Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Scuola di Dottorato in
Scienze Statistiche
Ciclo XXI

Forecasting Next-Day Electricity Prices: From Different Models to Combination

Fany Nan

Direttore: Prof.ssa Alessandra Salvan

Supervisore: Prof. Silvano Bordignon

Co-supervisori: Prof. Derek W. Bunn, Prof. Francesco Lisi

31 luglio 2009

Acknowledgements

This thesis is the results of advice and support provided by many people. Firstly, I am very grateful to prof. Silvano Bordignon and prof. Francesco Lisi for supervising my work during these three years, for their guidance, suggestions, support and for having always time.

I would like to express my deep thanks to prof. Derek Bunn for his helpful advice, enthusiastic support and for providing me with the UKPX dataset. It was a pleasure to work with him at London Business School, where I could appreciate the kindness of all his staff, the discussions with the Energy Group and the friendship of many people with whom I spent time both for work and for fun. There, I had my best life experience ever.

Many thanks to prof. Luisa Bisaglia for suggestions and moral support. I am deeply indebted to Padova University, the PhD director prof. Alessandra Salvan and my PhD professors.

I cannot forget in these acknowledgements all my colleagues and friends at the department of Statistical Sciences in Padova and my historical friends. Finally, I am greatly indebted to my parents for their support, understanding and encouragement.

To my mother
To my father

Two little mice fell into a pail full of cream. The first little mouse surrendered soon and it drowned. On the contrary, the second one busied itself so much that it turned the cream into butter, thus it could climb up and save itself. I am the second little mouse.

From the movie “Catch Me If You Can”

Abstract

As a result of deregulation of most power markets around the world electricity price modeling and forecasting have obtained increasing importance in recent years. Large number of models has been studied on a wide range of power markets, from linear time series and multivariate regression models to more complex non linear models with jumps, but results are mixing and there is no single model that provides convincing superior performance in forecasting spot prices.
This study considers whether combination forecasts of spot electricity prices are statistically superior to a wide range of single model based forecasts. To this end we focus on one-day ahead forecasting of half-hourly spot data from the British UK Power Exchange electricity market. In this work we focus on modeling data corresponding to some load periods of the day in order to evaluate the forecasting performance of prices representative of different moment of the day.
Several forecasting models for power spot prices are estimated on the basis of expanding and/or rolling estimation windows of different sizes. Included are linear ARMAX models, different specifications of multiple regression models, non linear Markov switching regression models and time-varying parameter regression models. One-day ahead forecasts are obtained for each model and evaluated according to different statistical criteria as prediction error statistics and the Diebold and Mariano test for equal predictive accuracy. Forecasting results highlight that no model globally outperforms the others: differences in forecasting accuracy depend on several factors, such as model specification, sample realization and forecasting period. Since different forecasting models seem to capture different features of spot price dynamics, we propose a forecasting approach based on the combination of forecasts. This approach has been useful to improve forecasting accuracy in several empirical situations, but it is novel in the spot electricity price forecasting context.
In this work different strategies have been employed to construct combination forecasts. The simplest approach is an equally weighted combination of the forecasts. An alternative is the use of adaptive forecast combination procedures, which allows for time-varying combination coefficients. Methods from Bates & Granger (1969) are considered. Models entering the combination are chosen for each forecasting season using the model confidence set method (MCS) described in Hansen et al. (2003, 2005) and then screened with the forecasts encompassing method of Fair & Shiller (1990). For each load period, our findings underline that models behave differently in each season. For this reason we propose a combination applied at a seasonal level. In this thesis some promising results in this direction are presented. The combination results are compared with the best results obtained from the single models in each forecasting period and for different prediction error statistics. Our

findings illustrate the usefulness of the procedure, showing that combining forecasts at a seasonal level have the potential to produce predictions of superior or equal accuracy relative to the individual forecasts.

Riassunto

Con la liberalizzazione dei mercati dell'elettricità, il problema della modellazione e previsione dei prezzi elettrici è diventato di fondamentale importanza. In letteratura sono stati studiati e applicati ad un gran numero di mercati molti tipi di modelli, come modelli per serie storiche, regressione lineare e modelli non lineari a salti molto più complessi. I risultati però sono contrastanti e finora nessun modello ha mostrato una capacità previsiva dei prezzi elettrici superiore rispetto agli altri.

L'obiettivo di questa tesi è capire se i modelli di combinazione di previsioni possano dare risultati statisticamente superiori rispetto alle previsioni ottenute da singoli modelli. In particolare, viene affrontato il problema della previsione dei prezzi elettrici del giorno dopo applicato al mercato elettrico britannico UK Power Exchange. In questo mercato, i prezzi hanno frequenza semioraria: al fine di valutare il comportamento previsivo dei modelli, relativamente all'andamento dei prezzi nei diversi momenti della giornata, sono state scelte specifiche fasce orarie.

I modelli usati per la previsione dei prezzi sono stati stimati sulla base di finestre di dati espandibili e/o mobili di diverse misure fissate. I modelli considerati includono modelli lineari di tipo ARMAX e diverse specificazioni di modelli di regressione multipla. Inotre sono stati considerati modelli di regressione non lineare a regimi Markov switching e modelli di regressione a parametri non costanti. Le previsioni a un passo ottenute dai modelli specificati sono state confrontate secondo diversi criteri statistici come le statistiche basate sull'errore di previsione e il test di Diebold e Mariano.

Dai risultati emerge che, globalmente, nessun modello considerato supera gli altri per abilità previsiva: vari fattori, tra cui specificazione del modello, realizzazione campionaria e periodo di previsione, influenzano l'accuratezza previsiva. Dal momento che modelli di previsione diversi sembrano evidenziare caratteristiche diverse della dinamica dei prezzi elettrici, viene proposto un approccio basato sulla combinazione di previsioni. Questo metodo, finalizzato a migliorare l'accuratezza previsiva, si è dimostrato utile in molti studi empirici, ma finora non è stato usato nel contesto della previsione dei prezzi elettrici.

In questa tesi sono state usate diverse tecniche di combinazione. L'approccio più semplice consiste nel dare lo stesso peso a tutte le previsioni ottenute dai singoli modelli. Altre procedure di combinazione di previsioni sono di tipo adattivo, poiché utilizzano coefficienti non costanti. In questo contesto, sono stati considerati i metodi di Bates & Granger (1969). I modelli usati nella combinazione sono stati scelti, per ciascuna stagione di previsione, con il metodo model confidence set (MCS) descritto in Hansen et al. (2003, 2005) e successivamente ridotti con il metodo forecasts encompassing di Fair & Shiller (1990). Per ciascuna ora considerata, i risultati sottolineano che i modelli si comportano in modo diverso a seconda della stagione di previsione. Questa caratteristica giustifica l'applicazione dei modelli di combinazione di previsioni ad un livello stagionale. In questa tesi vengono presentati risultati promettenti in questa direzione. Considerando le statistiche basate sull'errore di previsione, i risultati delle combinazioni sono stati confrontati con i migliori risultati ottenuti dai singoli modelli in ciascun periodo previsivo. Il vantaggio della procedura proposta deriva dal fatto che combinando le previsioni ad un livello stagionale, si ottengono previsioni di accuratezza superiore o uguale rispetto alle previsioni individuali.

Contents

Chapter 1

Introduction

1.1 Overview

After the liberalization of the electricity market, electricity prices have changed their behavior. Restructuring removes price controls and openly encourages market entry. In this context producers, retailers and consumers interact through a market in which the common target is to maximize their respective profits but new features appeared in electricity price series as high volatility and the presence of unexpected jumps. So, a new topic has emerged: developing forecasting models which can well describe electricity price dynamics.
The question is not trivial because of the characteristics the price series show: among them seasonality, heteroskedasticity and spikes. In literature many models have been proposed: models derived directly from the statistical techniques of demand forecasting are generally preferred like multivariate regression, time series models and smoothing techniques. But in the last years also more complex models have been tested for forecasting. Nonlinear models with jumps, Markov regime switching and time-varying parameter models have been proposed. Despite this seemingly large number of models and related empirical analysis on a wide range of power markets, results are mixing and there is no single model that provides convincing superior performance in forecasting spot prices.
In load forecasting some attempts of multiple modeling and combining were done (Smith, 1989; Taylor and Majithia, 2000), but there is no research about the use of combining techniques in electricity price forecasting. The reason for using combined forecasts is the opinion that identifying the true statistical representation of the time-series is elusive and that, statistically, more accurate forecasts can be obtained through a combination of the outputs of several good, but quite different models. In fact, different modeling strategies, even if applied with the same information set, may capture different parts of the characteristics of the data. In electricity price series, jump intensities are not constant over the time and a time-varying volatility is present. It is difficult to model them exactly: model combining provides a potential capability of better adaptation to such features.

This thesis considers whether combined forecasts of spot electricity prices are statistically superior or equal to a wide range of single model based forecasts. To this end the work focus on one-day ahead forecasting (since this is of greater concern when dealing with spot prices) of the half-hourly spot prices from the UK Power Exchange electricity market.

This work consists of two parts. In the first part, a brief introduction to the electricity market deregulation is presented, with a description of the electricity price features (seasonality, non-normality, high volatility and jumps) and their formation process. Chapter 3 reviews the relevant price forecasting techniques in literature, with particular attention to statistical methods. Among others, articles about time series techniques, Markov regime-switching and time-varying parameter models are discussed.

The second part of the thesis includes the comparative study of different forecasting models. In chapter 4 the UKPX market is described. The study is based on data from April, 1st 2005 (after the inclusion in the market of Scotland) to September, 30th 2006. As it is customary, the 48 half-hourly time series are considered separately: the modeling is focused on data corresponding to some periods of the day in order to evaluate the forecasting performance of prices representative of different moment of the day. The forecasting design is presented, in particular we choose the period January-September 2006 for out-of sample forecasting. This period is divided into three parts to best compare differences in forecasting accuracy caused by different seasons.

In chapter 5, several forecasting models for power spot prices are estimated on the basis of expanding and/or rolling estimation windows of different sizes. Included are different specifications (based on different regressors) of multiple regression models, non linear Markov switching regression models, time-varying parameter regression models, and short (estimated on a short rolling window) linear ARMAX models. The choice of developing these forecasting models was taken cause the different characteristics of these models: the Markov regime-switching regression model allows for discontinuities in pricing due to temporal irregularities, the time-varying parameter regression model allows for a continuously adaptive price structure and the short ARMAX can adjust rapidly to the fast changes in the market conditions. The linear regression model is used as baseline comparison. One-day ahead forecasts are obtained for each model and evaluated according to different statistical criteria, as prediction error statistics and the Diebold and Mariano test for equal predictive accuracy (Diebold & Mariano, 1995; Harvey et al., 1997).

Chapter 6 introduces the combination techniques. When we are dealing with many forecasting models, the choice of the models to be included in the combination is an important issue. The models are selected using for each forecasting season the model confidence set method method (MCS) described in Hansen et al. (2003, 2005), screened with the forecasts encompassing method of Fair & Shiller (1990). The chosen models are combined with different techniques. The simplest approach sets the combination to be the mean of the constituent forecasts and is, thus, an equally weighted combination of the forecasts. An alternative is the use of adaptive forecast combination procedures, which allows for time-varying combination coefficients. Methods from Bates & Granger (1969) are considered.

The combination results are then compared with the best results obtained from the single models in each forecasting period and for different prediction error statistics. The results illustrate the usefulness of the procedure, showing that the choice of combining forecasts at a seasonal level have the potential to produce predictions of superior or equal accuracy relative to the individual forecasts. At the same time, this technique allows to avoid the choice of a single model.

Chapter 7 concludes with some ideas for future research.

1.2 Contributions of the thesis

An overview of the original results presented in the thesis is listed below.

- An extensive review of the literature about the electricity price forecasting issue is presented, with particular attention to time series models.

- The heart of the thesis is the systematic comparative forecasting study that is conducted for four representative load periods: period 6 (2:30-3:00am), period 19 (9:00-9:30am), period 28 (13:30-14:00pm) and period 38 (18:30-19:00pm). These periods are representative of different moments of the day, as the contrast between day and night and the working habits of the population. Three sets of regressors are proposed. The difference among them is the specification of the annual periodic component of the prices. On the basis of these three regressor specifications, several forecasting models are estimated and one-day ahead forecasts are calculated. The forecasting period is divided into three parts, corresponding to different seasons: January-March, April-June and July-September. The reason for doing this is to better compare forecasting performances.

- Four different models are proposed for comparison: a multiple regression model, a Markov regime-switching (r-s) model, a time-varying parameter regression model and an ARMAX model. The usefulness of the Markov r-s models for power market applications has been already recognized. In particular this kind of models are useful for modeling several consecutive price jumps. However, their adequacy for forecasting has been only vaguely tested (see Misiorek et al., 2006 and Karakatsani & Bunn, 2008b). In this thesis, the parameters of this model were estimated using a daily expanding dataset and rolling windows with fixed length of 3, 6 and 9 months (66, 132 and 186 days). Time-varying parameter models are the most recent topic of research for electricity price forecasting, so it is interesting to compare forecasts obtained with this kind of models with that obtained from other specifications. The ARMAX model is estimated on rolling windows of fixed lengths, which depend on the load period. In an in-sample study we compare specifications of AR, ARX, ARMA and ARMAX models with two different exogenous variables and different window lengths: the ARMAX model with 'margin' as exogenous variable has given the best results, so it has been used for out-of-sample forecasting.

- Forecasting results highlight that no model globally outperforms the others: differences in forecasting accuracy depend on several factors, such as model specification, sample realization and forecasting period. Since different forecasting models seem to capture different features of spot price dynamics, we propose a forecasting approach based on the combination of forecasts. This approach has been useful to improve forecasting accuracy in several empirical situations (see, for example, Becker & Clements, 2008; Sánchez, 2008), but it is novel in the spot electricity price forecasting context.

- The choice of the models entering the combination is an important issue. We choose the models for each forecasting season using the MCS method (Hansen et al., 2005) and the forecasts encompassing method (Fair & Shiller, 1990). For each load period, our findings underline that models behave differently in each season. For this reason we propose a combination applied at a seasonal level. This is a first study in this direction. Combination results highlight that combinations made at seasonal level with few models produce better results than combinations of all the 19 models. The reason is the high variability that is introduced in the combinations using models that have poorly performance in determined seasons.

- Combination results are compare also with the best results obtained from the single models in each forecasting period and for different prediction error statistics. Results show that combination forecasts have the potential to produce forecasts of superior or equal accuracy relative to the best individual model.

To illustrate combination performance, in table 1.1 a summary of the results is presented. For each load period, we compare, in terms of Mean Squared Percentage Error (MSPE) and of Mean Absolute Percentage Error (MAPE), the forecast results obtained from the Bates and Granger combination model estimated on a rolling window of 10 days, with the best MSPEs and MAPEs obtained using the single 19 models. In the table, these relative measure statistics are called respectively R_{MSPE} and R_{MAPE}. Rows of the table represent respectively the forecasting periods January-March, April-June and July-September. The last row contains results of the whole period (January-September). Results show that the combination model outperforms or matches the best model among the single ones. In particular, over the whole forecasting period, improvements range from 4% to 21%. P-values obtained with the Diebold-Mariano test for equal predictive accuracy (numbers in parenthesis) underline the improvement of the combined models in comparison with the best single models.

Table 1.1: *Summary of the combination results. Load periods 6, 19, 28 and 38 (F.P.= Forecasting Period).*

	Load period 6		**Load period 19**		**Load period 28**		**Load period 38**	
F.P.	R_{MSPE}	R_{MAPE}	R_{MSPE}	R_{MAPE}	R_{MSPE}	R_{MAPE}	R_{MSPE}	R_{MAPE}
Jan-Mar	1.016	1.041	0.967	1.011	0.997	0.993	0.870	0.957
Apr-Jun	0.800	0.932	0.738	0.822	1.005	1.013	0.925	0.979
Jul-Sept	0.959	0.970	1.005	0.997	0.976	1.012	0.909	0.952
Whole	0.813	0.887	0.783	0.896	0.927	0.956	0.864	0.898
	(0.012)	(0.036)	(0.045)	(0.027)	(0.232)	(0.152)	(0.036)	(0.027)

Part I

Electricity price features and literature review

Chapter 2

Electricity market liberalization and characteristics of electricity prices

Over the past twenty years the electric power industry has undertaken significant restructuring. In many countries, the old concept of centralizing electric power industry seen as public service has been replaced by the idea that a competitive market is the most appropriate mechanism to take energy to consumers with high reliability and low costs. Restructuring removes price controls and openly encourages market entry. In this context producers, retailers and consumers interact through a market in which the common target is to maximize their respective profits. There is an extensive literature on the changes that have taken place and the main characteristics of various competitive power markets, see among others Bunn (2004) and Weron (2006).

Before deregulation, public commissions were in charge of the management of all the electric sector, which included for example tariff designs and investment decisions. Price change over time was minimal because tariffs were kept fixed for long periods (for price dynamics after deregulation see section 2.1). The chain of electric energy production was based on five principal components:

- *generation* of electricity through different technologies like hydroelectric stations, nuclear plants and steam power stations (activated with coal, natural gas or oil);
- *transmission* networks that from generators transfer electricity for long distances, disguised as alternating current with very high voltage;
- *distribution* of low voltage electricity to homes and factories;
- *system operations* to monitoring the system and to synchronize production and consumption to avoid electric grid blackouts;
- *retail* of electricity to consumers.

After deregulation, the two functions of generation and retail came out of the monopoly system: generators, industrial consumers and retailers became part of

organized markets in which electricity can be traded through a one-sided or a two-sided auction. This did not happen for transmission, distribution and system operator components. Their particular structure and the fact that all competitors need nondiscriminatory access to them are reasons to keep them into monopoly.
Out of the monopoly system, two main kind of market for electricity have been emerged: power pools and power exchange (Weron, 2006). Power pools can be divided into two categories: technical and economic. In *technical pools*, the power plants are ranked on merit order. Generation costs and network constraints are the determining factors to optimize generation with respect to cost minimization and optimal technical dispatch. *Economic pools* are one-sided auction markets, in which only generators can participate and participation is mandatory (indeed no trade is allowed outside the pool). Generators bid based on the prices at which they are willing to run their power plants. Then, the market clearing price, i.e. the price to be paid by retailers and to be charged by producers, is established as the intersection of the supply curve (constructed from aggregated supply bids) and the estimated demand (market clearing volume).
Power exchanges, or wholesale electricity markets, are commonly launched on a private initiative. Participants include generators, distribution companies, traders and large consumers and the participation can be mandatory or voluntary (in this case bilateral contracts are also allowed), depending on the market. Generally, the market clearing price is established in the form of a conducted once per day two-sided auction. Among others, OMEL (Spain) and PJM (Pennsylvania-New Jersey-Maryland) markets follow this scheme, whilst some markets trade electricity closer to the delivery through two-sided auctions conducted each period at a time: UK and Ontario markets are two examples. Producers submit to the Market Operator production bids that typically consist of a set of energy blocks and their corresponding minimum selling prices for every hour (or half-hour) of the next day. Analogously, retailers and large consumers submit consumption bids that consist of a set of energy blocks and their corresponding maximum buying prices. Each hourly (half-hourly) market clearing price is given by the Market Operator by the intersection of the supply curve and the demand curve, constructed respectively from aggregated supply and demand bids. Finally, the Independent System Operator checks if the schedule is feasible or if it needs some changes. For the pricing rule, there are two relevant variants. First, the *uniform pricing* provides the same price for every accepted bid, i.e. buyers (suppliers) with bids (offers) above (below) or equal to the clearing price are paid that price. In contrast, the transaction can be priced in a discriminatory manner called *pay-as-bid* pricing: a supplier would be paid the price he bid for the quantity transacted (as example, the UK market under NETA followed this pricing procedure, see section 4.1).

With the introduction of these electricity market systems, producers, retailers and large consumers need information gotten through day-ahead price forecasts to optimally self-schedule and to derive its bidding strategy. For this reason, it is very important to develop forecasting models which can good describe dynamics of electricity prices both in the short and medium-term. Clearly, this is not trivial to do, cause the complexities that these time series show. The main peculiarity is high volatility. There are several elements that explain this characteristic. Probably the most important one is the non-storability of electricity. Electricity cannot be physically stored in a direct way, and production and consumption have to be continuously balanced. Therefore, supply and demand shocks cannot easily be smoothed out and they will have a direct effect on equilibrium prices. This and other features characterize the electricity price series, so that electricity price forecasting is up till now a heated topic for researchers. In literature, many kinds of models have been studied but there is not one forecasting technique that prevails against the others. The next section describes the main characteristic factors of electricity prices and how these complexities have been tackled in literature. Then, the main forecasting models are presented (chapter 3), with particular attention to statistical techniques.

2.1 Electricity prices: formation process and features

For proposing adequate modeling and forecasting methods, it is necessary to know the features which characterize electricity price series and it is not a trivial aspect how to capture them in the model. Despite a few distributional similarities, electricity prices are dramatically different from those found in the financial or other commodity markets. Specifically, electricity prices display the following distinct characteristics:

high frequency: price series have hourly or half-hourly frequency;

multi-scale seasonality (intra-day, weekly, annual): electricity price dynamics are strongly influenced by human activities and by seasons;

calendar effects: demand is low in weekend days and holidays, and this influences prices;

dependence on atmospherical factors: temperature is one factor strongly correlated with price dynamics. In very cold (hot) periods, electricity demand increases cause heating (air-conditioned) systems;

non-normality: this feature manifests as leptokurtosis and positive skewness which is more elevate when there is high variability in demand;

high volatility: orders of magnitude higher than other commodities and financial assets. Electricity prices contain an inverse leverage effect: volatility tends to rise more so with positive shocks than negative shocks (Knittel & Roberts, 2005);

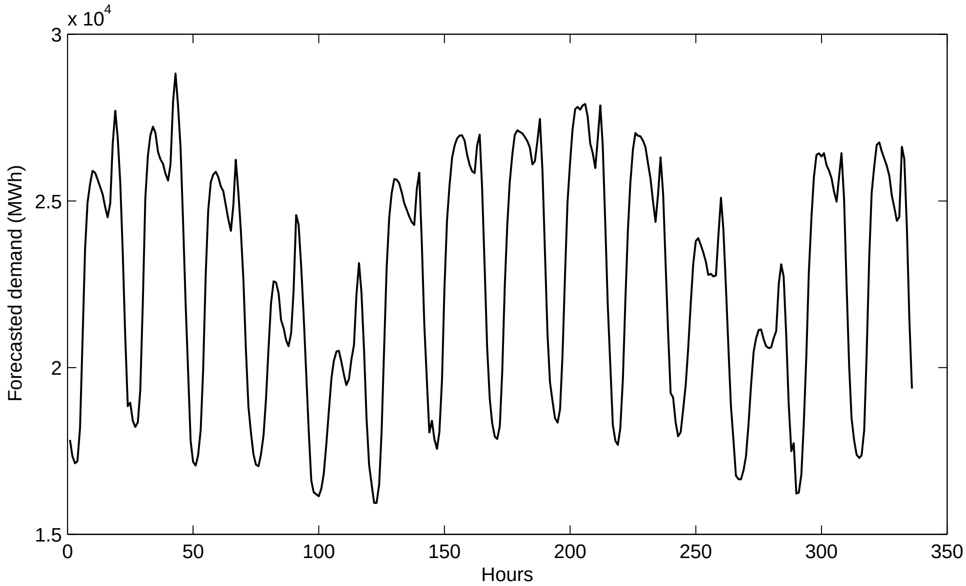
presence of jumps: one of the characteristics of evolution of these jumps is that the price does not stay in the new level, to which it jumps, but reverts to the previous level rapidly (mean reversion).

Many of the presented features are shared by most electricity spot markets in the world.

2.1.1 Seasonality and calendar effects

The main factor affecting electricity price is the demand (load) dynamics. Electricity demand is heavily influenced by economic and business activities and by the weather conditions, so that electricity markets across the world exhibit three different types of seasonality: daily, weekly and annual. The daily cycle refers to variations between day and night and during the different moments of the day. This variation follows the working habits of the population. For each day shown in figure 2.1, the daily cycle is clear. During the night, the demand is very low, while in the morning and in the evening there is a peak which reflects working activity and atmospherical conditions.

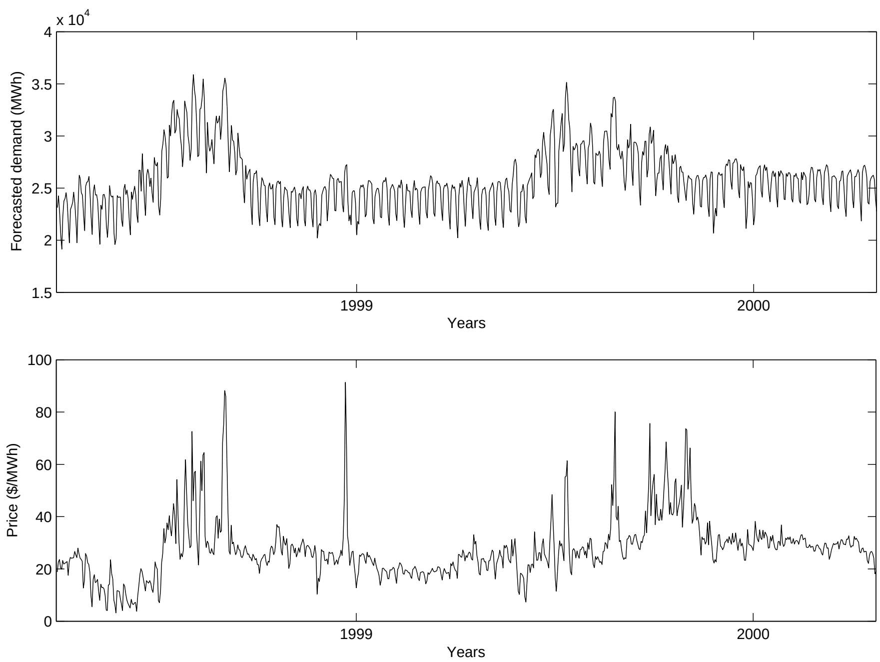
Figure 2.1: *Hourly forecasted demand in the California power exchange from Wednesday 01/04/1998 to Tuesday 14/04/1998.*



The intra-week variability is also non-negligible. Saturday and Sunday load profile is lower than during the weekdays. This feature is called "weekend effect" and it is present also during national holidays (in this case it is called "calendar effect", see for example figure 4.1 for the effect in the prices).

Seasonal fluctuations (annual cycle) mostly arise due to the use of artificial light and heating in winter and to the growing use of air conditioning in summer. Figure 2.2 shows the daily average electricity demand and the corresponding prices in the Californian market during the first two years of activity (till April, 23th 2000). In this market, the peaking season is the summer caused by high temperatures. This reflects to the prices, but the relation demand-price is not simple. A very high variability is present in the price series when demand is high (see section 2.1.3 for an explanation of the problem). The other three seasons are characterized by mild temperatures, so the mean demand is lower. Figure 2.3 shows another example of demand and price annual behaviour. The market is the PJM and the series have hourly frequency. In this case the seasons with high demand are two: summer and winter. Demand is very unstable especially during the summer season. The effect on prices is evident: high peaks are characteristic of the high demand periods. In general, electricity load and atmospheric temperature have a non-linear relationship: very low and very high temperatures correspond to high levels of electricity demand (see, for instance, figure 2.4).

Figure 2.2: *Daily mean forecasted demand (MWh) and daily mean price ($/MWh) in the California power exchange from April, 1998 to April 23th, 2000.*



A tool for modeling seasonal fluctuations is the sinusoidal approach with a linear trend, successfully applied by Pilipovic (1998) and Geman & Roncoroni (2006) (see figure 2.5). The modeling of intra-week and intra-day seasonalities may be approached analogously or with other tools, like differencing technique to remove the weekly seasonal component and the introduction of dummies for particular days. To remove intra-day seasonality, it is possible to model each hour (half-hour) of the day separately, like different commodities (see for example Ramanathan et al., 1997, and Guthrie & Videbeck, 2002, for an intra-day approach). Many authors are agree to assert that this approach can improve forecasts.

2.1.2 Non-normality of electricity prices

In general, like other financial asset returns, electricity price series are not normally distributed. This feature emerges by the fact that the series manifest positive skewness and leptokurtosis (or heavy-tailed character). As example, figure 2.6 shows these price features in the European Energy Exchange (EEX), the spot market in Germany (data are daily prices from June 2002 to May 2004). The histogram

Figure 2.3: *Hourly demand (MWh) and price ($/MWh) traded on the PJM market, from April, 2002 to August, 2003.*

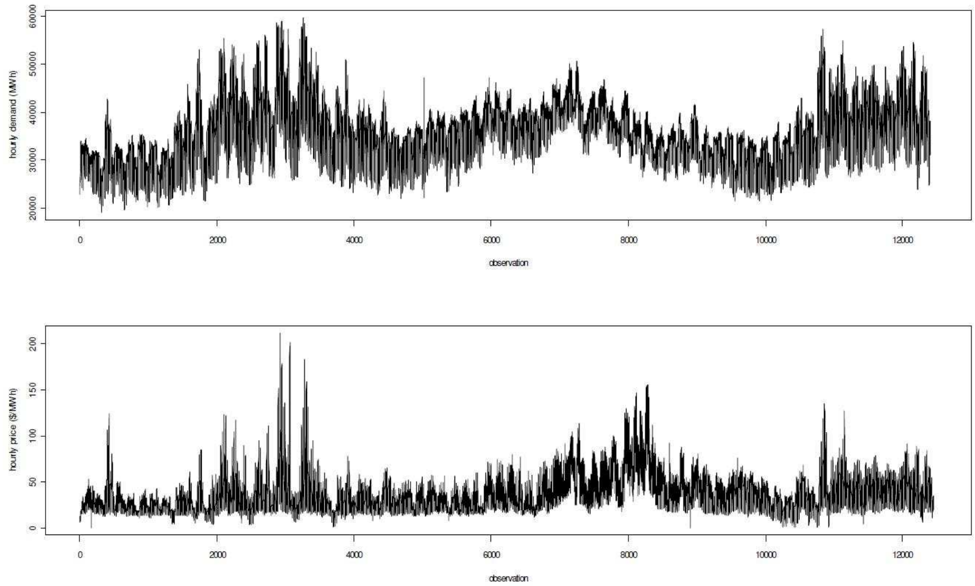


Figure 2.4: *Scatterplot between PJM daily mean demand (MWh) and atmospheric temperature (F^0), from April, 2002 to August, 2003 (source: Fezzi, 2007a).*

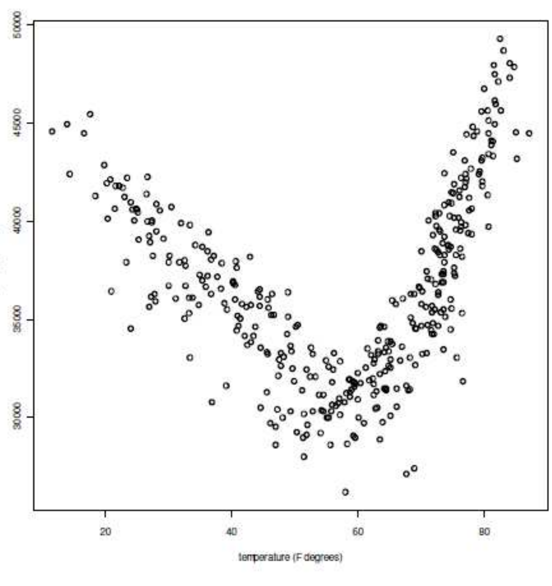
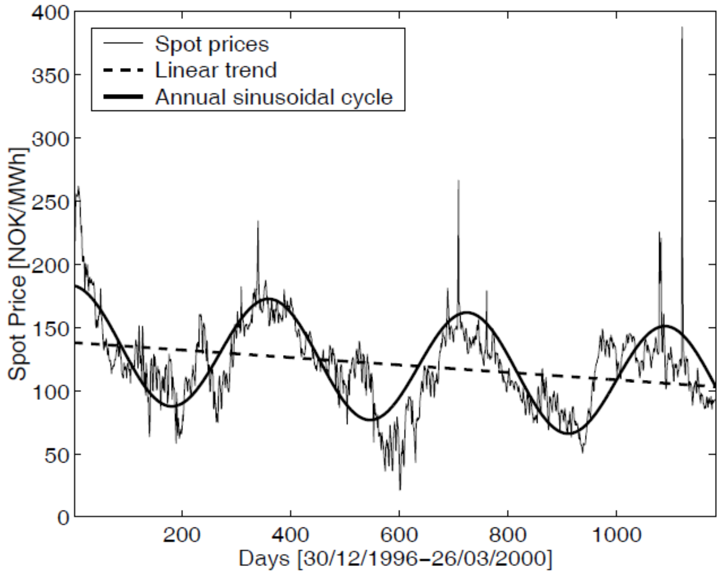


Figure 2.5: *Nord Pool market daily mean price from December 30th, 1996 until March 26th, 2000. Superimposed on the plot is an approximation of the annual seasonality by a sinusoid with a linear trend (source: Weron, 2005).*

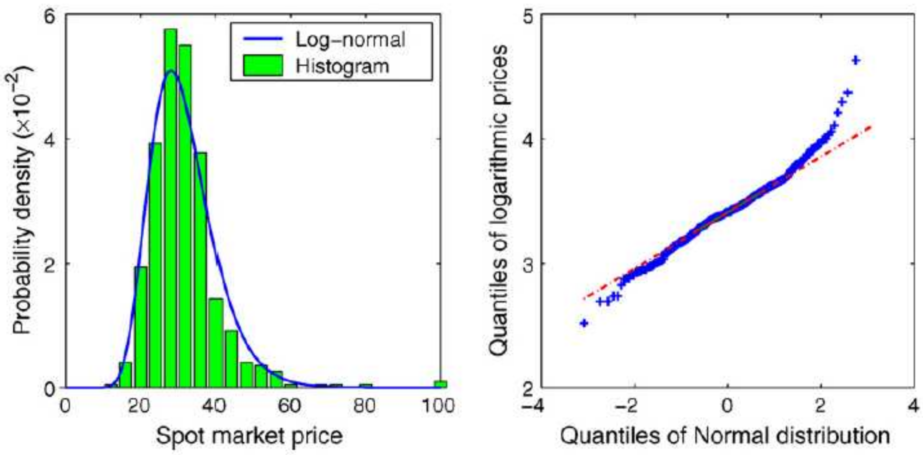


indicates that these particular prices may be well represented by a log-normal distribution. Knittel & Roberts (2005) underline that this kind of asymmetry suggests the presence of an inverse leverage effect. Thus, positive shocks to prices amplify the conditional variance of the process more so than negative shocks. Some authors have studied how to model the leptokurtic behavior and which probability distributions best describe the data. Weron (2005), for example, model electricity prices with distributions from two heavy-tailed families: α-stable and generalized hyperbolic distributions. In the case studied, heavy-tailed distributions, and the α-stable in particular, show very good fitting performance not only visually but also in terms of the goodness-of-fit statistics (just comparing the values obtained from Anderson-Darling and Kolmogorov tests). This results suggest that a model, with well-specified seasonal structure and which can control the intensity of the jumps, amended with heavy-tailed innovations could lead to improve performance.

