aimed primarily at describing the ex post properties of the market prices, the conduct of participants and questions of market efficiency (Bunn, 2004; Weron, 2006). How well these models perform out-of-sample for predictive purposes is therefore still incompletely understood. Furthermore, the forecasting models which have been published offer limited insights into their comparative performances, often being conducted in very different contexts (see Aggarwal et al., 2009), and to the extent that comparisons are included, no consistent conclusions have been emerging.

The published work is methodologically quite wide and includes purely time series forecasting models (Crespo Cuaresma et al., 2004; Conejo, Contreras, Espínola and Plazas, 2005; Conejo, Plazas, Espínola and Molina, 2005; Weron and Misiorek, 2005), the performances of which are generally improved with exogenous variables (eg Nogales et al., 2002; Contreras et al., 2003; Knittel and Roberts, 2005; Weron and Misiorek, 2005 and Misiorek et al., 2006). Power prices often show heteroskedasticity (Guirguis and Felder, 2004; Garcia et al., 2005; Knittel and Roberts, 2005; Misiorek et al., 2006; Bowden and Payne, 2008) and irregular spikes (Duffie et al., 1998) motivating price specifications involving GARCH (Koopman et al., 2007), jump components (Escribano et al., 2002; Karakatsani and Bunn, 2004), or regime switching (Kosater and Mosler, 2006; Misiorek et al., 2006). Time-varying parameter models (Pedregal and Trapero, 2007; Karakatsani and Bunn, 2008) appear particularly attractive (Granger, 2008) in capturing the evolutionary nature of power markets. Cross-comparisons within this body of work are difficult, however, because basic specifications vary (working with prices or log prices are equally common; unit root tests sometimes indicate mean reversion, sometimes the need to model returns; the deterministic components are incorporated in different ways and whether spikes are modelled or excluded are also discretionary modelling choices). Moreover markets are idiosyncratic (eg PJM, EEX, and NordPool have many different characteristics), and in terms of experimental designs, the out-of sample forecasting periods are sometimes insufficient to compare conclusions, recursive re-estimations are often not undertaken and the reported prediction error statistics also vary. Furthermore, if comparisons are undertaken, their statistical significances are often untested. More fundamentally, however, in addition to these general issues of methodological comparability, which are common to most areas of applied forecasting research, as it is becoming clear that the price formation processes for spot electricity may vary by time of day and by season (Karakatsani and Bunn, 2008), so it is plausible that there may be systematic, periodic, variations in the relative appropriateness of different forecasting methods as well. To the extent that such an effect exists, it may well confound simple cross-comparisons if it were not part of their experimental designs.

In this paper, we report a detailed forecasting comparison that involves four different classes of models, each one characterised by specific generalisable features. Since it would be possible to suggest a very wide range of different models, we focus instead upon some critical characteristics. One approach follows from the proposition that a well specified ARMA model, applied for short term forecasting, should be able to incorporate indirectly the time series effects of many exogenous variables, particularly the slowly moving ones. We do not explore this approach in its simplest form, however, since it is very well-known that the reserve margin is a crucial variable that determines the competitiveness of pricing, and to the extent that the market operator may provide timely forecasts of this variable, it will be valuable. So, we propose an ARMAX, with this extra variable, but the approach is essentially one of using ARMA to indirectly reflect the many other driving influences. The second approach is a constant parameter linear regression model (LR), with explicit representation of some of the possible exogenous driving variables. The linear model has often been advocated in forecasting out-of-sample, even if nonlinear models have been shown to fit better in-sample (Kosater and Mosler, 2006; Misiorek et al., 2006). The third approach relaxes the constant parameter aspect of the explicit regression approach to use time-varying coefficients. We specified this with random-walk coefficients, allowing for price dynamics that continuously adapt as the price formation process evolves. Finally the stylised fact of power prices being spiky suggests that a Markov regime-switching approach would be most suitable for the irregular, but repeated, discontinuities in the price series, distinguishing between normal and high-price regimes. We undertake a direct comparison of these four distinct approaches for both fitting and forecasting day ahead UK power prices, with a clear focus upon time of day and seasonal characterisation.

The paper is organized as follows. In section 2 we present a preliminary data analysis of the UK electricity prices and the methods that we use for processing the price series. We describe also the price drivers included in the models as regressors. Section 3 focuses on the description of the models, while section 4 includes the experimental design. Section 5 presents the comparative results obtained through various prediction error statistics and the Diebold-Mariano significance tests. Section 6 concludes.

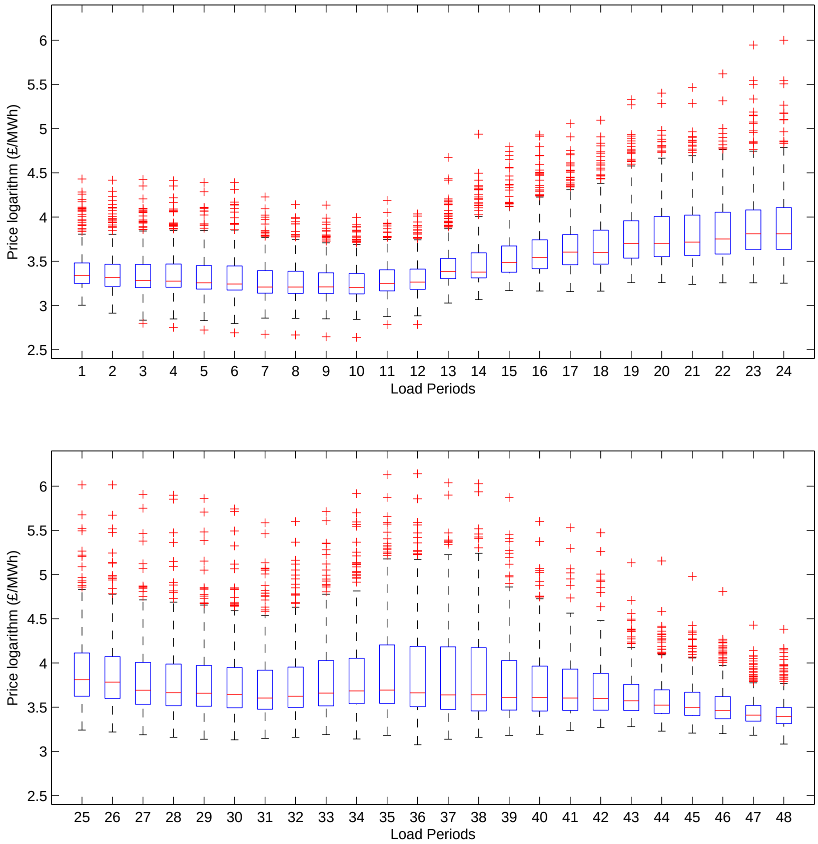
# 2 Data Analysis of the UKPX electricity market

## 2.1 Preliminary data analysis

This work considers the prices for the half-hourly trading periods in the British wholesale power market, which is generally regarded as the most mature and competitive in Europe. These are the reference prices from the UKPX, a power exchange which offered continuous bilateral trading from day ahead to an hour ahead of real-time. The data start from April 1st, 2005, when the market had just been extended to include Scotland, to September 30th, 2006. Each day consists of 48 load periods: period 1 is defined as 00:00-00:30am, period 2 as 00:30-01:00am, and so on up to period 48 (23:30-00.00pm). We denote by $P_{jt}$ the spot price at time $t$ and load period $j$ ($t = 1, 2, ..., N$, $j = 1, 2, ..., 48$). All weekends and holidays were removed from the data, yielding 380 days for each load period. The load profiles are quite different for those days, and in dropping them from the analysis, there was no significant loss of information (Ramanathan et al., 1997; Karakatsani and Bunn, 2008). The prices are spiky, mainly due to occasional supply shocks, and to stabilise variance, logarithmic transformations were used.

Figure 1 shows the distributions of the 48 half-hourly log price series. The usual daily cycle appears quite clearly. In particular, we observe very high prices at about 17:00-19:00pm (peak hours), corresponding to load periods 35-38: this evening peak is more pronounced in winter (see figure 2). Moreover, not only the average level of prices but also their variability depends on the load period. Power markets op-

**Figure 1:** *Boxplots of price logarithms for the 48 load periods.*




erate with low marginal cost generators providing "baseload power" throughout the day, while flexible plants, typically with high marginal costs, are used only during peak hours. Consequently, prices show an extremely high volatility on a daily basis and because of nonstorability, electricity products traded in different hours really constitute separate commodities. Formulating the time series problem as a sequence of day-by-day observations for a particular trading period, rather than as a sequence of trading periods throughout the day, has become well-established for electricity loads and prices (Ramanathan et al., 1997; Bunn, 2000; Bunn and Karakatsani, 2003). The improvement in fitting and prediction accuracy is a result of the increase in homogeneity of the day-by-day time series for a particular period in comparison with the contiguous period-by-period sequence. For next-day price forecasting, 48 one-step-ahead forecasts calculated everyday, contain less noise than forecasts with prediction horizons varying from 1 to 48. Thus, our models were estimated separately for each load period. In particular, we used five representative periods of the day: load periods 6 (02:30-03:00am), 18 (08:30-09:00am), 28 (13:30-14:00pm), 38 (18:30-19:00pm) and 44 (21:30-22:00pm).