2.1.3 Jumps and volatility

An important aspect of electricity prices is the existence of high, time-varying volatility and volatility clustering. This is a direct consequence of the electricity market liberalization, and it is one of the principal factors that make electricity prices modeling and forecasting so important in the short term and so difficult at the same time. One of the most popular approaches for modeling conditional volatility is the GARCH model and its extensions, but, as Duffie et al. (1998) pointed out, the application of these kind of models to electricity prices has its limitations, deriving erroneous results. The presence of unanticipated extreme changes in prices called

Figure 2.6: *Normalized histogram (left) and quantile-quantile plot (right) of daily averaged EEX spot market prices (source: Swider & Weber, 2007).*



jumps or spikes are the cause of bias in the estimation of the GARCH process. Escribano et al. (2002) show that one can improve GARCH models by working with a model that simultaneously takes into account volatility behavior and jumps (see section 3.1.1).

The feature of these spikes is that they are normally quite short-lived. When the weather phenomenon or outage that caused the peak is over, prices fall back to a normal level. Figure 2.7 shows an example of this characteristic from the Nord Pool market (the Scandinavian power exchange). So, energy spot prices are in general regarded to be mean reverting or anti-persistent. Among other financial time series spot electricity prices are perhaps the best example of anti-persistent data. In Weron & Przybyłowicz (2000) and Weron (2002) the R/S analysis, detrended fluctuation analysis and periodogram regression methods were used to verify this claim. Moreover spikes intensity is bigger during on-peak hours (around 9h and 18h on business days), and during high consumption periods (see for instance figure 4.2).

The volatile and spiky nature is a peculiar characteristic of electricity spot prices. This is mainly due to the fact that electricity cannot be stored in an economic feasible way. Serati et al. (2007) assert that to manage the creation of electricity, water reserves can be considered the only substitute method. This is explained by electricity price dynamics in that countries where reserves are abundant (in Scandinavia and United States): peaks are lower than in other markets, due to the great flexibility in the creation phase. Supply and demand must be balanced continuously cause the instantaneous nature of electricity. Shocks occur when there are extreme load fluctuations caused by severe weather conditions often in combination with generation outages or transmission failures, and they can turn into extremely high prices. The increase in demand is balanced using further power plants for electricity production. During peak hours, power plants that utilize fossil fuels (coal, oil and gas plants) are called to generate electricity with the more efficient nuclear or hydro power plants that operate most of the time.

Figure 2.7: *Hourly prices for the spot market (Elspot) at the Nordic power exchange Nord Pool from May, 1992 until December, 2004 (source: Weron, 2005).*

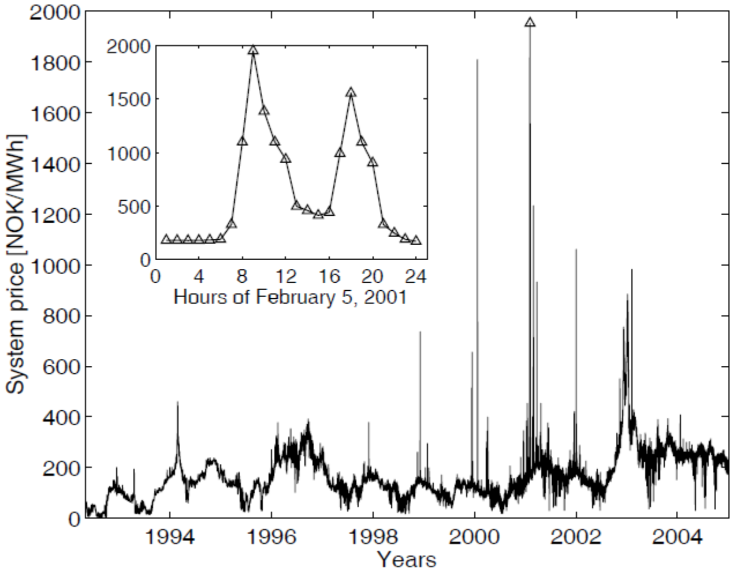


Figure 2.8: *Schematic marginal costs of production with two hypothetical demand curves superimposed on it. Source: Weron et al. (2004b).*

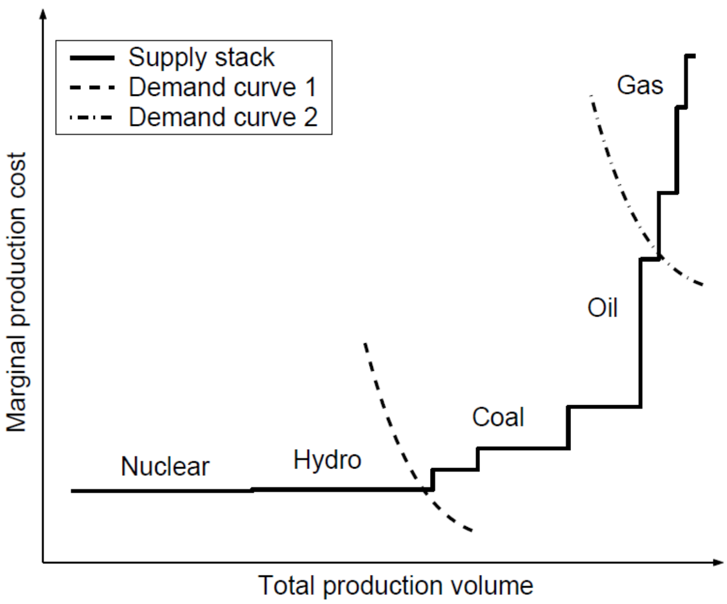


Figure 2.8 show that when demand is high, a larger fraction of power comes from expensive sources. This increases the marginal costs of production, and even a small unexpected positive shock in consumption can force the prices to rise substantially, hence we have a spike. Prices are therefore determined by the excess of capacity available on the system, i.e. the amount of electricity that power plants are able to produce in a specific hour or day (it is called “margin”). If demand is high, margin diminishes and the price rises. So margin is a substantial feature that affect price during peak hours and it is one of the reasons which may determine jumps.

Chapter 3

The electricity price forecasting issue in literature

In literature many different approaches have been presented to resolve the forecasting problem for electricity prices. A great part of these has been used at the beginning for load forecasting with good results. Forecasting electricity prices is not a trivial topic: prices are similar to loads in the sense that both series show seasonality and are dependent on atmospherical factors, but prices show more difficulties. Unanticipated jumps and high volatility are two examples. So, it is necessary to be very careful with the model to use, in fact there are many aspects to consider.

The most studied models to analyze and predict the behavior of electricity prices belong to two main classes: statistical methods and artificial intelligent-based models. Artificial intelligence-based (AI-based) techniques model price processes via non-parametric tools such as artificial neural networks (ANNs), expert systems, fuzzy logic and support vector machines. AI-based models tend to be flexible and can handle complexity and non-linearity. This makes them promising for short-term predictions. Many authors have reported their excellent performance in short-term price forecasting (Catalão et al., 2007; Pino et al., 2008; Vahidinasab et al., 2008), but the advocated models have generally been compared only to other AI-based techniques or very simple statistical methods. For instance, Arciniegas & Arciniegas Rueda (2008) used a Takagi-Sugeno-Kang (TSK) fuzzy inference system in forecasting the one-day ahead real-time peak price of the Ontario Electricity Market. Comparison with time series (ARMAX) and ANN models showed that their model was the more accurate in terms of forecasting. In the other hand, Conejo et al. (2005a) compared different methods for forecasting electricity prices for a day-ahead pool-based electric energy market: three time series specifications (transfer function, dynamic regression and ARIMA), a wavelet multivariate regression technique and a multilayer perceptron with one hidden layer. From results, the ANN technique was the worst out of the five tested models. The last example indicates that there might be serious problems with the efficiency of ANNs and AI-based methods if compared with more sophisticate methods. This is still topic of research.
Statistical approaches are in general preferred for finding the optimal model for electricity prices in terms of its forecasting capabilities. The efficiency and usefulness of some of these methods, like multivariate regression, time series models and

smoothing techniques, in financial markets is often questioned. On the contrary, in power markets these methods do stand a better chance, because of the seasonality prevailing in electricity price processes during non-spiky periods. This makes the electricity prices more predictable than those of very randomly fluctuating financial assets. To enhance their efficiency many of the statistical approaches incorporate fundamental factors, like loads or fuel prices. In the next section the use of statistical models for electricity forecasting purpose is discussed in detail.

3.1 Statistical models

One important feature of electricity prices is the presence of jumps. This behavior can be captured by introducing a Poisson process as in a jump-diffusion model. Its main drawback is that it ignores another fundamental characteristic of electricity prices: the mean reversion to the normal regime. When a jump occurs, the price does not stay in the new level, but reverts to the previous level rapidly. Instead in the jump-diffusion model, if a price spike occurs, the new price level would be assumed as a normal event. The model would proceed randomly via continuous diffusion process with no consideration of the prior price level, and a small chance of returning to the pre-spike level. So, mean-reverting models (arithmetic Ornstein-Uhlenbeck processes) with or without jumps were introduced. Some examples of mean-reverting jump-diffusion models can be find in Deng (2000), Lucia & Schwartz (2002) and Cartea & Figueroa (2005). Deng (2000) studied three models with additional non-linearities, such as regime-switching and stochastic volatility. These aspects allow richer dynamics to emerge, although they are not captured simultaneously in a single specification. His work has been drawn on in the article written by Escribano et al. (2002). The suggested model takes into account the possibility of mean-reversion, volatility clustering (in the form of GARCH effects), jumps (with the possibility of time-dependent intensity) and seasonality (deterministic). The last one is an aspect whose importance is emphasized by Lucia & Schwartz (2002).

In spite of the advantages of introducing jumps in the model there are some limitations in modeling electricity prices by jump-diffusion processes. In fact in this kind of models it is assumed that all the shocks affecting the price series die out at the same rate. Simple economic intuition would argue that this is not a likely case. Larger shocks are push back quite fast by forces of demand and supply, while, when shocks are smaller it is more likely that prices will revert slowly to the previous level due to the existence of adjustment costs. Jump diffusion process tends to capture the smallest and more frequents jumps in the data. As emphasised by Huisman & Mahieu (2003), stochastic jump-models do not disentangle mean-reversion from the reversal of spikes to normal levels. Furthermore, the jump-diffusion modeling approach does not capture the fact that jumps will probably appear in periods of small excess capacity. So, model assumptions for jump intensity (constant or seasonal) are convenient for simulating the distribution of prices over several periods of time, but restrictive for actual short-term predictions for a particular time. For this reason, Escribano et al. (2002) introduced in their model jumps with time-dependent intensity. To capture the rapid decline of electricity prices after a spike, Weron et al. (2004b) postulated that a positive jump should always be followed by a negative jump of approximately the same size. When analyzing average daily prices, spikes typically do not last more than a day so this approach seems to be a good approximation. Using this approach for hourly (or half-hourly) prices is not a good idea. The models developed for daily average prices cannot directly be applied to describe dynamics in hourly prices. For example, if hourly prices revert to an hourly specific mean prices level, then the daily average model with a daily mean will not suffice.

Regime-switching is a good alternative model to jump-diffusion (see also section 3.1.2). This model may be more suitable for actual price forecasting, in fact this can replicate the price discontinuities, observed in practice, and could detach the effects of mean-reversion and spike reversal, aliased in jump-diffusion. To resolve the problems of jump-diffusion approaches described above, Huisman & Mahieu (2003) suggested a model that allows an isolation of the two effects of jumps assuming three market regimes: a regular state with mean-reverting price, a jump regime that creates the spike and finally, a jump reversal regime that ensures with certainty reversion of prices to their previous normal level. This regime-transition structure is however restrictive, as it does not allow for consecutive irregular prices. This constraint is relaxed in de Jong & Huisman (2002) and de Jong (2006). The two-state model proposed assumes a stable mean-reverting regime and an independent spike regime of log-normal prices. Regime independence allows for multiple consecutive regimes, closed-form solutions and translates to a Kalman Filter algorithm in the implementation stage. De Jong (2006) found also that regime-switching models are better able to capture the market dynamics than a GARCH model or a Poisson jump-model. Regime-switching models with two regimes are also developed in Weron et al. (2004a) and Bierbrauer et al. (2004), who coped with the heavy-tailed nature of spike severities allowing log-normal and Pareto distributed spike regimes.

3.1.1 Time series models

Models derived directly from the statistical techniques of load forecasting are generally preferred to predict electricity prices. This is because what is an advantage of stochastic models in derivatives evaluation, such as simplicity and analytical tractability, in forecasting electricity prices it becomes a serious limitation. Just for their simplicity, this kind of models do not take into account a specific feature of electricity prices: time correlations between prices.

In the engineering context the standard model that takes into account the random nature and time correlations of the phenomenon under study is the autoregressive moving average (ARMA) model. In the ARMA model the current value of the process (say, the price) is expressed linearly in terms of its past values (autoregressive part) and in terms of previous values of the noise (moving average part). The ARMA modeling approach assumes that the time series under study is (weakly) stationary. If it is not, then a transformation of the series to the stationary form has to be done first. This can be performed by differencing. The resulting ARIMA model contains autoregressive as well as moving average parts, and explicitly includes differencing in the formulation. If differencing is performed at a larger lag

than 1 then the obtained model is known as seasonal ARIMA or SARIMA.
The forecasting ability of these particular models have been studied in many articles. Conejo et al. (2005a) compared different techniques, which include time series analysis (comprised ARMA), neural networks and wavelets. Using data from the PJM Interconnection in the year 2002, time series techniques reveal themselves as more efficacious than wavelet-transform or neural network techniques, but among time series techniques, the dynamic regression and transfer function algorithms are more effective than ARIMA models. In a subsequent article, Conejo et al. (2005b) proposed a novel technique to forecast day-ahead electricity prices based on the wavelet transform and ARIMA models. The price series is decomposed into three parts using a discrete wavelet transform: the resulting series are modeled with ARIMA processes to obtain 24 hourly predicted values. Then the inverse wavelet transform is applied to yield the forecasted prices for the next 24 hours. The authors concluded that their techniques outperforms the direct use of ARIMA models.

Crespo Cuaresma et al. (2004) studied autoregressive models and autoregressive-moving average models (including ARMA with jumps). They concluded that specifications, where each hour of the day is modeled separately present uniformly better forecasting properties than specifications for the whole time-series. Moreover, the inclusion of simple probabilistic processes for the arrival of extreme price events seems to lead to improvements in the forecasting abilities of univariate models for electricity spot prices. In this article the arrival of shocks is modeled using a binomial process. This implies that the shock arrival process is constant over time (over days at a given hour for the models treating with 24 time series). This choice, made for simplicity, is not so pertinent. In fact, many studies on the dynamics of jumps have underlined that spikes have not constant intensity.
In a related study, Weron & Misiorek (2005) observed that an AR model, where each hour of the day is modeled separately, perform better than a single for all hours, but large (S)ARIMA specification proposed by Contreras et al. (2003). From these results it seems pertinent to underline that modeling each hour of the day separately improves forecasts.

ARIMA-type models relate the signal under study to its own past and do not explicitly use the information contained in other pertinent time series. Electricity prices are not only related to their own past, but may also be influenced by the present and past values of various exogenous factors. These includes for example:

- historical and forecasted loads: load fluctuations translate into variations in electricity prices;
- time factors: the time of the year, the day of the week and the hour of the day influence price patterns;
- fuel prices: in the short-term horizon, the variable cost of power generation is essentially just the cost of the fuel.

So, we can generalizing ARMA model to an autoregressive moving average model with exogenous variables or ARMAX. Time series models with exogenous variables have been extensively applied to short-term price forecasting (see Nogales et al., 2002; Contreras et al., 2003; Knittel & Roberts, 2005; Weron & Misiorek, 2005 and

Misiorek et al., 2006, for some examples) and it is demonstrated that the inclusion in the model of exogenous variables improves forecasts.

Another important characteristic of electricity prices that we cannot ignore is heteroskedasticity. This is one of the non-linearities presented in these series, that show a non-constant conditional variance and clustering of large shocks. Generalized autoregressive conditional heteroskedastic GARCH models consider the moments of a time series as variant (i.e. the error term, real value minus forecasted value, does not have zero mean and constant variance as with an ARIMA process). The error term is now assumed to be serially correlated and can be modeled by an Auto Regressive (AR) process. Thus, a GARCH process can measure the implied volatility of a time series due to price spikes. In literature, this model is often used coupled with autoregression (or a more general (S)AR(I)MA model) but there are cases when modeling prices with GARCH models is advantageous and cases when it is not. For example, Guirguis & Felder (2004) utilized GARCH method to forecast the electricity prices in two regions of New York: New York City and Central New York State. The model is compared to dynamic regression, transfer function models, and exponential smoothing in terms of forecasting accuracy. They found that accounting for the extreme values and the heteroskedactic variance in the electricity price time-series can significantly improve the accuracy of the forecasting. On the other hand, Misiorek et al. (2006) found that models with the additional GARCH component (AR/ARX-G), fail to outperform in point forecasting the relatively simple ARX approach. Knittel & Roberts (2005) evaluated an AR-EGARCH specification on data of California. They found it superior to five other models during the crisis period (May 1, 2000 to August 31, 2000), whereas it yielded the worst forecasts of all models examined during the pre-crisis period. In another article (Garcia et al., 2005), an ARIMA-GARCH model and a general ARIMA model are compared. The GARCH model outperforms the ARIMA model, but only when volatility and price spikes are present.
So, it seems that GARCH models work good in spiky periods, but it is necessary to be careful. As Duffie et al. (1998) pointed out, this kind of models derive erroneous results for electricity prices due the bias introduced by extreme values. These complications are indeed resolved in the presence of a richer price specification with a jump component. The most general one (Escribano et al., 2002) postulates mean-reversion, jump-diffusion and seasonality both in the deterministic price component and the jump intensity. The coexistence of jumps and GARCH dynamics recovers the desired stationarity of the volatility process, as small jumps are often captured by the GARCH components (instead of dominating the estimation of the jump process). Karakatsani & Bunn (2004) resolved the limitations of GARCH models due to extreme values using a regression model with the assumptions of an implicit jump component for prices and a leptokurtic distribution for innovations.
Recently, Bowden & Payne (2008) compared ARIMA, ARIMA-GARCH and ARIMA-GARCH-M models examining in- and out-of-sample forecasting performances. The last specification, that considers also the impact of conditional volatility upon the mean electricity prices, produced better out-of-sample forecasts, but in-sample no one model dominated the others.

3.1.2 Regime switching models

The presence of jumps in electricity price series suggests that there exist a non-linear mechanism switching between normal and high-price state or regimes. For this reason, research suggests the study of regime switching models. The available specifications of regime switching models differ in the way the regime evolves over the time. The regime can be determined in two main ways: by an observable variable or by an unobservable, latent variable. The variety of this kind of models is due to the possibility of choosing both the number of regimes and the different stochastic process for the price in each regime. Especially for the spike regime it may be interesting to choose alternative distributions, like heavy-tailed distributions (this is because spikes happen very rarely but usually are of great magnitude).

The most prominent member of the first class is the threshold autoregressive TAR model with its generalizations due to the presence of exogenous variables (TARX model) or to allowing for specifications of the threshold variable or a gradual transition between the regimes (smooth transition AR model).
Weron & Misiorek (2006) studied various time series specifications, including TAR and TARX (with the system-wide load as exogenous variable) models, and evaluated their predictive capabilities in the California power market. Considering out-of-sample forecasting, the regime-switching approach provides only moderate results during normal (calm) periods, while, during spiky periods TAR-type models perform better but well below acceptable levels as well. In a related study, Misiorek et al. (2006) expanded the range of tested threshold variables. They found that the threshold variable equal to the difference in mean prices for yesterday and eight days ago lead to a much better forecasting performance. The resulting threshold autoregressive models give the best overall results. Both for point and interval forecasting the model outperform most of its competitors and is the best in several of the considered criteria.

Non-linear regime-switching time series models might provide us with good models of electricity price dynamics. However, it is not simple to understand which process governs the regime-switching mechanism. The spot electricity price is the outcome of a vast number of variables including fundamentals (like loads and network constraints) but also the unquantifiable psycho- and sociological factors that can cause an unexpected and irrational buy-out of certain commodities or contracts leading to pronounced price spikes. For this reason, the Markov regime-switching models, where the regime is determined by an unobservable, latent variable, have been studied.

In the literature, mean-reverting processes with Gaussian innovations are typically suggested for the regimes (Huisman & Mahieu, 2003). Other model specifications are also possible and straightforward. The usefulness of Markov regime-switching models for power market applications has been already recognized, in particular their capability of modeling several consecutive price jumps or spikes as opposed to jump-diffusion models. However, their adequacy for forecasting has been only vaguely tested. Only recently this issue has been tackled in the literature. Misiorek et al. (2006) investigated the forecasting power of various time series models, a non-linear Markov regime-switching model with AR(1)-type processes and threshold regime-switching models (TAR and TARX). The models were tested on a time

series of hourly system prices and loads from the California power market. The best results were obtained using a non-linear TARX model and a relatively simple ARX model, with the day-ahead load forecast as exogenous/fundamental variable, while the Markov regime-switching model, failed to outperform the relatively simple ARX approach.
Haldrup & Nielsen (2006a) developed a regime switching model which can generate long memory (fractional integration) in each of the regime states (ARFIMA). The model was adapted to data for the Nordic electricity spot market: electricity spot prices in Nordic countries are characterized by a high degree of long memory, because of the use of hydropower. They found that regime switching and long memory are empirically relevant to co-exist, because of the presence or absence of bottlenecks in electricity transmission that changes the price behaviour. Moreover, from Monte Carlo forecasting results, the regime switching model appears to be especially attractive in forecasting relative prices.
Markov regime-switching models are also considered in the work of Kosater & Mosler (2006) and compared with ordinary linear autoregressive specifications. The obtained results of the forecast study suggest that there is a benefit by taking the non-linear model at least for long-run forecasting.

3.1.3 Time-varying parameter models

In a recent paper, Granger (2008) discussed about the increasing use of time-varying parameter linear models in econometrics. As the author pointed out, the reason of this success is that time-varying parameter linear models can approximate any non-linear model and they have the advantage to be more readily interpretable and to easily produce multi-step forecast. In recent years, the use of this kind of models has been tested also in the electricity context.
Pedregal & Trapero (2007) proposed a univariate dynamic harmonic regression model set up in a state space framework for forecasting prices in the PJM and Spanish markets. Their results highlighted the rapid adaptability of the model to changes in the data and the competitive forecast performance of their method with respect to other results published in the literature by means of ARIMA models.
Karakatsani & Bunn (2008b) studied the forecasting performance of various specifications of time-varying parameter regression models. Their findings underlined that is important to include in the model the time-varying effect of market fundamentals on the electricity price formation. The authors found that this kind of models exhibited the best predictive performance for day-ahead horizons and intra-day trading periods among various alternatives, including autoregressive models with similar coefficient dynamics.

In the light of the advantages that time-varying parameter models show, especially the high adaptability to price structure changes, and the importance of a regime-switching specification, that allows for changes in the price level, some attempts have been done to merge these methods. A recent paper written by Mount et al. (2006) shows that a stochastic regime-switching model with time-varying parameters can capture the type of volatile price behavior observed in many deregulated spot markets for electricity. The structure of the model is very flexible: the mean prices in the two regimes and the two transition probabilities are functions of the load and/or the implicit reserve margin. Correct market information allow to predict price spikes. However, the authors underlined that the accuracy of the prediction is sensitive to the accuracy of the explanatory variables. Following this work, Kanamura & Ōhashi (2008) analyzed the transition probabilities of regime switching in electricity prices by explicitly incorporating the demand/supply structure. In contrast with the usual assumption of constant transition probabilities, their findings show that the transition probabilities depend on both the current demand level relative to the supply capacity and the trends of demand fluctuation. These results open a new issue of research.

Part II

Half-hourly price forecasting of the UK electricity market

Chapter 4

Data Analysis of the UKPX electricity market

4.1 The UK market

The UK electricity market is the oldest organized market for wholesale electricity in Europe. It started the operations in 1990 when, after the reform, the England and Wales Electricity Pool was established and three companies were created. Only two of these companies, National Power (50% of share) and Powergen (30%) were able to set the price, while the third company (Nuclear Electric) was providing baseload, nuclear power which is price-taking. The pool was a compulsory day-ahead one-sided auction market where the electricity was bought and sold on a half-hourly basis.

In march 2001, the pool was replaced by fully liberalized bilateral contracting and voluntary spot trading with the introduction of the New Electricity Trading Arrangements (NETA). Subsequently, three independent power exchanges, the UK Power Exchange (UKPX), the UK Automated Power Exchange (APX UK) and the International Petroleum Exchange (IPE, today called Intercontinental Exchange, ICE) began operations. At the beginning the UKPX was an electricity futures market as IPE, but at the closure of the Electricity Pool it added a spot market that traded half-hourly spot contracts. The APX UK launched a spot market too and in 2003 it was acquired by the Dutch APX. A year later, UKPX and Dutch APX merged into the APX Group.

The UK market, that includes Scotland from March 2005, is currently fully competitive, and perhaps the most mature market in the world. As Karakatsani & Bunn (2008a) and Karakatsani & Bunn (2008b) pointed out studying the market during the period June 2001 - April 2002, there is a strong linkage between price 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 (see section 4.3).

4.2 Preliminary data analysis

All the models proposed in this work are empirically estimated on UK Power Exchange half-hourly market data. This study focuses on the period from 1st of April 2005 to 30th of September 2006: the starting date is important because it refers to the market that included also Scotland from March 2005. All weekend days and Bank holidays were removed from the data (see table 4.1), yielding 380 days for each load period. Each day consists of 48 load periods: period 1 is defined as 00:00-00:30am, period 2 as 00:30-01:00am, and similarly the other periods up to 48 (23:30-00.00pm).

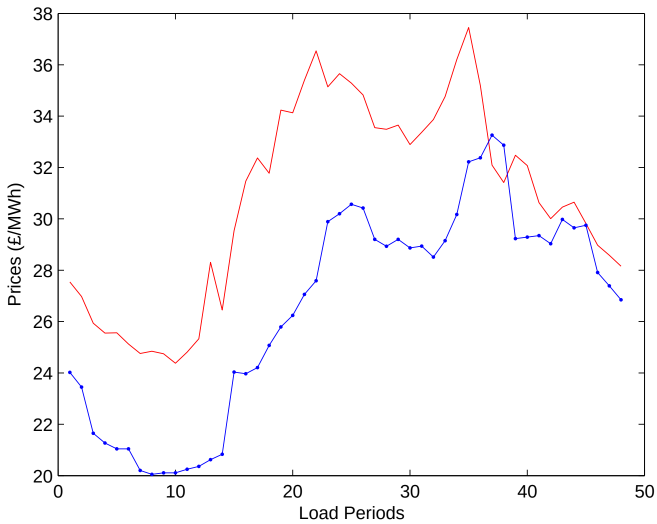
Table 4.1: *Bank holidays removed from the dataset*

02 May 2005	Early May Bank Holiday
30 May 2005	Spring Bank Holiday
29 August 2005	Summer Bank Holiday
26 December 2005	Boxing Day
27 December 2005	Bank Holiday
02 January 2006	New Year's Day
14 April 2006	Good Friday
17 April 2006	Easter Monday
01 May 2006	Early May Bank Holiday
29 May 2006	Spring Bank Holiday
28 August 2006	Summer Bank Holiday

As pointed out in chapter 2.1.1, the reason for removing weekend days and Bank holidays is the different profile that these days show. Figure 4.1 is an example that highlights this characteristic: during non-working days the electricity demand is very low and this affects prices. This feature could affect price forecasting increasing noise and inducing a worsening in the prediction accuracy. Moreover, dropping the weekends from the analysis, the weekly cycle is eliminated with no significant loss of information. This approach has been implemented, for instance, in Ramanathan et al. (1997) and in Karakatsani & Bunn (2008b).

In figure 4.2 price and day-ahead forecasted demand are compared. The two graphs show the respective values for all the 48 load periods during the whole analysis interval, from April 2005 to September 2006. From the demand profile the daily cycle appears clear. During the central hours of the day the demand is higher than during the night, reflecting human activities. In particular we observe a point of very high demand (peak hour) at about 17:00-19:00pm, corresponding to load periods 35-38: this peak is more evident during the winter season. The reason of this increase in the electricity demand is the night lighting. During the winter, sunset occurs earlier so the night lighting coincides with the last working hours of the day.
The seasonal cycle is evident too (see also figure 4.5). The load peak occurs during the winter that corresponds to low temperatures and therefore higher electricity demand for heating purpose.
The prices strongly reflect the demand features with very high peaks in winter and

Figure 4.1: *Half-hourly mean price of working days from the 31st of May to the 3rd of June 2005 (straight line) and half-hourly price during the Spring Bank Holiday of the 30th May 2005 (line with markers).*



summer, especially during high demand load periods. As explained in chapter 2.1.3, this spiky behaviour is strictly connected with supply shocks: the same quantity of electricity can reach very different prices when the available capacity is not the same. To attain a more stable variance, logarithmic transformations of the price series are studied in this work.

Figure 4.3 show the distribution of the 48 half-hourly price logarithms. It is immediately clear that not only the level of the prices but also their variability depends on the corresponding load period. This comment can be extended to other markets since it is a consequence of the instantaneous effect of demand on price. Power markets are organized so that low marginal cost generators are operative during all the day, while flexible plants, typically with high marginal cost, are used only during peak hours. Consequently prices show an extremely high volatility on daily basis. This observation induced authors to consider electricity traded in different hours as different commodities. The firsts were Ramanathan et al. (1997), who applied the method for demand forecasting. Owing to their work, this approach became rather established in load and price modeling and forecasting, as Bunn (2000) and Bunn & Karakatsani (2003) pointed out. The improvement in fitting and prediction accuracy is a result of the increase in homogeneity of the hourly (half-hourly) time series in comparison with the complete one. In particular, in the case of the next-day price forecasting issue, 24 (48) one-step-ahead forecasts that are calculated everyday contain less noise than 24 (48) forecasts with prediction horizons varying from 1 to 24 (48).

In the light of these considerations, in our work models were estimated separately for each load period. In particular, we used four representative periods of the day: period 6 (2:30-3:00am), period 19 (9:00-9:30am), period 28 (13:30-14:00pm) and pe-

Figure 4.2: *Electricity prices and demand forecasts for the 48 load periods during the analysis interval.*

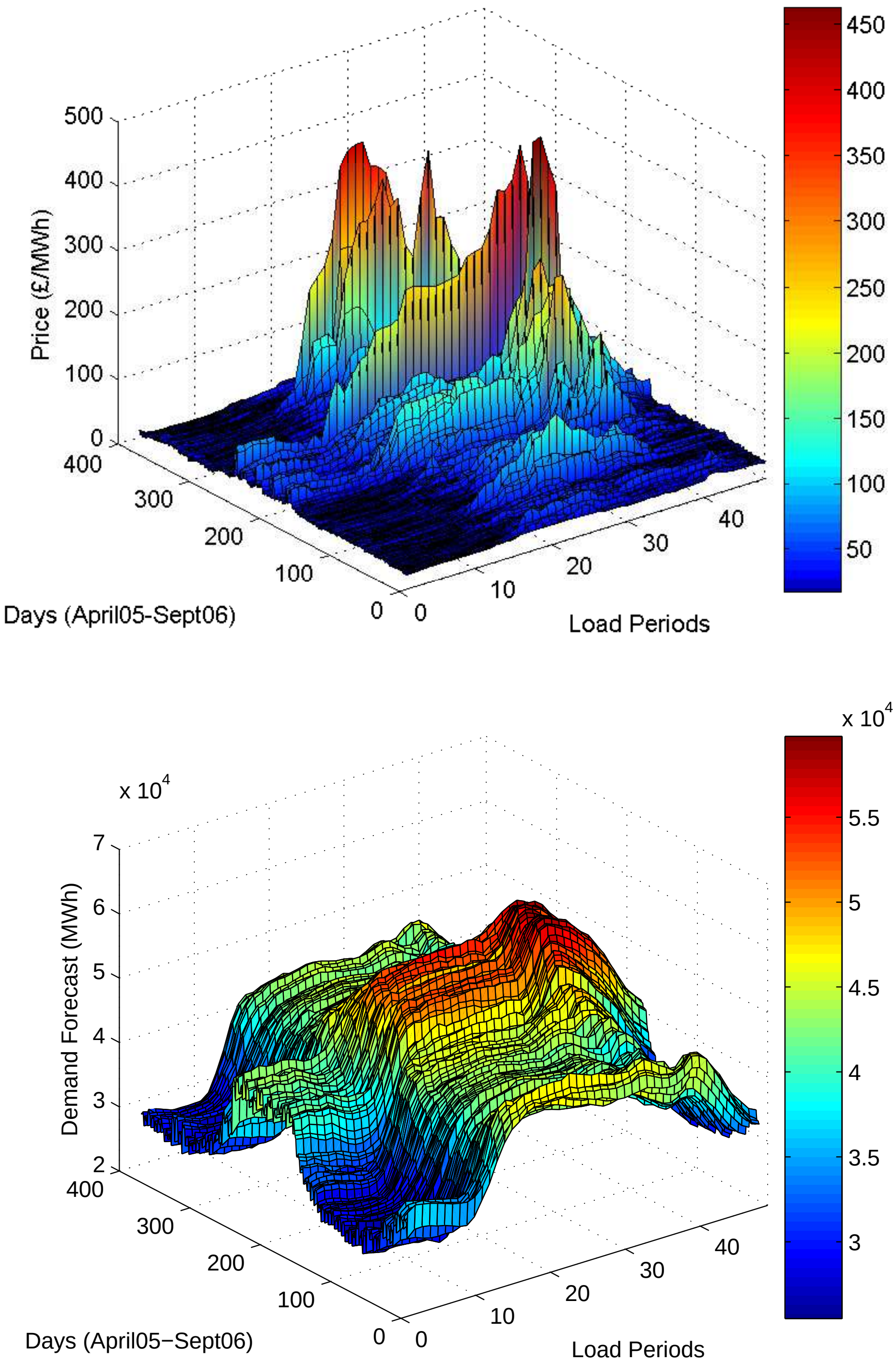
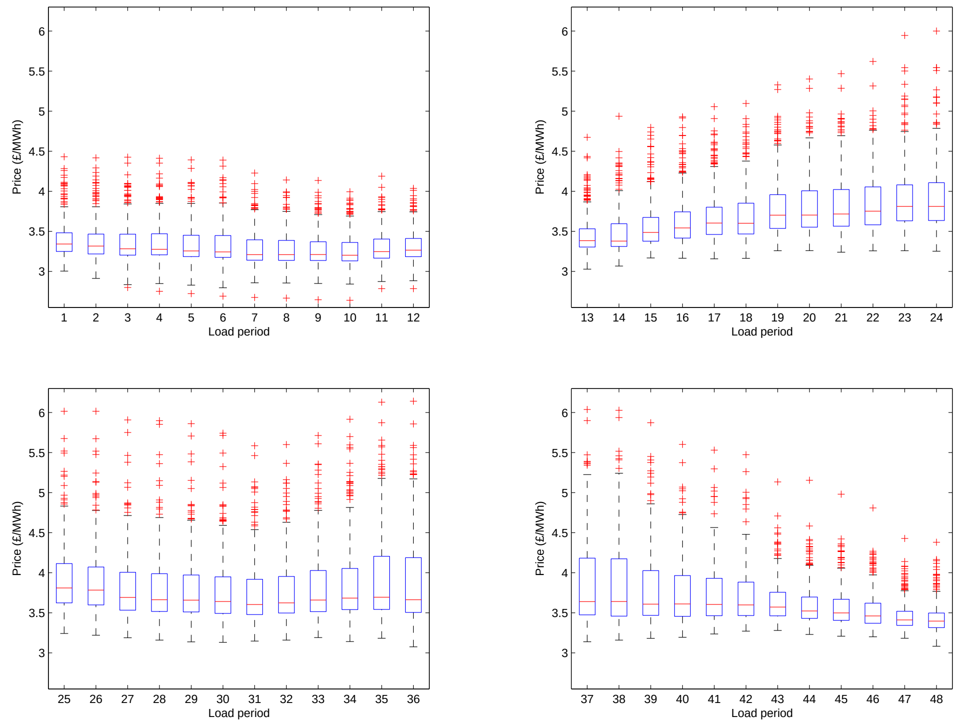


Figure 4.3: *Boxplots of price logarithms for the 48 load periods.*



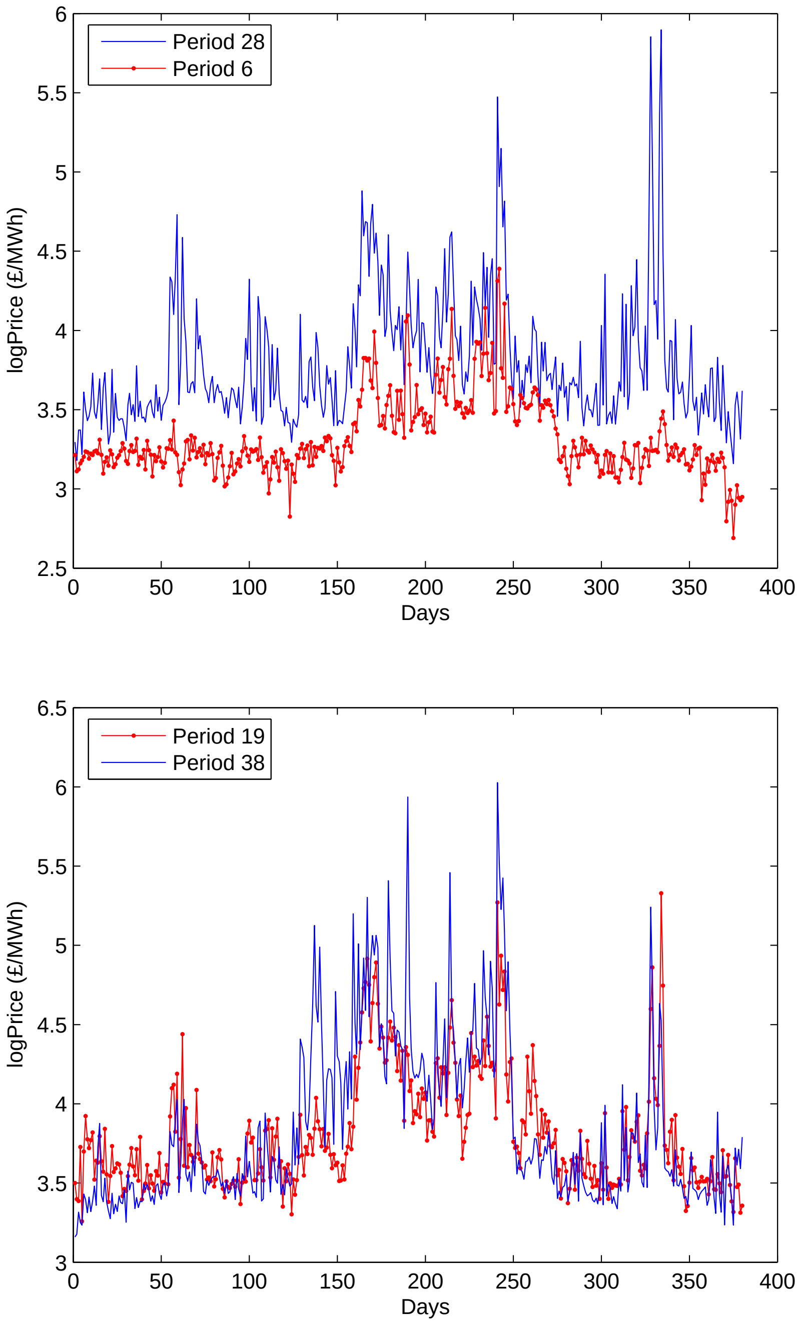
riod 38 (18:30-19:00pm). As figure 4.4 shows, the considered price series are linked by a clear annual trend connected with the variation of the demand (see figure 4.5), with an increase in pricing during the winter season. Differently from the other periods, load period 6 seems more stable but it is still complex to model. In the other three periods volatility is very high, with sudden peaks during winter and summer in both 2005 and 2006. This is because load period 6 is a baseload hour, while the others are peak hours, characterized by supply shocks and high cost generation units.
Unit root tests (Augmented Dickey-Fuller e Phillips-Perron) for prices across this four periods, after adjusting for annual seasonality, rejected the unit root null hypothesis at the 5% significance level. In the same way, KPSS stationarity test cannot reject the null hypothesis at the 5% significance level. A descriptive analysis of these load periods is presented in section 4.4.

4.3 Analysis of the market outcomes

With electricity prices, we considered also the following variables:

Demand Forecast ($DemF$). This is the national day-ahead demand forecast published by the system operator.

Figure 4.4: *Electricity price logarithms for period 6 (2:30-3:00am), period 28 (13:30-14:00pm), period 19 (9:00-9:30am) and period 38 (18:30-19:00pm) from 1 April 2005 to 30 September 2006.*



Indicated Margin (*Margin*). 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 (*DemF*).

Fuel Prices. Our dataset included three different fuel price series:

- daily UK natural gas one-day forward price (*GasF*), from the main National Balancing Point (NBP) hub;
- the Daily Steam Coal Europe-ARA (Amsterdam, Rotterdam and Antwerp) Index (*Coal*);
- the London Brent Crude Oil Index (*Oil*).

For the last two price series we took into account the US dollar to UK sterling exchange rate.

Carbon Emission Price (Co_2). In this study it was used EEX-EU daily emission price (we took into account the EURO to UK sterling exchange rate).

All the series are considered with the logarithmic scale (the logarithmic transformation is underlined by writing the variable names in small letters).
Figure 4.5 shows the dynamics of all the time series described above. From the graphs of the demand forecast, in each load period it is clear the annual seasonality, feature that affects price series. However, demand dynamics do not explain the high variability and jumps showed by prices. The attention must dwell upon the margin series and the gas price series. The gas price trend is very interesting during the winter season: prices are higher than during the rest of the year with jumps reaching values about six times larger than the base price. This is the effect of the increase in gas demand: gas prices are seasonal since demand is very much dependent on temperature. The price series reflect the gas price dynamics, especially in the peak hours. In addition to this relation, connection with margin must be considered. Comparing margin and price, a negative correlation is evident (see also figure 4.7). A decrease in margin induces an increase in prices and sudden positive price peaks appear to be consequence of negative margin jumps.

The variables *coal*, *oil* and *co*2 were omitted from this study: unit root tests and stationary test (Said & Dickey, 1984; Phillips & Perron, 1988; Kwiatkowski et al., 1992) showed the non-stationary nature of these series (see table 4.2) caused by a stochastic trend, while regression fitting showed the spurious relation between price series and these variables.
We followed a different approach for the variable *gasF*. Despite it showed the same characteristics of the other fuel prices, its trend is not of stochastic nature. The strong relation with price series, especially during winter spikes, induced us to include it into the models. This is in accord with Serletis & Shahmoradi (2006) findings: the relation between gas and electricity price is strong, non only on the mean but also on the variance. To avoid possible wrong specification of the models, the *gasF* series was decomposed into two components: the annual trend (*super*)

Figure 4.5: *Half-hourly forecasted demand and indicated margin, UK Power Exchange market, load period 6, 28 and 19, 38. Daily Uk natural gas one-day forward (National Balancing Point), steam coal (Europe-ARA index), London brent crude oil and carbon emission (EEX-EU) prices. Time span: 01/04/2005-30/09/2006 working days only.*

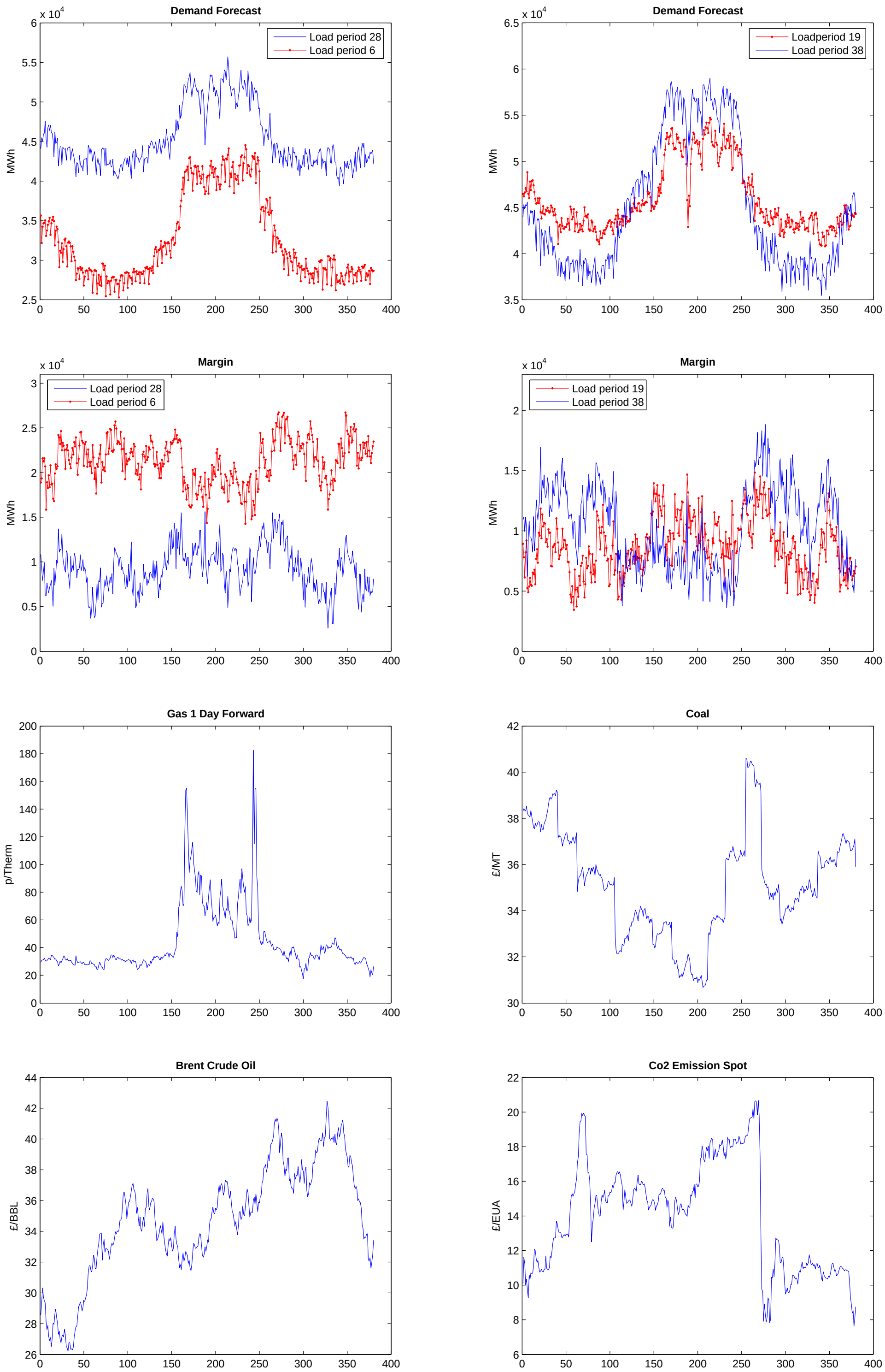


Table 4.2: *Unit root tests results for fuel price series and carbon emission prices (logarithmic scale).*

	ADF	**PP**	**KPSS**
gasF	-1.642	-2.475	1.1224*
coal	-2.036	-2.078	0.9027*
oil	-1.657	-1.860	4.0798*
co2	-1.864	-2.469	0.8726*

Note: *, **, *** *represent respectively* 1%, 5% *and* 10% *significance level. ADF and PP stand for augmented Dickey-Fuller test and Phillips-Perron test respectively. Lag lengths are chosen following Ng & Perron (1995) method.*

obtained from the *gasF* series with the Friedman's Supersmoother (Friedman, 1984), and the series of deviations from the annual trend (*gasF.res*). These series were recalculated at every estimation step because the inclusion of new data (see figure 4.6).

Figure 4.6: *Gas price logarithms one-day forward from 1 April 2005 to 30 September 2006 (line with markers), with the annual trend obtained with the Friedman's Supersmoother.*

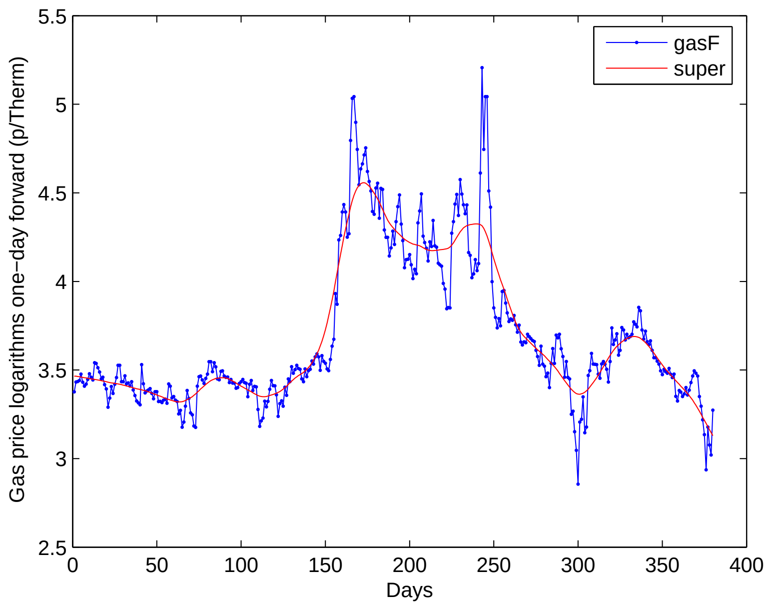
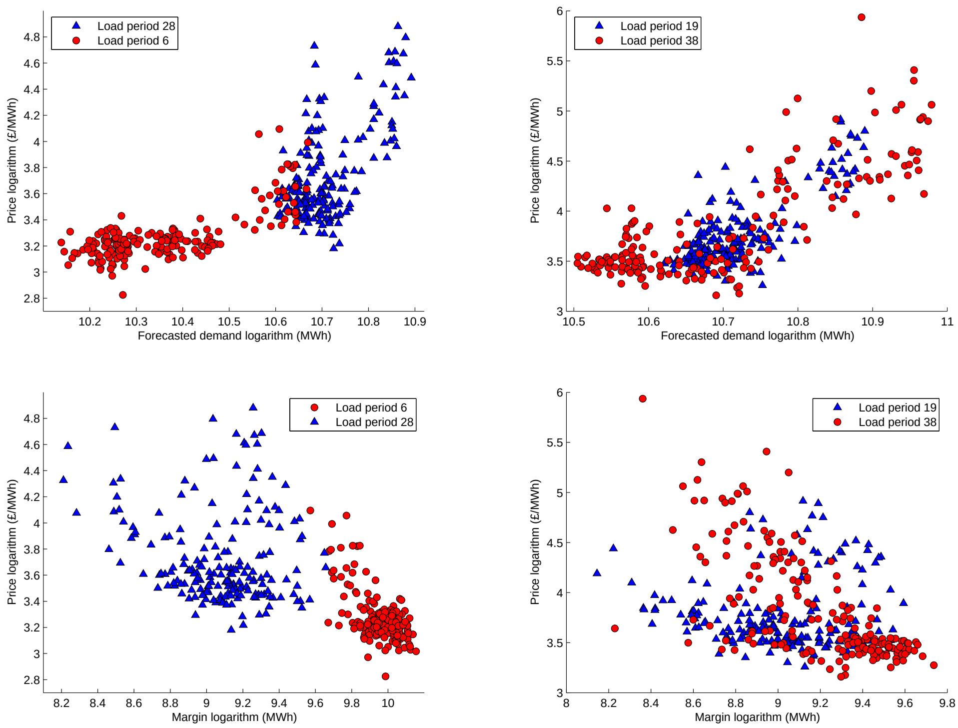


Figure 4.7 show scatterplots between demand and price, and margin and price during the in-sample period from April to December 2005 (see section 4.4 for details). Since relations are not linear, we decided to introduce a quadratic polynomial for both demand and margin, that appear to be the most adequate representation. To resolve collinearity, at every estimation step we used to subtract the mean from the variables $demF$ and $margin$ and then we calculated the quadratic components, denoted as $demF^2$ and $margin^2$.

Figure 4.7: *Scatterplots demand-price and margin-price for load periods 6, 19, 28 and 38 over the period April-December 2005 (all the series are in the logarithmic scale).*



To improve the models we included also the following variables:

Past Prices. Two past values of spot price logarithms were considered: *price* in the same trading period on the previous day (p_{t-1}), and *price* in the same trading period and day in the previous week (p_{t-5}, Bank holidays included to avoid incorrect time lag).

Volatilities. An indicator of instability and risk was defined for electricity price logarithm series ($priceVol$) and for demand forecast logarithm series ($demVol$). This is the coefficient of variation (standard deviation/mean) in a 5 days moving window.

Period Trend ($superP$). We used the Friedman's Supersmoother to obtain the annual trend of the price series. This series was recalculated at every estimation step because the inclusion of new data.

4.4 Forecasting design and prediction error statistics

In order to be able to make out-of-sample predictions, the last 189 observations of the dataset were not considered for model building. Therefore, the sample period can be divided in an in-sample period (April 1st 2005 - December, 31th 2005) and an out-of-sample period (January 1st, 2006 - September 30th, 2006). In our analysis, moreover, the out-of-sample period was divided also in three sub-periods, associated to the different seasons (January-March, 64 data, April-June, 61 data and July-September, 64 data). The scope was to compare performances also in each season to better understand how much forecasting accuracy of each considered model is influenced by a particular period of the year.

Table 4.3: *Descriptive statistics for load periods 6, 19, 28 and 38 over the period April-December 2005 (price logarithms).*

	Period 6	**Period 19**	**Period 28**	**Period 38**
mean	3.2704	3.7760	3.7358	3.8345
st.deviation	0.1899	0.3497	0.3580	0.5385
skewness	1.9190	1.3905	1.2263	1.2478
kurtosis	7.6259	4.2356	3.8107	3.9183

Table 4.4: *Descriptive statistics for load periods 6, 19, 28 and 38 over the out-of-sample period January-September 2006 (price logarithms).*

	Period 6	**Period 19**	**Period 28**	**Period 38**
mean	3.3666	3.8478	3.8580	3.8633
st.deviation	0.2775	0.3772	0.4410	0.5085
skewness	0.8484	1.1863	1.9106	1.4211
kurtosis	4.2165	4.6002	8.0285	5.0560

Table 4.3 displays some statistical moments of the logarithm transformed price series for the considered four load periods in the fitting sample. For comparison, table 4.4 displays the summary statistics in the out-of-sample set. In both in-sample and out-of-sample period, load period 6 has the lowest mean value, followed by load period 28, 19 and 38. This is reasonable considering that period 6 is an off-peak hour, while the others are peak hours (super-peak in the case of period 38). Volatility follows the same scheme, with periods 6 and 28 that reach the standard deviation values of 0.28 and 0.44 respectively considering the year 2006. These increases in

variance depend on the winter profile in the case of period 6, while for load period 28 the cause is the presence of very high jumps affecting prices during summer 2006. The values of skewness and kurtosis show that all the period are characterized by positive asymmetry and fat tails, even if values change a lot depending on the sample: period 6 is the most skewed and fat-tailed in the fitting set, while in the out-of-sample period load periods 28 and 38 deviate considerably from normality.

In general, from the preliminary analysis it is clear that this kind of data is affected by a lot of complexities. For forecasting purpose the main difficulties are the high volatility and jumps. Chapter 5 is dedicated to this issue: forecasting results are presented for different classes of models.

In order to formulate on day t a price forecast for a certain trading period on day $t+1$, the parameters of the models were estimated at each step from a daily expanding dataset and/or from rolling windows of specified lengths. To compare forecasting results we used 4 prediction error statistics:

$$\begin{aligned}
\text{MSE} &= \frac{1}{N}\sum_{i=1}^{N}(P_i - F_i)^2 \\
\text{MSPE} &= \frac{1}{N}\sum_{i=1}^{N}\left(100 \times \frac{P_i - F_i}{P_i}\right)^2 \\
\text{MAE} &= \frac{1}{N}\sum_{i=1}^{N}|P_i - F_i| \\
\text{MAPE} &= \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - F_i}{P_i}\right| \times 100
\end{aligned}$$

where N is the size of the forecasting period, P_i is the real price at time i and $F_i = \exp(f_i)$ is our forecast at time i.

The Mean Squared Error (MSE) is popular, largely because of its theoretical relevance in statistical modeling. 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 is important. If we have, for instance, a normal price forecast and a peak price forecast that give the same MSE value, in the case of the peak price the prediction could be more accurate. This depends on the height of the spike. Statistics based on percentage errors can distinguish between the two situations.

Chapter 5

Predictive models and forecasting results

In this chapter, the forecasting performances provided by four different classes of models are considered. The models are a constant parameter linear regression model (LR), a time-varying parameter regression model (TVR), a Markov regime switching model (MRS) and an ARMAX model. Different specifications of the models and estimation schemes, i.e. expanding and/or rolling windows of different sizes, are also compared. The set of model specifications that we studied includes 19 models. Moreover, the comparison is conducted not only in the whole forecasting period, but also in three sub-samples which reflect three seasons: winter, spring and summer. The scope is to understand how the forecasting accuracy of these different models changes on the grounds of the considered season and if there is a particular specification (class of models, set of regressors and/or estimation window) that outperforms the others in terms of predictive performance.
Forecasting performances will be compared both in terms of descriptive error statistics and in terms of test (Diebold and Mariano test for equal predictive accuracy).

5.1 Multiple linear regression model

The first model that we have considered is a multiple linear regression (LR). This simple model helps to elucidate the average characteristics of price formation and will be the benchmark model in forecasting comparison. For a given load period j, the model is specified as:

$$p_{jt} = \mathbf{X}'_{jt}\boldsymbol{\beta}_j + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim \text{i.i.d}(0, \sigma_j^2) \tag{5.1}$$

where p_{jt} denotes the price logarithm on day t and load period j ($t = 6, 7, ..., T$, $j = 6, 19, 28, 38$), $\boldsymbol{\beta}_j$ a $k \times 1$ vector of constant coefficients and ε_{jt} an i.i.d. error term. $\mathbf{X}_{jt}$ is a $k \times 1$ vector of regressors defined as:

$$\begin{aligned}
\mathbf{X}1_{jt} &= (1, p_{j(t-1)}, p_{j(t-5)}, demF_{jt}, demF_{jt}^2, margin_{jt}, margin_{jt}^2, gasF.res_t, \\
&\quad demVol_{jt}, priceVol_{jt}, super_t)' \\
\mathbf{X}2_{jt} &= (1, p_{j(t-1)}, p_{j(t-5)}, demF_{jt}, demF_{jt}^2, margin_{jt}, margin_{jt}^2, gasF.res_t, \\
&\quad demVol_{jt}, priceVol_{jt})' \\
\mathbf{X}3_{jt} &= (1, p_{j(t-1)}, p_{j(t-5)}, demF_{jt}, demF_{jt}^2, margin_{jt}, margin_{jt}^2, gasF.res_t, \\
&\quad demVol_{jt}, priceVol_{jt}, superP_{j(t-1)})' \ ,
\end{aligned}$$

where $\mathbf{X}1, \mathbf{X}2, \mathbf{X}3$ are three different sets of regressors, referred to "model 1", "model 2" and "model 3" respectively. The difference among them is mainly the specification of the annual trend component. In model 1 the seasonal trend is given by $super$, in model 3 by $superP$, while in model 2 the period trend is given by the forecasted demand. The scope of using three sets of regressors is to understand if one of the specifications gives better forecasts than the others because of the inclusion of the annual trend profile. As pointed out in section 2.1.1, the sinusoidal approach is a common technique used by many authors to capture the annual seasonality (see, for example, Misiorek et al., 2006; Karakatsani & Bunn, 2008a). We preferred to not use this approach. The reason is that this seasonal function has to be estimated on historical data. Our data concern the UK market straight after the inclusion of Scotland, so we do not have information about the seasonal component of this new market. Our annual trend is calculated day by day using only the available information by the gas or the price series, as explained in section 4.3.
Standard assumptions for model 5.1 that contains stochastic regressors include:

1. $\{p_{jt}, \mathbf{X}_{jt}\}$ is jointly stationary and ergodic: regressors with stochastic trends are not included;

2. the regressors are predetermined or exogenous: endogenous variables are excluded but the lagged response is allowed. The variables x_t used in this analysis are known to the market at time $t-1$.

Under these assumptions, the OLS estimates are consistent and asymptotically normally distributed even under non-normal, i.i.d. errors.

Linear regression models were derived in-sample (period April-December 2005) for trading periods 6, 19, 28 and 38 with stepwise backward techniques (AIC criterion) and are as follow:
Period 6 ($j = 6$):

$$\begin{aligned}
\mathbf{X}1_{jt} &= (1, p_{j(t-1)}, margin_{jt}, margin_{jt}^2, gasF.res_t, demVol_{jt}, super_t)' \\
\mathbf{X}2_{jt} &= (1, p_{j(t-1)}, demF_{jt}, demF_{jt}^2, margin_{jt}, margin_{jt}^2, demVol_{jt})' \\
\mathbf{X}3_{jt} &= (1, p_{j(t-1)}, margin_{jt}, margin_{jt}^2, gasF.res_t, demVol_{jt}, superP_{j(t-1)})' \ ;
\end{aligned}$$

Period 19 ($j = 19$):

$$\begin{aligned}
\mathbf{X}1_{jt} &= (1, p_{j(t-1)}, margin_{jt}, margin_{jt}^2, gasF.res_t, super_t)' \\
\mathbf{X}2_{jt} &= (1, p_{j(t-1)}, p_{j(t-5)}, demF_{jt}, demF_{jt}^2, margin_{jt}, margin_{jt}^2, demVol_{jt},)' \\
\mathbf{X}3_{jt} &= (1, demF_{jt}, margin_{jt}, margin_{jt}^2, gasF.res_t, superP_{j(t-1)})' \ ;
\end{aligned}$$

Period 28 ($j = 28$):

$$\begin{aligned}
\mathbf{X}1_{jt} &= (1, p_{j(t-1)}, demF^2_{jt}, margin_{jt}, margin^2_{jt}, super_t)' \\
\mathbf{X}2_{jt} &= (1, p_{j(t-1)}, demF_{jt}, demF^2_{jt}, margin_{jt}, margin^2_{jt})' \\
\mathbf{X}3_{jt} &= (1, demF_{jt}, demF^2_{jt}, margin_{jt}, margin^2_{jt}, superP_{j(t-1)})' \;;
\end{aligned}$$

Period 38 ($j = 38$):

$$\begin{aligned}
\mathbf{X}1_{jt} &= (1, p_{j(t-1)}, p_{j(t-5)}, margin_{jt}, super_t)' \\
\mathbf{X}2_{jt} &= (1, p_{j(t-1)}, demF_{jt}, demF^2_{jt}, margin_{jt})' \\
\mathbf{X}3_{jt} &= (1, p_{j(t-1)}, margin_{jt}, margin^2_{jt}, superP_{j(t-1)})' \;.
\end{aligned}$$

These final sets of regressors are used also for models 5.2 and 5.4.
For all these multiple regression models, the inspection of the variance inflation factors (VIF) doesn't show collinearity problems. The graphs of the autocorrelation function for the residuals don't show any particular problem. Table 5.1 contains estimation results for the three specificated models in each load period on the period April-December 2005 (in-sample). From the results, models do not appear to be misspecified (see the Durbin-Watson statistic values). Only for period 38, the Ljung-Box statistic is not significative at 1% significance level: this underlines that this trading period is very difficult to model with a linear regression. There is something that this kind of models cannot explain.
Regressor *margin* is significant in all the models and this underlines the importance of this factor in price formation. As expected from, the sign for the variable is negative: the lower reserve margin, the higher the price becomes. This effect is particulary evident during peak hours where the *margin* coefficient values are higher (absolute values). The positive coefficient of $demF$ reflects the demand-price relationships from the increasing supply function. In general this regressor is not present in models 1 and 3. Only in load periods 19 and 28, $demF$ and $superP$ are both present. An explanation could be the absence of the lagged price p_{-1}.
The adjusted R squared values range from 69% to 87%. It is interesting to observe that for each load period the highest adjusted R squared values and the lowest standard deviations are obtained with model 3.

For each load period j and model specification, parameters $\boldsymbol{\beta}_j$ are estimated at every step from a daily expanding dataset, then out-of-sample one-day ahead spot price forecasts are obtained as:

$$\begin{aligned}
F_{j(t+1)} &= \exp(f_{j(t+1)}) \\
&= \exp(\mathbf{X}'_{j(t+1)}\hat{\boldsymbol{\beta}}^t_j).
\end{aligned}$$

Tables 5.2, 5.3, 5.4 and 5.5 summarize the performance of the three models for period 6, 19, 28 and 38 respectively under the three forecasting periods (January-March, April-June, July-September) and the whole period from January to September 2006. From prediction results for the whole forecasting period, it is clear that even if model 3 gives best results in fitting, it gives in general worst results in forecasting. During

Table 5.1: *In-sample parameter estimates (April-December 2005) for the linear regression models (LR1=Linear Regression model 1, etc.). Load periods 6, 19, 28 and 38.*

	Load period 6			**Load period 19**			**Load period 28**			**Load period 38**		
	LR1	LR2	LR3	LR1	LR2	LR3	LR1	LR2	LR3	LR1	LR2	LR3
intercept	2.289*	1.934*	2.718*	3.087*	2.085*	3.764*	3.076*	2.655*	3.706*	1.954*	2.278*	3.353*
p_{-1}	0.284*	0.386*	0.154**	0.179*	0.301*	—	0.162*	0.257*	—	0.355*	0.379*	0.121***
p_{-5}	—	—	—	—	0.109**	—	—	—	—	0.140**	—	—
demF	—	0.349*	—	—	1.796*	0.940*	—	1.483*	1.022*	—	0.688*	—
$demF^2$	—	1.346*	—	—	12.02*	—	7.534**	18.13*	4.510	—	5.914*	—
margin	−0.34*	−0.17***	−0.23*	−0.42*	−0.13*	−0.20*	−0.58*	−0.35*	−0.33*	−0.51*	−0.50*	−0.33*
$margin^2$	1.210**	0.999**	1.260*	0.251*	0.438*	0.254*	0.403*	0.571*	0.264**	—	—	0.319**
gasF.res	0.216**	—	0.256*	0.200***	—	0.327*	—	—	—	—	—	—
demVol	13.60***	12.53***	11.75***	—	46.37*	—	—	—	—	—	—	—
super	0.192*	—	—	0.712*	—	—	0.538*	—	—	0.326*	—	—
$superP_{-1}$	—	—	0.701*	—	—	0.910*	—	—	0.807*	—	—	0.750*
σ_ε	0.096	0.100	0.091	0.131	0.145	0.128	0.184	0.197	0.177	0.247	0.245	0.225
AdjR^2	0.748	0.727	0.772	0.860	0.828	0.865	0.732	0.691	0.752	0.788	0.792	0.824
D-W stat	1.874	1.928	1.808	2.049	2.037	1.894	1.936	1.966	1.850	1.806	1.745	1.797
J-B stat	392.8*	404.9*	83.27*	6.592**	7.933**	1.240	50.36*	36.00*	35.53*	172.5*	153.5*	98.23*
Ljung-Box	15.86	19.58	15.91	20.27	21.54	19.21	35.98**	31.85***	27.32	48.84*	45.57*	43.52*

*Note: *, **, *** means significance at 1%, 5% and 10% level respectively. D-W stat represents Durbin-Watson statistic, while J-B stat is the Jarque-Bera Normality test. Lag for the Ljung-Box statistics depends on the sample length: it is* $10\log_{10}(n) \approx 22$.

the winter season it is difficult to obtain accurate predictions for all the considered load periods, but it seems that model 1 gives better forecasts for periods 19, 28 and 38 while model 2 overcomes model 1 in load period 6. A possible explanation is that peak hours are highly influenced by gas price and margin in this season, as pointed out in section 4.3, so the variables *super* and *margin* help to forecast better the spikes that characterize these load periods. Prediction error statistics values are high also during summer for the peak hours because of very high spikes.

Table 5.2: *Prediction error statistics for the multiple regression models (LRi=Linear Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 6.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	LR1	60.72	231.78	5.02	11.17
	LR2	47.97	174.55	4.48	10.01
	LR3	55.01	302.96	5.39	13.23
Apr-Jun	LR1	8.97	94.72	2.36	8.14
	LR2	8.42	97.52	2.30	8.11
	LR3	6.26	82.32	2.03	7.45
Jul-Sept	LR1	7.16	162.27	2.00	8.93
	LR2	5.20	132.04	1.72	7.79
	LR3	5.28	131.03	1.69	7.62
Whole	LR1	25.88	164.00	3.14	9.44
	LR2	**20.72**	**135.29**	**2.84**	**8.65**
	LR3	22.44	173.53	3.05	9.47

Note: in bold the better statistics for the whole forecasting period (from January to September).