As figure 2 shows, the log price series have a deterministic component linked to variations in demand. The night-time load period 6 is more stable. In periods 18, 28 and 38 volatility is very high, with sudden peaks during winter and summer in both 2005 and 2006. The deterministic component changes according to the load period. Some authors prefer to include it in the models (Karakatsani and Bunn, 2008), others prefer to remove it and to work on the adjusted price series (Geman and Roncoroni, 2006; Misiorek et al., 2006). We followed this second approach. Thus, we removed the deterministic component from the log price series and we estimated the models on the filtered series. Common tools for modelling deterministic components include functions with dummies (Haldrup and Nielsen, 2006), functions of time (Weron et al., 2004; Cartea and Figueroa, 2005) using sinusoidal approaches (Pilipovic, 1998) or a combination of both (Kosater and Mosler, 2006; Misiorek et al., 2006). However, these are idiosyncratic to the data. In fact our prices show the effects of two components: one affected by the seasonal use of lighting and heating in winter and to a lesser extent by the increasing use of air conditioning in summer, and the other caused by long run market behaviour. Instead of using a parametric method we pursued the Friedman's Supersmoother (Friedman, 1984), which is a very flexible method to estimate a deterministic component of a time series. It is a nonparametric technique based on the nearest neighbor method characterised by specific procedures for the selection of the smoothing parameter.

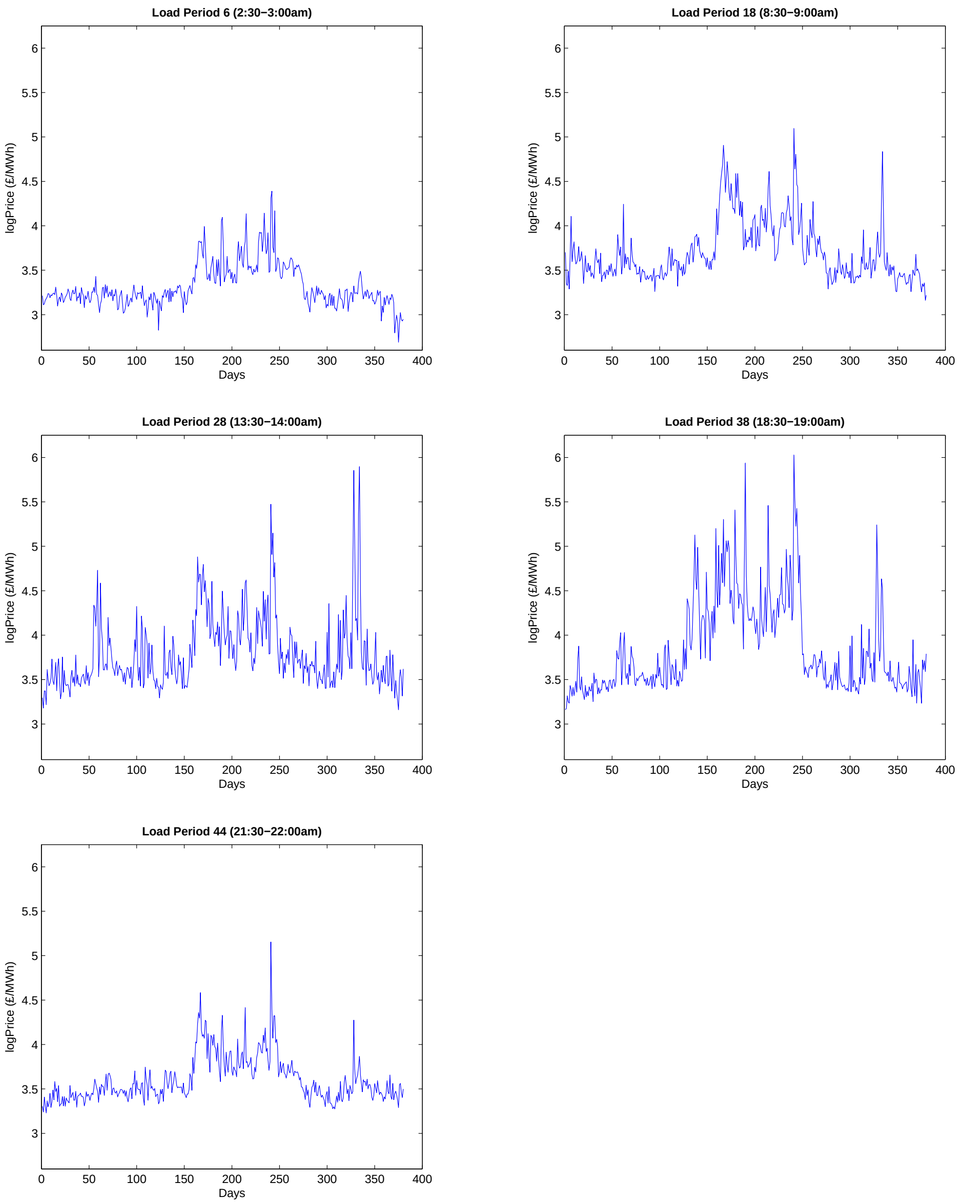
Let $D_{jt}$ be the estimate of the deterministic component obtained applying the Friedman's Supersmoother to the log price series $\log P_{jt}$. Then, the adjusted series of log prices is given by $p_{jt} = \log P_{jt} - D_{jt}$ ($t = 1, 2, ..., N$ and $j = 6, 18, 28, 38, 44$). Table 1 displays some descriptive statistics of the adjusted log price series $p_{jt}$ for the

**Table 1:** *Descriptive statistics for adjusted log price series, load periods 6, 18, 28, 38 and 44 over the period April 2005 - September 2006.*

| | **Period 6** | **Period 18** | **Period 28** | **Period 38** | **Period 44** |
|---|---|---|---|---|---|
| **mean** | 0.002 | 0.004 | 0.002 | <0.001 | 0.001 |
| **st.deviation** | 0.128 | 0.191 | 0.291 | 0.285 | 0.141 |
| **skewness** | 1.119 | 1.694 | 1.926 | 1.808 | 2.342 |
| **kurtosis** | 7.250 | 10.075 | 9.967 | 8.902 | 17.185 |

five load periods. Load period 6 has the lowest standard deviation value, followed by load period 44, 18, 38 and 28. This is reasonable considering that period 6 is an off-peak hour, while load periods 28 and 38 are super-peak hours. The high variance of the peak hours is due to spikes. The values of skewness and kurtosis show that all the periods are characterized by positive asymmetry and fat tails, and they deviate considerably from normality. Augmented Dickey-Fuller (Said and Dickey, 1984) and KPSS (Kwiatkowski et al., 1992) tests are applied to the log prices before and after removing the deterministic component to assess the stationarity of the series. Table 2 shows results for the unit root tests. If the tests applied to the log price series show non stationarity, after adjusting for the deterministic component, Augmented Dickey-Fuller unit root test for $p_{jt}$ across the five load periods rejected the unit root

**Figure 2:** *Electricity price logarithms for the considered load periods from April 2005 to September 2006.*



null hypothesis at the 1% significance level. In the same way, KPSS stationarity test did not reject the null hypothesis at the 5% significance level.

**Table 2:** *Unit root tests results for log prices and adjusted log price series in the considered load periods.*

| | Log Prices | | Adj Log Prices | |
|---|---|---|---|---|
| **Load Period** | **ADF** | **KPSS** | **ADF** | **KPSS** |
| 6 (02:30-03:00am) | $-1.981$ | $0.958^{*}$ | $-7.795^{*}$ | 0.015 |
| 18 (08:30-09:00am) | $-2.973$ | $0.829^{*}$ | $-6.917^{*}$ | 0.017 |
| 28 (13:30-14:00pm) | $-3.537^{**}$ | $0.417^{*}$ | $-6.372^{*}$ | 0.015 |
| 38 (18:30-19:00pm) | $-2.442$ | $1.002^{*}$ | $-7.309^{*}$ | 0.014 |
| 44 (21:30-22:00pm) | $-2.455$ | $0.914^{*}$ | $-7.555^{*}$ | 0.016 |

*Note: *, ** means that the null hypothesis is rejected at 1% and 5% significance level respectively. ADF stands for augmented Dickey-Fuller test. Lag lengths are chosen following Ng and Perron (1995) method.*

As figure 2 shows, our series present a number of very high, sudden spikes, which will affect model estimation and, consequently, the forecasting experiments. A characteristic feature of these spikes is that prices fall back to normal levels almost immediately when the weather condition or outage that caused the peak is over. In Weron and Przybyłowicz (2000) and Weron (2002) the R/S analysis, detrended fluctuation analysis and periodogram regression methods were used to verify anti-persistence in electricity prices. One approach is to treat spikes as outliers and use some procedure to preprocess the data, as in Conejo, Contreras, Espínola and Plazas (2005) and Weron and Misiorek (2008). We decided to conduct our analyseis on the series with and without spikes. For our despiked analyses, we followed Weron (2006), and did not cut the spikes at a specified threshold, but dampened them by differentiating between jumps and extreme jumps. The method is iterative (until a stop criterion was satisfied) on the spot price series $P_{jt}$. At each iteration a threshold is set. 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. For period $j$, the procedure at the $i$-th iteration is as follows:

1. remove the deterministic component from the spot price series $P^i/\exp(D^i)$;
2. set the threshold $T^i = \mu^i + 3\sigma^i$ where $\mu^i$ and $\sigma^i$ are respectively the mean and standard deviation calculated from the adjusted price series;
3. for $t = 1, ..., N$, if $P_t^i > T^i$, set $P_t^{(i+1)} = T^i + T^i \log_{10}(P_t^i/T^i)$, else $P_t^{(i+1)} = P_t^i$;
4. insert $\exp(D^i)$ in the new price series $P_t^{(i+1)}$ and recalculate the deterministic component as $D^{(i+1)}$;
5. if the maximum difference between the prices of the old series and the new one is bigger than 0.1, restart from point 1 ($i = i + 1$), else stop.

Section 5 contains the forecasting results for the models estimated on the adjusted logarithmic price series with and without spikes. Moreover, results from a direct comparison between the two studies are presented.

### 2.2 Market data

As Karakatsani and Bunn (2008) observed on the UK market, there is a strong linkage between prices and market fundamentals. In our research we considered the following variables (in logs):

**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** ($margin_t$).This 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 $margin$. To resolve collinearity, at every estimation step we demeaned the variables and then we calculated the quadratic components, denoted as $demF^2$ and $margin^2$.