5.1.1 Tests for parameter stability

The time series regression model 5.1 assumes for each load period that the parameters of the model, $\boldsymbol{\beta}$, are constant over the estimation sample. A way to investigate parameter constancy is to compute recursive estimates of the parameters. The model is estimated by least squares recursively for $t = 2, ..., T$ giving $T - 1$ recursive least squares (RLS) estimates. If $\boldsymbol{\beta}$ is really constant then the recursive estimates $\hat{\boldsymbol{\beta}}_t$ should quickly settle down near a common value. If some of the elements in $\boldsymbol{\beta}$ are not constant then the corresponding RLS estimates should show instability.
Starting from RLS, Brown et al. (1975) proposed two simple tests for parameter instability. These tests are known as the CUSUM and CUSUMSQ tests. CUSUM test is based on the cumulated sum of the standardized recursive residuals, while the CUSUMSQ test is based on the cumulative sum of the squared standardized recursive residuals. For both methods, Brown et al. (1975) gave approximate 95% confidence bands: if CUSUM or CUSUMSQ value for some t lies outside of these bands, then there is evidence of some form of parameter instability.
As Coutts et al. (1997) pointed out, the CUSUMSQ test have good properties against heteroscedasticity. If the parameters of equation 5.1 are time varying but are estimated by OLS as being constant, then the residuals will be heteroscedastic: this type of misspecification can be detected by the CUSUMSQ test.

Table 5.3: *Prediction error statistics for the multiple regression models (LRi=Linear Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 19.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	LR1	423.25	526.47	11.95	16.86
	LR2	491.10	736.07	13.71	20.12
	LR3	426.06	730.57	13.11	19.35
Apr-Jun	LR1	68.37	218.71	5.42	11.59
	LR2	52.60	178.93	4.52	10.02
	LR3	76.07	257.71	5.80	12.63
Jul-Sept	LR1	385.80	489.33	8.63	15.55
	LR2	387.43	412.85	8.26	14.26
	LR3	499.10	1213.37	11.15	21.80
Whole	LR1	**296.03**	**414.56**	**8.72**	**14.72**
	LR2	314.47	446.80	8.90	14.88
	LR3	337.83	741.44	10.09	18.01

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.4: *Prediction error statistics for the multiple regression models (LRi=Linear Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 28.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	LR1	791.50	676.46	15.54	21.08
	LR2	829.23	839.31	16.25	22.88
	LR3	849.31	1222.28	17.39	26.01
Apr-Jun	LR1	96.39	391.70	7.31	16.40
	LR2	84.58	360.33	6.31	14.30
	LR3	115.12	403.16	7.39	15.82
Jul-Sept	LR1	614.51	576.32	12.67	19.31
	LR2	693.97	585.29	13.18	18.95
	LR3	1203.54	1026.22	17.25	26.16
Whole	LR1	**507.22**	**550.64**	**11.91**	18.97
	LR2	543.09	598.70	12.00	**18.78**
	LR3	732.30	891.52	14.12	22.77

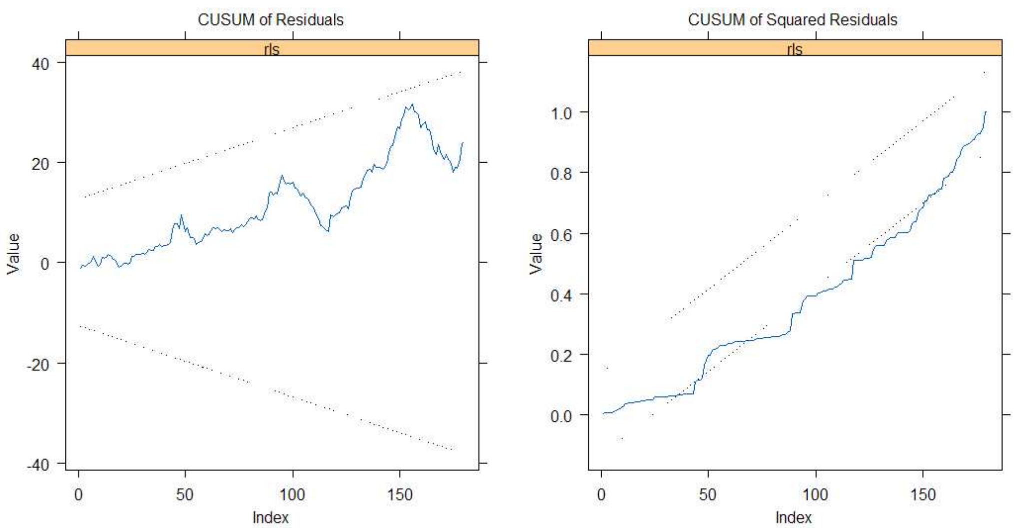
Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.5: *Prediction error statistics for the multiple regression models (LRi=Linear Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 38.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	LR1	2312.75	844.11	24.09	23.10
	LR2	2231.77	1050.08	25.02	25.41
	LR3	2011.45	1247.55	25.96	28.06
Apr-Jun	LR1	49.56	256.81	5.00	12.76
	LR2	55.31	306.48	5.37	13.73
	LR3	62.59	310.78	5.39	13.52
Jul-Sept	LR1	326.11	676.26	9.49	19.90
	LR2	378.79	764.69	10.02	20.67
	LR3	392.21	554.42	10.14	19.46
Whole	LR1	909.58	**597.72**	**12.99**	**18.68**
	LR2	901.85	713.44	13.60	20.04
	LR3	**834.14**	710.49	13.96	20.46

Note: in bold the better statistics for the whole forecasting period (from January to September).

Figure 5.1: *Example of CUSUM and CUSUMSQ tests. Load period 28, model 2.*



We calculated RLS estimates of the parameters of the three model specifications in each load period, and then we conducted the two tests for stability. Our findings show that there is strong evidence of instability in the parameters for all the models. Figure 5.1 is an example of the tests results for load period 28, model 2. CUSUMSQ values lie out of the bands, underlining that residuals are affected by heteroscedasticity. Even if the CUSUM values do not cross the bands, the trend and the structural changes in the value series are effects of instability in the parameters. Given our conclusions, it is important to model the form of parameter variation. We choose to consider here two types of parameter variation: one in which the change in the parameters is determined by a discrete variable which evolves according to a Markovian process (Markov regime-switching model, see section 5.2) and one in which the change is stochastic and assumed to be generated by a random walk (time-varying multiple regression model, see section 5.3).

5.2 Markov regime-switching model

Markov regime-switching models are frequently discussed in the literature that deals with electricity spot-prices. Even so, in the studies that have been done the focus goes principally to the mere estimation. Forecasting ability of this kind of non-linear models is an open issue tackled in literature only in the last years (see, for instance, Kosater & Mosler, 2006; Misiorek et al., 2006; Karakatsani & Bunn, 2008b).
In this section, forecasting results obtained with different specification of Markov regime-switching regression models are presented. Specifications include the use of three different sets of regressors, as pointed out in section 5.1, and different sample lengths for estimation.

The Markov regime-switching model (MRS) is defined as:

$$p_{jt} = \mathbf{X}'_{jt}\boldsymbol{\beta}_{jS_t} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim N(0, \sigma^2_{jS_t}), \tag{5.2}$$

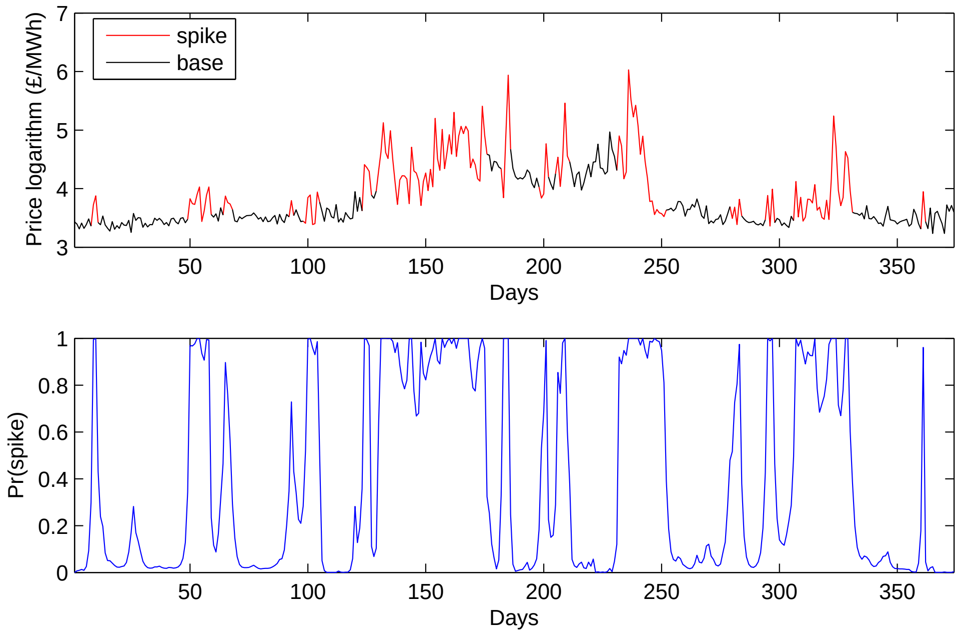
$$\Pr(S_t = i|S_{t-1} = h) = \pi_{ih}, \ \forall i, h \in S \tag{5.3}$$

where p_{jt} denotes the price logarithm on day t and load period j ($t = 6, 7, ..., T$, $j = 6, 19, 28, 38$), 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 (as specified in section 5.1), $\sigma^2_{jS_t}$ the error variance in regime S_t and π_{ih} the transition probability between states i and h.
In this model maximum likelihood estimates of $\boldsymbol{\beta}_{jS_t}$ and $\sigma^2_{jS_t}$ were calculated by the EM algorithm while, to calculate smoothed inferences of regimes, Kim's algorithm was used (see Kim, 1994 and Hamilton, 1994, chap. 22 for a complete explanation about the estimation procedure).
The model assumes that the market at each time point is in one of the 2 possible states, indexed by an unobservable discrete variable, S_t, which evolves according to a first-order, homogeneous Markovian process. At each time point the model parameters are a function of the prevailing state S_t, so that each market regime is characterized by a distinct regression price model. Prices are not classified into regimes a priori, but endogenously through the latent state estimation and probabilistic inference.

Figure 5.2 shows an example of regime classification during the estimation of the MRS model for load period 38 using the set of regressors **X**3. Each observation is classified to be in the base regime or in the spike regime with a certain probability. From the figure it appears that the model works well, in fact spikes are correctly classified in the spike regime.

Tables 5.6 and 5.7 contain in-sample coefficients' estimates for the Markov regime-

Figure 5.2: *Example of regime classification with the Markov regime-switching model. Load period 38, model 3.*



switching models in load periods 6, 19, 28 and 38. All variables had a statistically significant effect in at least one regime (with the only exception of p_{-1} for period 38 model 3), but the magnitudes of their coefficients displayed great variation. The Ljung-Box statistics shows an improvement respect to the linear regression model: all the values are non significative. Also the adjusted R squared values confirm a better adaptation of the model to data $(84 - 93\%)$.

Given the Markov r-s model and load period j, we calculated the price forecast as the expected value, i.e. the linear combination of predicted price logarithm across regimes weighted by predicted regime probabilities:

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

Table 5.6: *In-sample parameter estimates (April-December 2005) for the Markov regime-switching models estimated on expanding dataset, load periods 6 and 19 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$).*

	Load period 6					
	MRS1		**MRS2**		**MRS3**	
	base $(i=1)$	**peak** $(i=2)$	**base** $(i=1)$	**peak** $(i=2)$	**base** $(i=1)$	**peak** $(i=2)$
$intercept$	2.577*	2.898*	1.987*	4.148*	2.347*	3.169*
p_{-1}	0.214*	0.079	0.380*	−0.22**	0.291*	0.016
p_{-5}	—	—	—	—	—	—
$demF$	—	—	0.289*	4.039*	—	—
$demF^2$	—	—	0.941**	−9.49*	—	—
$margin$	−0.01	−0.75*	−0.14***	−2.15*	0.147*	−0.34*
$margin^2$	0.224	1.490**	0.143	−1.61***	0.954*	0.530
$gasF.res$	0.054	0.413*	—	—	0.127***	0.297*
$demVol$	−5.99	29.35**	4.633	−53.8*	−14.6*	15.85**
$super$	0.121*	0.254*	—	—	—	—
$superP_{-1}$	—	—	—	—	0.236*	1.060*
σ_ε	0.049	0.106	0.073	0.044	0.035	0.086
π_{ii}	0.844	0.817	0.964	0.482	0.582	0.751
$\Pr(S=i)$	0.539	0.461	0.935	0.065	0.373	0.627
AdjR^2	0.838		0.863		0.872	
Ljung-Box	17.23		11.41		18.74	

	Load period 19					
	MRS1		**MRS2**		**MRS3**	
	base $(i=1)$	**peak** $(i=2)$	**base** $(i=1)$	**peak** $(i=2)$	**base** $(i=1)$	**peak** $(i=2)$
$intercept$	2.149*	3.362*	1.728*	3.928*	3.791*	3.735*
p_{-1}	0.424*	0.106***	0.416*	−0.13*	—	—
p_{-5}	—	—	0.104**	−0.06	—	—
$demF$	—	—	1.803*	0.315***	−0.714**	2.081*
$demF^2$	—	—	9.830*	37.90*	—	—
$margin$	−0.39*	−0.41*	−0.21*	0.151*	−0.36*	−0.10*
$margin^2$	0.274*	0.335*	0.460*	1.077*	0.330*	0.365*
$gasF.res$	−0.90*	0.471*	—	—	−0.46*	0.839*
$demVol$	—	—	28.05**	140.3*	—	—
$super$	0.412*	0.778*	—	—	—	—
$superP_{-1}$	—	—	—	—	1.221*	0.627*
σ_ε	0.029	0.132	0.119	0.041	0.098	0.099
π_{ii}	0.237	0.807	0.910	0.539	0.649	0.684
$\Pr(S=i)$	0.202	0.798	0.836	0.164	0.474	0.526
AdjR^2	0.888		0.904		0.933	
Ljung-Box	22.13		25.47		17.73	

*Note: *, **, *** means significance at 1%, 5% and 10% level respectively. π_{ii} is the probability of remaining in the same regime in the next time step, and $\Pr(S=i)$ is the unconditional probability of being in regime i (ergodic probability). Lag for the Ljung-Box statistics depends on the sample length: it is $10\log_{10}(n) \approx 22$.*

Table 5.7: *In-sample parameter estimates (April-December 2005) for the Markov regime-switching models estimated on expanding dataset, load period 28 and 38 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$).*

	Load period 28					
	MRS1		**MRS2**		**MRS3**	
	base ($i = 1$)	**peak** ($i = 2$)	**base** ($i = 1$)	**peak** ($i = 2$)	**base** ($i = 1$)	**peak** ($i = 2$)
intercept	2.781*	4.025*	2.732*	3.798*	3.511*	3.823*
p_{-1}	0.203*	−0.06	0.208*	−0.01	—	—
p_{-5}	—	—	—	—	—	—
demF	—	—	0.296**	3.400*	0.247***	1.328*
$demF^2$	15.33*	8.563***	20.39*	17.13*	18.42*	−0.14
margin	−0.27*	−0.87*	−0.22*	−0.60*	−0.20*	−0.43*
$margin^2$	0.274*	0.235	0.332*	0.360**	0.380*	0.093
gasF.res	—	—	—	—	—	—
demVol	—	—	—	—	—	—
super	0.167*	0.719*	—	—	—	—
$superP_{-1}$	—	—	—	—	0.312*	0.899*
σ_ε	0.072	0.185	0.077	0.184	0.066	0.175
π_{ii}	0.744	0.731	0.753	0.711	0.668	0.686
$\Pr(S = i)$	0.512	0.488	0.539	0.461	0.486	0.514
AdjR^2	0.874		0.879		0886	
Ljung-Box	14.12		13.14		18.02	

	Load period 38					
	MRS1		**MRS2**		**MRS3**	
	base ($i = 1$)	**peak** ($i = 2$)	**base** ($i = 1$)	**peak** ($i = 2$)	**base** ($i = 1$)	**peak** ($i = 2$)
intercept	2.167*	3.161*	2.265*	3.097*	3.822*	3.704*
p_{-1}	0.291*	0.176**	0.345*	0.192**	−0.027	0.042
p_{-5}	0.118*	0.022	—	—	—	—
demF	—	—	0.716*	0.835**	—	—
$demF^2$	—	—	7.234*	3.936**	—	—
margin	−0.13*	−1.10*	−0.14*	−0.99*	−0.07*	−0.82*
$margin^2$	—	—	—	—	0.041	0.526***
gasF.res	—	—	—	—	—	—
demVol	—	—	—	—	—	—
super	0.473*	0.352*	—	—	—	—
$superP_{-1}$	—	—	—	—	0.749*	0.589*
σ_ε	0.076	0.270	0.072	0.269	0.070	0.250
π_{ii}	0.917	0.922	0.911	0.917	0.905	0.908
$\Pr(S = i)$	0.483	0.517	0.483	0.517	0.491	0.509
AdjR^2	0.868		0.868		0.889	
Ljung-Box	23.89		25.32		29.93	

Note: *, **, *** *means significance at* 1%, 5% *and* 10% *level respectively.* π_{ii} *is the probability of remaining in the same regime in the next time step, and* $\Pr(S = i)$ *is the unconditional probability of being in regime i (ergodic probability). Lag for the Ljung-Box statistics depends on the sample length: it is* $10 \log_{10}(n) \approx 22$.

We had also considered the predicted price from the regime with the highest predicted probability of occurrence as price forecast, i.e.

$$f_{j(t+1)} = f^i_{j(t+1)} \quad \text{where } i = \underset{s\in\{1,2\}}{\operatorname{argmax}} \left[\hat{\mathrm{P}}(S_{t+1} = s|I_t)\right],$$

but results were worst than those obtained with the expected values.
The spot price forecast is then calculated as

$$F_{j(t+1)} = \exp(f_{j(t+1)}).$$

Tables 5.8, 5.9, 5.10 and 5.11 summarize the one-day ahead forecasting performance of the three Markov regime switching models estimated on daily expanding dataset for period 6, 19, 28 and 38 respectively under the three forecasting periods and the whole period from January to September 2006. Prediction error statistics for load period 38 show an improvement in accuracy respect to the linear regression models, probably because of the high variability that characterizes this hour. For the other trading periods, comparisons are not so clear, but from period 6 we can see that MRS models give better results than the respective LR models during spring and summer (calm periods), and worst results during the winter season.

The parameters of the Markov regime-switching models were also estimated on rolling windows with fixed length of 3, 6 and 9 months (66, 132 and 186 days). The scope is to test if an MRS model estimated on a sliding window can produce better forecasts than its expanding counterpart. Results are summarize in tables 5.12, 5.13, 5.14 and 5.15. For load period 6, it is clear an improvement during the winter season (that affect prediction error statistics also on the whole period) especially in MSE: some peaks are better forecasted. The same happen for load period 38, in which improvements concern also the summer season with rolling windows of 6 and 9 months. A rolling window is helpful also to better predict the two high peaks affecting load period 28 during summer.
In general, it seems that Markov regime-switching models estimated on rolling windows can help to forecast during spiky periods.

Table 5.8: *Prediction error statistics for the Markov regime-switching regression models estimated on a daily expanding dataset, period 6 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$; F.P.=Forecasting Period).*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	MRS1	69.61	251.45	5.31	11.75
	MRS2	65.20	234.28	4.97	11.03
	MRS3	73.92	316.23	5.53	12.80
Apr-Jun	MRS1	8.11	80.97	2.22	7.59
	MRS2	5.98	68.99	1.95	6.89
	MRS3	4.50	62.81	1.63	6.09
Jul-Sept	MRS1	5.34	133.78	1.70	7.69
	MRS2	4.31	115.69	1.52	6.97
	MRS3	4.55	111.35	1.50	6.68
Whole	MRS1	28.00	156.58	3.09	9.03
	MRS2	**25.47**	**140.77**	**2.83**	**8.32**
	MRS3	28.02	165.06	2.91	8.56

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.9: *Prediction error statistics for the Markov regime-switching regression models estimated on a daily expanding dataset, period 19 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$; F.P.=Forecasting Period).*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	MRS1	420.85	541.37	11.93	16.77
	MRS2	432.22	623.61	13.41	19.38
	MRS3	404.34	690.72	12.33	18.20
Apr-Jun	MRS1	68.10	216.98	5.32	11.41
	MRS2	47.63	157.23	4.30	9.42
	MRS3	67.43	232.03	5.50	12.05
Jul-Sept	MRS1	402.43	491.47	8.68	15.36
	MRS2	364.86	528.30	8.71	15.60
	MRS3	513.84	1239.14	10.15	19.32
Whole	MRS1	300.76	**419.78**	**8.70**	**14.56**
	MRS2	**285.28**	440.81	8.88	14.88
	MRS3	332.68	728.39	9.39	16.59

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.10: *Prediction error statistics for the Markov regime-switching regression models estimated on a daily expanding dataset, period 28 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$; F.P.=Forecasting Period).*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	MRS1	814.78	616.46	15.08	19.78
	MRS2	819.26	759.89	15.99	21.88
	MRS3	811.24	864.71	15.86	22.19
Apr-Jun	MRS1	99.29	394.82	7.41	16.59
	MRS2	80.62	350.51	6.51	14.92
	MRS3	111.02	388.52	7.24	15.60
Jul-Sept	MRS1	998.07	521.68	13.82	18.44
	MRS2	938.12	496.45	13.41	17.42
	MRS3	1438.69	770.06	16.93	22.66
Whole	MRS1	645.92	**512.83**	12.18	18.30
	MRS2	**621.11**	538.56	**12.06**	**18.12**
	MRS3	797.71	678.97	13.44	20.22

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.11: *Prediction error statistics for the Markov regime-switching regression models estimated on a daily expanding dataset, period 38 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$; F.P.=Forecasting Period).*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	MRS1	2034.27	807.60	23.45	21.86
	MRS2	2035.11	970.12	24.32	23.44
	MRS3	1906.58	956.03	23.92	24.04
Apr-Jun	MRS1	41.91	203.91	4.24	10.70
	MRS2	47.27	241.04	4.64	11.61
	MRS3	59.20	297.84	5.33	13.38
Jul-Sept	MRS1	268.39	616.21	8.50	17.24
	MRS2	302.99	723.69	8.76	17.54
	MRS3	300.37	516.39	9.06	17.46
Whole	MRS1	793.26	**547.95**	**12.18**	**16.69**
	MRS2	807.00	651.36	12.70	17.63
	MRS3	**766.44**	594.73	12.89	18.37

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.12: *Prediction error statistics for the Markov regime-switching regression models estimated on fixed rolling samples, period 6 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$, RSj= model estimated on fixed rolling samples of $j = 3, 6, 9$ months).*

		MRS1			**MRS2**			**MRS3**		
Forecasting Period	**Model**	RS3	RS6	RS9	RS3	RS6	RS9	RS3	RS6	RS9
January-March	MSE	55.41	59.50	65.37	58.31	49.91	57.09	52.53	55.57	67.11
	MSPE	235.44	237.10	254.97	208.93	164.29	203.47	255.51	292.16	322.20
	MAE	5.25	5.03	5.15	4.56	4.34	4.55	5.45	5.38	5.56
	MAPE	12.29	11.42	11.60	10.17	9.50	10.02	12.91	12.91	13.11
April-June	MSE	7.02	8.03	6.91	8.17	6.45	7.07	6.95	5.95	5.09
	MSPE	102.93	111.11	75.23	114.22	85.17	90.89	105.72	88.40	78.90
	MAE	2.12	2.29	2.10	2.12	2.05	2.08	1.95	1.94	1.83
	MAPE	7.97	8.51	7.33	8.00	7.52	7.54	7.40	7.38	7.04
July-September	MSE	6.30	8.12	8.92	4.99	7.40	5.89	4.81	7.57	6.35
	MSPE	149.41	165.84	196.92	127.87	160.59	138.98	119.99	170.67	152.19
	MAE	1.84	1.98	2.14	1.66	2.08	1.83	1.62	1.95	1.85
	MAPE	8.18	8.71	9.51	7.55	9.14	8.18	7.31	8.73	8.30
Whole Period	MSE	23.16	25.49	27.39	24.07	**21.49**	23.61	21.66	23.30	26.52
	MSPE	163.54	172.31	177.30	150.92	**137.50**	145.30	161.28	185.26	186.11
	MAE	3.09	3.11	3.15	**2.79**	2.84	2.83	3.02	3.11	3.10
	MAPE	9.50	9.56	9.52	**8.58**	8.74	8.60	9.23	9.71	9.52

Note: in bold the better statistics for the whole forecasting period (from January to September), while the underlined numbers are the better statistics for each model among different sample lengths.

Table 5.13: *Prediction error statistics for the Markov regime-switching regression models estimated on fixed rolling samples, period 19 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$, RSj= model estimated on fixed rolling samples of $j = 3, 6, 9$ months).*

		MRS1			**MRS2**			**MRS3**		
Forecasting Period	**Model**	RS3	RS6	RS9	RS3	RS6	RS9	RS3	RS6	RS9
January-March	MSE	268.88	427.14	403.45	320.43	522.50	450.00	252.73	392.03	394.46
	MSPE	546.51	500.75	499.90	650.70	706.06	646.97	599.54	694.56	736.00
	MAE	11.45	11.62	11.68	13.05	13.87	13.45	11.51	12.72	12.50
	MAPE	17.30	16.34	16.57	19.31	19.76	19.33	17.72	18.72	18.74
April-June	MSE	78.06	62.59	68.04	62.78	67.76	49.54	66.51	93.30	86.40
	MSPE	431.59	248.98	231.86	305.73	347.33	195.81	300.93	353.30	314.43
	MAE	6.63	5.53	5.48	5.77	5.95	4.72	5.39	5.71	5.72
	MAPE	16.18	12.75	12.09	13.48	14.48	10.89	12.60	12.85	12.70
July-September	MSE	436.86	360.65	442.30	513.84	400.15	377.50	517.08	526.64	627.13
	MSPE	689.78	750.81	720.97	1214.76	773.28	681.14	1346.80	1010.75	1800.28
	MAE	9.41	9.63	10.13	11.35	10.09	10.22	11.27	10.39	12.06
	MAPE	16.94	18.61	19.00	23.06	19.78	19.64	22.15	19.56	23.43
Whole Period	MSE	<u>**264.18**</u>	286.97	308.35	302.77	334.30	<u>296.20</u>	<u>282.14</u>	341.20	373.82
	MSPE	557.93	504.17	<u>**488.25**</u>	730.36	613.04	<u>512.93</u>	756.20	<u>691.49</u>	960.33
	MAE	9.20	<u>**8.98**</u>	9.15	10.12	10.03	<u>9.54</u>	<u>9.45</u>	9.67	10.16
	MAPE	16.81	<u>**15.95**</u>	15.95	18.70	18.07	<u>16.71</u>	17.57	<u>17.11</u>	18.38

Note: in bold the better statistics for the whole forecasting period (from January to September), while the underlined numbers are the better statistics for each model among different sample lengths.

Table 5.14: *Prediction error statistics for the Markov regime-switching regression models estimated on fixed rolling samples, period 28 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$, RSj= model estimated on fixed rolling samples of $j = 3, 6, 9$ months).*

		MRS1			**MRS2**			**MRS3**		
Forecasting Period	**Model**	RS3	RS6	RS9	RS3	RS6	RS9	RS3	RS6	RS9
January-March	MSE	869.79	870.54	811.71	875.41	831.80	841.50	842.99	829.70	810.58
	MSPE	692.19	720.37	636.78	884.50	804.60	808.81	1000.62	1051.77	892.35
	MAE	15.74	15.87	15.17	17.07	16.24	16.67	17.26	16.83	16.18
	MAPE	20.56	21.07	20.03	23.75	22.39	22.81	24.19	24.26	22.77
April-June	MSE	104.59	81.72	87.42	291.45	89.64	80.46	82.70	89.80	93.07
	MSPE	492.3	506.56	399.85	1859.11	565.23	405.96	414.08	427.01	401.19
	MAE	7.63	7.18	7.29	9.16	7.24	6.48	6.66	6.93	7.18
	MAPE	18.05	18.06	16.85	23.19	18.25	15.19	16.13	16.40	16.26
July-September	MSE	19337.17	388.75	460.66	386.87	424.81	332.73	590.43	580.21	513.27
	MSPE	2267.31	550.96	602.51	734.48	1140.84	598.66	687.72	796.66	816.34
	MAE	30.15	11.26	11.84	11.37	12.37	11.01	12.83	13.19	13.45
	MAPE	25.08	18.92	19.93	20.50	22.47	19.29	20.63	21.44	22.80
Whole Period	MSE	6876.33	452.80	459.07	521.50	454.45	**423.59**	512.08	506.41	478.33
	MSPE	1161.05	594.00	**548.71**	1148.26	841.20	607.63	705.36	763.74	708.09
	MAE	18.00	11.50	11.50	12.59	12.02	**11.47**	12.34	12.40	12.35
	MAPE	21.28	19.37	18.97	22.47	21.08	**19.16**	20.39	20.77	20.68

Note: in bold the better statistics for the whole forecasting period (from January to September), while the underlined numbers are the better statistics for each model among different sample lengths.

Table 5.15: *Prediction error statistics for the Markov regime-switching regression models estimated on fixed rolling samples, period 38 (MRSi=Markov regime-switching model i, $i = 1, 2, 3$, RSj= model estimated on fixed rolling samples of $j = 3, 6, 9$ months).*

		MRS1			**MRS2**			**MRS3**		
Forecasting Period	**Model**	RS3	RS6	RS9	RS3	RS6	RS9	RS3	RS6	RS9
January-March	MSE	2491.72	1646.25	2020.82	2005.19	2015.92	2019.74	1728.24	2000.17	1289.41
	MSPE	932.10	715.27	831.94	801.92	845.43	985.49	972.05	1316.52	904.82
	MAE	25.75	21.73	23.82	23.06	22.50	24.32	22.38	26.16	21.73
	MAPE	24.35	21.00	22.46	22.72	22.04	23.57	23.27	28.34	23.09
April-June	MSE	50.16	40.38	43.56	49.52	51.87	38.84	43.17	47.87	54.38
	MSPE	307.08	259.24	253.22	333.66	361.39	214.43	264.74	259.47	297.94
	MAE	5.26	4.76	4.43	4.80	5.48	4.20	4.93	5.17	5.21
	MAPE	13.90	12.88	11.51	13.13	14.99	10.68	13.09	13.38	13.29
July-September	MSE	449.28	326.28	288.98	455.27	335.72	254.98	392.10	339.68	282.77
	MSPE	632.43	653.18	728.67	568.35	816.84	572.17	799.99	573.67	497.95
	MAE	10.09	8.80	9.57	9.67	8.90	8.85	9.82	9.80	9.14
	MAPE	19.35	17.19	20.21	17.53	17.80	18.39	18.93	19.30	18.28
Whole Period	MSE	1012.08	680.98	796.21	849.15	813.06	782.81	731.93	807.78	**549.93**
	MSPE	628.90	**547.06**	610.19	571.69	679.52	596.67	685.50	723.81	571.17
	MAE	13.83	**11.88**	12.74	12.63	12.40	12.59	12.49	13.85	12.14
	MAPE	19.28	**17.09**	18.17	17.87	18.33	17.66	18.52	20.45	18.29

Note: in bold the better statistics for the whole forecasting period (from January to September), while the underlined numbers are the better statistics for each model among different sample lengths.

5.3 Time-varying multiple regression model

In section 5.1.1 we underlined the presence of parameter instability. To model parameters' dynamics is important in this case because the response of price to the various market fundamentals may change continuously. So, in our study we considered also a time-varying parameter regression model (TVR).
It is specified as:

$$p_{jt} = \mathbf{X}'_{jt}\boldsymbol{\beta}_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim GWN(0, \sigma^2_{\varepsilon_j}), \tag{5.4}$$

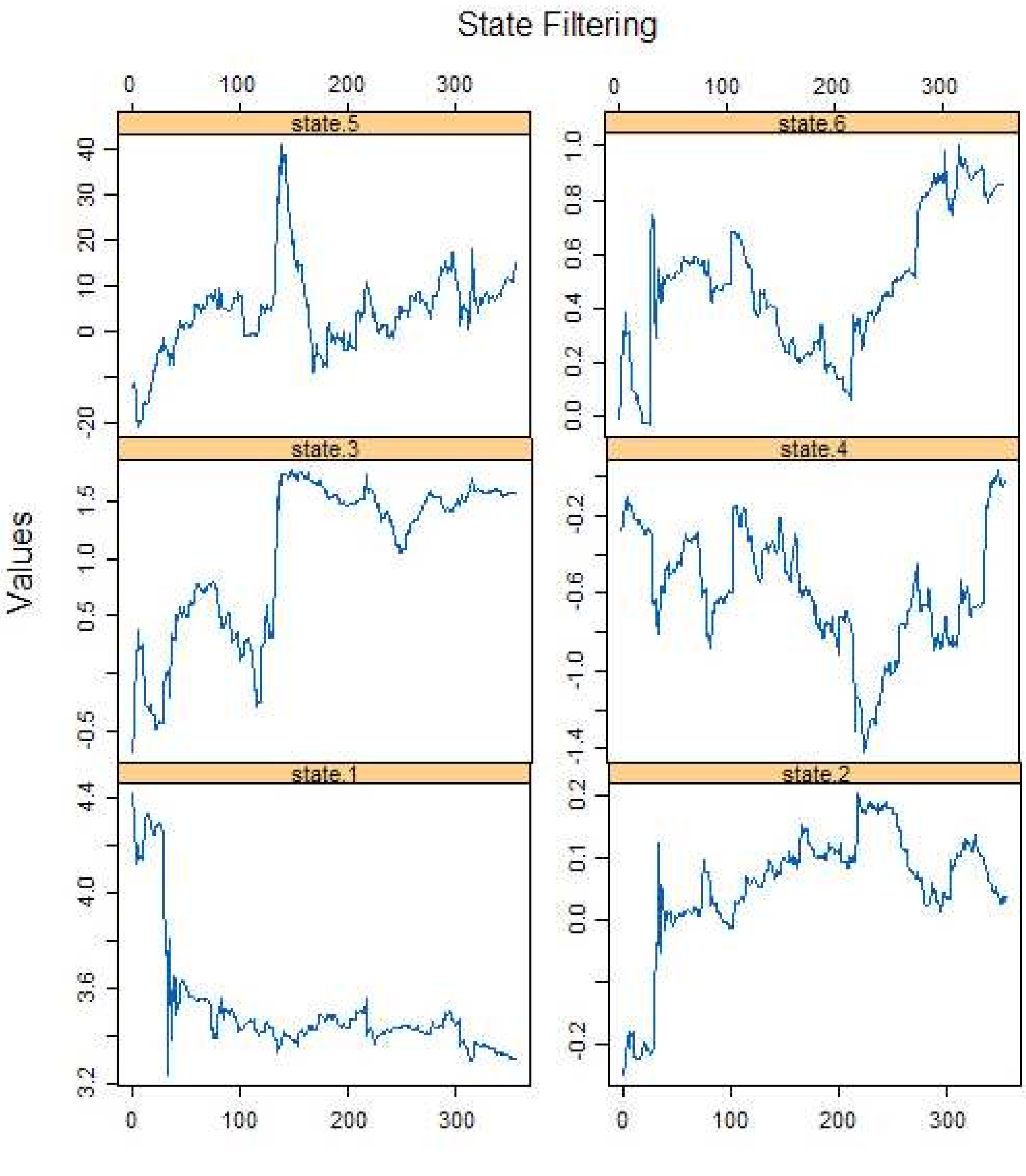
$$\boldsymbol{\beta}_{j(t+1)} = \boldsymbol{\beta}_{jt} + \boldsymbol{\nu}_{jt}, \quad \boldsymbol{\nu}_{jt} \sim GWN_k(0, \mathbf{H}_j), \tag{5.5}$$

where p_{jt} denotes the price logarithm on day t and load period j ($t = 6, 7, ..., T$, $j = 6, 19, 28, 38$), $\boldsymbol{\beta}_{jt}$ a $k \times 1$ vector of coefficients on day t and $\mathbf{X}_{jt}$ the $k \times 1$ vector of regressors. ε_{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 $\mathbf{H}_j = \mathrm{diag}\{\sigma^2_{\nu_{jk}}\}$.
Differently from linear regression model (5.1), here the regression coefficients are not unknown constants but latent, stochastic variables that a follow random walk. The estimation of this model was performed using state space methods and the Kalman Filter (Hamilton, 1994 and Durbin & Koopman, 2001).
As state space model, the above formulation can be written as:

$$\begin{pmatrix} \boldsymbol{\beta}_{j(t+1)} \\ p_{jt} \end{pmatrix} = \boldsymbol{\Phi}_{jt}\boldsymbol{\beta}_{jt} + \boldsymbol{\mu}_{jt}, \tag{5.6}$$

where $\boldsymbol{\Phi}_{jt} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{X}'_{jt} \end{pmatrix}$, $\boldsymbol{\mu}_{jt} = \begin{pmatrix} \boldsymbol{\nu}_{jt} \\ \varepsilon_{jt} \end{pmatrix} \sim N_{k+1}(0, \boldsymbol{\Omega}_j)$ and $\boldsymbol{\Omega}_j = \begin{pmatrix} \mathbf{H}_j & 0 \\ 0 & \sigma^2_{\varepsilon_j} \end{pmatrix}$.
As initial values $\boldsymbol{\beta}_{j1} \sim N_k(\mathbf{a}, \mathbf{P})$. Since $\boldsymbol{\beta}_{jt}$ is $I(1)$ the initial state vector does not have finite variance so the Kalman filter has to be initialized using a method called 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 κ is large ($\kappa = 10^6$).
Tables 5.16 and 5.16 show in-sample parameter estimates for the time-varying parameter regression models for each considered load period. Adjusted R squared values range from 84% to 99% confirming the high adaptability to data that is characteristic of time-varying models. It is interesting to note that the variables *margin* and *demF* (or their corresponding squared versions) are important in the adaptation of the models for all the load periods, showing a high variability in the coefficient estimates on the fitting sample. Figure 5.3 is an example of time-varying coefficients' evolution, for load period 28 with the set of regressors $\mathbf{X}2$. While the intercept coefficient reaches a constant level, the other coefficients show a dynamical evolution. In particular, the effects on price formation of the variables *demF* and *margin* increase strongly, respectively with a positive and a negative impact, during the winter season when prices are higher.

Figure 5.3: *Example of time-varying coefficients' evolution. Estimations for load period 28 model 2.*



Note: from state 1 to 6 estimates of coefficients corresponding to intercept, p_{t-1}, $demF_t$, $margin_t$, $demF_t^2$, $margin_t^2$.

Table 5.16: *In-sample parameter estimates (April-December 2005) for the time-varying regression models (TVRi=Time-Varying Regression model i, $i = 1, 2, 3$). Load periods 6 and 19.*

	Load period 6		
	TVR1	TVR2	TVR3
$\sigma_{intercept}$	6.25×10^{-08}	1.95×10^{-07}	1.29×10^{-06}
$\sigma_{p_{-1}}$	7.09×10^{-07}	2.18×10^{-09}	1.09×10^{-06}
$\sigma_{p_{-5}}$	—	—	—
σ_{demF}	—	9.76×10^{-02}	—
σ_{demF^2}	—	7.42×10^{-01}	—
σ_{margin}	1.39×10^{-01}	1.52×10^{-01}	1.03×10^{-01}
σ_{margin^2}	1.39×10^{-07}	1.66×10^{-06}	1.03×10^{-07}
$\sigma_{gasF.res}$	4.76×10^{-07}	—	3.13×10^{-16}
σ_{demVol}	1.76×10^{-03}	2.13×10^{-04}	1.07×10^{-04}
σ_{super}	9.95×10^{-02}	—	—
$\sigma_{superP_{-1}}$	—	—	2.40×10^{-01}
σ_{ε}	0.067	0.077	0.068
AdjR^2	0.938	0.901	0.933
Ljung-Box	21.77	15.67	22.72

	Load period 19		
	TVR1	TVR2	TVR3
$\sigma_{intercept}$	2.71×10^{-07}	3.21×10^{-07}	9.78×10^{-03}
$\sigma_{p_{-1}}$	4.41×10^{-03}	7.40×10^{-03}	—
$\sigma_{p_{-5}}$	—	3.34×10^{-14}	—
σ_{demF}	—	3.78×10^{-01}	1.89×10^{-01}
σ_{demF^2}	—	7.27×10^{-13}	—
σ_{margin}	1.18×10^{-10}	1.61×10^{-09}	1.79×10^{-06}
σ_{margin^2}	4.15×10^{-02}	3.59×10^{-02}	3.67×10^{-02}
$\sigma_{gasF.res}$	3.12×10^{-06}	—	1.40×10^{-06}
σ_{demVol}	—	1.48×10^{-03}	—
σ_{super}	6.35×10^{-02}	—	—
$\sigma_{superP_{-1}}$	—	—	5.00×10^{-02}
σ_{ε}	0.115	0.118	0.114
AdjR^2	0.928	0.931	0.922
Ljung-Box	9.646	12.40	10.66

*Note: *, **, *** means significance at 1%, 5% and 10% level respectively. Lag length for the Ljung-Box statistics depends on the sample length: it is $10\log_{10}(n) \approx 22$.*

Table 5.17: *In-sample parameter estimates (April-December 2005) for the time-varying regression models (TVRi=Time-Varying Regression model i, $i = 1, 2, 3$). Load periods 28 and 38.*

	Load period 28		
	TVR1	TVR2	TVR3
$\sigma_{intercept}$	2.54×10^{-06}	3.00×10^{-02}	7.93×10^{-03}
$\sigma_{p_{-1}}$	1.15×10^{-03}	3.14×10^{-07}	—
$\sigma_{p_{-5}}$	—	—	—
σ_{demF}	—	5.63×10^{-12}	3.18×10^{-04}
σ_{demF^2}	2.01×10^{-00}	2.98×10^{-00}	1.74×10^{-00}
σ_{margin}	1.05×10^{-01}	7.52×10^{-02}	4.09×10^{-03}
σ_{margin^2}	1.00×10^{-30}	7.73×10^{-06}	8.36×10^{-02}
$\sigma_{gasF.res}$	—	—	—
σ_{demVol}	—	—	—
σ_{super}	7.29×10^{-20}	—	—
$\sigma_{superP_{-1}}$	—	—	4.14×10^{-02}
σ_{ε}	0.160	0.159	0.163
AdjR^2	0.856	0.869	0.836
Ljung-Box	34.92**	28.04	32.20***

	Load period 38		
	TVR1	TVR2	TVR3
$\sigma_{intercept}$	1.98×10^{-06}	1.91×10^{-06}	1.83×10^{-06}
$\sigma_{p_{-1}}$	1.87×10^{-02}	2.87×10^{-02}	3.24×10^{-03}
$\sigma_{p_{-5}}$	3.14×10^{-08}	—	—
σ_{demF}	—	2.72×10^{-01}	—
σ_{demF^2}	—	4.00×10^{-00}	—
σ_{margin}	7.99×10^{-02}	7.91×10^{-02}	5.75×10^{-02}
σ_{margin^2}	—	—	8.49×10^{-02}
$\sigma_{gasF.res}$	—	—	—
σ_{demVol}	—	—	—
σ_{super}	2.94×10^{-01}	—	—
$\sigma_{superP_{-1}}$	—	—	4.15×10^{-02}
σ_{ε}	0.134	0.104	0.193
AdjR^2	0.978	0.992	0.903
Ljung-Box	28.46	31.73***	49.19*

Note: *, **, *** *means significance at* 1%, 5% *and* 10% *level respectively. Lag length for the Ljung-Box statistics depends on the sample length: it is* $10 \log_{10}(n) \approx 22$.

For each load period j and model specification, the out-of-sample one-day ahead spot price forecasts are obtained as:

$$\begin{aligned} F_{j(t+1)} &= \exp(f_{j(t+1)}) \\ &= \exp(\mathbf{X}'_{j(t+1)}\hat{\boldsymbol{\beta}}_{j(t+1|t)}). \end{aligned}$$

Tables 5.18, 5.19, 5.20 and 5.21 summarize the performance of the three models for period 6, 19, 28 and 38 respectively under the three forecasting periods (January-March, April-June, July-September) and the whole period from January to September 2006.

In general, the time-varying parameter models appear to outperform Markov regime-switching models for all the load periods if we consider statistics based on absolute errors. Even so, MRS models estimated on rolling windows predict better the jumps that occur during winter and summer in load periods 19 and 28. Best results are collected from TVR models also respect to the linear models, but it is interesting to underline that during the summer season linear models 1 and 2 outperform the time-varying models for load periods 19 and 28.

Table 5.18: *Prediction error statistics for the time-varying parameter regression models (TVRi=Time-Varying Regression model* i, $i = 1, 2, 3$*; F.P.=Forecasting Period), period 6.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	TVR1	45.78	230.52	4.69	11.25
	TVR2	40.99	204.74	4.65	11.18
	TVR3	46.84	245.78	4.88	11.83
Apr-Jun	TVR1	5.54	79.53	1.90	7.08
	TVR2	5.73	82.72	1.99	7.47
	TVR3	6.04	77.47	1.93	6.97
Jul-Sept	TVR1	5.44	127.79	1.65	7.40
	TVR2	4.23	106.98	1.51	6.77
	TVR3	5.57	133.21	1.68	7.58
Whole	TVR1	19.14	147.00	2.76	8.60
	TVR2	**17.16**	**132.26**	**2.73**	**8.49**
	TVR3	19.7	153.34	2.84	8.82

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.19: *Prediction error statistics for the time-varying parameter regression models (TVRi=Time-Varying Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 19.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	TVR1	460.44	492.14	11.92	15.78
	TVR2	380.97	392.94	11.00	14.83
	TVR3	388.15	569.51	12.19	17.19
Apr-Jun	TVR1	46.18	176.58	4.49	10.21
	TVR2	40.62	184.29	4.50	10.49
	TVR3	47.83	170.12	4.23	9.49
Jul-Sept	TVR1	441.23	624.15	9.18	15.31
	TVR2	431.37	588.34	9.14	16.01
	TVR3	450.81	909.46	10.23	18.75
Whole	TVR1	320.23	435.00	8.59	**13.82**
	TVR2	**288.19**	**391.76**	**8.27**	13.83
	TVR3	299.53	555.72	8.96	15.23

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.20: *Prediction error statistics for the time-varying parameter regression models (TVRi=Time-Varying Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 28.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	TVR1	857.51	661.06	15.01	18.69
	TVR2	762.68	573.55	14.21	18.14
	TVR3	824.09	824.50	16.31	21.80
Apr-Jun	TVR1	70.76	370.74	6.17	14.78
	TVR2	66.11	355.08	6.06	14.61
	TVR3	62.58	292.14	5.60	13.15
Jul-Sept	TVR1	1154.99	531.59	13.88	18.05
	TVR2	872.62	528.78	13.01	18.29
	TVR3	1024.49	603.13	14.54	19.87
Whole	TVR1	704.32	523.51	11.77	17.21
	TVR2	**575.09**	**487.88**	**11.17**	**17.05**
	TVR3	646.17	577.72	12.25	18.35

Note: in bold the better statistics for the whole forecasting period (from January to September).

Table 5.21: *Prediction error statistics for the time-varying parameter regression models (TVRi=Time-Varying Regression model i, $i = 1, 2, 3$; F.P.=Forecasting Period), period 38.*

F.P.	**Model**	MSE	MSPE	MAE	MAPE
Jan-Mar	TVR1	1700.77	679.66	20.82	20.15
	TVR2	1888.24	823.23	21.91	20.94
	TVR3	1710.29	786.49	21.68	20.90
Apr-Jun	TVR1	38.78	245.41	4.77	12.79
	TVR2	35.70	223.91	4.33	11.65
	TVR3	40.52	237.57	4.60	12.06
Jul-Sept	TVR1	254.76	415.04	8.14	15.46
	TVR2	227.48	427.67	8.18	16.38
	TVR3	149.25	492.79	7.34	15.66
Whole	TVR1	674.71	**449.90**	11.34	**16.19**
	TVR2	727.95	495.85	11.59	16.40
	TVR3	**642.76**	509.87	**11.31**	16.27

Note: in bold the better statistics for the whole forecasting period (from January to September).

5.4 Short ARMAX models

We decided to consider also the time series modeling approach. Following Fezzi (2007b), we conducted a study comparing AR, ARX, ARMA and ARMAX models in day-ahead forecasting, estimated using different sample lengths (we called these models "short" because they are not estimated on expanding dataset, but on short, rolling windows). Our scope was to find the best model and the best sample length to forecast our data.

In the AutoRegressive Moving Average ARMA(p, q) models, made popular by Box & Jenkins (1976), the current value y_t of the time series under study is expressed linearly in terms of its p past values (autoregressive part) and in terms of q previous values of the noise (moving average part):

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t, \quad \varepsilon_t \sim GWN(0, \sigma^2), \tag{5.7}$$

where ε_t is the error term. In the formula 5.7, y_t is a stationary time series, B is the backward shift operator, i.e. $B^h y_t = y_{t-h}$, $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$. For $q = 0$ the model reduces to an autoregressive AR(p) model.
In the ARMA models the only information used to forecast is the one embedded in the past values of the series itself. Nevertheless, the information contained in other pertinent time series might be important in shape characteristic dynamics. Exogenous variables can be included in the models, giving ARMAX models. A general formula for the ARMAX$(p, q, m_1, ..., m_k)$ model can be compactly written

as:

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t + \sum_{i=1}^{k} \beta^i(B)v_t^i, \quad \varepsilon_t \sim WN(0, \sigma^2), \tag{5.8}$$

where $v^1, ..., v^k$ are the exogenous variables and $\beta^i(B) = \beta_0^i + \beta_1^i B + \cdots + \beta_{m_i}^i B^{m_i}$, with the β_j^i's and m_i's being the corresponding coefficients and orders, respectively. In our forecasting study, y_t denotes the first difference of the price logarithms on day t and load period j ($t = T - l + 1, ..., T, j = 6, 19, 28, 38$). The forecasting performances of each model are evaluated producing 90 one step ahead forecasts over the period 24th August 2005 - 30th December 2005 (in-sample). At every step, the models are estimated on rolling samples of fixed lengths ($l = 10, 11, ..., 90$), then the forecasting performances are evaluated using the MSE and MAE prediction error statistics. The forecasting models considered are AR, ARMA, ARX1, ARX2, ARMAX1 and ARMAX2 where the exogenous variable is the forecasted demand change (X1) or the margin change (X2).

For comparison, figure 5.4 show forecasting performances for all the models with orders equal to 1 and with different sample lengths for load periods 6, 19, 28 and 38. All the models reach an equilibrium level when the sample length increases. For periods 6, the stabilization of the models occurs earlier: this is because period 6 is less volatile than the other models, so it needs a smaller sample to produce good forecasts. Among all the models, the ARMAX2 shows the best forecasting performances, followed by the ARX2 model. It is clear that the variable *margin* is determinant for shaping price dynamics. In period 28, also the forecasted demand has an important role: the ARMAX1 model comes off from the group with the worst performances clearly respect to, for example, period 6. In general, the introduction of an exogenous factor (see the difference among AR, ARX1 and ARX2 and among ARMA, ARMAX1 and ARMAX2) and of the moving average part (see the difference between ARX1 and ARMAX1 and between ARX2 and ARMAX2) improves the forecasts.

For our dataset the selected model was the ARMAX2(1,1) (we will call this model ARMAX for simpliciy): this model is then used to forecast one day ahead the out-of-sample period from January 2006 to the end of September 2006. The model is estimated at every step on a rolling window of length 37 for period 6, 44 for period 19, 58 for period 28 and 45 for period 38: these are the points where the model reaches the lowest values (we decided these points after an evaluation between RMSE values and MAE values). ARMAX model can be written as:

$$\triangle p_{jt} = \phi_j \triangle p_{j(t-1)} + \varepsilon_{jt} + \theta_j \varepsilon_{j(t-1)} + \beta_j \triangle v_{jt}, \quad \varepsilon_{jt} \sim WN(0, \sigma_j^2), \tag{5.9}$$

where v_{jt} is the exogenous factor (margin change). For each load period j, the out-of-sample one-day ahead spot price forecasts are obtained as:

$$\begin{aligned} F_{j(t+1)} &= \exp(f_{j(t+1)}) \\ &= \exp(\hat{\triangle} p_{j(t+1)} + p_{jt}). \end{aligned}$$

Forecasting performances for spot prices are shown in table 5.22.

Figure 5.4: *MSE and MAE for different models and for different values of sample length. Load Period 6, 19, 28 and 38.*

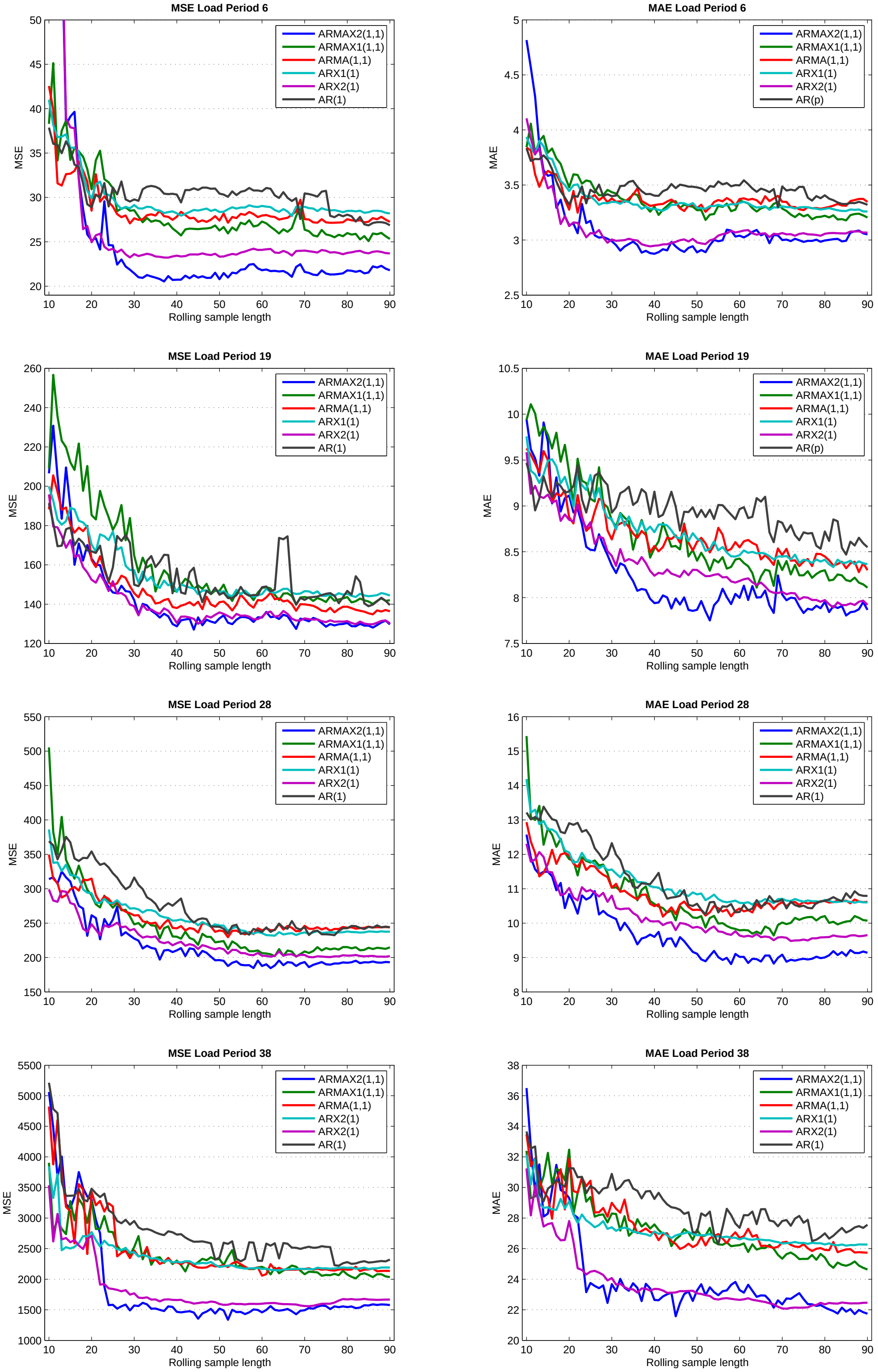


Table 5.22: *Prediction error statistics for the ARMAX2 model, period 6, 19, 28 and 38.*

	Load period 6			
Forecasting Period	MSE	MSPE	MAE	MAPE
January-March	37.66	211.12	4.70	11.53
April-June	7.25	82.94	2.09	7.35
July-September	5.29	126.71	1.69	7.51
Whole Period	16.88	141.17	2.84	8.82

	Load period 19			
Forecasting Period	MSE	MSPE	MAE	MAPE
January-March	372.01	462.10	11.43	15.62
April-June	34.84	157.89	4.20	9.75
July-September	488.06	758.41	9.71	16.48
Whole Period	302.48	464.25	8.51	14.02

	Load period 28			
Forecasting Period	MSE	MSPE	MAE	MAPE
January-March	719.39	595.04	14.04	17.96
April-June	81.90	432.40	7.10	17.09
July-September	1264.31	1225.73	16.96	25.85
Whole Period	698.16	756.11	12.79	20.35

	Load period 38			
Forecasting Period	MSE	MSPE	MAE	MAPE
January-March	1846.12	736.35	20.09	19.07
April-June	41.07	253.69	4.66	12.43
July-September	320.56	650.23	9.31	19.10
Whole Period	746.95	551.41	11.46	16.93

From the results, it appears that the ARMAX model has similar predictive performance as the time-varying regression model on the whole forecasting period. A thorough analysis highlights a particular behaviour: compared to the TVR models, the ARMAX model gives better results in terms of MSE during the winter season (periods 6, 19 and 28), while it gives worst forecasts during spring for all the load periods. A possible explanation could be the impact of variable *margin* on price formation. Prices are linked to this market fundamental particularly in high demand periods during the year if we consider the singular hours. Winter is a high demand season, spring not.

Comparing the ARMAX model with the Markov regime-switching models, in general the former gives better results if we consider absolute error statistics, but it cannot forecast the high peaks as MRS models estimated on rolling samples do (especially during summer for load period 28). In period 6, ARMAX outperforms MRS models in MSE because of the very good performance in winter.

5.5 The Diebold-Mariano test for equal predictive accuracy

Suppose that a pair of h-steps ahead forecasts have produce errors $(e_t^{(1)}, e_t^{(2)})$ with $t = 1, ..., n$. The accuracy of each forecast is measured by a particular loss function $L(e_t^{(i)})$ for $i = 1, 2$. Some popular loss functions are

- squared error loss: $L(e_t^{(i)}) = (e_t^{(i)})^2$,
- absolute error loss: $L(e_t^{(i)}) = |e_t^{(i)}|$.

Then, the null hypothesis of the equality of expected forecast performance is

$$\mathrm{E}[L(e_t^{(1)}) - L(e_t^{(2)})] = 0.$$

Defining

$$d_t = L(e_t^{(1)}) - L(e_t^{(2)}) \qquad t = 1, ..., n, \tag{5.10}$$

it is natural to base the Diebold and Mariano test (see Diebold & Mariano, 1995) on the observed sample mean:

$$\bar{d} = n^{-1} \sum_{t=1}^{n} d_t$$

A difficulty is that the series d_t is likely to be autocorrelated. Indeed, for optimal h-steps ahead forecasts, the sequence of forecast errors follows a moving average process of order $(h - 1)$. This result can be expected to hold approximately for any reasonably well-conceived set of forecasts. Therefore, in what follows it will be assumed for h-steps ahead forecasts that all autocorrelations of order h or higher of the sequence d_t are zero. In that case, it can be shown that the variance of $\bar{d}$ is, asymptotically,

$$V(\bar{d}) \approx n^{-1} \left[\gamma_0 + 2 \sum_{k=1}^{h-1} \gamma_k \right], \tag{5.11}$$

where γ_k is the kth autocovariance of d_t. This autocovariance can be estimated by

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^{n} (d_t - \bar{d})(d_{t-k} - \bar{d}). \tag{5.12}$$

The Diebold-Mariano test statistic is then

$$\mathrm{DB} = \frac{\bar{d}}{\sqrt{(\hat{V}(\bar{d}))}}, \tag{5.13}$$

where $\hat{V}(\bar{d})$ is obtained by substituting the estimates (5.12) in (5.11). Under the null hypothesis, this statistic has an asymptotic standard normal distribution.
Harvey et al. (1997) proposed a modification of the Diebold and Mariano test, explaining that in general the original test was found to be quite seriously over-sized for moderate numbers of sample observations or in the case of two-steps ahead (or

more) prediction. In our case, the modified Diebold and Mariano test gave results that substantially are not different than those obtained with the original test. This is because we are considering one-step ahead forecasting in a sample of 189 observations that is a quite large sample. Diebold & Mariano (1995) reported that for moderately large samples, the test performance was satisfactory in a wide range of situations, including contemporaneously correlated and autocorrelated forecast errors, and heavy-tailed as well as normal error distributions. For these reasons in our forecasting accuracy study the original test is used.

Tables 5.24-5.31 show results of the Diebold and Mariano test for equal predictive accuracy applied to the forecasts obtained with the 19 models on the whole forecasting period (January-September 2006). The test was carried out for both squared error and absolute error loss functions following formula (5.13), where $e^{(1)}$ is the forecasting error series of models placed in the rows of the tables, while $e^{(2)}$ is the forecasting error series of models placed in the columns. Missing values in tables can be obtained considering that the tables are symmetric respect to the diagonal unless the sign that is opposite. The sign of the statistic values is important because it gives information about which model between the two is more accurate (even if the improvement could be non significative). A negative sign means that the first model gives better forecasts, while a positive sign means that the second model outperforms the first one.

5.6 Comments on the forecasting results

From the comparison study conducted in the previous sections, it seems that globally no model outperforms the others. This consideration is supported also by results obtained with the Diebold and Mariano test. In fact, test results in tables 5.24-5.31 highlight that not only globally but also in each load period there is not a model that produce significant better forecasts respect to all the other models. Examining results in load period 6, for TVR and ARMAX models compared with the others, there is an improvement in predictive accuracy: this can be deduced from the sign of the test statistic values. In particular, for TVR model 2 the improvement is significative at 10% significance level in most of the comparisons considering a squared error loss function. It is interesting also to note that the group of MRS models 2 (estimated on expanding and fixed rolling samples with $\mathbf{X}2$) performs better than linear models and than the group of MRS models 1 in terms of absolute errors.
In load period 19, TVR and ARMAX models presents significative better results compared to MRS models 2 and 3. Markov regime-switching models 2 estimated on rolling windows are not very effective: linear models 1 and 2 beat them.
Linear regression model 3 is the worst in load period 28, while in period 38 all the linear models do not produce good forecasts. On the contrary, time-varying parameter models and ARMAX model outperform significatively the others in most of the cases.
The foregoing comments regard model predictive accuracy on the whole forecasting period from January to September 2006, considering three seasons together. From the previous sections, considering each season separately gives more information

Table 5.23: *Models with the best prediction error statistics at each forecasting period (F.P.= forecasting period, J-M=January-March, A-J=April-June, J-S=July-September). Load periods 6, 19, 28 and 38.*

	Load period 6			
F.P.	MSE	MSPE	MAE	MAPE
Jan-Mar	ARMAX	MRS2RS6	MRS2RS6	MRS2RS6
Apr-Jun	MRS3	MRS3	MRS3	MRS3
Jul-Sept	TVR2	TVR2	MRS3	MRS3
Whole	ARMAX	TVR2	TVR2	MRS2

	Load period 19			
F.P.	MSE	MSPE	MAE	MAPE
Jan-Mar	MRS3RS3	TVR2	TVR2	TVR2
Apr-Jun	ARMAX	MRS2	ARMAX	MRS2
Jul-Sept	MRS1RS6	LR2	LR2	LR2
Whole	MRS1RS3	TVR2	TVR2	TVR2

	Load period 28			
F.P.	MSE	MSPE	MAE	MAPE
Jan-Mar	ARMAX	TVR2	ARMAX	ARMAX
Apr-Jun	TVR3	TVR3	TVR3	TVR3
Jul-Sept	MRS2RS9	MRS2	MRS2RS9	MRS2
Whole	MRS2RS9	TVR2	TVR2	TVR2

	Load period 38			
F.P.	MSE	MSPE	MAE	MAPE
Jan-Mar	MRS3RS9	TVR1	ARMAX	ARMAX
Apr-Jun	TVR2	MRS1	MRS2RS9	MRS2RS9
Jul-Sept	TVR3	TVR1	TVR3	TVR1
Whole	MRS3RS9	TVR1	TVR3	TVR1

about the forecasting performance of the models, but it does not resolve the main issue of the study: which forecasting model specification performs better?
Table 5.23 shows the models with the best prediction error statistics at each forecasting period. It is clear that the answer to the question is that no model outperforms all the others. Differences in forecasting accuracy depend on several factors, such as model specification, sample realization and forecasting period. Since different forecasting models seem to capture different features of spot price dynamics, we propose a forecasting approach based on the combination of forecasts. This approach has been useful to improve forecasting accuracy in several empirical situations (see, for example, Becker & Clements, 2008 and Sánchez, 2008), but it is novel in electricity price forecasting. In chapter 6 a first study in this direction is presented.