**Gas Price** . This is the daily UK natural gas one-day forward price, from the main National Balancing Point (NBP) hub. At the time, the UK power market was widely recognised as being a "spark spread" market, i.e. driven by underlying gas prices. This is also in accord with Serletis and Shahmoradi (2006) who observed that the general relation between gas and electricity price is strong, not only on the mean but also on the variance. For consistency, our models include the series of deviations ($gasF.res_t$) of gas prices from their deterministic components calculated with the Friedman's Supersmoother.

For autocorrelation and heteroskedasticity, we included the following variables:

**Past Prices** ($p_{t-j}$). They are the lagged spot prices. In particular, lags 1 and 5, corresponding to a daily and weekly lag were considered.

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

## 3 Predictive models

The first general model class that we have considered is the AutoRegressive Moving Average model with exogenous variables ARMAX$(p, q, m_1, ..., m_k)$. The current value $y_t$ of the time series is expressed linearly in terms of its $p$ past values, of

$q$ past shocks and of some past values of $k$ exogenous variables. A general formula for the ARMAX$(p, q, m_1, ..., m_k)$ model can be compactly written as:

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t + \sum_{i=1}^{k} \beta^i(B)z_t^i, \quad \varepsilon_t \sim WN(0, \sigma^2), \tag{1}$$

where $\varepsilon_t$ is the error term, $z^1, ..., z^k$ are the exogenous variables, $B$ is the lag operator, $\beta^i(B) = \beta_0^i + \beta_1^i B + \cdots + \beta_{m_i}^i B^{m_i}$, $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$.
For our dataset the identified model is the ARMAX(1,1,1)

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

where the exogenous variable $z_{jt}$ is the indicated margin. $\phi_j, \theta_j, \beta_j$ are constant coefficients. This model, estimated through maximum likelihood, is the simplest among our models.
For each load period $j$, the out-of-sample one-day ahead price forecast is given by:

$$f_{j(t+1)} \quad = \quad \hat{\phi}_j p_{jt} + \hat{\theta}_j \hat{\varepsilon}_{jt} + \hat{\beta}_j z_{j(t+1)}.$$

The second class of model is linear regression (LR), which explicitly accounts for the relation between prices and price drivers. The model is specified as:

$$p_{jt} = \beta_j' \mathbf{X}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim \text{i.i.d}(0, \sigma_j^2) \tag{3}$$

where $\beta_j$ is a $k \times 1$ vector of constant coefficients and $\varepsilon_{jt}$ is an i.i.d. error term. $\mathbf{X}_{jt}$ is a $k \times 1$ vector of regressors selected in-sample with stepwise backward identification (AIC criterion). The regressors are listed in Table 3. These final sets of regressors are also used for models (4) and (7). Only regressors $p_{t-1}$ and $margin_t$ were significant

**Table 3:** *Final sets of regressors obtained with stepwise backward techniques.*

| | **Period 6** | **Period 18** | **Period 28** | **Period 38** | **Period 44** |
|---|---|---|---|---|---|
| $intercept$ | √ | √ | √ | √ | √ |
| $p_{t-1}$ | √ | √ | √ | √ | √ |
| $demF_t$ | — | √ | √ | √ | — |
| $demF_t^2$ | — | √ | — | √ | — |
| $margin_t$ | √ | √ | √ | √ | √ |
| $margin_t^2$ | √ | — | — | √ | — |
| $gasF.res_t$ | √ | √ | — | — | √ |
| $demVol_t$ | √ | — | — | — | — |
| $priceVol_t$ | — | — | — | — | √ |

for all the load periods. The demand (and its quadratic term) is not significant in load periods 6 and 44, as it affects prices mainly during peak hours.
At each time $t$ parameters $\boldsymbol{\beta}_j$ are estimated by OLS using an expanding dataset,

then out-of-sample one-day ahead price forecasts for the adjusted series are obtained as:

$$f_{j(t+1)} \quad = \quad \hat{\boldsymbol{\beta}}_j' \mathbf{X}_{j(t+1)}.$$

Model (3) assumes that the parameters, $\boldsymbol{\beta}$, are constant over the estimation sample. However, the application of the CUSUM and CUSUMSQ tests for stability (see Brown et al., 1975) shows that there is strong evidence of instability in the parameters for all the models. Thus, it may be useful to model the dynamics of parameter evolutions. To consider also nonlinearity and heteroskedasticity issues, we used two methods: one in which the changes in the parameters are assumed to be generated by a random walk (time-varying parameter, TVP, regression model) and one in which the changes are determined by a discrete variable which evolves according to a Markov process (a Markov regime-switching, MS, model).

The TVP approach would appear to be most suited to the situations where the response of the prices to the various market fundamentals may change continuously. This is specified as:

$$p_{jt} = \boldsymbol{\beta}_{jt}' \mathbf{X}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim WN(0, \sigma^2_{\varepsilon_j}), \tag{4}$$

$$\boldsymbol{\beta}_{j(t+1)} = \boldsymbol{\beta}_{jt} + \boldsymbol{\nu}_{jt}, \quad \boldsymbol{\nu}_{jt} \sim WN_k(0, \mathbf{H}_j), \tag{5}$$

where $\boldsymbol{\beta}_{jt}$ is a $k \times 1$ vector of coefficients and $\mathbf{X}_{jt}$ the $k \times 1$ vector of regressors. $\varepsilon_{jt}$ is the error term of the measurement equation, while $\boldsymbol{\nu}_{jt}$ is the error term vector of the transition equation, $\mathrm{E}(\varepsilon_{jt}, \boldsymbol{\nu}_{jt}) = 0$ and $\mathbf{H}_j = \mathrm{diag}\{\sigma^2_{\nu_{jk}}\}$.
The estimation of this model was performed using state space methods and the Kalman filter (Hamilton, 1994 and Durbin and Koopman, 2001).
The above formulation can be written in a state space form:

$$\begin{pmatrix} \boldsymbol{\beta}_{j(t+1)} \\ p_{jt} \end{pmatrix} = \boldsymbol{\Phi}_{jt} \boldsymbol{\beta}_{jt} + \boldsymbol{\mu}_{jt}, \tag{6}$$

where $\boldsymbol{\Phi}_{jt} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{X}_{jt}' \end{pmatrix}$, $\boldsymbol{\mu}_{jt} = \begin{pmatrix} \boldsymbol{\nu}_{jt} \\ \varepsilon_{jt} \end{pmatrix} \sim WN_{k+1}(0, \boldsymbol{\Omega}_j)$ and $\boldsymbol{\Omega}_j = \begin{pmatrix} \mathbf{H}_j & 0 \\ 0 & \sigma^2_{\varepsilon_j} \end{pmatrix}$.
We choose $\boldsymbol{\beta}_{j1} \sim WN_k(\mathbf{a}, \mathbf{P})$ as initial values. Since $\boldsymbol{\beta}_{jt}$ is $I(1)$ the initial state vector does not have finite variance and so the Kalman filter has to be initialised using diffuse priors. This procedure assigns very large initial value to the covariance matrix while the initial values of the time varying coefficients are arbitrarily chosen. We set $\mathbf{a} = \mathbf{0}$ and $\mathbf{P} = \kappa \mathbf{I}_k$ where $\kappa$ is large ($\kappa = 10^6$).
For each load period $j$, the out-of-sample one-day ahead spot price forecasts are obtained as:

$$f_{j(t+1)} \quad = \quad \hat{\boldsymbol{\beta}}_{j(t+1|t)}' \mathbf{X}_{j(t+1)}.$$

The presence of jumps in electricity price series suggests that, distinct from a continuously evolving structure, there could more appropriately be a discontinuous non-linear functionality switching between normal and high-price regimes. The most common modelling approach to this is the Markov regime-switching model (MS) defined as:

$$p_{jt} = \boldsymbol{\beta}'_{jS_t}\mathbf{X}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim WN(0, \sigma^2_{jS_t}), \tag{7}$$

$$\Pr(S_t = i|S_{t-1} = h) = \pi_{ih}, \ \forall i, h \in S \tag{8}$$

where $S_t$ the latent regime at time $t$, $S = \{1, 2\}$ the set of possible states (say, base and peak), $\boldsymbol{\beta}_{jS_t}$ a $k \times 1$ vector of coefficients in regime $S_t$, $\mathbf{X}_{jt}$ a $k \times 1$ vector of regressors, $\sigma^2_{jS_t}$ the error variance in regime $S_t$ and $\pi_{ih}$ the transition probability between states $i$ and $h$.

This class of models assumes that the market at each time point is in one of the 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 characterised by a distinct regression price model, i.e. the model parameters are a function of the prevailing state $S_t$ at each time point. Prices are classified into regimes endogenously through the latent state estimation and probabilistic inference. Maximum likelihood estimates of $\boldsymbol{\beta}_{jS_t}$ and $\sigma^2_{jS_t}$ are performed using the EM algorithm while for smoothed inferences of regimes, Kim's algorithm was used (Hamilton, 1994; Kim, 1994).

Parameters $\boldsymbol{\beta}_{jS_t}$ and $\sigma^2_{jS_t}$ are estimated both on a daily expanding dataset (MS) and on a rolling window of 6 months (MS6).

Once a MS model has been estimated, price forecasts are the linear combinations of predicted prices across regimes weighted by predicted regime probabilities:

$$\begin{aligned} f_{j(t+1)} &= \sum_{i=1}^{2} f^i_{j(t+1)} \cdot \hat{\mathrm{P}}(S_{t+1} = i|I_t) \\ &= \sum_{i=1}^{2} f^i_{j(t+1)} \cdot \left[\sum_{h=1}^{2} (\Pr(S_{t+1} = i|S_t = h)\Pr(S_t = h|I_t))\right]. \end{aligned}$$

## 4 Experimental design

In order to make out-of-sample predictions, the data have been split into an in-sample period (April 1st, 2005 - December 31th, 2005) and an out-of-sample period (January 1st, 2006 - September 30th, 2006). Moreover, the out-of-sample period was divided in three sub-periods, associated with the different seasons (January-March, 64 data, April-June, 61 data and July-September, 64 data), in order to understand how much the forecasting accuracy of our models is influenced by the period of the year.