Table 5.24: *Diebold-Mariano statistic values with squared error loss function. Load period 6.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	*-3.05	—	—	—	—	—	—	—	—
LR3	-0.83	0.47	—	—	—	—	—	—	—
MRS1	0.92	**2.09	1.04	—	—	—	—	—	—
M1RS3	-1.75	1.39	0.31	-1.31	—	—	—	—	—
M1RS6	-0.31	**2.05	0.84	-1.07	1.02	—	—	—	—
M1RS9	0.64	**2.11	1.05	-0.49	1.32	0.98	—	—	—
MRS2	-0.13	1.35	0.51	-1.00	0.57	-0.01	-0.76	—	—
M2RS3	-0.63	1.33	0.29	-1.08	0.27	-0.39	-0.87	-0.49	—
M2RS6	*-6.25	0.74	-0.24	**-2.21	-0.76	**-2.11	**-2.24	-1.36	-1.27
M2RS9	-1.03	1.25	0.22	-1.31	0.13	-0.62	-1.16	-0.79	-0.26
MRS3	0.69	**2.33	1.25	0.02	**2.11	1.25	0.57	***1.68	1.01
M3RS3	-0.99	0.24	-0.49	-1.11	-0.49	-0.91	-1.10	-0.61	-0.44
M3RS6	-0.54	0.58	0.56	-0.79	0.04	-0.50	-0.75	-0.33	-0.12
M3RS9	0.16	1.42	1.33	-0.32	1.12	0.28	-0.20	0.21	0.44
TVR1	-1.22	-0.32	-1.36	-1.35	-1.34	-1.31	-1.37	-0.86	-0.69
TVR2	***-1.65	-0.80	***-1.79	***-1.65	***-1.89	***-1.71	***-1.71	-1.18	-1.07
TVR3	-1.10	-0.20	-1.01	-1.19	-1.04	-1.12	-1.19	-0.75	-0.61
ARMAX	-1.54	-0.74	***-1.78	-1.50	-1.49	-1.53	-1.54	-1.08	-1.01

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	1.22	—	—	—	—	—	—	—	—
MRS3	*3.01	1.63	—	—	—	—	—	—	—
M3RS3	0.04	-0.35	-1.32	—	—	—	—	—	—
M3RS6	0.38	-0.05	-0.98	0.61	—	—	—	—	—
M3RS9	1.30	0.61	-0.47	1.15	1.22	—	—	—	—
TVR1	-0.45	-0.67	-1.40	-1.37	-1.65	***-1.56	—	—	—
TVR2	-0.88	-1.03	***-1.72	**-2.27	***-1.73	***-1.71	-1.08	—	—
TVR3	-0.33	-0.58	-1.20	-1.46	-1.20	-1.27	0.34	1.36	—
ARMAX	-0.81	-0.95	-1.60	***-1.76	***-1.78	***-1.68	-1.27	-0.16	***-1.83

*Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** before a number represent significance at* 1%, 5% *and* 10% *level respectively.*

Table 5.25: *Diebold-Mariano statistic values with absolute error loss function. Load period 6.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	*-2.60	—	—	—	—	—	—	—	—
LR3	-0.40	1.14	—	—	—	—	—	—	—
MRS1	-0.41	1.49	0.16	—	—	—	—	—	—
M1RS3	-0.30	1.50	0.15	-0.02	—	—	—	—	—
M1RS6	-0.19	1.56	0.28	0.16	0.14	—	—	—	—
M1RS9	0.05	***1.71	0.42	0.66	0.30	0.27	—	—	—
MRS2	**-1.91	-0.09	-1.10	***-1.66	-1.48	-1.46	***-1.88	—	—
M2RS3	***-1.71	-0.34	-1.21	-1.32	***-1.85	-1.35	-1.46	-0.23	—
M2RS6	**-2.21	-0.04	-1.10	-1.42	-1.42	***-1.72	*** -1.81	0.06	0.30
M2RS9	**-2.20	-0.10	-1.17	-1.42	-1.54	-1.64	***-1.76	0.02	0.27
MRS3	-1.08	0.34	-1.21	-0.97	-0.94	-1.06	-1.31	0.50	0.56
M3RS3	-0.50	0.93	-0.22	-0.25	-0.30	-0.35	-0.47	0.93	1.26
M3RS6	-0.14	1.41	0.51	0.08	0.12	-0.01	-0.16	1.32	1.54
M3RS9	-0.18	1.19	0.36	0.03	0.06	-0.06	-0.21	1.29	1.38
TVR1	-1.57	-0.38	-1.49	-1.25	**-1.91	-1.47	-1.49	-0.28	-0.13
TVR2	-1.72	-0.57	-1.61	-1.36	**-2.17	-1.58	-1.56	-0.43	-0.31
TVR3	-1.20	0.02	-1.12	-0.87	-1.17	-0.99	-1.06	0.07	0.24
ARMAX	-1.20	-0.02	-0.96	-0.91	-1.16	-1.01	-1.11	0.04	0.22

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	-0.07	—	—	—	—	—	—	—	—
MRS3	0.37	0.42	—	—	—	—	—	—	—
M3RS3	0.85	0.92	0.68	—	—	—	—	—	—
M3RS6	1.40	1.38	1.52	0.58	—	—	—	—	—
M3RS9	1.25	1.30	***1.65	0.42	-0.14	—	—	—	—
TVR1	-0.34	-0.31	-0.65	-1.46	***-1.81	-1.57	—	—	—
TVR2	-0.52	-0.48	-0.77	***-1.70	***-1.89	-1.62	-0.33	—	—
TVR3	0.04	0.06	-0.25	-1.21	-1.39	-1.08	0.63	0.88	—
ARMAX	0.00	0.03	-0.28	-0.89	-1.23	-1.06	0.50	0.66	-0.05

*Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** before a number represent significance at* 1%, 5% *and* 10% *level respectively.*

Table 5.26: *Diebold-Mariano statistic values with squared error loss function. Load period 19.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	0.83	—	—	—	—	—	—	—	—
LR3	0.97	0.46	—	—	—	—	—	—	—
MRS1	1.16	-0.62	-0.86	—	—	—	—	—	—
M1RS3	-0.49	-0.74	-1.16	-0.55	—	—	—	—	—
M1RS6	-0.35	-0.79	-1.07	-0.48	0.30	—	—	—	—
M1RS9	1.60	-0.28	-0.66	1.04	0.68	0.78	—	—	—
MRS2	-0.46	-1.53	-1.22	-0.59	0.38	-0.05	-0.95	—	—
M2RS3	0.10	-0.18	-0.51	0.03	1.42	0.20	-0.08	0.31	—
M2RS6	1.16	0.66	-0.05	0.97	0.77	1.60	0.79	1.26	0.35
M2RS9	0.01	-0.97	-1.00	-0.18	0.53	0.34	-0.52	1.28	-0.11
MRS3	0.77	0.31	-0.33	0.67	1.01	0.84	0.48	0.92	0.42
M3RS3	-0.17	-0.37	-0.98	-0.23	0.39	-0.06	-0.32	-0.04	-0.41
M3RS6	1.01	0.49	0.08	0.92	1.28	0.89	0.73	1.10	0.65
M3RS9	1.01	0.69	0.76	0.95	1.29	1.05	0.84	1.12	0.82
TVR1	1.22	0.22	-0.42	1.01	0.83	0.93	0.63	1.30	0.25
TVR2	-0.43	-1.25	-1.15	-0.71	0.38	0.03	-1.31	0.14	-0.24
TVR3	0.10	-0.36	**-2.20	-0.04	0.66	0.32	-0.25	0.47	-0.06
ARMAX	0.22	-0.35	-0.93	0.06	0.72	0.33	-0.21	0.54	-0.01

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	-1.07	—	—	—	—	—	—	—	—
MRS3	-0.02	0.73	—	—	—	—	—	—	—
M3RS3	-0.49	-0.19	-0.95	—	—	—	—	—	—
M3RS6	0.10	0.89	0.21	0.97	—	—	—	—	—
M3RS9	0.41	0.99	1.27	***1.65	0.56	—	—	—	—
TVR1	-0.38	0.87	-0.27	0.47	-0.43	-0.73	—	—	—
TVR2	-1.25	-0.36	-0.90	0.08	-1.15	-1.12	-1.99	—	—
TVR3	-0.60	0.11	-1.36	0.32	-1.04	-1.36	-0.62	0.35	—
ARMAX	-0.60	0.19	-0.79	0.32	-1.02	-1.17	-0.70	0.69	0.11

Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** *before a number represent significance at* 1%, 5% *and* 10% *level respectively.*

Table 5.27: *Diebold-Mariano statistic values with absolute error loss function. Load period 19.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	0.32	—	—	—	—	—	—	—	—
LR3	**2.15	1.46	—	—	—	—	—	—	—
MRS1	-0.17	-0.37	**-2.30	—	—	—	—	—	—
M1RS3	0.75	0.39	-1.01	0.73	—	—	—	—	—
M1RS6	0.65	0.11	-1.55	0.64	-0.35	—	—	—	—
M1RS9	***1.69	0.42	-1.36	1.64	-0.08	0.58	—	—	—
MRS2	0.31	-0.07	-1.67	0.35	-0.47	-0.17	-0.51	—	—
M2RS3	**2.01	***1.89	0.04	***1.93	1.36	1.46	1.33	**2.31	—
M2RS6	**2.18	***1.99	-0.06	**2.18	0.92	**1.96	1.61	**1.97	-0.11
M2RS9	1.60	***1.90	-0.81	1.64	0.46	0.97	0.70	*2.72	-1.06
MRS3	1.16	0.59	*-3.75	1.24	0.22	0.61	0.36	0.69	-0.84
M3RS3	0.91	0.56	-0.99	0.90	0.36	0.59	0.35	0.68	-0.91
M3RS6	1.56	1.00	-0.95	1.62	0.56	1.08	0.78	1.25	-0.56
M3RS9	***1.85	1.25	0.16	***1.92	1.02	1.41	1.25	1.37	0.04
TVR1	-0.31	-0.47	**-2.42	-0.28	-0.89	-0.70	-1.36	-0.53	**-2.06
TVR2	-1.07	-1.15	*-2.93	-1.07	-1.41	-1.38	**-2.10	-1.47	*-2.92
TVR3	0.48	0.09	*-2.59	0.54	-0.36	-0.04	-0.37	0.15	-1.60
ARMAX	-0.42	-0.59	**-2.46	-0.39	-1.01	-0.75	-1.23	-0.65	**-2.32

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	-0.88	—	—	—	—	—	—	—	—
MRS3	-0.71	-0.21	—	—	—	—	—	—	—
M3RS3	-0.53	-0.11	0.11	—	—	—	—	—	—
M3RS6	-0.42	0.22	0.63	0.34	—	—	—	—	—
M3RS9	0.12	0.70	**2.13	1.22	0.85	—	—	—	—
TVR1	**-2.19	***-1.70	-1.37	-1.05	***-1.75	**-2.02	—	—	—
TVR2	*-2.85	*-3.05	***-1.85	-1.54	**-2.32	**-2.30	-0.94	—	—
TVR3	-1.37	-1.15	-1.06	-0.79	***-1.66	**-1.96	0.88	***1.77	—
ARMAX	**-2.05	***-1.83	-1.54	-1.32	***-1.75	**-2.37	-0.21	0.75	-0.99

*Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** before a number represent significance at 1%, 5% and 10% level respectively.*

Table 5.28: *Diebold-Mariano statistic values with squared error loss function. Load period 28.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	1.47	—	—	—	—	—	—	—	—
LR3	***1.65	1.58	—	—	—	—	—	—	—
MRS1	1.40	1.14	-0.90	—	—	—	—	—	—
M1RS3	1.03	1.03	1.01	1.01	—	—	—	—	—
M1RS6	-0.87	-1.11	-1.50	-1.23	-1.03	—	—	—	—
M1RS9	-0.76	-1.01	-1.41	-1.37	-1.03	0.09	—	—	—
MRS2	1.25	1.03	-1.95	-0.31	-1.03	1.21	1.06	—	—
M2RS3	0.16	-0.21	-1.01	-0.72	-1.02	0.99	0.81	-0.60	—
M2RS6	-0.72	-1.03	-1.44	-1.12	-1.03	0.04	-0.06	-1.13	-0.99
M2RS9	-1.12	-1.35	-1.54	-1.29	-1.04	-0.82	-0.50	-1.29	-1.53
MRS3	1.48	1.42	0.96	1.19	-1.01	1.40	1.35	1.59	1.03
M3RS3	0.07	-0.41	-1.29	-1.15	-1.02	0.60	1.01	-0.78	-0.09
M3RS6	-0.01	-0.48	-1.29	-1.04	-1.02	0.67	0.89	-0.81	-0.18
M3RS9	-0.52	-0.95	-1.46	-1.19	-1.03	0.44	0.37	-1.04	-0.59
TVR1	0.94	0.82	-0.23	0.30	-1.03	1.06	0.91	0.65	0.69
TVR2	0.52	0.27	-1.67	-0.48	-1.04	0.78	0.60	-0.66	0.29
TVR3	1.04	0.88	-1.37	0.00	-1.03	1.13	0.97	0.46	0.64
ARMAX	1.28	1.18	-0.72	0.45	-1.02	1.25	1.14	1.37	0.82

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	-1.31	—	—	—	—	—	—	—	—
MRS3	1.34	1.43	—	—	—	—	—	—	—
M3RS3	0.58	0.90	-1.27	—	—	—	—	—	—
M3RS6	0.65	1.04	-1.24	-0.15	—	—	—	—	—
M3RS9	0.43	0.96	-1.36	-0.65	-0.98	—	—	—	—
TVR1	1.02	1.11	-0.71	0.73	0.76	0.91	—	—	—
TVR2	0.75	0.90	-1.55	0.33	0.38	0.57	-1.48	—	—
TVR3	1.08	1.21	-1.36	0.74	0.79	0.97	-0.64	1.47	—
ARMAX	1.21	1.33	-1.24	0.96	0.98	1.15	-0.06	1.63	1.27

Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** *before a number represent significance at* 1%, 5% *and* 10% *level respectively.*

Table 5.29: *Diebold-Mariano statistic values with absolute error loss function. Load period 28.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	0.19	—	—	—	—	—	—	—	—
LR3	*2.55	**2.17	—	—	—	—	—	—	—
MRS1	0.54	0.33	**-2.41	—	—	—	—	—	—
M1RS3	1.15	1.14	0.80	1.16	—	—	—	—	—
M1RS6	-0.72	-0.57	**-2.33	-0.71	-1.19	—	—	—	—
M1RS9	-0.73	-0.64	**-2.11	-0.78	-1.12	0.00	—	—	—
MRS2	0.22	0.12	**-2.31	-0.23	-1.23	0.57	0.52	—	—
M2RS3	0.69	0.56	-0.99	0.32	-0.93	1.20	1.27	0.42	—
M2RS6	0.13	0.03	-1.53	-0.14	-1.07	0.72	0.59	-0.04	-0.70
M2RS9	-0.59	-0.71	**-1.96	-0.65	-1.17	-0.06	-0.05	-0.63	-1.42
MRS3	1.61	1.51	-1.46	***1.92	-0.97	1.49	1.45	***1.81	0.52
M3RS3	0.55	0.38	***-1.76	0.18	-1.00	0.88	1.22	0.26	-0.22
M3RS6	0.70	0.47	***-1.83	0.24	-1.00	1.23	1.38	0.34	-0.19
M3RS9	0.71	0.43	**-1.95	0.19	-1.02	1.40	1.38	0.30	-0.25
TVR1	-0.17	-0.24	*-2.71	-0.53	-1.36	0.28	0.21	-0.44	-0.56
TVR2	-1.00	-1.06	*-3.17	-1.28	-1.41	-0.41	-0.29	-1.57	-1.17
TVR3	0.42	0.30	**-2.48	0.11	-1.19	0.77	0.65	0.29	-0.25
ARMAX	1.25	1.22	***-1.64	1.06	-1.05	1.30	1.23	1.40	0.16

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	-1.27	—	—	—	—	—	—	—	—
MRS3	0.95	1.35	—	—	—	—	—	—	—
M3RS3	0.30	0.85	-0.96	—	—	—	—	—	—
M3RS6	0.42	1.08	-0.90	0.11	—	—	—	—	—
M3RS9	0.40	1.17	-0.97	0.01	-0.17	—	—	—	—
TVR1	-0.21	0.26	***-1.94	-0.45	-0.54	-0.54	—	—	—
TVR2	-0.95	-0.34	**-2.37	-1.02	-1.22	-1.31	-1.41	—	—
TVR3	0.20	0.69	-1.52	-0.09	-0.16	-0.10	1.05	**2.11	—
ARMAX	0.69	1.27	-0.86	0.43	0.38	0.46	1.40	**2.73	0.9

*Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** before a number represent significance at* 1%, 5% *and* 10% *level respectively.*

Table 5.30: *Diebold-Mariano statistic values with squared error loss function. Load period 38.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	-0.17	—	—	—	—	—	—	—	—
LR3	-0.60	-0.73	—	—	—	—	—	—	—
MRS1	-1.26	***-1.76	-0.48	—	—	—	—	—	—
M1RS3	1.44	1.10	1.05	1.47	—	—	—	—	—
M1RS6	-1.41	***-1.79	**-2.38	-1.30	-1.57	—	—	—	—
M1RS9	-1.19	***-1.67	-0.46	0.27	-1.43	1.41	—	—	—
MRS2	-0.91	-1.26	-0.35	0.50	-1.22	***1.88	0.44	—	—
M2RS3	-0.63	-0.83	0.29	0.65	-1.21	***1.76	0.63	0.49	—
M2RS6	-1.35	**-2.09	-0.21	0.39	***-1.70	1.16	0.31	0.09	-0.44
M2RS9	-1.07	-1.44	-0.64	-0.33	-1.31	1.57	-0.44	-1.49	-0.72
MRS3	-1.15	-1.51	-1.07	-0.50	-1.45	1.38	-0.56	-0.79	-0.98
M3RS3	-1.03	-1.26	***-1.69	-0.56	-1.30	1.09	-0.62	-0.81	-1.17
M3RS6	-0.68	-0.83	-0.48	0.16	-1.08	2.11	0.13	0.01	-0.45
M3RS9	-1.17	-1.31	-1.48	-1.06	-1.30	-0.87	-1.10	-1.24	-1.29
TVR1	-1.32	-1.55	**-2.12	-1.04	-1.45	-0.12	-1.09	-1.31	-1.55
TVR2	-1.13	-1.33	-1.49	-0.63	-1.34	0.67	-0.67	-0.83	-1.18
TVR3	-1.47	***-1.71	**-1.96	-1.42	-1.55	-0.60	-1.48	***-1.73	-1.52
ARMAX	-1.27	-1.61	**-2.06	-0.58	-1.51	1.05	-0.62	-0.80	-1.45

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	-0.42	—	—	—	—	—	—	—	—
MRS3	-0.61	-0.34	—	—	—	—	—	—	—
M3RS3	-0.63	-0.57	-0.48	—	—	—	—	—	—
M3RS6	-0.05	0.33	0.85	1.43	—	—	—	—	—
M3RS9	-1.01	-1.15	-1.09	-1.22	-1.44	—	—	—	—
TVR1	-0.95	-1.14	-0.99	-0.82	-1.55	0.91	—	—	—
TVR2	-0.65	-0.62	-0.45	-0.05	-0.93	1.07	1.27	—	—
TVR3	-1.19	***-1.64	-1.43	-1.12	-1.89	0.64	-0.61	-1.37	—
ARMAX	-0.70	-0.51	-0.35	0.23	-1.12	1.04	1.09	0.33	1.29

*Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** before a number represent significance at 1%, 5% and 10% level respectively.*

Table 5.31: *Diebold-Mariano statistic values with absolute error loss function. Load period 38.*

$e^1 \downarrow e^2 \rightarrow$	LR1	LR2	LR3	MRS1	M1RS3	M1RS6	M1RS9	MRS2	M2RS3
LR1	—	—	—	—	—	—	—	—	—
LR2	1.41	—	—	—	—	—	—	—	—
LR3	1.38	0.46	—	—	—	—	—	—	—
MRS1	-1.44	**-2.13	***-1.81	—	—	—	—	—	—
M1RS3	1.33	0.28	-0.16	**1.94	—	—	—	—	—
M1RS6	***-1.74	*-2.51	*-2.61	-0.62	**-2.35	—	—	—	—
M1RS9	-0.43	-1.25	-1.27	**1.99	-1.33	***1.77	—	—	—
MRS2	-0.45	-1.54	-1.20	1.54	-1.18	1.43	-0.09	—	—
M2RS3	-0.55	-1.57	**-2.11	0.50	-1.56	1.04	-0.11	-0.07	—
M2RS6	-0.89	**-1.94	-1.59	0.38	-1.49	0.95	-0.55	-0.58	-0.29
M2RS9	-0.60	***-1.69	-1.32	0.96	-1.31	1.30	-0.37	-0.36	-0.05
MRS3	-0.11	-0.80	-1.46	1.05	-1.06	***1.80	0.23	0.24	0.29
M3RS3	-0.47	-1.10	**-2.08	0.26	-1.28	0.67	-0.22	-0.17	-0.18
M3RS6	0.92	0.26	-0.19	***1.70	0.01	*2.57	1.19	1.11	1.34
M3RS9	-0.79	-1.42	*-2.14	-0.05	-1.35	0.33	-0.62	-0.54	-0.49
TVR1	**-2.12	*-2.76	*-3.60	-1.02	*-2.58	-0.90	***-1.64	-1.44	***-1.85
TVR2	***-1.90	*-2.63	*-3.21	-0.75	**-2.42	-0.46	-1.40	-1.24	-1.58
TVR3	**-1.93	*-2.44	*-3.09	-1.07	**-2.41	-0.79	***-1.78	-1.45	-1.37
ARMAX	-1.63	**-2.13	*-2.89	-0.72	**-2.33	-0.51	-1.24	-1.11	-1.30

$e^1 \downarrow e^2 \rightarrow$	M2RS6	M2RS9	MRS3	M3RS3	M3RS6	M3RS9	TVR1	TVR2	TVR3
LR1	—	—	—	—	—	—	—	—	—
LR2	—	—	—	—	—	—	—	—	—
LR3	—	—	—	—	—	—	—	—	—
MRS1	—	—	—	—	—	—	—	—	—
M1RS3	—	—	—	—	—	—	—	—	—
M1RS6	—	—	—	—	—	—	—	—	—
M1RS9	—	—	—	—	—	—	—	—	—
MRS2	—	—	—	—	—	—	—	—	—
M2RS3	—	—	—	—	—	—	—	—	—
M2RS6	—	—	—	—	—	—	—	—	—
M2RS9	0.35	—	—	—	—	—	—	—	—
MRS3	0.56	0.38	—	—	—	—	—	—	—
M3RS3	0.08	-0.08	-0.44	—	—	—	—	—	—
M3RS6	1.35	1.28	1.52	***1.82	—	—	—	—	—
M3RS9	-0.25	-0.47	-1.19	-0.44	**-2.43	—	—	—	—
TVR1	-1.21	-1.42	**-2.20	-1.51	*-3.28	-1.22	—	—	—
TVR2	-0.98	-1.22	***-1.79	-1.16	*-2.76	-0.75	0.89	—	—
TVR3	-1.09	-1.47	**-2.28	-1.47	*-3.03	-1.16	-0.05	-0.51	—
ARMAX	-0.95	-1.10	***-1.66	-1.16	*-2.84	-0.81	0.25	-0.25	0.21

Note: M1RS3 in the table stands for MRS1RS3, etc. *, **, *** *before a number represent significance at* 1%, 5% *and* 10% *level respectively.*

Chapter 6

Forecast combinations

Reid (1968, 1969) and Bates & Granger (1969) were the first to develop a general analytical model specifically for combining forecasts in an optimal way and to apply their techniques in real world situations. From their seminal works, forecast combination methods have been intensely studied Bunn (1975, 1977, 1978) as reported by the reviews of Clemen (1989), de Menezes et al. (2000) and more recently Timmermann (2006).
The idea behind combining forecasting techniques is straightforward: no forecasting model is appropriate for all situations. This means that single forecasting models may capture different parts of the characteristics of the data, i.e. they may be optimal only conditional on a given sample realization, information set, model specification or forecasting periods. By implementing a combination of forecasts obtained by different models, we can compensate the weakness of the singular forecasting model under particular conditions.
Forecast combinations have frequently been found in empirical studies to produce better forecasts, on average, than methods based on the ex-ante best individual forecasting model (recent studies comprise, for instance, Altavilla & De Grauwe, 2006; Clark & McCracken, 2007; Clemens & Hendry, 1998 ;Riedel & Gabrys, 2005). Although the theoretical literature suggests that appropriate combinations of individual forecasts often have superior performance, such methods have not been exploited in the electricity price literature.
Based on empirical evidences and theoretical considerations, Chen & Yang (2007) advocate the use of forecast combining when there is considerable instability in model selection by testing procedures. From findings in chapter 5, this is our case. In this chapter we applied forecast combination methods to electricity price data. Our results underline that the application at a seasonal level of these techniques produces significative improvements in prediction accuracy with respect to singular models (see section 6.3).

6.1 Combining models

Let p_t be the variable of interest at time $t = 1, 2, ...$, and let $f_t^{(1)}, ..., f_t^{(K)}$ be the set of K competing predictors of p_t made with information available at time $t-1$. The simplest way to obtain a combination of forecasts is to consider a linear combination:

$$p_t^C = \sum_{k=1}^{K} \omega_t^{(k)} f_t^{(k)}, \tag{6.1}$$

for $t = 1, 2, ...,$ where $\omega_t^{(k)}$ are suitable weights. When $\omega = (\omega_t^{(1)}, ..., \omega_t^{(K)})'$ is a vector of constants, it can be estimated through the model:

$$p_t = p_t^C + e_t^C = \sum_{k=1}^{K} \omega^{(k)} f_t^{(k)} + e_t^C, \tag{6.2}$$

where e_t^C is an error term. In our study we considered the simplest approach with constant weights that is the equally weighted combination of the forecasts, where $\omega^{(k)} = 1/K$, so the combination is the mean of the constituent forecasts.
For time-varying weights Bates & Granger (1969) (see also Newbold & Granger, 1974) proposed several adaptive estimation schemes. They are based on the simple idea to assign larger weights to models that performed best most recently. In our work, we have tested one of the methods proposed by Bates and Granger with two different specifications. Let $e_t^{(k)} = p_t - f_t^{(k)}$ the forecasting error of model k, the Bates and Granger weights are:

$$\omega_t^{(k)} = \frac{\left(\sum_{\tau=t-l+1}^{t-1} \left(e_\tau^{(k)}\right)^2\right)^{-1}}{\sum_{j=1}^{K} \left(\sum_{\tau=t-l+1}^{t-1} \left(e_\tau^{(j)}\right)^2\right)^{-1}}, \tag{6.3}$$

where $0 \leq \omega_t^{(k)} \leq 1$ and $\sum_{k=1}^{K} \omega_t^{(k)} = 1$. This method uses a rolling window of the most recent l observations based on the forecasting models' relative performances. A second specification is obtained when an expanding window is used, i.e. $l = t$. Weights in (6.3) depend on the inverse of MSEs i.e. larger weights are given to models with the smaller forecasting error.
Correlations between forecast errors are ignored in (6.3). Accounting for correlations, the weights are estimated by

$$\omega_t^{(k)} = \boldsymbol{\Sigma}_t^{-1} \iota / \left(\iota' \boldsymbol{\Sigma}_t^{-1} \iota\right), \tag{6.4}$$

$$\boldsymbol{\Sigma}_t[i, j] = \frac{1}{l} \sum_{\tau=t-l+1}^{t-1} e_\tau^{(i)} e_\tau^{(j)}, \tag{6.5}$$

where ι is a vector of ones and $\boldsymbol{\Sigma}$ is the matrix of forecast error correlations estimated using a rolling window of the most recent l observations. We also considered correlations between forecast errors, but our findings underline a worsening of the results. This is in line with the empirical findings in the literature on forecast combinations (see for instance, Bunn, 1985; Clemen & Winkler, 1986; Dunis et al., 2001; Makridakis & Winkler, 1983; Newbold & Granger, 1974): combinations that do not require estimating many parameters, such as that considered in our work, do better than more sophisticated methods which require the estimation of weights depending

on the full variance-covariance matrix of forecast errors.
In a recent paper, Smith & Wallis (2009) presented a formal explanation of this behaviour. Following the works of West (1996) and Clark & West (2006), they pointed out that the cause lies in the effect of finite-sample error in estimating the combining weights. In particular, they showed that, in terms of MSE, a simple mean of the forecasts is expected to be more accurate than a combination based on estimated weights when the optimal combining weights are equal or closed to equality. This explains for example the results of Gupta & Wilton (1987): they found that equally weighted combinations depends strongly on the relative size of the variance of the forecast errors associated with different forecasting methods. When these are similar, equal weights perform well. In the case of instability in the forecast errors variances, some time-variation or adaptive adjustment in the weights can improve combination forecasting performance (Stock & Watson, 2004; Winkler & Makridakis, 1983). For this reason, we decided to use Bates and Granger method (formula (6.3)): Smith & Wallis (2009) provided with a firmer foundation, based on an asymptotic approximation to the variance of the estimated weights, the recommendation to neglect any covariances between the forecasting errors.

To obtain a first indication about the behaviour of combination with our data, for each load period we calculated equally weighted combination ('mean') and Bates and Granger combination with a rolling window of 10 data and an expanding window ('BG10' and 'BGExp' respectively) with all the 19 models. To diminish forecast error variability, combinations are carried out considering the logarithm transformation of all the series and then the combination prediction is re-transformed with the exponential function. With this method, we observed a little improvement in the combination results.
Tables 6.1-6.4 contain results of the combinations expressed as relative measures that are calculated for each forecasting period as the ratio between the combination prediction error statistic and the best value of the statistic obtained among all the singular models:

$$\begin{aligned} \mathrm{R}_{\mathrm{MSE}} &= \frac{\mathrm{MSE}_C}{\mathrm{MSE}_{Best}} \\ \mathrm{R}_{\mathrm{MSPE}} &= \frac{\mathrm{MSPE}_C}{\mathrm{MSPE}_{Best}} \\ \mathrm{R}_{\mathrm{MAE}} &= \frac{\mathrm{MAE}_C}{\mathrm{MAE}_{Best}} \\ \mathrm{R}_{\mathrm{MAPE}} &= \frac{\mathrm{MAPE}_C}{\mathrm{MAPE}_{Best}}. \end{aligned}$$

As Hyndman & Koehler (2006) pointed out, an advantage of these measures is their interpretability. A value less than 1 means that the combination is better than the best singular model, while a value greater than 1 means that the combination does not outperform the best singular model.
Results are not very appealing. We obtained some slight improvement in forecasting accuracy only considering the whole forecasting period. If we consider the subperiods, combinations work better only during spring for load period 6, that is a very

calm period. The reason of a so disappointing results is that using all the 19 models we introduced too much variability in the combinations. This variability is given by models that perform very poorly during particular seasons and/or for particular hours. In literature it has been established that rather than combining the full set of forecasts, it is often advantageous to discard the models with the worst performance (see, for instance, Aiolfi & Favero, 2005; Granger & Jeon, 2004). This is called 'trimming'. The works of Stock & Watson (2001, 2004) and Marcellino (2004) opened a whole set of new issues concerning the number of models and the types of forecasts entering the combination. This depends on the type of misspecification or instability the model combination can hedge against.
In the next section, we propose a study conducted at a seasonal level to highlight that few models, characterized by good forecasting accuracy in the considered season, can improve combination performances.

Table 6.1: *Results of the combinations considering all the 19 models. Load period 6.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.257	1.140	1.008	1.044
April-June	**0.934**	**0.902**	**0.989**	**0.977**
July-September	1.080	1.088	1.051	1.062
Whole Period	1.122	**0.916**	**0.929**	**0.924**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.257	1.158	1.021	1.061
April-June	**0.919**	**0.913**	**0.993**	**0.987**
July-September	1.063	1.072	1.041	1.052
Whole Period	1.119	**0.921**	**0.935**	**0.930**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.257	1.144	1.012	1.049
April-June	**0.934**	**0.908**	**0.989**	**0.978**
July-September	1.069	1.080	1.045	1.057
Whole Period	1.121	**0.916**	**0.930**	**0.924**

6.2 Selection of forecasting models

To select the models entering the combination, we used two procedures: the Model Confidence Set (MCS) method (Hansen et al., 2005) is used to obtain in each season the set of models with best forecasting accuracy. Then, the chosen models are screened with the forecasts encompassing method (Fair & Shiller, 1990) to determine which models contain useful information about the price dynamics during each season. Results highlight that the set of models change with seasons, for this reason we propose to apply combination methods at seasonal level. Our work represents a

Table 6.2: *Results of the combinations considering all the 19 models. Load period 19.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.337	1.177	**0.985**	1.049
April-June	1.253	**0.952**	**0.928**	**0.922**
July-September	1.117	1.505	1.074	1.141
Whole Period	1.003	1.060	**0.959**	**0.982**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.342	1.115	**0.964**	1.013
April-June	1.167	**0.891**	**0.906**	**0.900**
July-September	1.107	1.404	1.057	1.099
Whole Period	**0.996**	**0.995**	**0.940**	**0.949**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.371	1.136	**0.972**	1.019
April-June	1.232	**0.946**	**0.919**	**0.913**
July-September	1.118	1.486	1.071	1.132
Whole Period	1.013	1.039	**0.950**	**0.966**

Table 6.3: *Results of the combinations considering all the 19 models. Load period 28.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.046	1.066	1.044	1.098
April-June	1.135	1.067	1.096	1.077
July-September	1.263	1.006	**0.981**	1.031
Whole Period	**0.992**	**0.977**	**0.949**	1.016

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.048	1.043	1.038	1.082
April-June	1.090	1.047	1.061	1.047
July-September	1.139	**0.954**	**0.939**	1.002
Whole Period	**0.958**	**0.946**	**0.927**	**0.993**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.048	1.0306	1.029	1.070
April-June	1.124	1.0637	1.089	1.072
July-September	1.250	1.0017	**0.978**	1.028
Whole Period	**0.989**	**0.961**	**0.941**	1.004

Table 6.4: *Results of the combinations considering all the 19 models. Load period 38.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.386	**0.979**	1.061	1.068
April-June	1.090	1.004	1.031	1.033
July-September	1.991	1.097	1.161	1.059
Whole Period	1.306	**0.990**	1.017	**0.989**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.381	**0.948**	1.056	1.053
April-June	1.067	**0.997**	1.020	1.029
July-September	1.927	1.058	1.123	1.011
Whole Period	1.296	**0.961**	1.004	**0.966**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.377	**0.986**	1.065	1.071
April-June	1.076	**0.997**	1.028	1.033
July-September	1.944	1.090	1.152	1.053
Whole Period	1.295	**0.991**	1.017	**0.988**

first study to understand if combining methods could be helpful for the electricity price forecasting issue. Combination results are shown in section 6.3.

6.2.1 The MCS method (EPA test)

The model confidence set (MCS) is a method developed in Hansen et al. (2003) and Hansen et al. (2005) that allows forecasting model selection. In fact the MCS is a set of models constructed to contain the best models in terms of equal predictive ability (EPA). Moreover, it is analogous to the confidence interval of a parameter in the sense that it contains the best forecasting models with a certain probability.
The construction of an MCS requires an iterative procedure based on a sequence of test for EPA. The procedure starts with the set of all the K models $\mathcal{M}_0$. At each step the the worst performing model is deleted and the set of candidate models $\mathcal{M} = \{1, ..., m\}$ with $m \leq K$ is trimmed. The models in the final set of the MCS include the optimal models not significantly different in terms of forecasting accuracy.
Each step begins with the computation of loss differentials between models i and j:

$$d_{ij,t} = L(e_t^{(i)}) - L(e_t^{(j)}) \tag{6.6}$$

where $i, j = 1, ..., m$, $t = 1, ..., T$ and L is the squared error or the absolute error loss function (as explained in section 5.5). Then the EPA hypothesis

$$\mathrm{H}_0 \ : \ E(d_{ij,t}) = 0, \quad \forall i > j \in \mathcal{M} \tag{6.7}$$

is tested. To establish EPA, we used the two test statistics propose by Hansen et al. (2005): a range statistic

$$T_r = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{ij}|}{\sqrt{\widehat{var}(\bar{d}_{ij})}}, \tag{6.8}$$

and a semi-quadratic statistic

$$T_{sq} = \sum_{\substack{i,j \in \mathcal{M} \\ i<j}} \frac{(\bar{d}_{ij})^2}{\widehat{var}(\bar{d}_{ij})}. \tag{6.9}$$

In the equations above $\bar{d}_{ij} = \frac{1}{T}\sum_{t=1}^{T} d_{ij,t}$ and $\widehat{var}(\bar{d}_{ij})$ is an estimate of $var(\bar{d}_{ij})$ obtained from a block bootstrap procedure with constant block fully described in Hansen et al. (2003) . At the core of the bootstrap procedure is the generation of bootstrap replications of $d_{ij,t}$.

Both test statistics (6.8) and (6.9) indicate a rejection of the EPA hypothesis for large values: p-values are obtained from the bootstrap distributions of the test statistics. When H_0 is rejected at the significance level α, the worst performing model, identified by

$$i = \underset{i \in \mathcal{M}}{\operatorname{argmax}} \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}} \tag{6.10}$$

with $\bar{d}_i = \frac{1}{m-1}\sum_{j \in \mathcal{M}} \bar{d}_{ij}$, is removed and the process re-starts until non-rejection occurs. The set of surviving models is the MCS, $\widehat{\mathcal{M}}^*_\alpha$. For a given significance level α, $\widehat{\mathcal{M}}^*_\alpha$ contains the best model from $\mathcal{M}_0$ with $(1-\alpha)$ confidence. Despite the testing procedure involving multiple hypothesis tests, Hansen et al. (2005) reported a detailed discussion about the statistically correctness of this interpretation.

A very useful feature of the procedure is that it yields MCS p-values for all models under consideration. Suppose that at the kth step model i is deleted from $\mathcal{M}$: we denote the (bootstrapped) p-value of the EPA test with (6.8) and (6.9) as $p(k)$ (with the convention $p(K) = 1$ for the model that survives all $K-1$ tests). The MCS p-value of model i is defined by

$$\hat{p}_i = \max_{j \leq k} p(j). \tag{6.11}$$

A model belongs to $\widehat{\mathcal{M}}^*_\alpha$ (i.e. it belongs to the set of best forecast models) when its MCS p-value exceeds α.

Tables 6.5-6.16 report the MCS results in each season (in our work, we used 5000 bootstrap replicates) of all the individual forecasts based on both squared error loss and absolute error loss function, for load periods 6, 19, 28 and 38. The first row in the tables represents the first model removed, down to the best performing model in the last row. ‘p.r’ and ‘MCS.r’ denote respectively the p-value of the EPA test and the MCS p-value with the range statistic. The use of the semi-quadratic statistic is denoted with ‘sq’.

Results confirm the findings in the previous chapter: in general there is not only one model with best forecasting accuracy. Nevertheless, rankings highlight relative predictive performance of the models in each season. Moreover, with the MCS p-values it is possible to select the set of models with better forecasting performance.

We decided to not choose a particular significance level for trimming the sets of models. The reason is that with a fixed α the risk is to obtain, in some cases, a set containing all the models (if α is low) or only one model (if α is high), and we wanted to avoid this situations. For each load period and each season, the choice of the models has been done starting from the best model in the ranking and adding models to the set, until a large decrease in the MCS p-value happens for at least one of the statistics. This procedure was conducted for both the considered loss functions: in this way we obtained two sets of models that have to be merge into one. The scope is to insert in the sets the best models for both the loss functions. The chosen models are highlighted in yellow.

Table 6.5: *MCS results for the forecasting period January-March 2006. Load period 6.*

Squared Error Loss					**Absolute Error Loss**				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS1	0.217	0.217	0.157	0.157	LR3	0.020	0.020	0.058	0.058
MRS3	0.211	0.217	0.169	0.169	MRS3RS6	0.029	0.029	0.067	0.067
MRS1RS9	0.206	0.217	0.198	0.198	MRS3RS3	0.026	0.029	0.077	0.077
MRS3RS9	0.197	0.217	0.199	0.199	MRS3	0.051	0.051	0.105	0.105
LR1	0.189	0.217	0.213	0.213	MRS3RS9	0.047	0.051	0.127	0.127
MRS1RS6	0.275	0.275	0.226	0.226	MRS1RS3	0.043	0.051	0.169	0.169
MRS1RS3	0.260	0.275	0.235	0.235	MRS1	0.044	0.051	0.228	0.228
LR3	0.233	0.275	0.221	0.235	MRS1RS9	0.108	0.108	0.310	0.310
MRS2	0.215	0.275	0.208	0.235	LR1	0.102	0.108	0.339	0.339
MRS3RS6	0.210	0.275	0.247	0.247	MRS1RS6	0.136	0.136	0.480	0.480
MRS2RS9	0.184	0.275	0.228	0.247	MRS2	0.275	0.275	0.567	0.567
MRS3RS3	0.168	0.275	0.282	0.282	TVR3	0.849	0.849	0.853	0.853
MRS2RS3	0.546	0.546	0.507	0.507	TVR1	0.831	0.849	0.915	0.915
MRS2RS6	0.520	0.546	0.553	0.553	ARMAX	0.767	0.849	0.885	0.915
TVR3	0.436	0.546	0.450	0.553	TVR2	0.677	0.849	0.818	0.915
TVR1	0.581	0.581	0.588	0.588	MRS2RS9	0.548	0.849	0.685	0.915
LR2	0.700	0.700	0.598	0.598	MRS2RS3	0.732	0.849	0.696	0.915
TVR2	0.422	0.700	0.422	0.598	LR2	0.493	0.849	0.493	0.915
ARMAX	—	1.000	—	1.000	MRS2RS6	—	1.000	—	1.000

Table 6.6: *MCS results for the forecasting period April-June 2006. Load period 6.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR2	0.426	0.426	0.485	0.485	LR2	0.053	0.053	0.421	0.421
MRS2RS3	0.433	0.433	0.531	0.531	MRS1RS6	0.051	0.053	0.436	0.436
LR1	0.413	0.433	0.545	0.545	LR1	0.046	0.053	0.448	0.448
MRS1RS6	0.407	0.433	0.600	0.600	MRS2RS3	0.046	0.053	0.470	0.470
MRS2RS9	0.629	0.629	0.653	0.653	MRS1RS3	0.042	0.053	0.490	0.490
MRS1RS3	0.608	0.629	0.672	0.672	MRS2RS9	0.036	0.053	0.525	0.525
MRS1	0.646	0.646	0.706	0.706	MRS2RS6	0.033	0.053	0.489	0.525
ARMAX	0.644	0.646	0.741	0.741	MRS1	0.031	0.053	0.456	0.525
MRS2RS6	0.623	0.646	0.711	0.741	ARMAX	0.030	0.053	0.438	0.525
MRS3RS3	0.582	0.646	0.678	0.741	LR3	0.027	0.053	0.365	0.525
LR3	0.559	0.646	0.693	0.741	TVR2	0.024	0.053	0.316	0.525
MRS1RS9	0.519	0.646	0.671	0.741	MRS1RS9	0.019	0.053	0.382	0.525
MRS3RS6	0.494	0.646	0.637	0.741	MRS3RS6	0.015	0.053	0.309	0.525
TVR3	0.635	0.646	0.672	0.741	TVR1	0.205	0.205	0.424	0.525
TVR2	0.545	0.646	0.637	0.741	MRS3RS3	0.379	0.379	0.443	0.525
MRS2	0.644	0.646	0.680	0.741	TVR3	0.366	0.379	0.431	0.525
TVR1	0.502	0.646	0.481	0.741	MRS2	0.242	0.379	0.343	0.525
MRS3RS9	0.423	0.646	0.423	0.741	MRS3RS9	0.133	0.379	0.133	0.525
MRS3	—	1.000	—	1.000	MRS3	—	1.000	—	1.000

Table 6.7: *MCS results for the forecasting period July-September 2006. Load period 6.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS1RS9	0.098	0.098	0.134	0.134	MRS2RS6	0.253	0.253	0.138	0.138
MRS3RS6	0.094	0.098	0.158	0.158	MRS1RS9	0.252	0.253	0.178	0.178
MRS2RS6	0.088	0.098	0.168	0.168	MRS3RS6	0.429	0.429	0.233	0.233
MRS1RS3	0.083	0.098	0.172	0.172	LR1	0.412	0.429	0.256	0.256
MRS3RS9	0.079	0.098	0.199	0.199	MRS1RS3	0.402	0.429	0.284	0.284
LR1	0.073	0.098	0.220	0.220	MRS3RS9	0.383	0.429	0.300	0.300
MRS1RS6	0.069	0.098	0.231	0.231	MRS2RS9	0.355	0.429	0.342	0.342
MRS2RS9	0.069	0.098	0.248	0.248	MRS1RS6	0.333	0.429	0.401	0.401
TVR1	0.065	0.098	0.237	0.248	LR2	0.328	0.429	0.452	0.452
TVR3	0.276	0.276	0.287	0.287	TVR3	0.311	0.429	0.449	0.452
LR2	0.295	0.295	0.312	0.312	ARMAX	0.534	0.534	0.449	0.452
ARMAX	0.267	0.295	0.307	0.312	MRS1	0.593	0.593	0.460	0.456
MRS1	0.263	0.295	0.330	0.330	TVR1	0.532	0.593	0.429	0.456
LR3	0.231	0.295	0.301	0.330	LR3	0.474	0.593	0.471	0.471
MRS2RS3	0.180	0.295	0.274	0.330	MRS2RS3	0.396	0.593	0.467	0.471
MRS3RS3	0.541	0.541	0.572	0.572	MRS3RS3	0.670	0.670	0.645	0.645
MRS3	0.837	0.837	0.852	0.852	MRS2	0.991	0.991	0.989	0.989
MRS2	0.851	0.851	0.851	0.852	TVR2	0.922	0.991	0.922	0.989
TVR2	—	1.000	—	1.000	MRS3	—	1.000	—	1.000

Table 6.8: *MCS results for the forecasting period January-March 2006. Load period 19.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR2	0.395	0.395	0.492	0.492	MRS2	0.040	0.040	0.262	0.262
MRS2RS9	0.452	0.452	0.566	0.566	MRS2RS9	0.038	0.040	0.325	0.325
MRS2RS6	0.431	0.452	0.590	0.590	MRS2RS6	0.035	0.040	0.377	0.377
LR3	0.424	0.452	0.643	0.643	LR3	0.034	0.040	0.502	0.502
MRS2	0.405	0.452	0.648	0.648	LR2	0.630	0.630	0.656	0.656
TVR1	0.500	0.500	0.652	0.652	MRS3RS6	0.726	0.726	0.757	0.757
TVR3	0.778	0.778	0.673	0.673	MRS2RS3	0.697	0.726	0.795	0.795
MRS3	0.762	0.778	0.661	0.673	TVR3	0.647	0.726	0.827	0.827
LR1	0.748	0.778	0.654	0.673	MRS3RS9	0.695	0.726	0.846	0.846
MRS3RS6	0.725	0.778	0.654	0.673	MRS3	0.671	0.726	0.864	0.864
MRS3RS9	0.697	0.778	0.642	0.673	LR1	0.621	0.726	0.867	0.867
MRS1RS6	0.658	0.778	0.633	0.673	MRS1	0.576	0.726	0.878	0.878
MRS1	0.734	0.778	0.654	0.673	TVR1	0.519	0.726	0.861	0.878
MRS1RS9	0.871	0.871	0.683	0.683	MRS1RS9	0.844	0.844	0.919	0.919
ARMAX	0.814	0.871	0.705	0.705	MRS1RS6	0.791	0.844	0.902	0.919
TVR2	0.803	0.871	0.734	0.734	MRS3RS3	0.718	0.844	0.871	0.919
MRS2RS3	0.603	0.871	0.531	0.734	ARMAX	0.520	0.844	0.664	0.919
MRS1RS3	0.790	0.871	0.790	0.790	MRS1RS3	0.796	0.844	0.796	0.919
MRS3RS3	—	1.000	—	1.000	TVR2	—	1.000	—	1.000

Table 6.9: *MCS results for the forecasting period April-June 2006. Load period 19.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS3RS6	0.006	0.006	0.117	0.117	MRS1RS3	0.000	0.000	0.031	0.031
MRS3RS9	0.006	0.006	0.109	0.117	MRS1RS6	0.000	0.000	0.036	0.036
LR3	0.006	0.006	0.100	0.117	MRS2RS6	0.000	0.000	0.049	0.049
MRS2RS6	0.006	0.006	0.096	0.117	LR3	0.022	0.022	0.077	0.077
MRS1RS3	0.006	0.006	0.119	0.119	MRS2RS3	0.021	0.022	0.081	0.081
MRS1RS6	0.005	0.006	0.113	0.119	MRS3	0.090	0.090	0.130	0.130
MRS3	0.005	0.006	0.118	0.119	MRS1RS9	0.080	0.090	0.160	0.160
MRS1RS9	0.005	0.006	0.120	0.120	MRS3RS9	0.076	0.090	0.164	0.164
MRS2RS3	0.005	0.006	0.106	0.120	MRS3RS6	0.070	0.090	0.189	0.189
LR1	0.018	0.018	0.219	0.219	MRS3RS3	0.107	0.107	0.239	0.239
MRS1	0.016	0.018	0.217	0.219	LR1	0.507	0.507	0.356	0.356
MRS3RS3	0.015	0.018	0.236	0.236	MRS1	0.485	0.507	0.447	0.447
LR2	0.013	0.018	0.240	0.240	MRS2RS9	0.543	0.543	0.622	0.622
MRS2RS9	0.595	0.595	0.594	0.594	LR2	0.705	0.705	0.797	0.797
MRS2	0.632	0.632	0.619	0.619	TVR2	0.629	0.705	0.823	0.823
TVR1	0.499	0.632	0.520	0.619	TVR1	0.931	0.931	0.906	0.906
TVR3	0.366	0.632	0.421	0.619	MRS2	0.966	0.966	0.964	0.964
TVR2	0.199	0.632	0.199	0.619	TVR3	0.956	0.966	0.956	0.964
ARMAX	—	1.000	—	1.000	ARMAX	—	1.000	—	1.000

Table 6.10: *MCS results for the forecasting period July-September 2006. Load period 19.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS2RS3	0.157	0.157	0.211	0.211	MRS3RS3	0.022	0.022	0.071	0.071
MRS3RS3	0.201	0.201	0.267	0.267	MRS3RS9	0.021	0.022	0.100	0.100
MRS3RS9	0.188	0.201	0.278	0.278	LR3	0.020	0.022	0.132	0.132
LR3	0.187	0.201	0.296	0.296	MRS2RS3	0.019	0.022	0.148	0.148
MRS3RS6	0.171	0.201	0.316	0.316	MRS3RS6	0.019	0.022	0.161	0.161
MRS1RS9	0.162	0.201	0.318	0.318	MRS2RS9	0.017	0.022	0.186	0.186
MRS3	0.423	0.423	0.364	0.364	MRS1RS9	0.015	0.022	0.263	0.263
ARMAX	0.412	0.423	0.373	0.373	TVR3	0.014	0.022	0.289	0.289
MRS1RS3	0.403	0.423	0.395	0.395	MRS3	0.200	0.200	0.420	0.420
TVR3	0.676	0.676	0.441	0.441	MRS2RS6	0.620	0.620	0.528	0.528
TVR1	0.619	0.676	0.466	0.466	ARMAX	0.600	0.620	0.568	0.568
TVR2	0.562	0.676	0.531	0.531	MRS1RS3	0.572	0.620	0.592	0.592
MRS1	0.741	0.741	0.677	0.677	MRS1RS6	0.491	0.620	0.586	0.592
MRS2RS6	0.676	0.741	0.736	0.736	TVR2	0.445	0.620	0.554	0.592
LR1	0.678	0.741	0.759	0.759	TVR1	0.357	0.620	0.630	0.630
LR2	0.544	0.741	0.676	0.759	MRS2	0.317	0.620	0.544	0.630
MRS2RS9	0.458	0.741	0.618	0.759	MRS1	0.673	0.673	0.638	0.638
MRS2	0.926	0.926	0.926	0.926	LR1	0.441	0.673	0.441	0.638
MRS1RS6	—	1.000	—	1.000	LR2	—	1.000	—	1.000

Table 6.11: *MCS results for the forecasting period January-March 2006. Load period 28.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS2RS3	0.239	0.239	0.579	0.579	MRS2RS3	0.018	0.018	0.164	0.164
MRS1RS3	0.311	0.311	0.662	0.662	LR3	0.105	0.105	0.186	0.186
MRS1RS6	0.398	0.398	0.700	0.700	MRS3RS3	0.099	0.105	0.202	0.202
MRS2RS9	0.397	0.398	0.736	0.736	MRS3RS6	0.092	0.105	0.224	0.224
TVR1	0.359	0.398	0.739	0.739	MRS2RS9	0.211	0.211	0.265	0.265
LR3	0.356	0.398	0.708	0.739	MRS3RS9	0.192	0.211	0.274	0.274
MRS3RS3	0.374	0.398	0.741	0.741	LR2	0.369	0.369	0.292	0.292
MRS2RS6	0.335	0.398	0.721	0.741	TVR3	0.333	0.369	0.262	0.292
LR2	0.305	0.398	0.697	0.741	MRS1RS6	0.318	0.369	0.262	0.292
MRS3RS6	0.256	0.398	0.655	0.741	MRS2RS6	0.350	0.369	0.289	0.292
MRS2	0.356	0.398	0.724	0.741	MRS1RS3	0.321	0.369	0.261	0.292
TVR3	0.310	0.398	0.677	0.741	MRS3	0.278	0.369	0.283	0.292
MRS3	0.574	0.574	0.644	0.741	MRS2	0.422	0.422	0.313	0.313
MRS1RS9	0.541	0.574	0.620	0.741	LR1	0.540	0.540	0.339	0.339
MRS3RS9	0.506	0.574	0.543	0.741	MRS1RS9	0.528	0.540	0.402	0.402
LR1	0.699	0.699	0.551	0.741	MRS1	0.500	0.540	0.416	0.416
MRS1	0.629	0.699	0.518	0.741	TVR1	0.419	0.540	0.442	0.442
TVR2	0.357	0.699	0.357	0.741	TVR2	0.716	0.716	0.716	0.716
ARMAX	—	1.000	—	1.000	ARMAX	—	1.000	—	1.000

Table 6.12: *MCS results for the forecasting period April-June 2006. Load period 28.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS2RS3	0.153	0.153	0.148	0.148	MRS2RS3	0.049	0.049	0.167	0.167
LR3	0.145	0.153	0.149	0.149	MRS1RS3	0.047	0.049	0.172	0.172
MRS3	0.144	0.153	0.198	0.198	LR1	0.043	0.049	0.180	0.180
MRS1	0.142	0.153	0.247	0.247	MRS1	0.041	0.049	0.178	0.180
LR1	0.145	0.153	0.293	0.293	MRS1RS9	0.040	0.049	0.181	0.181
MRS1RS3	0.299	0.299	0.352	0.352	LR3	0.037	0.049	0.170	0.181
MRS3RS9	0.282	0.299	0.394	0.394	MRS2RS6	0.034	0.049	0.154	0.181
MRS2RS6	0.269	0.299	0.373	0.394	MRS3	0.029	0.049	0.186	0.186
MRS1RS9	0.256	0.299	0.393	0.394	MRS1RS6	0.026	0.049	0.163	0.186
MRS3RS6	0.237	0.299	0.381	0.394	ARMAX	0.023	0.049	0.149	0.186
LR2	0.217	0.299	0.333	0.394	MRS3RS9	0.220	0.220	0.236	0.236
MRS2	0.197	0.299	0.329	0.394	MRS3RS6	0.204	0.220	0.213	0.236
ARMAX	0.166	0.299	0.302	0.394	MRS3RS3	0.182	0.220	0.169	0.236
MRS3RS3	0.404	0.404	0.386	0.394	MRS2RS9	0.150	0.220	0.163	0.236
MRS2RS9	0.341	0.404	0.325	0.394	MRS2	0.115	0.220	0.154	0.236
MRS1RS6	0.348	0.404	0.387	0.394	LR2	0.081	0.220	0.124	0.236
TVR1	0.240	0.404	0.316	0.394	TVR1	0.050	0.220	0.070	0.236
TVR2	0.473	0.473	0.473	0.473	TVR2	0.105	0.220	0.105	0.236
TVR3	——	1.000	——	1.000	TVR3	——	1.000	——	1.000

Table 6.13: *MCS results for the forecasting period July-September 2006. Load period 28.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS1RS3	0.254	0.254	0.253	0.253	ARMAX	0.100	0.100	0.191	0.191
ARMAX	0.254	0.254	0.255	0.255	LR3	0.091	0.100	0.314	0.314
LR3	0.382	0.382	0.296	0.296	MRS3	0.396	0.396	0.504	0.504
MRS3	0.375	0.382	0.322	0.322	MRS1RS3	0.395	0.396	0.606	0.606
MRS2	0.375	0.382	0.382	0.382	MRS3RS9	0.395	0.396	0.673	0.673
MRS1	0.374	0.382	0.385	0.385	TVR3	0.846	0.846	0.705	0.705
LR2	0.372	0.382	0.399	0.399	MRS3RS6	0.835	0.846	0.754	0.754
TVR3	0.359	0.382	0.413	0.413	LR2	0.790	0.846	0.744	0.754
TVR1	0.359	0.382	0.427	0.427	MRS1	0.768	0.846	0.750	0.754
LR1	0.359	0.382	0.441	0.441	TVR1	0.762	0.846	0.743	0.754
TVR2	0.349	0.382	0.455	0.455	MRS3RS3	0.753	0.846	0.768	0.768
MRS3RS9	0.348	0.382	0.515	0.515	MRS2	0.695	0.846	0.749	0.768
MRS3RS6	0.341	0.382	0.518	0.518	LR1	0.674	0.846	0.764	0.768
MRS3RS3	0.317	0.382	0.523	0.523	TVR2	0.621	0.846	0.749	0.768
MRS1RS9	0.310	0.382	0.583	0.583	MRS2RS6	0.566	0.846	0.726	0.768
MRS2RS6	0.276	0.382	0.457	0.583	MRS1RS9	0.921	0.921	0.909	0.909
MRS1RS6	0.651	0.651	0.615	0.615	MRS2RS3	0.899	0.921	0.922	0.922
MRS2RS3	0.392	0.651	0.392	0.615	MRS1RS6	0.838	0.921	0.838	0.922
MRS2RS9	——	1.000	——	1.000	MRS2RS9	——	1.000	——	1.000

Table 6.14: *MCS results for the forecasting period January-March 2006. Load period 38.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR2	0.260	0.260	0.253	0.253	MRS3RS6	0.003	0.003	0.142	0.142
MRS1RS3	0.253	0.260	0.278	0.278	LR3	0.044	0.044	0.226	0.226
LR3	0.251	0.260	0.305	0.305	MRS1RS3	0.384	0.384	0.345	0.345
MRS2RS9	0.253	0.260	0.333	0.333	LR2	0.419	0.419	0.383	0.383
MRS2	0.246	0.260	0.326	0.333	LR1	0.496	0.496	0.445	0.445
LR1	0.230	0.260	0.321	0.333	MRS3	0.477	0.496	0.458	0.458
MRS3RS6	0.226	0.260	0.346	0.346	MRS2	0.463	0.496	0.494	0.494
MRS2RS3	0.595	0.595	0.421	0.421	MRS2RS9	0.494	0.496	0.540	0.540
MRS1	0.566	0.595	0.418	0.421	MRS1RS9	0.620	0.620	0.604	0.604
MRS1RS9	0.538	0.595	0.420	0.421	MRS2RS3	0.607	0.620	0.666	0.666
TVR2	0.501	0.595	0.431	0.431	MRS1	0.567	0.620	0.708	0.708
MRS3	0.527	0.595	0.504	0.504	MRS3RS3	0.532	0.620	0.763	0.763
ARMAX	0.554	0.595	0.536	0.536	MRS2RS6	0.471	0.620	0.732	0.763
MRS2RS6	0.790	0.790	0.601	0.601	TVR2	0.431	0.620	0.661	0.763
TVR3	0.795	0.795	0.684	0.684	TVR3	0.746	0.746	0.808	0.808
MRS3RS3	0.706	0.795	0.633	0.684	MRS3RS9	0.821	0.821	0.786	0.808
TVR1	0.517	0.795	0.538	0.684	MRS1RS6	0.740	0.821	0.712	0.808
MRS1RS6	0.486	0.795	0.486	0.684	TVR1	0.606	0.821	0.606	0.808
MRS3RS9	—	1.000	—	1.000	ARMAX	—	1.000	—	1.000

Table 6.15: *MCS results for the forecasting period April-June 2006. Load period 38.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR3	0.004	0.004	0.062	0.062	MRS2RS6	0.007	0.007	0.111	0.111
MRS3	0.004	0.004	0.085	0.085	LR2	0.007	0.007	0.177	0.177
LR2	0.004	0.004	0.124	0.124	LR3	0.008	0.008	0.254	0.254
MRS2RS6	0.004	0.004	0.181	0.181	MRS3	0.007	0.008	0.276	0.276
MRS1RS3	0.165	0.165	0.257	0.257	MRS1RS3	0.007	0.008	0.294	0.294
LR1	0.157	0.165	0.267	0.267	MRS3RS6	0.007	0.008	0.310	0.310
MRS3RS9	0.238	0.238	0.329	0.329	LR1	0.006	0.008	0.323	0.323
MRS2	0.234	0.238	0.380	0.380	TVR1	0.377	0.377	0.505	0.505
MRS2RS3	0.342	0.342	0.433	0.433	MRS3RS9	0.455	0.455	0.531	0.531
MRS3RS6	0.422	0.422	0.584	0.584	MRS3RS3	0.439	0.455	0.575	0.575
MRS1RS9	0.654	0.654	0.790	0.790	MRS1RS6	0.404	0.455	0.640	0.640
MRS3RS3	0.592	0.654	0.793	0.793	MRS2RS3	0.359	0.455	0.713	0.713
MRS1	0.839	0.839	0.868	0.868	MRS2	0.311	0.455	0.677	0.713
ARMAX	0.793	0.839	0.817	0.868	ARMAX	0.583	0.583	0.754	0.754
MRS1RS6	0.698	0.839	0.732	0.868	TVR3	0.510	0.583	0.776	0.776
TVR3	0.693	0.839	0.689	0.868	MRS1RS9	0.402	0.583	0.684	0.776
TVR1	0.561	0.839	0.609	0.868	TVR2	0.942	0.942	0.939	0.939
MRS2RS9	0.532	0.839	0.532	0.868	MRS1	0.919	0.942	0.919	0.939
TVR2	—	1.000	—	1.000	MRS2RS9	—	1.000	—	1.000

Table 6.16: *MCS results for the forecasting period July-September 2006. Load period 38.*

Squared Error Loss					Absolute Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS3RS3	0.677	0.677	0.259	0.259	MRS1RS3	0.434	0.434	0.369	0.369
LR2	0.641	0.677	0.297	0.297	MRS3RS6	0.417	0.434	0.390	0.390
MRS2RS3	0.706	0.706	0.356	0.356	MRS1RS9	0.400	0.434	0.416	0.416
MRS2RS6	0.674	0.706	0.374	0.374	LR3	0.380	0.434	0.437	0.437
MRS1RS3	0.659	0.706	0.383	0.383	LR2	0.369	0.434	0.491	0.491
LR3	0.659	0.706	0.403	0.403	MRS2RS3	0.355	0.434	0.588	0.588
MRS3RS6	0.657	0.706	0.427	0.427	LR1	0.327	0.434	0.578	0.588
MRS2	0.639	0.706	0.410	0.427	ARMAX	0.291	0.434	0.626	0.626
MRS3	0.621	0.706	0.413	0.427	MRS3RS3	0.647	0.647	0.697	0.697
MRS1RS9	0.600	0.706	0.406	0.427	MRS3RS9	0.613	0.647	0.711	0.711
MRS1RS6	0.552	0.706	0.426	0.427	MRS3	0.649	0.649	0.752	0.752
LR1	0.488	0.706	0.396	0.427	MRS2RS9	0.602	0.649	0.781	0.781
MRS3RS9	0.434	0.706	0.374	0.427	MRS2RS6	0.858	0.858	0.804	0.804
MRS2RS9	0.385	0.706	0.360	0.427	MRS1RS6	0.879	0.879	0.780	0.804
ARMAX	0.512	0.706	0.361	0.427	MRS2	0.833	0.879	0.762	0.804
MRS1	0.413	0.706	0.329	0.427	MRS1	0.770	0.879	0.762	0.804
TVR1	0.290	0.706	0.311	0.427	TVR2	0.563	0.879	0.610	0.804
TVR2	0.486	0.706	0.486	0.486	TVR1	0.548	0.879	0.548	0.804
TVR3	—	1.000	—	1.000	TVR3	—	1.000	—	1.000

6.2.2 Forecasts encompassing

Even with the choice of subsets of models conducted with the MCS method, in some cases the number of models is high. The idea behind forecast combinations is that the less-performing forecasts may provide some marginal information that is not contained in the better forecast. In such a case, the combination will perform better then either forecast alone. The question is if all the selected models contain useful information for improving combination results. For this reason, we applied the forecasts encompassing method to screen the models in each set.
Let $f_t^{(1)}$ and $f_t^{(2)}$ be two competing predictors of the variable of interest. Assume that one of the two sets, say $f_t^{(1)}$, performs better by some criteria. If $f_t^{(2)}$ contains no useful marginal information, than it is said that $f_t^{(1)}$ encompasses $f_t^{(2)}$.
A simple methodology that helps to show up if a set of forecasts encompasses an other one has been developed by Fair & Shiller (1990). The testing procedure is based on the following equation:

$$p_t - p_{t-1} = \alpha + \beta_1(f_t^{(1)} - p_{t-1}) + \beta_2(f_t^{(2)} - p_{t-1}) + \varepsilon_t. \qquad (6.12)$$

The intuition behind this testing procedure is straightforward. If both forecasts contain useful and independent information concerning p_t, then the estimates of both the slope coefficients β_1 and β_2 should be significant. In contrast, if the information in one forecast is completely contained in the other, then the coefficient of the second forecast should be nonzero while that of the first one should be zero.

Applying this method on the sets of models obtained for each season with the MCS method, we obtained the following reduced sets of models:
Period 6
- winter: ARMAX MRS2RS6;
- spring: MRS2 MRS3;
- summer: MRS2 MRS3 TVR2;
Period 19
- winter: MRS2RS3 MRS3RS3 TVR2;
- spring: MRS2 ARMAX;
- summer: LR1 LR2 MRS1;
Period 28
- winter: TVR2 ARMAX;
- spring: LR2 TVR3;
- summer: MRS1RS6 MRS2RS9;
Period 38
- winter: MRS1RS6 TVR1;
- spring: MRS1 MRS2RS9 TVR2 TVR3;
- summer: TVR1 TVR3.
It is interesting to note that now the larger set contains only four models. Each load period is characterized by particular dynamics that change with the seasons. This is underlined by the pattern of models that is not the same during the year. Load period 38 is the more volatile one: its dynamics is predicted by the combined action of models with changing regimes (MRS) and time-adaptive models (TVR). Markov regime-switching models estimated on rolling windows are useful during spiky periods, as summer for period 28 and during periods characterized by structural changes, as winter season for load periods 6 and 19. In general, estimation based on moving window performs better than that based on expanding window if there is dynamic evolution in the specification. If we consider the different sets of regressors, it doesn't appear that one specification is better than the others.
For each season, the forecasts obtained with the selected models are combined with the methods described in section 6.1. The next section contains comments on the combination results.

6.3 Results of the combinations

Tables 6.17-6.20 contains relative measure values for the combinations carried out at a seasonal level. Results highlight significant improvements respect to the combinations with all the models, in fact the combination models outperforms or matches also in each season the best model among the single ones. In particular, over the whole forecasting period, improvements range from 4% to 21%.
To better understand the increasing in forecasting accuracy obtained by combination models, we carried out for the whole forecasting period the procedure for MCS considering all the singular models and the combination models. We used loss functions based on squared errors, squared percentage errors, absolute errors and absolute percentage errors. Rankings and MCS p-values shown in tables 6.21-6.24 underline that combinations are superior respect to the singular models. In particu-

lar, combinations with the Bates and Granger weights estimates on a rolling window of 10 observations results to be the best models in terms of forecasting accuracy for load periods 19, 28 and 38. In load period 6 best results are obtained with the equally weighted combination or the BG expanding model. The reason is that load period 6 is a base hour with a more stable behaviour than the other periods. So, reasonably the models involved in the combination have similar forecast error variances. The other load periods are characterized by high variability and unexpected jumps. This produces instability in the variance of the forecast errors, so the little time-variation in the weights of the BG10 model becomes important. This is in line with comments in section 6.1.
Only for load period 38, combination models are not superior to the singular model with the best MSE value, but this result is not significative as pointed out by the test of Diebold and Mariano (table 6.25).
In the light of our findings, the use of forecast combinations can be a good solution also for the electricity price forecasting issue. However, because of the strong link between singular model predictive accuracy and forecasting period, we recommend to develop these methods at a seasonal level.