To formulate on day $t$ a price forecast for period $j$ on day $t + 1$, the parameters of the models were estimated at each step from a daily expanding dataset or from rolling windows of specified lengths.

The forecasting experiment is carried out for all the models described in section 3 estimated on both the basic filtered log price series and the filtered log price series without spikes. However, predictions are always made in terms of the original spot

prices.
To compare forecasting results we used 4 prediction error statistics:

$$\text{MSE} = \frac{1}{m}\sum_{t=1}^{m}(P_t - F_t)^2 \qquad \text{MSPE} = \frac{1}{m}\sum_{t=1}^{m}\left(100 \times \frac{P_t - F_t}{P_t}\right)^2$$

$$\text{MAE} = \frac{1}{m}\sum_{t=1}^{m}|P_t - F_t| \qquad \text{MAPE} = \frac{1}{m}\sum_{t=1}^{m}\left|100 \times \frac{P_t - F_t}{P_t}\right|$$

where $m$ is the size of the forecasting period, $P_t$ is the observed spot price at time $t$ and $F_t$ is the forecast at time $t$. We obtain the forecasts for the spot prices using the inverse transformation from the filtered log price, that is through $F_t = \exp(f_t + D_t)$. The Mean Squared Error (MSE) is popular, largely because of its closeness to variance and its consequent theoretical relevance in statistical modelling. However, it is more sensitive to jumps than the Mean Absolute Error (MAE). Percentage errors have the advantage of being scale-independent and, in our case with very high spikes, this could be important.
One of the contributions of this paper is to test if there are significant differences in forecasting accuracy among models. For this we used the Diebold-Mariano test (Diebold and Mariano, 1995), whose null hypothesis is of no differences in the accuracy of two competing forecasts. It works under quite general loss functions and errors are allowed to be non-Gaussian as well as serially and cross-correlated. In our case, the test was carried out using squared error (adjusted Diebold-Mariano test, see Harvey et al., 1997) and absolute error loss functions.

## 5 Forecasting results

The first step is the evaluation of the in-sample performances (fitting) for the four classes of models. At this step, we omit MS6, because it is a variant of MS and it gave similar results. All of our results, fitting and forecasting, refer to the original data, not the smoothed series. With respect to despiking, Table 4 shows that the performance indicators are not dramatically affected by adjusting for spikes and, rather surprisingly, they appear to be negatively affected, with despiking caused a deterioration in fit. The reason is that when the models are fitted on the despiked series, they systematically underestimate peaks and this causes large differences in the in-sample errors corresponding to peaks. This is particularly true for TVR which, being very flexible, is able to adapt itself to follow the peaks. However, only in period 6 for the TVR models, was despiking significantly worse at the 5% level.
As regards the comparison among models, in terms of descriptive statistics (MSE, MSPE, MAE and MAPE) nonlinear models, i.e. Markov regime-switching and time-varying parameter regression models, always give better results than linear ones. At 5% significance level, the Diebold and Mariano test indicates that, whether spikes are present or not, nonlinear models significantly outperform linear models in terms of fit (see Table 5). This is not surprising given the extra parameterisation and the inclusive nature of linear within the nonlinear specifications. This question is whether this represents over-fitting, and the forecasting insights in the next section

will address that. Note also that TVR performs substantially better than MS, even with the spikes, and significantly so in periods 6 and 38. Again this is not surprising as the coefficients are modelled as random walks, and in these specifications, it is well known that the model fit can appear high precisely because the parametric noise components incorporate a lot of the randomness from the price series. Again the out-of-sample forecasting may reveal whether this outperformance generalises.

Thus, with the out-of-sample predictions, the results are quite revealing. Firstly, we performed a direct comparison between forecasting accuracy of the best models estimated on the filtered log price series with and without spikes. Despiking the data produces better performance indicators in the 49% of the cases, and similar or worst (especially during peak hours) in the remaining cases. The Diebold and Mariano test shows that only in very few cases, clustered in load periods 6 and 44, are the improvements statistically significant (see Table 6). Thus, in our case, preprocessing data through a despiking procedure does not seem to be a critical issue in general for producing better forecasts.

Turning to the out-of-sample predictive accuracy of the models. Tables 7-9 show the results for both the with- and without-spikes cases, this time including also MS6. For each descriptive error statistic, season and load period, they display the best model and the ratio between the error statistics of the best model and the ARMAX model:

$$\mathrm{R_{MSE}} = \frac{\mathrm{MSE_{Best}}}{\mathrm{MSE_{ARMAX}}} \qquad \mathrm{R_{MSPE}} = \frac{\mathrm{MSPE_{Best}}}{\mathrm{MSPE_{ARMAX}}}$$
$$\mathrm{R_{MAE}} = \frac{\mathrm{MAE_{Best}}}{\mathrm{MAE_{ARMAX}}} \qquad \mathrm{R_{MAPE}} = \frac{\mathrm{MAPE_{Best}}}{\mathrm{MAPE_{ARMAX}}}.$$

We used ARMAX as the base comparator, as it was the best in most cases. We used the Diebold and Mariano test for equal prediction accuracy, using both squared and absolute error loss functions. At 5% of significance level, the results in Tables 10-14 indicate that in our study, no model is significantly more accurate than the ARMAX. Instead, the only model that produces significantly better forecasts for some of the periods with respect to all the others is the ARMAX.

Overall, we observe therefore that, at least for our data, the overwhelmingly better in-sample fit of the MS and TVR models did not lead to similarly better out-of-sample forecasting and that the ability of linear time series to perform well was endorsed. However, looking at some of the individual trading periods and seasonal results, it is possible to see indications of plausible model discrimination. Looking at the evening peak, for example, period 38, this has the largest spikes, and in winter these will be most serious, caused by extreme weather conditions and occasional gas price spikes. The MS model appears to perform best in these situations on all four error criteria (but not significantly so). In the low demand, period 6 in winter, price variations will be influenced mainly by the weather and transitions will not be so abrupt. The TVR would be expected to pick that up, and indeed TVR appears to perform best over all criteria. The MSE criterion will give most weight to extreme errors, and on this criterion, MS appears to perform best in 3, and TVR in two, out of the 5 trading periods (with spikes) for winter. Thus, there are indications that winter conditions lead to more volatile and extreme prices which may be effectively modelled by the nonlinear methods, even if pairwise significance tests do not provide convincing evidence.

## 6 Summary and Conclusions

In summary, we compared the modelling and forecasting performance of four classes of linear and nonlinear models, using data from the UKPX prompt market. Particular attention was paid in order to obtain comparability among models across different load periods and seasons. The classes of models were Constant and Time-Varying Parameter Regressions, ARMAX and Markov Regime-Switching models. The results lead us to conclude that: (i) despiking does not seem to be a critical issue in modelling and forecasting, (ii) when in-sample fit is considered, non linear models were significantly better than linear models, but they may overfit, and (iii) for out-of-sample forecasting, no single model completely outperforms all other models. Forecasting accuracy depends on load period, season and performance indicator, as well as the methods. However, the model which in most cases leads to better performance is the ARMAX model. Moreover, when the ARMAX is not the best, no model significantly outperforms it. There is however some indicative, but not statistically significant, evidence that the nonlinear methods of regime switching and time-varying parameters can forecast the more spiky and volatile winter prices better. In winter, extreme weather conditions, supply interruptions and gas price spikes can cause scarcity pricing to emerge suddenly - hence the value of these approaches. For the rest of the year, price formations are less subject to abrupt shocks and the weather effects are less extreme, such that the an ARMAX may well adequately adjust to their more steady evolution. Overall, this work has contributed to the emerging body of research on power price formation, emphasising the delicate process of model specification, especially for forecasting. Price formation does vary by time of day and time of year, and this may suggest the use of different models. However, this work also links to the conventional wisdom in applied forecasting that, out-of-sample, it is hard to show convincingly that simple linear time series models can be outperformed in general and on average. This is in line with findings in Clements and Krolzig (1998): while nonlinear models are superior to linear ones in capturing certain features of the data series, from a forecasting perspective linear models appear to be robust for prediction, even when the data are generated by a nonlinear model.

**Table 4:** *In-sample results (fitting) with and without spikes.*

| Model | Estimation with spikes | | | | Estimation without spikes | | | |
|---|---|---|---|---|---|---|---|---|
| | MSE | MSPE | MAE | MAPE | MSE | MSPE | MAE | MAPE |
| | | | | **Load Period 6** | | | | |
| REG | 9.48 | 84.94 | 1.94 | 6.82 | 10.13 | 84.28 | 1.94 | 6.76 |
| TVR | 2.35 | 23.21 | 0.98 | 3.47 | 9.34 | 77.91 | 1.77 | 6.15 |
| MS | 6.55 | 58.73 | 1.70 | 6.00 | 8.52 | 69.02 | 1.73 | 6.01 |
| ARMAX | 10.78 | 89.79 | 2.03 | 6.99 | 10.95 | 88.23 | 2.02 | 6.94 |
| | | | | **Load Period 18** | | | | |
| REG | 44.24 | 127.68 | 4.30 | 8.82 | 45.73 | 125.88 | 4.29 | 8.72 |
| TVR | 21.58 | 78.69 | 2.97 | 6.61 | 23.47 | 78.28 | 2.97 | 6.54 |
| MS | 23.35 | 80.83 | 2.97 | 6.51 | 20.37 | 68.39 | 2.78 | 6.13 |
| ARMAX | 72.83 | 191.46 | 5.00 | 9.98 | 74.09 | 191.65 | 5.01 | 9.96 |
| | | | | **Load Period 28** | | | | |
| REG | 120.42 | 308.48 | 7.01 | 13.76 | 121.13 | 305.65 | 7.01 | 13.71 |
| TVR | 76.06 | 224.44 | 5.53 | 11.29 | 76.86 | 216.89 | 5.47 | 11.08 |
| MS | 68.99 | 171.48 | 4.79 | 9.35 | 71.53 | 173.64 | 4.88 | 9.47 |
| ARMAX | 159.90 | 374.01 | 7.95 | 15.41 | 159.07 | 369.48 | 7.89 | 15.24 |
| | | | | **Load Period 38** | | | | |
| REG | 674.25 | 501.08 | 11.46 | 16.07 | 738.73 | 471.85 | 11.64 | 15.81 |
| TVR | 67.37 | 63.58 | 3.88 | 5.60 | 390.23 | 108.06 | 5.80 | 6.87 |
| MS | 449.26 | 371.15 | 9.80 | 13.63 | 449.74 | 303.56 | 8.88 | 12.39 |
| ARMAX | 820.45 | 519.24 | 12.41 | 16.38 | 841.22 | 499.89 | 12.36 | 16.10 |
| | | | | **Load Period 44** | | | | |
| REG | 22.89 | 104.96 | 3.14 | 7.75 | 23.69 | 103.18 | 3.11 | 7.65 |
| TVR | 12.90 | 64.70 | 2.40 | 5.96 | 14.28 | 66.13 | 2.46 | 6.05 |
| MS | 15.75 | 75.37 | 2.51 | 6.21 | 17.16 | 76.56 | 2.59 | 6.33 |
| ARMAX | 24.09 | 103.04 | 3.13 | 7.75 | 24.79 | 102.39 | 3.14 | 7.74 |

**Table 5:** *In sample p-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function.*

| | | **Adj D-M Test (se)** | | | **D-M Test (ae)** | | |
|---|---|---|---|---|---|---|---|
| | | REG | MS | TVR | REG | MS | TVR |
| | | **Estimation with spikes** | | | | | |
| **Load Period 6** | MS | 0.047 | — | — | 0.026 | — | — |
| | TVR | 0.009 | 0.002 | — | 0.000 | 0.000 | — |
| | ARMAX | 0.221 | 0.079 | 0.025 | 0.188 | 0.013 | 0.000 |
| **Load Period 18** | MS | 0.000 | — | — | 0.000 | — | — |
| | TVR | 0.004 | 0.665 | — | 0.000 | 0.999 | — |
| | ARMAX | 0.044 | 0.004 | 0.011 | 0.096 | 0.000 | 0.000 |
| **Load Period 28** | MS | 0.000 | — | — | 0.000 | — | — |
| | TVR | 0.007 | 0.468 | — | 0.000 | 0.026 | — |
| | ARMAX | 0.037 | 0.000 | 0.005 | 0.033 | 0.000 | 0.000 |
| **Load Period 38** | MS | 0.203 | — | — | 0.029 | — | — |
| | TVR | 0.045 | 0.009 | — | 0.000 | 0.000 | — |
| | ARMAX | 0.008 | 0.026 | 0.017 | 0.000 | 0.000 | 0.000 |
| **Load Period 44** | MS | 0.006 | — | — | 0.000 | — | — |
| | TVR | 0.015 | 0.228 | — | 0.000 | 0.479 | — |
| | ARMAX | 0.729 | 0.103 | 0.082 | 0.962 | 0.004 | 0.009 |
| | | **Estimation without spikes** | | | | | |
| **Load Period 6** | MS | 0.049 | — | — | 0.000 | — | — |
| | TVR | 0.020 | 0.167 | — | 0.002 | 0.462 | — |
| | ARMAX | 0.187 | 0.071 | 0.056 | 0.159 | 0.000 | 0.000 |
| **Load Period 18** | MS | 0.000 | — | — | 0.000 | — | — |
| | TVR | 0.004 | 0.438 | — | 0.000 | 0.379 | — |
| | ARMAX | 0.037 | 0.002 | 0.009 | 0.078 | 0.000 | 0.000 |
| **Load Period 28** | MS | 0.000 | — | — | 0.000 | — | — |
| | TVR | 0.005 | 0.557 | — | 0.000 | 0.072 | — |
| | ARMAX | 0.036 | 0.000 | 0.004 | 0.035 | 0.000 | 0.000 |
| **Load Period 38** | MS | 0.038 | — | — | 0.001 | — | — |
| | TVR | 0.002 | 0.262 | — | 0.000 | 0.000 | — |
| | ARMAX | 0.156 | 0.039 | 0.007 | 0.152 | 0.002 | 0.000 |
| **Load Period 44** | MS | 0.010 | — | — | 0.000 | — | — |
| | TVR | 0.008 | 0.127 | — | 0.000 | 0.298 | — |
| | ARMAX | 0.732 | 0.111 | 0.069 | 0.855 | 0.005 | 0.004 |

**Table 6:** *P-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function between the best models for each indicator estimated with and without spikes.*

| | | Adjusted D-M Test (se) | | | | D-M Test (ae) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **F.P.** | $B_{MSE}$ | $B_{MSPE}$ | $B_{MAE}$ | $B_{MAPE}$ | $B_{MSE}$ | $B_{MSPE}$ | $B_{MAE}$ | $B_{MAPE}$ |
| **Period 6** | Jan-Mar | 0.927 | 0.927 | 0.927 | 0.927 | 0.835 | 0.835 | 0.835 | 0.835 |
| | Apr-Jun | 0.622 | 0.111 | 0.111 | 0.111 | 0.474 | 0.229 | 0.229 | 0.229 |
| | Jul-Sept | 0.566 | 0.062 | 0.567 | 0.567 | 0.104 | 0.046 | 0.291 | 0.291 |
| | Whole | 0.998 | 0.975 | 0.998 | 0.766 | 0.765 | 0.002 | 0.765 | 0.999 |
| **Period 18** | Jan-Mar | 0.126 | 0.200 | 0.242 | 0.264 | 0.113 | 0.328 | 0.417 | 0.195 |
| | Apr-Jun | 0.504 | 0.504 | 0.504 | 0.504 | 0.753 | 0.753 | 0.753 | 0.753 |
| | Jul-Sept | 0.338 | 0.361 | 0.972 | 0.394 | 0.548 | 0.560 | 0.535 | 0.496 |
| | Whole | 0.213 | 0.361 | 0.219 | 0.219 | 0.765 | 0.762 | 0.336 | 0.336 |
| **Period 28** | Jan-Mar | 0.544 | 0.236 | 0.236 | 0.236 | 0.834 | 0.357 | 0.357 | 0.357 |
| | Apr-Jun | 0.346 | 0.070 | 0.346 | 0.070 | 0.067 | 0.340 | 0.067 | 0.340 |
| | Jul-Sept | 0.243 | 0.243 | 0.243 | 0.243 | 0.272 | 0.272 | 0.272 | 0.272 |
| | Whole | 0.237 | 0.188 | 0.237 | 0.188 | 0.241 | 0.278 | 0.241 | 0.278 |
| **Period 38** | Jan-Mar | 0.375 | 0.164 | 0.164 | 0.164 | 0.804 | 0.202 | 0.202 | 0.202 |
| | Apr-Jun | 0.423 | 0.763 | 0.970 | 0.763 | 0.081 | 0.525 | 0.592 | 0.525 |
| | Jul-Sept | 0.286 | 0.286 | 0.286 | 0.295 | 0.335 | 0.335 | 0.335 | 0.409 |
| | Whole | 0.324 | 0.229 | 0.145 | 0.229 | 0.814 | 0.587 | 0.562 | 0.587 |
| **Period 44** | Jan-Mar | 0.387 | 0.399 | 0.399 | 0.399 | 0.657 | 0.354 | 0.354 | 0.354 |
| | Apr-Jun | 0.262 | 0.085 | 0.019 | 0.019 | 0.129 | 0.057 | 0.022 | 0.022 |
| | Jul-Sept | 0.949 | 0.949 | 0.456 | 0.526 | 0.280 | 0.280 | 0.248 | 0.415 |
| | Whole | 0.461 | 0.490 | 0.490 | 0.490 | 0.183 | 0.031 | 0.031 | 0.031 |

*Note: the best models for each indicator, obtained with and without spikes, are presented in Tables 7-9.*

**Table 7:** *Out-of-sample forecasting results: ratios between the prediction error statistics of the best models and the statistics of ARMAX.*

**Load Period 6**

| **Forecasting Period** | | $R_{MSE}$ | $R_{MSPE}$ | $R_{MAE}$ | $R_{MAPE}$ |
|---|---|---|---|---|---|
| | | **Estimation with spikes** | | | |
| January-March | Statistics Value | 0.732 | 0.866 | 0.829 | 0.840 |
| | Best Model | TVR | TVR | TVR | TVR |
| April-June | Statistics Value | 0.812 | 1.000 | 1.000 | 1.000 |
| | Best Model | MS | ARMAX | ARMAX | ARMAX |
| July-September | Statistics Value | 1.000 | 0.966 | 1.000 | 1.000 |
| | Best Model | ARMAX | REG | ARMAX | ARMAX |
| Whole Period | Statistics Value | 0.793 | 1.000 | 0.920 | 0.976 |
| | Best Model | TVR | ARMAX | TVR | MS |
| | | **Estimation without spikes** | | | |
| January-March | Statistics Value | 0.735 | 0.923 | 0.896 | 0.911 |
| | Best Model | TVR | TVR | TVR | TVR |
| April-June | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |
| July-September | Statistics Value | 1.000 | 0.924 | 0.942 | 0.967 |
| | Best Model | ARMAX | REG | TVR | TVR |
| Whole Period | Statistics Value | 0.795 | 1.000 | 0.962 | 1.000 |
| | Best Model | TVR | ARMAX | TVR | ARMAX |