Table 6.17: *Results of the combinations considering the subsets of models. Load period 6.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.030	**0.977**	**0.965**	1.018
April-June	**0.872**	**0.814**	**0.992**	**0.968**
July-September	**0.897**	**0.935**	**0.942**	**0.957**
Whole Period	**0.929**	**0.792**	**0.886**	**0.883**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.037	1.016	**0.981**	1.041
April-June	**0.811**	**0.800**	**0.939**	**0.932**
July-September	**0.925**	**0.959**	**0.957**	**0.971**
Whole Period	**0.931**	**0.813**	**0.887**	**0.887**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.043	**0.992**	**0.970**	1.025
April-June	**0.839**	**0.821**	**0.947**	**0.938**
July-September	**0.914**	**0.951**	**0.947**	**0.963**
Whole Period	**0.937**	**0.804**	**0.881**	**0.880**

Table 6.18: *Results of the combinations considering the subsets of models. Load period 19.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	**0.876**	1.025	**0.957**	1.041
April-June	**0.962**	**0.764**	**0.826**	**0.829**
July-September	1.067	1.038	**0.999**	1.013
Whole Period	**0.818**	**0.817**	**0.904**	**0.914**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	**0.906**	**0.967**	**0.947**	1.011
April-June	**0.887**	**0.738**	**0.816**	**0.822**
July-September	1.070	1.005	**0.997**	**0.997**
Whole Period	**0.826**	**0.783**	**0.897**	**0.896**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	**0.960**	**0.946**	**0.947**	1.001
April-June	**0.898**	**0.766**	**0.834**	**0.845**
July-September	1.070	1.003	**0.997**	**0.998**
Whole Period	**0.844**	**0.779**	**0.900**	**0.898**

Table 6.19: *Results of the combinations considering the subsets of models. Load period 28.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.016	**0.989**	**0.992**	**0.991**
April-June	1.090	**0.982**	1.029	1.000
July-September	**0.816**	1.026	**0.895**	1.037
Whole Period	**0.854**	**0.937**	**0.887**	**0.961**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.023	**0.997**	**0.994**	**0.993**
April-June	1.099	1.005	1.036	1.013
July-September	**0.783**	**0.976**	**0.868**	1.012
Whole Period	**0.849**	**0.927**	**0.881**	**0.956**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.018	**0.993**	**0.993**	**0.992**
April-June	1.091	**0.992**	1.030	1.004
July-September	**0.796**	1.019	**0.878**	1.029
Whole Period	**0.849**	**0.938**	0.882	**0.960**

Table 6.20: *Results of the combinations considering the subsets of models. Load period 38.*

	mean			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.236	**0.872**	**0.976**	**0.958**
April-June	**0.986**	**0.922**	**0.970**	**0.980**
July-September	1.251	**0.954**	1.042	**0.982**
Whole Period	1.117	**0.879**	**0.932**	**0.908**

	BG10			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.256	**0.870**	**0.981**	**0.957**
April-June	**0.981**	**0.925**	**0.966**	**0.979**
July-September	1.250	**0.909**	1.022	**0.952**
Whole Period	1.132	**0.864**	**0.931**	**0.898**

	BGExp			
Forecasting Period	R_{MSE}	R_{MSPE}	R_{MAE}	R_{MAPE}
January-March	1.240	**0.901**	**0.982**	**0.969**
April-June	**0.985**	**0.915**	**0.959**	**0.968**
July-September	1.262	**0.971**	1.046	**0.987**
Whole Period	1.121	**0.898**	**0.936**	**0.912**

Table 6.21: *MCS results for the whole forecasting period with the combinations. Load period 6.*

Squared Error Loss					Squared Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS1RS9	0.169	0.169	0.072	0.072	MRS3RS6	0.002	0.002	0.007	0.007
MRS1RS6	0.163	0.169	0.073	0.073	MRS1RS9	0.002	0.002	0.007	0.007
LR1	0.158	0.169	0.074	0.074	LR3	0.002	0.002	0.007	0.007
MRS1	0.154	0.169	0.075	0.075	MRS1RS3	0.002	0.002	0.005	0.007
MRS1RS3	0.142	0.169	0.074	0.075	MRS3RS9	0.008	0.008	0.006	0.007
MRS3RS6	0.135	0.169	0.074	0.075	MRS1RS6	0.008	0.008	0.005	0.007
LR3	0.127	0.169	0.074	0.075	LR1	0.008	0.008	0.005	0.007
MRS2RS9	0.116	0.169	0.073	0.075	TVR3	0.007	0.008	0.005	0.007
MRS3RS9	0.113	0.169	0.074	0.075	MRS3RS3	0.008	0.008	0.005	0.007
MRS3	0.103	0.169	0.072	0.075	TVR1	0.007	0.008	0.006	0.007
MRS3RS3	0.102	0.169	0.071	0.075	MRS1	0.005	0.008	0.006	0.007
MRS2RS6	0.094	0.169	0.074	0.075	MRS2RS3	0.004	0.008	0.004	0.007
MRS2	0.087	0.169	0.078	0.075	MRS2RS9	0.004	0.008	0.003	0.007
MRS2RS3	0.081	0.169	0.078	0.078	MRS3	0.006	0.008	0.005	0.007
TVR3	0.079	0.169	0.072	0.078	ARMAX	0.004	0.008	0.001	0.007
TVR1	0.066	0.169	0.099	0.099	LR2	0.007	0.008	0.001	0.007
LR2	0.301	0.301	0.267	0.267	MRS2	0.024	0.024	0.004	0.007
TVR2	0.682	0.682	0.559	0.559	MRS2RS6	0.019	0.024	0.002	0.007
ARMAX	0.659	0.682	0.588	0.588	TVR2	0.017	0.024	0.010	0.010
BGExp	0.580	0.682	0.526	0.588	**BG10**	0.313	0.313	0.221	0.221
BG10	0.836	0.836	0.836	0.836	**BGExp**	0.191	0.313	0.191	0.221
mean	—	1.000	—	1.000	**mean**	—	1.000	—	1.000

Absolute Error Loss					Absolute Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS1RS3	0.003	0.003	0.010	0.010	MRS3RS6	0.000	0.000	0.000	0.000
MRS3RS6	0.003	0.003	0.009	0.010	MRS1RS3	0.000	0.000	0.000	0.000
MRS1RS9	0.003	0.003	0.009	0.010	LR3	0.000	0.000	0.001	0.001
LR3	0.003	0.003	0.010	0.010	MRS3RS9	0.000	0.000	0.001	0.001
LR1	0.003	0.003	0.009	0.010	MRS1RS9	0.000	0.000	0.002	0.002
MRS1RS6	0.002	0.003	0.009	0.010	MRS1RS6	0.000	0.000	0.001	0.002
MRS3RS9	0.002	0.003	0.012	0.012	LR1	0.000	0.000	0.002	0.002
MRS1	0.002	0.003	0.011	0.012	MRS3RS3	0.000	0.000	0.001	0.002
MRS3RS3	0.002	0.003	0.010	0.012	TVR3	0.000	0.000	0.001	0.002
MRS2RS9	0.002	0.003	0.009	0.012	ARMAX	0.000	0.000	0.002	0.002
TVR3	0.008	0.008	0.011	0.012	MRS1	0.001	0.001	0.002	0.002
ARMAX	0.007	0.008	0.011	0.012	TVR1	0.001	0.001	0.001	0.002
LR2	0.006	0.008	0.010	0.012	MRS2RS9	0.003	0.003	0.003	0.003
MRS2RS6	0.006	0.008	0.006	0.012	LR2	0.022	0.022	0.004	0.004
MRS2	0.006	0.008	0.011	0.012	MRS2RS6	0.041	0.041	0.004	0.004
TVR1	0.005	0.008	0.006	0.012	MRS2RS3	0.049	0.049	0.004	0.004
MRS3	0.089	0.089	0.024	0.024	TVR2	0.037	0.049	0.009	0.009
MRS2RS3	0.071	0.089	0.021	0.024	MRS3	0.046	0.049	0.021	0.021
TVR2	0.052	0.089	0.039	0.039	MRS2	0.034	0.049	0.039	0.039
BG10	0.495	0.495	0.636	0.636	**BG10**	0.394	0.394	0.545	0.545
mean	0.687	0.687	0.687	0.687	**mean**	0.798	0.798	0.798	0.798
BGExp	—	1.000	—	1.000	**BGExp**	—	1.000	—	1.000

Table 6.22: *MCS results for the whole forecasting period with the combinations. Load period 19.*

Squared Error Loss					Squared Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR3	0.014	0.014	0.155	0.155	MRS3RS6	0.020	0.020	0.046	0.046
MRS3RS6	0.012	0.014	0.165	0.165	LR3	0.019	0.020	0.049	0.049
MRS3RS9	0.011	0.014	0.172	0.172	MRS3RS3	0.015	0.020	0.047	0.049
MRS3	0.011	0.014	0.186	0.186	MRS3RS9	0.015	0.020	0.041	0.049
MRS2RS6	0.009	0.014	0.190	0.190	MRS2RS6	0.014	0.020	0.036	0.049
TVR3	0.009	0.014	0.191	0.191	MRS2RS3	0.014	0.020	0.049	0.049
MRS1RS9	0.009	0.014	0.174	0.191	MRS1RS3	0.012	0.020	0.051	0.051
TVR1	0.007	0.014	0.156	0.191	MRS3	0.012	0.020	0.068	0.068
LR2	0.007	0.014	0.152	0.191	MRS2RS9	0.011	0.020	0.062	0.068
ARMAX	0.006	0.014	0.144	0.191	MRS1RS9	0.072	0.072	0.098	0.098
MRS1	0.006	0.014	0.125	0.191	TVR3	0.067	0.072	0.104	0.104
MRS2RS3	0.005	0.014	0.106	0.191	MRS1RS6	0.064	0.072	0.102	0.104
MRS2RS9	0.049	0.049	0.194	0.194	MRS2	0.052	0.072	0.107	0.107
TVR2	0.042	0.049	0.164	0.194	ARMAX	0.327	0.327	0.143	0.143
LR1	0.039	0.049	0.136	0.194	MRS1	0.305	0.327	0.126	0.143
MRS2	0.032	0.049	0.106	0.194	LR2	0.279	0.327	0.109	0.143
MRS1RS3	0.030	0.049	0.075	0.194	TVR1	0.244	0.327	0.100	0.143
MRS3RS3	0.134	0.134	0.227	0.227	LR1	0.209	0.327	0.080	0.143
MRS1RS6	0.589	0.589	0.517	0.517	TVR2	0.208	0.327	0.119	0.143
BGExp	0.612	0.612	0.621	0.621	**mean**	0.278	0.327	0.257	0.257
BG10	0.711	0.711	0.711	0.711	**BG10**	0.597	0.597	0.597	0.597
mean	—	1.000	—	1.000	**BGExp**	—	1.000	—	1.000

Absolute Error Loss					Absolute Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR3	0.000	0.000	0.005	0.005	MRS3RS6	0.001	0.001	0.003	0.003
MRS3RS6	0.000	0.000	0.006	0.006	LR3	0.002	0.002	0.003	0.003
MRS2RS6	0.000	0.000	0.008	0.008	MRS2RS6	0.002	0.002	0.003	0.003
MRS3RS9	0.000	0.000	0.009	0.009	MRS3RS3	0.002	0.002	0.004	0.004
MRS2RS3	0.000	0.000	0.010	0.010	MRS3RS9	0.003	0.003	0.006	0.006
MRS2RS9	0.001	0.001	0.016	0.016	MRS2RS3	0.002	0.003	0.008	0.008
MRS3	0.007	0.007	0.023	0.023	MRS3	0.002	0.003	0.009	0.009
MRS1RS9	0.013	0.013	0.034	0.034	MRS2RS9	0.002	0.003	0.013	0.013
MRS3RS3	0.011	0.013	0.032	0.034	MRS1RS3	0.061	0.061	0.024	0.024
MRS1RS3	0.062	0.062	0.038	0.038	MRS1RS9	0.054	0.061	0.037	0.037
TVR3	0.054	0.062	0.045	0.045	MRS1RS6	0.050	0.061	0.044	0.044
MRS2	0.047	0.062	0.054	0.054	TVR3	0.042	0.061	0.053	0.053
MRS1RS6	0.133	0.133	0.069	0.069	MRS2	0.037	0.061	0.062	0.062
LR1	0.193	0.193	0.090	0.090	LR1	0.177	0.177	0.085	0.085
LR2	0.176	0.193	0.078	0.090	MRS1	0.196	0.196	0.085	0.085
ARMAX	0.150	0.193	0.077	0.090	LR2	0.162	0.196	0.092	0.092
MRS1	0.122	0.193	0.081	0.090	TVR2	0.141	0.196	0.098	0.098
TVR2	0.086	0.193	0.097	0.097	ARMAX	0.191	0.196	0.125	0.125
TVR1	0.168	0.193	0.173	0.173	TVR1	0.417	0.417	0.224	0.224
mean	0.760	0.760	0.711	0.711	**mean**	0.313	0.417	0.293	0.293
BGExp	0.536	0.760	0.536	0.711	**BGExp**	0.697	0.697	0.697	0.697
BG10	—	1.000	—	1.000	**BG10**	—	1.000	—	1.000

Table 6.23: *MCS results for the whole forecasting period with the combinations. Load period 28.*

Squared Error Loss					Squared Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
MRS1RS3	0.099	0.099	0.118	0.118	LR3	0.009	0.009	0.010	0.010
LR3	0.099	0.099	0.122	0.122	MRS3RS6	0.009	0.009	0.014	0.014
MRS3	0.096	0.099	0.123	0.123	MRS2RS6	0.034	0.034	0.022	0.022
MRS1	0.096	0.099	0.126	0.126	MRS2RS3	0.031	0.034	0.027	0.027
LR2	0.093	0.099	0.119	0.126	MRS3RS9	0.027	0.034	0.034	0.034
MRS2	0.091	0.099	0.117	0.126	ARMAX	0.028	0.034	0.046	0.046
ARMAX	0.087	0.099	0.117	0.126	MRS3RS3	0.028	0.034	0.051	0.051
TVR3	0.085	0.099	0.114	0.126	MRS3	0.025	0.034	0.078	0.078
TVR1	0.085	0.099	0.111	0.126	MRS1RS3	0.023	0.034	0.094	0.094
MRS3RS6	0.085	0.099	0.101	0.126	MRS1RS6	0.020	0.034	0.093	0.094
MRS2RS3	0.078	0.099	0.083	0.126	MRS2RS9	0.018	0.034	0.117	0.117
MRS3RS9	0.062	0.099	0.100	0.126	LR1	0.016	0.034	0.108	0.117
LR1	0.056	0.099	0.084	0.126	MRS1RS9	0.021	0.034	0.165	0.165
MRS3RS3	0.054	0.099	0.066	0.126	TVR3	0.584	0.584	0.333	0.333
MRS2RS6	0.048	0.099	0.053	0.126	LR2	0.512	0.584	0.384	0.384
TVR2	0.060	0.099	0.080	0.126	MRS1	0.484	0.584	0.447	0.447
MRS1RS6	0.056	0.099	0.052	0.126	MRS2	0.570	0.584	0.494	0.494
MRS2RS9	0.045	0.099	0.039	0.126	TVR1	0.778	0.778	0.578	0.578
MRS1RS9	0.268	0.268	0.208	0.208	TVR2	0.702	0.778	0.606	0.606
mean	0.651	0.651	0.664	0.664	**BGExp**	0.786	0.786	0.748	0.748
BGExp	0.985	0.985	0.985	0.985	**mean**	0.553	0.786	0.553	0.748
BG10	——	1.000	——	1.000	**BG10**	——	1.000	——	1.000

Absolute Error Loss					Absolute Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR3	0.010	0.010	0.052	0.052	LR3	0.002	0.002	0.004	0.004
MRS3	0.010	0.010	0.062	0.062	MRS2RS3	0.002	0.002	0.005	0.005
MRS1RS3	0.010	0.010	0.061	0.062	MRS3RS6	0.001	0.002	0.006	0.006
MRS3RS9	0.010	0.010	0.049	0.062	MRS3RS9	0.001	0.002	0.008	0.008
MRS3RS6	0.010	0.010	0.048	0.062	MRS2RS6	0.001	0.002	0.009	0.009
ARMAX	0.009	0.010	0.048	0.062	MRS3RS3	0.014	0.014	0.017	0.017
MRS3RS3	0.009	0.010	0.043	0.062	MRS1RS3	0.045	0.045	0.028	0.028
MRS2RS3	0.008	0.010	0.041	0.062	MRS3	0.039	0.045	0.025	0.028
LR1	0.008	0.010	0.053	0.062	MRS1RS6	0.035	0.045	0.027	0.028
LR2	0.007	0.010	0.046	0.062	ARMAX	0.035	0.045	0.036	0.036
MRS1	0.006	0.010	0.039	0.062	MRS1RS9	0.032	0.045	0.049	0.049
MRS2RS6	0.006	0.010	0.033	0.062	LR1	0.060	0.060	0.074	0.074
MRS2	0.027	0.027	0.048	0.062	MRS2RS9	0.074	0.074	0.116	0.116
TVR3	0.024	0.027	0.040	0.062	MRS1	0.343	0.343	0.175	0.175
MRS1RS6	0.021	0.027	0.030	0.062	LR2	0.364	0.364	0.190	0.190
MRS1RS9	0.019	0.027	0.030	0.062	TVR3	0.444	0.444	0.227	0.227
MRS2RS9	0.014	0.027	0.039	0.062	MRS2	0.387	0.444	0.322	0.322
TVR1	0.449	0.449	0.208	0.208	TVR2	0.446	0.446	0.447	0.447
TVR2	0.453	0.453	0.318	0.318	TVR1	0.800	0.800	0.676	0.676
mean	0.566	0.566	0.499	0.499	**mean**	0.767	0.800	0.737	0.737
BGExp	0.669	0.669	0.669	0.669	**BGExp**	0.605	0.800	0.605	0.737
BG10	——	1.000	——	1.000	**BG10**	——	1.000	——	1.000

Table 6.24: *MCS results for the whole forecasting period with the combinations. Load period 38.*

Squared Error Loss					Squared Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR3	0.164	0.164	0.070	0.070	MRS3RS6	0.015	0.015	0.041	0.041
LR2	0.193	0.193	0.071	0.071	LR2	0.013	0.015	0.047	0.047
MRS2RS3	0.198	0.198	0.081	0.081	LR3	0.012	0.015	0.047	0.047
MRS1RS3	0.280	0.280	0.082	0.082	MRS3RS3	0.012	0.015	0.054	0.054
MRS3RS6	0.280	0.280	0.095	0.095	MRS1RS3	0.011	0.015	0.054	0.054
MRS2	0.260	0.280	0.093	0.095	LR1	0.022	0.022	0.076	0.076
LR1	0.250	0.280	0.094	0.095	MRS1RS9	0.021	0.022	0.072	0.076
ARMAX	0.249	0.280	0.107	0.107	MRS2RS6	0.019	0.022	0.072	0.076
MRS2RS6	0.239	0.280	0.115	0.115	MRS3	0.017	0.022	0.063	0.076
MRS2RS9	0.211	0.280	0.119	0.119	MRS2RS3	0.016	0.022	0.064	0.076
MRS1RS9	0.202	0.280	0.114	0.119	MRS2	0.155	0.155	0.104	0.104
MRS3RS3	0.193	0.280	0.111	0.119	MRS2RS9	0.149	0.155	0.096	0.104
TVR2	0.167	0.280	0.113	0.119	MRS3RS9	0.136	0.155	0.089	0.104
MRS3	0.216	0.280	0.147	0.147	ARMAX	0.215	0.215	0.112	0.112
MRS1	0.208	0.280	0.154	0.154	MRS1RS6	0.287	0.287	0.140	0.140
TVR1	0.203	0.280	0.192	0.192	MRS1	0.350	0.350	0.145	0.145
MRS1RS6	0.432	0.432	0.386	0.386	TVR3	0.288	0.350	0.151	0.151
BG10	0.602	0.602	0.526	0.526	TVR2	0.300	0.350	0.170	0.170
TVR3	0.484	0.602	0.505	0.526	TVR1	0.266	0.350	0.158	0.170
BGExp	0.398	0.602	0.463	0.526	**BGExp**	0.258	0.350	0.211	0.211
mean	0.551	0.602	0.551	0.551	**mean**	0.251	0.350	0.251	0.251
MRS3RS9	—	1.000	—	1.000	**BG10**	—	1.000	—	1.000

Absolute Error Loss					Absolute Percentage Error Loss				
Model	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**	**Model**	**p.r**	**MCS.r**	**p.sq**	**MCS.sq**
LR3	0.020	0.020	0.009	0.009	MRS3RS6	0.002	0.002	0.007	0.007
MRS3RS6	0.028	0.028	0.013	0.013	MRS1RS3	0.001	0.002	0.013	0.013
MRS1RS3	0.026	0.028	0.015	0.015	LR3	0.016	0.016	0.018	0.018
LR1	0.025	0.028	0.018	0.018	LR2	0.044	0.044	0.026	0.026
LR2	0.023	0.028	0.019	0.019	LR1	0.042	0.044	0.027	0.027
MRS2RS3	0.022	0.028	0.023	0.023	MRS3RS3	0.037	0.044	0.028	0.028
MRS3	0.020	0.028	0.026	0.026	MRS3	0.034	0.044	0.031	0.031
MRS2	0.018	0.028	0.024	0.026	MRS1RS9	0.034	0.044	0.035	0.035
MRS3RS3	0.017	0.028	0.021	0.026	MRS3RS9	0.030	0.044	0.036	0.036
MRS1RS9	0.042	0.042	0.029	0.029	MRS2RS3	0.027	0.044	0.054	0.054
MRS2RS6	0.040	0.042	0.024	0.029	MRS2RS6	0.024	0.044	0.065	0.065
MRS2RS9	0.044	0.044	0.034	0.034	MRS1RS6	0.022	0.044	0.073	0.073
MRS3RS9	0.042	0.044	0.033	0.034	MRS2RS9	0.131	0.131	0.092	0.092
MRS1RS6	0.036	0.044	0.056	0.056	MRS2	0.118	0.131	0.084	0.092
MRS1	0.196	0.196	0.091	0.091	ARMAX	0.108	0.131	0.063	0.092
TVR2	0.169	0.196	0.106	0.106	TVR2	0.085	0.131	0.063	0.092
TVR1	0.140	0.196	0.153	0.153	MRS1	0.108	0.131	0.089	0.092
ARMAX	0.424	0.424	0.291	0.291	TVR1	0.081	0.131	0.083	0.092
TVR3	0.522	0.522	0.382	0.382	TVR3	0.306	0.306	0.169	0.169
BGExp	0.552	0.552	0.624	0.624	**BGExp**	0.378	0.378	0.358	0.358
mean	0.829	0.829	0.829	0.829	**mean**	0.294	0.378	0.294	0.358
BG10	—	1.000	—	1.000	**BG10**	—	1.000	—	1.000

Table 6.25: *P-values obtained with the test of Diebold and Mariano with four loss functions between the B-G combination (*$l = 10$*) and the best single model on the whole forecasting period. Load periods 6, 19, 28 and 38.*

	SE	SPE	AE	APE
Load period 6	0.518	0.012	0.018	0.036
Load period 19	0.021	0.045	0.016	0.027
Load period 28	0.031	0.232	0.204	0.152
Load period 38	0.608	0.036	0.079	0.027

Chapter 7

Conclusions and further research

During the last twenty years electricity price modeling and forecasting has become a heated topic of research. Many techniques have been proposed from the use of linear models to the developing of more complex nonlinear models with jumps and time-varying parameters. Despite this seemingly large number of models and related empirical analysis on a wide range of power markets, results are mixing and there is no single model that provides convincing superior performance in forecasting spot prices.

In our work a systematic comparative study of different forecasting techniques is proposed. The scope is to evaluate the relative forecasting performance and to understand if there is a particular class of models that outperformes all the others. Included in the study are linear ARMAX models, different specifications of multiple regression models, non linear Markov switching regression models and time-varying parameter regression models. One-day ahead forecasts are obtained for each model and evaluated according to prediction error statistics and the Diebold and Mariano test for equal predictive accuracy. Forecasting results in chapter 5 highlight that no model globally outperforms the others. Different forecasting models capture different features of spot price dynamics, so no model produce good forecasts in all the situations. This is confirmed also by the Diebold and Mariano testing procedure applied to the models in section 5.6.

So, we propose a forecasting approach based on the combination of forecasts (chapter 6). The use of combination models is novel in the electricity price forecasting context.

Methods from Bates & Granger (1969) and the equally weighted combination are considered and applied to all the models. Results are not very appealing: too much variability is introduced in the combinations because of models that perform very poorly during particular seasons and/or for particular hours. In literature it has been established that it is often advantageous to discard the models with the worst performance. Considering each season separately, subsets of models are choosen applying the model confidence set procedure (Hansen et al., 2005) and the forecasts encompassing method (Fair & Shiller, 1990). In this way, poorly-performing models and models that do not contain useful information about price dynamics are discarded. Findings highlight that each load period is characterized by particular dynamics that change with the seasons and this reflect to the models' forecasting

performance. For this reason we propose a combination applied at a seasonal level. Results highlight the usefulness of the procedure, showing that combining forecasts at a seasonal level have the potential to produce predictions of superior or equal accuracy relative to the best individual forecasts.
In the light of our findings, forecast combinations can work well also in the electricity price forecasting context if the forecasting period is considered. There is a strong link between singular model predictive accuracy and forecasting period. Our recommendation is to develop these methods at a seasonal level.

Aspects of further development are not absent. The articles of Mount et al. (2006) and Kanamura & Ōhashi (2008) opened a new issue of research about regime switching models with time varying parameters. It would be interesting to develop this kind of models with both the parameters of the regimes and the transition probabilities with a time-varying structure, and to applied them in the forecasting context.
In the time-varying regression model developed in our work, we specified the parameter process as a random walk. Other specifications are also possible: the parameters can be specified as depending on an autoregressive process with some important fundamentals as exogenous variables. Moreover, since the models are estimated separately each hour of the day, it does not embed the possibility of intra-daily effects among different hourly prices, especially in a market like in the UK, where electricity is traded each period at a time.
Future research is also addressed to the study of different specifications of combination models. Yang (2004) proposed a combination for adaptation method called Aggregated Forecast Through Exponential Re-weighting (AFTER). Sánchez (2008) pointed out that this method is not adaptive in a nonstationary situation and introduced a forgetting factor in the estimation of the weights of the combination. It would be interesting to study the behaviour of this combination model applied to electricity price data.

Bibliography

AIOLFI, M. & FAVERO, C. A. (2005). Model uncertainty, thick modeling and the predictability of stock returns. *Journal of Forecasting* **24**, 233–254.

ALTAVILLA, C. & DE GRAUWE, P. (2006). Forecasting and combining competing models of exchange rate determination. CESifo Working Paper Series No.5, CESifo GmbH.

ARCINIEGAS, A. I. & ARCINIEGAS RUEDA, I. E. (2008). Forecasting short-term power prices in the ontario electricity market (oem) with a fuzzy logic based inference system. *Utilities Policy* **16**, 39–48.

BATES, J. M. & GRANGER, C. W. J. (1969). The combination of forecasts. *OR* **20**, 451–468.

BECKER, R. & CLEMENTS, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting* **24**, 122–133.

BIERBRAUER, M., TRÜCK, S. & WERON, R. (2004). Modeling electricity prices with regime switching models. *Lecture Notes in Computer Science* **3039**, 859–867.

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

BOX, J. E. P. & JENKINS, G. M. (1976). *Time Series Analysis, Forecasting and Control. Revised Edition.* Holden Day, San Francisco.

BROWN, R. L., DURBIN, J. & EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)* **37**, 149–192.

BUNN, D. W. (1975). A bayesian approach to the linear combination of forecasts. *Operational Research Quarterly (1970-1977)* **26**, 325–329.

BUNN, D. W. (1977). A comparative evaluation of the outperformance and minimum variance procedures for the linear synthesis of forecasts. *Operational Research Quarterly (1970-1977)* **28**, 653–662.

BUNN, D. W. (1978). A simplification of the matrix beta distribution for combining estimators. *Journal of the Operational Research Society* **29**, 1013–1016.

Bunn, D. W. (1985). Statistical efficiency on the linear combination of forecasts. *International Journal of Forecasting* **1**, 151–163.

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

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

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

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

Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. S. & Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research* **77**, 1297–1304.

Chen, Z. & Yang, Y. (2007). Time series models for forecasting: Testing or combining? *Studies in Nonlinear Dynamics & Econometrics* **11**, 1385–1385.

Clark, T. E. & McCracken, M. W. (2007). Combining forecasts from nested models. Finance and Economics Discussion Series 2007-43, Board of Governors of the Federal Reserve System (U.S.).

Clark, T. E. & West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics* **135**, 155–186.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* **5**, 559–583.

Clemen, R. T. & Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics* **4**, 39–46.

Clemens, M. P. & Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge University Press.

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

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

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

Coutts, J. A., Roberts, J. & Mills, T. C. (1997). Parameter stability in the market model: tests and time varying parameter estimation with uk data. *The Statistician* **46**, 57–70.

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

de Jong, C. (2006). The nature of power spikes: A regime-switch approach. *Studies in Nonlinear Dynamics & Econometrics* **10**, Article 3.

de Jong, C. & Huisman, R. (2002). Option formulas for mean-reverting power prices with spikes. Research paper, Erasmus Research Institute of Management (ERIM), Erasmus University Rotterdam.

de Menezes, L. M., Bunn, D. W. & Taylor, L. W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research* **120**, 190–204.

Deng, S. (2000). Stochastic models of energy commodity prices and their applications: mean-reversion with jumps and spikes. *Power Working Paper* 73, University of California Energy Institute.

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

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

Dunis, C., Laws, J. & Chauvin, S. (2001). The use of market data and model combinations to improve forecast accuracy. In *Development in Forecasts Combination and Portfolio Choice*, Dunis, Timmermann & Moody, eds. Wiley, Oxford.

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

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

Fair, R. C. & Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review* **80**, 375–389.

Fezzi, C. (2007a). *Econometric models for the analysis of electricity markets*. Phd thesis in statistical methodology for scientific research, XIX cycle, Department of Statistics, University of Bologna.

Fezzi, C. (2007b). Forecasting electricity prices: the importance of being unstable. 4^{th} International conference: The European Electricity Market. EEM-07. Challenge of the Unification.

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

Garcia, R. C., Contreras, J., van Akkeren, M. & 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. & 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**, Article 1.

Granger, C. W. J. & Jeon, Y. (2004). Thick modeling. *Economic Modelling* **21**, 323–343.

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

Gupta, S. & Wilton, P. C. (1987). Combination of forecasts: An extension. *Management Science* **33**, 356–372.

Guthrie, G. & Videbeck, S. (2002). Hight frequency electricity spot price dynamics: an intra-day approach. *Working paper*, New Zeland Institute for the study of competition and regulation.

Haldrup, N. & Nielsen, M. O. (2006a). 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.

Hansen, P. R., Lunde, A. & Nason, L. M. (2003). Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics* **65**, 839–861. Supplement.

Hansen, P. R., Lunde, A. & Nason, L. M. (2005). Model confidence sets for forecasting models. Working Paper 2005-7, Federal Reserve Bank of Atlanta.

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

Huisman, R. & Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics* **25**, 425–34.

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**, 679–688.

Kanamura, T. & Ōhashi, K. (2008). On transition probabilities of regime switching in electricity prices. *Energy Economics* **30**, 1158–1172.

KARAKATSANI, N. & BUNN, D. W. (2004). Modelling stochastic volatility in high-frequency spot electricity prices. *Working paper*, London Business School.

KARAKATSANI, N. & BUNN, D. W. (2008a). Intra-day and regime-switching dynamics in electricity price formation. *Energy Economics* **30**, 1776–1797.

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

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

KNITTEL, C. R. & ROBERTS, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics* **27**, 791–817.

KOSATER, P. & 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. & SHIN, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **54**, 159–178.

LUCIA, J. J. & SCHWARTZ, E. S. (2002). Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of Derivatives Research* **5**, 5–50.

MAKRIDAKIS, S. & WINKLER, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science* **29**, 987–996.

MARCELLINO, M. (2004). Forecast pooling for short time series of macroeconomic variables. *Oxford Bulletin of Economic and Statistics* **6**, 91–112.

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

MOUNT, T. D., NING, Y. & CAI, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics* **28**, 62–80.

NEWBOLD, P. & GRANGER, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts (with discussion). *Journal of the Royal Statistical Society, Series A* **137**, 131–164.

NG, S. & 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. & PLAZAS, M. A. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems* **17**, 342–348.

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

PHILLIPS, P. C. B. & PERRON, P. (1988). Testing for unit roots in time series regression. *Biometrika* **75**, 335–346.

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

PINO, R., PARRENO, J., GOMEZ, A. & PRIORE, P. (2008). Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. *Engineering Applications of Artificial Intelligence* **21**, 53–62.

RAMANATHAN, R., ENGLE, R., GRANGER, C. W. J., VAHID-ARAGHI, F. & BRACE, C. (1997). Short-run forecasting of electricity loads and peaks. *International Journal of Forecasting* **13**, 161–174.

REID, D. J. (1968). Combining three estimates of gross domestic product. *Economica* **35**, 431–444.

REID, D. J. (1969). *A comparative study of time series prediction techniques on economic data*. Ph.D. thesis, University of Nottingham, Nottingham.

RIEDEL, S. & GABRYS, B. (2005). Evolving multilevel forecast combination models - an experimental study. In *Proceedings of NiSIS'2005 Symposium*. Albufeira, Portugal.

SAID, S. E. & DICKEY, D. (1984). Testing for unit roots in autoregressive moving-average models with unknown order. *Biometrika* **71**, 599–607.

SÁNCHEZ, I. (2008). Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting* **24**, 679–693.

SERATI, M., MANERA, M. & PLOTEGHER, M. (2007). Modelling electricity prices: from the state of the art to a draft of a new proposal. LIUC Papers in Economics 210, Cattaneo University (LIUC).

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

SMITH, J. & WALLIS, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics* **71**, 331–355.

STOCK, J. H. & WATSON, M. W. (2001). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In *Festschrift in Honour of Clive Granger*, R. F. Engle & H. White, eds. Cambridge University Press, Cambridge, pp. 1–44.

STOCK, J. H. & WATSON, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**, 405–430.

SWIDER, D. J. & WEBER, C. (2007). Extended ARMA models for estimating price developments on day-ahead electricity markets. *Electric Power Research* **77**, 583–593.

TIMMERMANN, A. G. (2006). Forecast combinations. In *Handbook of Economic Forecasting*, G. C. W. J. Elliot, G & A. Timmermann, eds., vol. 1. Amsterdam: Elsevier, pp. 135–196.

VAHIDINASAB, V., JADID, S. & KAZEMI, A. (2008). Day-ahead price forecasting in restructured power systems using artificial neural networks. *Electric Power Systems Research* **78**, 1332–1342.

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

WERON, R. (2005). Heavy tails and electricity prices. Invited paper presented at the Deutsche Bundesbank's 2005 Annual Fall Conference (Eltville, 10-12 November 2005).

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

WERON, R., BIERBRAUER, M. & TRÜCK, S. (2004a). Modeling electricity prices: jump diffusion and regime switching. *Physica A* **336**, 39–48.

WERON, R. & MISIOREK, A. (2005). Forecasting spot electricity prices with time series models. In *Proceedings Volume*. International Conference "The European Electricity Market EEM-05", May 10-12, 2005, Lodz, Poland.

WERON, R. & MISIOREK, A. (2006). Short-term electricity price forecasting with time series models: A review and evaluation. In *Complex Electricity Markets*, W. Mielczarski, ed., The European Power Industry Supply, chap. 10. Lodz, pp. 231–254.

WERON, R. & PRZYBYŁOWICZ, B. (2000). Hurst analysis of electricity price dynamics. *Physica A* **283**, 462–468.

WERON, R., SIMONSEN, I. & WILMAN, P. (2004b). Modeling highly volatile and seasonal markets: evidence from the nord pool electricity market. *Applications of Econophysics* , 182–191.

WEST, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* **64**, 1067–1084.

Winkler, R. L. & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society Series A* **146**, 150–157.

Yang, Y. (2004). Combining forecasting procedures: some theoretical results. *Econometric Theory* **20**, 176–222.

Fany Nan - Curriculum Vitae

Personal information

Address: Department of Statistics, University of Padova,
Via Cesare Battisti 241-243 - 35121 Padova (Italy)
E-Mail: fany.nan@stat.unipd.it

Studies

University

From February 2009 until present:
research fellow at the Department of Statistical Sciences, University of Padova, Italy.

From January 2006 until present (expected completion by July 2009):
enrolled at the Doctoral School in Statistical Sciences at the Department of Statistical Sciences, University of Padova, Italy.

16th March 2005:
Laurea Degree in Mathematics (4 years), at the Faculty of Mathematical, Physical and Natural Sciences, University of Udine, Italy. Thesis on "Goodness-of-fit tests", advisor: Prof. Luigi Pace.

Secondary school

July 1999:
Secondary School Diploma in Scientific Education at Liceo Statale "Ettore Majorana", San Vito al Tagliamento (Pordenone), Italy.

Study abroad periods

From June 2007 to March 2008:
I have worked on my thesis under the supervision of Prof. Derek Bunn, at the *Department of Management Science and Operations, Decision Sciences Group*, **London Business School, London, UK**.

Conferences and Presentations

---

30 May 2007:
Second Energy Markets Group Workshop, London Business School, Londra. Organizer: Prof. Derek Bunn.

23-25 October 2007:
London Energy Forum: Market Risk, Policy Uncertainty and Energy Investment. The Royal Society, Londra.

15-16 May 2008:
Spring School in Finance: Crash Course on Energy Markets. Department of Mathematics, Bologna. Speakers: Prof. Fred Espen Benth, Prof. Helyette Geman.

Professional experience: Teaching assistantship

September 2007 (12 hours):
tutoring and workshop in PC-Lab for the course Statistics Foundation, Master in Finance, London Business School. People in charge of the course: Prof. Derek Bunn and Paul Ellis.

January 2008 (12 hours):
tutoring and workshop in PC-Lab for the course Statistics Foundation, Master in Finance, London Business School. People in charge of the course: Prof. Derek Bunn and Paul Ellis.

1st March 2008 (2 hours)::
tutoring for the course Managerial Statistics, Executive MBA, London Business School. People in charge of the course: Prof. Catalina Stefanescu and Paul Ellis.