**Load Period 18**

| **Forecasting Period** | | $R_{MSE}$ | $R_{MSPE}$ | $R_{MAE}$ | $R_{MAPE}$ |
|---|---|---|---|---|---|
| | | **Estimation with spikes** | | | |
| January-March | Statistics Value | 0.821 | 0.998 | 0.923 | 1.000 |
| | Best Model | MS6 | MS6 | MS6 | ARMAX |
| April-June | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |
| July-September | Statistics Value | 0.982 | 0.725 | 0.952 | 0.910 |
| | Best Model | MS6 | REG | MS | MS |
| Whole Period | Statistics Value | 0.902 | 0.988 | 1.000 | 1.000 |
| | Best Model | MS6 | MS | ARMAX | ARMAX |
| | | **Estimation without spikes** | | | |
| January-March | Statistics Value | 0.824 | 0.989 | 0.961 | 1.000 |
| | Best Model | MS | TVR | MS6 | ARMAX |
| April-June | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |
| July-September | Statistics Value | 0.967 | 0.901 | 1.000 | 0.972 |
| | Best Model | TVR | REG | ARMAX | MS |
| Whole Period | Statistics Value | 0.891 | 1.000 | 1.000 | 1.000 |
| | Best Model | TVR | ARMAX | ARMAX | ARMAX |

**Table 8:** *Out-of-sample forecasting results: ratios between the prediction error statistics of the best models and the statistics of ARMAX.*

**Load Period 28**

| **Forecasting Period** | | $R_{MSE}$ | $R_{MSPE}$ | $R_{MAE}$ | $R_{MAPE}$ |
|---|---|---|---|---|---|
| | | **Estimation with spikes** | | | |
| January-March | Statistics Value | 0.995 | 1.000 | 1.000 | 1.000 |
| | Best Model | TVR | ARMAX | ARMAX | ARMAX |
| April-June | Statistics Value | 1.000 | 0.992 | 1.000 | 0.994 |
| | Best Model | ARMAX | TVR | ARMAX | TVR |
| July-September | Statistics Value | 0.561 | 0.633 | 0.738 | 0.779 |
| | Best Model | TVR | TVR | TVR | TVR |
| Whole Period | Statistics Value | 0.697 | 0.856 | 0.893 | 0.926 |
| | Best Model | TVR | TVR | TVR | TVR |
| | | **Estimation without spikes** | | | |
| January-March | Statistics Value | 0.931 | 1.000 | 1.000 | 1.000 |
| | Best Model | TVR | ARMAX | ARMAX | ARMAX |
| April-June | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |
| July-September | Statistics Value | 0.866 | 0.930 | 0.946 | 0.966 |
| | Best Model | TVR | TVR | TVR | TVR |
| Whole Period | Statistics Value | 0.890 | 1.000 | 0.992 | 1.000 |
| | Best Model | TVR | ARMAX | TVR | ARMAX |

**Load Period 38**

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

**Table 9:** *Out-of-sample forecasting results: ratios between the prediction error statistics of the best models and the statistics of ARMAX.*

**Load Period 44**

| **Forecasting Period** | | $R_{MSE}$ | $R_{MSPE}$ | $R_{MAE}$ | $R_{MAPE}$ |
|---|---|---|---|---|---|
| | | **Estimation with spikes** | | | |
| January-March | Statistics Value | 0.881 | 1.000 | 1.000 | 1.000 |
| | Best Model | MS | ARMAX | ARMAX | ARMAX |
| April-June | Statistics Value | 0.934 | 0.977 | 1.000 | 1.000 |
| | Best Model | MS | MS6 | ARMAX | ARMAX |
| July-September | Statistics Value | 0.989 | 0.984 | 0.996 | 1.000 |
| | Best Model | REG | REG | REG | ARMAX |
| Whole Period | Statistics Value | 0.908 | 1.000 | 1.000 | 1.000 |
| | Best Model | MS | ARMAX | ARMAX | ARMAX |
| | | **Estimation without spikes** | | | |
| January-March | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |
| April-June | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |
| July-September | Statistics Value | 1.000 | 1.000 | 0.995 | 0.977 |
| | Best Model | ARMAX | ARMAX | MS | MS |
| Whole Period | Statistics Value | 1.000 | 1.000 | 1.000 | 1.000 |
| | Best Model | ARMAX | ARMAX | ARMAX | ARMAX |

**Table 10:** *P-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function. Load period 6.*

| | | Adjusted D-M Test (se) | | | | D-M Test (ae) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **F.P.** | | REG | MS | MS6 | TVR | REG | MS | MS6 | TVR |
| | | **Estimation with spikes** | | | | | | | |
| Jan-Mar | MS | 0.725 | — | — | — | 0.843 | — | — | — |
| | MS6 | 0.945 | 0.395 | — | — | 0.321 | 0.583 | — | — |
| | TVR | 0.381 | 0.390 | 0.482 | — | 0.286 | 0.412 | 0.520 | — |
| | ARMAX | 0.380 | 0.759 | 0.571 | 0.324 | 0.280 | 0.389 | 0.250 | 0.107 |
| Apr-Jun | MS | 0.257 | — | — | — | 0.323 | — | — | — |
| | MS6 | 0.474 | 0.066 | — | — | 0.283 | 0.084 | — | — |
| | TVR | 0.870 | 0.219 | 0.837 | — | 0.881 | 0.419 | 0.500 | — |
| | ARMAX | 0.812 | 0.259 | 0.569 | 0.708 | 0.425 | 0.905 | 0.207 | 0.481 |
| Jul-Sept | MS | 0.875 | — | — | — | 0.635 | — | — | — |
| | MS6 | 0.319 | 0.349 | — | — | 0.801 | 0.567 | — | — |
| | TVR | 0.395 | 0.295 | 0.846 | — | 0.994 | 0.723 | 0.885 | — |
| | ARMAX | 0.603 | 0.730 | 0.265 | 0.218 | 0.496 | 0.853 | 0.464 | 0.640 |
| Whole | MS | 0.818 | — | — | — | 0.454 | — | — | — |
| | MS6 | 0.920 | 0.750 | — | — | 0.816 | 0.486 | — | — |
| | TVR | 0.495 | 0.524 | 0.496 | — | 0.353 | 0.719 | 0.392 | — |
| | ARMAX | 0.417 | 0.732 | 0.647 | 0.392 | 0.794 | 0.450 | 0.764 | 0.321 |
| | | **Estimation without spikes** | | | | | | | |
| Jan-Mar | MS | 0.753 | — | — | — | 0.306 | — | — | — |
| | MS6 | 0.493 | 0.611 | — | — | 0.739 | 0.689 | — | — |
| | TVR | 0.269 | 0.381 | 0.326 | — | 0.207 | 0.507 | 0.477 | — |
| | ARMAX | 0.356 | 0.053 | 0.542 | 0.303 | 0.778 | 0.180 | 0.677 | 0.199 |
| Apr-Jun | MS | 0.939 | — | — | — | 0.953 | — | — | — |
| | MS6 | 0.027 | 0.052 | — | — | 0.026 | 0.077 | — | — |
| | TVR | 0.263 | 0.367 | 0.905 | — | 0.463 | 0.606 | 0.563 | — |
| | ARMAX | 0.791 | 0.735 | 0.071 | 0.372 | 0.316 | 0.143 | 0.009 | 0.267 |
| Jul-Sept | MS | 0.465 | — | — | — | 0.735 | — | — | — |
| | MS6 | 0.433 | 0.946 | — | — | 0.904 | 0.622 | — | — |
| | TVR | 0.613 | 0.753 | 0.787 | — | 0.398 | 0.224 | 0.380 | — |
| | ARMAX | 0.951 | 0.552 | 0.496 | 0.494 | 0.592 | 0.350 | 0.636 | 0.480 |
| Whole | MS | 0.726 | — | — | — | 0.507 | — | — | — |
| | MS6 | 0.342 | 0.433 | — | — | 0.590 | 0.317 | — | — |
| | TVR | 0.419 | 0.478 | 0.339 | — | 0.303 | 0.536 | 0.273 | — |
| | ARMAX | 0.376 | 0.088 | 0.723 | 0.375 | 0.679 | 0.921 | 0.500 | 0.571 |

**Table 11:** *P-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function. Load period 18.*

| | | Adjusted D-M Test (se) | | | | D-M Test (ae) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **F.P.** | | REG | MS | MS6 | TVR | REG | MS | MS6 | TVR |
| | | **Estimation with spikes** | | | | | | | |
| Jan-Mar | MS | 0.297 | — | — | — | 0.850 | — | — | — |
| | MS6 | 0.223 | 0.784 | — | — | 0.154 | 0.046 | — | — |
| | TVR | 0.911 | 0.336 | 0.128 | — | 0.767 | 0.844 | 0.278 | — |
| | ARMAX | 0.461 | 0.347 | 0.276 | 0.440 | 0.710 | 0.777 | 0.553 | 0.855 |
| Apr-Jun | MS | 0.397 | — | — | — | 0.559 | — | — | — |
| | MS6 | 0.672 | 0.502 | — | — | 0.158 | 0.108 | — | — |
| | TVR | 0.322 | 0.391 | 0.213 | — | 0.755 | 0.671 | 0.550 | — |
| | ARMAX | 0.025 | 0.039 | 0.021 | 0.517 | 0.021 | 0.030 | 0.000 | 0.004 |
| Jul-Sept | MS | 0.250 | — | — | — | 0.642 | — | — | — |
| | MS6 | 0.669 | 0.608 | — | — | 0.121 | 0.114 | — | — |
| | TVR | 0.583 | 0.635 | 0.220 | — | 0.233 | 0.198 | 0.386 | — |
| | ARMAX | 0.762 | 0.692 | 0.683 | 0.131 | 0.696 | 0.593 | 0.222 | 0.100 |
| Whole | MS | 0.349 | — | — | — | 0.687 | — | — | — |
| | MS6 | 0.277 | 0.622 | — | — | 0.876 | 0.942 | — | — |
| | TVR | 0.750 | 0.464 | 0.124 | — | 0.410 | 0.297 | 0.317 | — |
| | ARMAX | 0.840 | 0.623 | 0.378 | 0.912 | 0.361 | 0.451 | 0.506 | 0.055 |
| | | **Estimation without spikes** | | | | | | | |
| Jan-Mar | MS | 0.473 | — | — | — | 0.959 | — | — | — |
| | MS6 | 0.688 | 0.478 | — | — | 0.082 | 0.332 | — | — |
| | TVR | 0.588 | 0.822 | 0.464 | — | 0.507 | 0.539 | 0.977 | — |
| | ARMAX | 0.353 | 0.361 | 0.310 | 0.316 | 0.906 | 0.937 | 0.654 | 0.693 |
| Apr-Jun | MS | 0.069 | — | — | — | 0.013 | — | — | — |
| | MS6 | 0.579 | 0.569 | — | — | 0.754 | 0.309 | — | — |
| | TVR | 0.450 | 0.918 | 0.646 | — | 0.688 | 0.111 | 0.520 | — |
| | ARMAX | 0.028 | 0.081 | 0.031 | 0.272 | 0.028 | 0.152 | 0.007 | 0.002 |
| Jul-Sept | MS | 0.292 | — | — | — | 0.957 | — | — | — |
| | MS6 | 0.647 | 0.555 | — | — | 0.089 | 0.159 | — | — |
| | TVR | 0.589 | 0.548 | 0.573 | — | 0.627 | 0.641 | 0.641 | — |
| | ARMAX | 0.538 | 0.485 | 0.432 | 0.664 | 0.986 | 0.977 | 0.158 | 0.298 |
| Whole | MS | 0.486 | — | — | — | 0.503 | — | — | — |
| | MS6 | 0.910 | 0.714 | — | — | 0.716 | 0.738 | — | — |
| | TVR | 0.389 | 0.694 | 0.316 | — | 0.912 | 0.748 | 0.927 | — |
| | ARMAX | 0.606 | 0.554 | 0.479 | 0.330 | 0.421 | 0.686 | 0.472 | 0.450 |

**Table 12:** *P-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function. Load period 28.*

| | | Adjusted D-M Test (se) | | | | D-M Test (ae) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **F.P.** | | REG | MS | MS6 | TVR | REG | MS | MS6 | TVR |
| | | | | | **Estimation with spikes** | | | | |
| Jan-Mar | MS | 0.400 | — | — | — | 0.574 | — | — | — |
| | MS6 | 0.469 | 0.433 | — | — | 0.384 | 0.865 | — | — |
| | TVR | 0.677 | 0.512 | 0.604 | — | 0.237 | 0.361 | 0.366 | — |
| | ARMAX | 0.193 | 0.198 | 0.000 | 0.968 | 0.034 | 0.040 | 0.039 | 0.323 |
| Apr-Jun | MS | 0.637 | — | — | — | 0.211 | — | — | — |
| | MS6 | 0.292 | 0.216 | — | — | 0.228 | 0.109 | — | — |
| | TVR | 0.039 | 0.034 | 0.019 | — | 0.011 | 0.011 | 0.000 | — |
| | ARMAX | 0.002 | 0.002 | 0.001 | 0.147 | 0.001 | 0.000 | 0.000 | 0.569 |
| Jul-Sept | MS | 0.343 | — | — | — | 0.940 | — | — | — |
| | MS6 | 0.703 | 0.619 | — | — | 0.394 | 0.434 | — | — |
| | TVR | 0.189 | 0.183 | 0.130 | — | 0.161 | 0.168 | 0.123 | — |
| | ARMAX | 0.266 | 0.242 | 0.558 | 0.162 | 0.902 | 0.949 | 0.329 | 0.071 |
| Whole | MS | 0.240 | — | — | — | 0.557 | — | — | — |
| | MS6 | 0.768 | 0.553 | — | — | 0.597 | 0.899 | — | — |
| | TVR | 0.175 | 0.150 | 0.108 | — | 0.046 | 0.064 | 0.017 | — |
| | ARMAX | 0.178 | 0.114 | 0.331 | 0.179 | 0.023 | 0.073 | 0.025 | 0.162 |
| | | | | | **Estimation without spikes** | | | | |
| Jan-Mar | MS | 0.296 | — | — | — | 0.572 | — | — | — |
| | MS6 | 0.682 | 0.327 | — | — | 0.615 | 0.902 | — | — |
| | TVR | 0.436 | 0.345 | 0.414 | — | 0.106 | 0.169 | 0.120 | — |
| | ARMAX | 0.715 | 0.233 | 0.608 | 0.496 | 0.037 | 0.029 | 0.035 | 0.935 |
| Apr-Jun | MS | 0.147 | — | — | — | 0.029 | — | — | — |
| | MS6 | 0.792 | 0.685 | — | — | 0.980 | 0.353 | — | — |
| | TVR | 0.081 | 0.181 | 0.009 | — | 0.088 | 0.239 | 0.016 | — |
| | ARMAX | 0.003 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |
| Jul-Sept | MS | 0.124 | — | — | — | 0.012 | — | — | — |
| | MS6 | 0.104 | 0.188 | — | — | 0.008 | 0.069 | — | — |
| | TVR | 0.143 | 0.138 | 0.134 | — | 0.168 | 0.078 | 0.059 | — |
| | ARMAX | 0.165 | 0.154 | 0.137 | 0.143 | 0.379 | 0.138 | 0.070 | 0.103 |
| Whole | MS | 0.092 | — | — | — | 0.767 | — | — | — |
| | MS6 | 0.111 | 0.957 | — | — | 0.371 | 0.165 | — | — |
| | TVR | 0.110 | 0.090 | 0.107 | — | 0.014 | 0.016 | 0.005 | — |
| | ARMAX | 0.121 | 0.076 | 0.108 | 0.131 | 0.001 | 0.000 | 0.000 | 0.724 |

**Table 13:** *P-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function. Load period 38.*

| | | Adjusted D-M Test (se) | | | | D-M Test (ae) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **F.P.** | | REG | MS | MS6 | TVR | REG | MS | MS6 | TVR |
| | | | | | **Estimation with spikes** | | | | |
| Jan-Mar | MS | 0.320 | — | — | — | 0.577 | — | — | — |
| | MS6 | 0.079 | 0.977 | — | — | 0.015 | 0.180 | — | — |
| | TVR | 0.647 | 0.292 | 0.287 | — | 0.463 | 0.310 | 0.072 | — |
| | ARMAX | 0.405 | 0.760 | 0.625 | 0.149 | 0.363 | 0.775 | 0.406 | 0.059 |
| Apr-Jun | MS | 0.003 | — | — | — | 0.018 | — | — | — |
| | MS6 | 0.078 | 0.029 | — | — | 0.006 | 0.001 | — | — |
| | TVR | 0.308 | 0.778 | 0.028 | — | 0.333 | 0.840 | 0.006 | — |
| | ARMAX | 0.107 | 0.356 | 0.002 | 0.404 | 0.259 | 0.762 | 0.000 | 0.899 |
| Jul-Sept | MS | 0.339 | — | — | — | 0.212 | — | — | — |
| | MS6 | 0.270 | 0.732 | — | — | 0.251 | 0.913 | — | — |
| | TVR | 0.255 | 0.235 | 0.233 | — | 0.166 | 0.384 | 0.147 | — |
| | ARMAX | 0.611 | 0.444 | 0.335 | 0.237 | 0.812 | 0.494 | 0.542 | 0.152 |
| Whole | MS | 0.257 | — | — | — | 0.304 | — | — | — |
| | MS6 | 0.070 | 0.919 | — | — | 0.081 | 0.685 | — | — |
| | TVR | 0.923 | 0.482 | 0.511 | — | 0.918 | 0.469 | 0.336 | — |
| | ARMAX | 0.340 | 0.726 | 0.638 | 0.505 | 0.257 | 0.872 | 0.889 | 0.241 |
| | | | | | **Estimation without spikes** | | | | |
| Jan-Mar | MS | 0.755 | — | — | — | 0.556 | — | — | — |
| | MS6 | 0.643 | 0.660 | — | — | 0.426 | 0.739 | — | — |
| | TVR | 0.546 | 0.518 | 0.484 | — | 0.583 | 0.708 | 0.810 | — |
| | ARMAX | 0.155 | 0.118 | 0.120 | 0.294 | 0.361 | 0.241 | 0.186 | 0.044 |
| Apr-Jun | MS | 0.001 | — | — | — | 0.001 | — | — | — |
| | MS6 | 0.046 | 0.002 | — | — | 0.001 | 0.000 | — | — |
| | TVR | 0.054 | 0.760 | 0.007 | — | 0.147 | 0.874 | 0.001 | — |
| | ARMAX | 0.043 | 0.530 | 0.003 | 0.813 | 0.096 | 0.879 | 0.000 | 0.822 |
| Jul-Sept | MS | 0.404 | — | — | — | 0.102 | — | — | — |
| | MS6 | 0.318 | 0.322 | — | — | 0.635 | 0.860 | — | — |
| | TVR | 0.273 | 0.279 | 0.371 | — | 0.398 | 0.641 | 0.801 | — |
| | ARMAX | 0.185 | 0.214 | 0.356 | 0.304 | 0.513 | 0.930 | 0.890 | 0.600 |
| Whole | MS | 0.822 | — | — | — | 0.287 | — | — | — |
| | MS6 | 0.806 | 0.758 | — | — | 0.299 | 0.129 | — | — |
| | TVR | 0.260 | 0.264 | 0.409 | — | 0.856 | 0.889 | 0.394 | — |
| | ARMAX | 0.106 | 0.084 | 0.324 | 0.993 | 0.132 | 0.289 | 0.043 | 0.205 |

**Table 14:** *P-values of the Diebold and Mariano test with squared error (adjusted version) and absolute error loss function. Load period 44.*

| | | Adjusted D-M Test (se) | | | | D-M Test (ae) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **F.P.** | | REG | MS | MS6 | TVR | REG | MS | MS6 | TVR |
| | | **Estimation with spikes** | | | | | | | |
| Jan-Mar | MS | 0.218 | — | — | — | 0.000 | — | — | — |
| | MS6 | 0.504 | 0.288 | — | — | 0.766 | 0.301 | — | — |
| | TVR | 0.047 | 0.078 | 0.055 | — | 0.050 | 0.017 | 0.059 | — |
| | ARMAX | 0.206 | 0.383 | 0.685 | 0.027 | 0.045 | 0.898 | 0.263 | 0.008 |
| Apr-Jun | MS | 0.112 | — | — | — | 0.186 | — | — | — |
| | MS6 | 0.065 | 0.622 | — | — | 0.051 | 0.944 | — | — |
| | TVR | 0.497 | 0.097 | 0.206 | — | 0.579 | 0.215 | 0.241 | — |
| | ARMAX | 0.212 | 0.625 | 0.749 | 0.306 | 0.133 | 0.700 | 0.623 | 0.237 |
| Jul-Sept | MS | 0.362 | — | — | — | 0.619 | — | — | — |
| | MS6 | 0.375 | 0.380 | — | — | 0.637 | 0.927 | — | — |
| | TVR | 0.308 | 0.313 | 0.258 | — | 0.552 | 0.809 | 0.809 | — |
| | ARMAX | 0.838 | 0.397 | 0.411 | 0.326 | 0.928 | 0.663 | 0.705 | 0.566 |
| Whole | MS | 0.273 | — | — | — | 0.042 | — | — | — |
| | MS6 | 0.978 | 0.275 | — | — | 0.724 | 0.324 | — | — |
| | TVR | 0.035 | 0.067 | 0.046 | — | 0.042 | 0.014 | 0.040 | — |
| | ARMAX | 0.183 | 0.454 | 0.426 | 0.021 | 0.027 | 0.631 | 0.220 | 0.006 |
| | | **Estimation without spikes** | | | | | | | |
| Jan-Mar | MS | 0.667 | — | — | — | 0.328 | — | — | — |
| | MS6 | 0.243 | 0.486 | — | — | 0.491 | 0.209 | — | — |
| | TVR | 0.131 | 0.508 | 0.832 | — | 0.361 | 0.099 | 0.799 | — |
| | ARMAX | 0.317 | 0.055 | 0.130 | 0.076 | 0.012 | 0.179 | 0.038 | 0.007 |
| Apr-Jun | MS | 0.209 | — | — | — | 0.198 | — | — | — |
| | MS6 | 0.846 | 0.231 | — | — | 0.792 | 0.519 | — | — |
| | TVR | 0.861 | 0.547 | 0.802 | — | 0.995 | 0.706 | 0.895 | — |
| | ARMAX | 0.087 | 0.265 | 0.080 | 0.372 | 0.069 | 0.162 | 0.087 | 0.312 |
| Jul-Sept | MS | 0.386 | — | — | — | 0.927 | — | — | — |
| | MS6 | 0.243 | 0.753 | — | — | 0.227 | 0.133 | — | — |
| | TVR | 0.300 | 0.271 | 0.340 | — | 0.567 | 0.476 | 0.963 | — |
| | ARMAX | 0.502 | 0.362 | 0.252 | 0.291 | 0.974 | 0.937 | 0.393 | 0.554 |
| Whole | MS | 0.541 | — | — | — | 0.210 | — | — | — |
| | MS6 | 0.164 | 0.436 | — | — | 0.380 | 0.091 | — | — |
| | TVR | 0.091 | 0.200 | 0.735 | — | 0.389 | 0.102 | 0.785 | — |
| | ARMAX | 0.174 | 0.030 | 0.064 | 0.039 | 0.005 | 0.111 | 0.009 | 0.012 |

## References

Aggarwal, S. K., Saini, L. M. and Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation, *Electrical Power and Energy Systems* **31**: 13–22.

Bowden, N. and Payne, J. E. (2008). Short term forecasting of electricity prices for miso hubs: Evidence from arima-egarch models, *Energy Economics* **30**: 3186–3197.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society. Series B (Methodological)* **37**(2): 149–192.

Bunn, D. W. (2000). Forecasting loads and prices in competitive power markets, *Proceedings of the IEEE* **88**(2): 163–169.

Bunn, D. W. (2004). *Modelling prices in competitive electricity markets*, Wiley.

Bunn, D. W. and Karakatsani, N. (2003). Forecasting electricity prices. Working Paper, London Business School.

Cartea, A. and Figueroa, M. G. (2005). Pricing in electricity markets: A mean reverting jump diffusion model with seasonality, *Applied Mathematical Finance* **12**(4): 313–335.

Clements, M. P. and Krolzig, H. M. (1998). A comparison of the forecast performance of markov-switching and threshold autoregressive models of us gnp, *Econometrics Journal* **1**(1): C47–C75.

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

Conejo, A. J., Plazas, M. A., Espínola, R. and Molina, A. B. (2005). Day-ahead electricity prices forecasting using the wavelet transform and arima models, *IEEE Transactions on Power Systems* **20**(2): 1035–1042.

Contreras, J., Espínola, R., Nogales, F. J. and Conejo, A. J. (2003). Arima models to predict next-day electricity prices, *IEEE Transactions on Power Systems* **18**(3): 1014–1020.

Crespo Cuaresma, J., Hlouskova, J., Kossmeier, S. and Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models, *Applied Energy* **77**: 87–106.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**: 253–263.

Duffie, D., Gray, S. and Hoang, P. (1998). Volatility in energy prices, *Managing Energy Price Risk*, Risk Publications.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press.

Escribano, A., Peña, J. I. and Villaplana, P. (2002). Modeling electricity prices: international evidence, Working paper *02-27*, University Carlos III, Madrid.

Friedman, J. H. (1984). A variable span scatterplot smoother, *Tecnical Report 5*, Laboratory for Computational Statistics, Stanford University.

Garcia, R. C., Contreras, J., van Akkeren, M. and Garcia, J. B. C. (2005). A garch forecasting model to predict day-ahead electricity prices, *IEEE Transactions on Power Systems* **20**: 867–874.

Geman, H. and Roncoroni, A. (2006). Understanding the fine structure of electricity prices, *Journal of Business* **79**.

Granger, C. W. J. (2008). Non-linear models: where do we go next - time varying parameter models?, *Studies in Nonlinear Dynamics & Econometrics* **12**(3): Article 1.

Guirguis, H. S. and Felder, F. A. (2004). Further advances in forecasting day-ahead electricity prices using time series models, *KIEE International Transactions on PE* **4-A**(3): 159–166.

Haldrup, N. and Nielsen, M. O. (2006). A regime switching long memory model for electricity prices, *Journal of Econometrics* **135**: 349–76.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.

Harvey, D., Leybourne, S. and Newbold, P. (1997). Testing the equality of prediction mean squared errors, *International Journal of Forecasting* **13**: 281–291.

Karakatsani, N. and Bunn, D. W. (2004). Modelling stochastic volatility in high-frequency spot electricity prices, Working paper, London Business School.

Karakatsani, N. and Bunn, D. W. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients., *International Journal of Forecasting* **24**: 764–785.

Kim, C. J. (1994). Dynamic linear models with markov-switching, *Journal of Econometrics* **60**: 1–22.

Knittel, C. R. and Roberts, M. R. (2005). An empirical examination of restructured electricity prices, *Energy Economics* **27**(5): 791–817.

Koopman, S. J., Ooms, M. and Carnero, M. A. (2007). Periodic seasonal reg-ARFIMA-GARCH models for daily electricity spot prices, *Journal of the American Statistical Association* **102**(477): 16–27.

Kosater, P. and Mosler, K. (2006). Can markov regime-switching models improve power-price forecasts? evidence from german daily power prices, *Applied Energy* **83**: 943–958.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* **54**: 159–178.

Misiorek, A., Trueck, S. and Weron, R. (2006). Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models, *Studies in Nonlinear Dynamics & Econometrics* **10**(3): Article 2.

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* **90**: 268–281.

Nogales, F., Contreras, J., Conejo, A. J., Espínola, R. and Plazas, M. A. (2002). Forecasting next-day electricity prices by time series models, *IEEE Transactions on Power Systems* **17**: 342–348.

Pedregal, D. J. and Trapero, J. R. (2007). Electricity prices forecasting by automatic dynamic harmonic regression models, *Energy Conversion and Management* **48**: 1710–1719.

Pilipovic, D. (1998). *Energy Risk: Valuing and Managing Energy Derivatives*, McGraw-Hill, New York.

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F. and Brace, C. (1997). Short-run forecasting of electricity loads and peaks, *International Journal of Forecasting* **13**: 161–174.

Said, S. E. and Dickey, D. (1984). Testing for unit roots in autoregressive moving-average models with unknown order, *Biometrika* **71**: 599–607.

Serletis, A. and Shahmoradi, A. (2006). Measuring and testing natural gas and electricity markets volatility: Evidence from alberta's deregulated markets, *Studies in Nonlinear Dynamics & Econometrics* **10**(3): Article 10.

Weron, R. (2002). Measuring long-range dependence in electricity prices, *Empirical Science of Financial Fluctuations* pp. 110–119.

Weron, R. (2006). *Modelling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, Chichester.

Weron, R., Bierbrauer, M. and Trück, S. (2004). Modeling electricity prices: jump diffusion and regime switching, *Physica A* **336**: 39–48.

Weron, R. and Misiorek, A. (2005). Forecasting spot electricity prices with time series models, *Proceedings Volume*, International Conference "The European Electricity Market EEM-05", May 10-12, 2005, Lodz, Poland, pp. 133–141.

Weron, R. and Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models, *International Journal of Forecasting* **24**: 744–763.

Weron, R. and Przybyłowicz, B. (2000). Hurst analysis of electricity price dynamics, *Physica A* **283**: 462–468.

## Working Paper Series
## Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it


**Department of Statistical Sciences**
*University of Padua*
*Italy*