

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE

CICLO XXVIII

Modeling and Forecasting Electricity Market Variables

Direttore della Scuola: Prof. Monica Chiogna

Supervisore: Prof. Francesco Lisi

Dottorando: ISMAIL SHAH

Data 31/01/2016

Abstract

In deregulated electricity markets, accurate modeling and forecasting of different variables, e.g. demand, prices, production etc. have obtained increasing importance in recent years. As in most electricity markets, the daily demand and prices are determined the day before the physical delivery by means of (semi-) hourly concurrent auctions, accurate forecasts are necessary for the efficient management of power systems. However, it is well known that electricity (demand/price) data exhibit some specific features, among which, daily, weekly and annual periodic patterns as well as non-constant mean and variance, jumps and dependency on calendar effects. Modeling and forecasting, thus, is a challenging task. This thesis tackles these two issues, and to do this, two approaches are followed.

In the first case, we address the issue of modeling and out-of-sample forecasting electricity demand and price time series. For this purpose, an additive component model was considered that includes some deterministic and a stochastic residual components. The deterministic components include a long-term dynamics, annual and weekly periodicities and calendar effects. The first three components were estimated using splines while the calendar effects were modeled using dummy variables. The residual component is instead treated as stochastic and different univariate and multivariate models have been considered with increasing level of complexity. In both cases, linear parametric and nonlinear nonparametric models, as well as functional based models, have been estimated and compared in a one day-ahead out-of-sample forecast framework.

The class of univariate models includes parametric autoregressive models (AR), nonparametric and nonlinear regression models based on splines (NPAR) and scalar-response functional models, that in turns can be formulated parametrically (FAR) or non parametrically (NPFAR). The multivariate models are vector autoregressive models (VAR) and functional-response, parametric (FFAR) and nonparametric (NPFAR), models. For this issue, five different electricity markets, namely, British electricity market (APX Power UK), Nord

Pool electricity market (NP), Italian electricity market (IPEX), Pennsylvania-New Jersey-Maryland electricity market (PJM) and Portuguese electricity market (OMIE(Po)) were considered for the period 2009 to 2014. The first five years were used for model estimation while the year 2014 was left for one-day-ahead forecasts. Predictive performances are first evaluated by means of descriptive indicators and then through a test to assess the significance of the differences. The analyses suggest that the multivariate approach leads to better results than the univariate one and that, within the multivariate framework, functional models are the most accurate, with VAR being a competitive model in some cases. The results also lead to another important finding concerning to the performance of parametric and nonparametric approach that showed strong linkage with underlying process. Finally the obtained results were compared with other works in the literature that suggest our forecasting errors are smaller compared with the state-of-art prediction techniques used in the literature.

In the second part of this thesis the issue of electricity price forecasting is revisited following a completely different approach. The main idea of this approach is that of modeling the daily supply and demand curves, predicting them and finding the intersection of the predicted curves in order to find the predicted market clearing price and volume. In this approach, the raw bids/offers data for demand and supply, corresponding to each (half-) hour is first aggregated in a specific order. The functional approach converts the resulted piece wise curves into smooth functions. For this issue, parametric functional model (FFAR) and the nonlinear nonparametric counterpart (NPFFAR) were considered. As benchmark, an ARIMA model was fitted to the scalar time series corresponding to the market clearing prices obtained from the crossing points of supply and demand curves. Data from Italian electricity market were used for this issue and the results are summarized by different descriptive indicators. As in the first case, results show superior forecasting performance of our functional approach compare to ARIMA. Among different models, the nonparametric functional model produces better results compared to parametric models.

Apart from the improvement in forecasting accuracy, it is important to stress that this approach can be used for optimizing bidding strategies. As forecasting the whole curves gives deep insight into the market, our analysis showed that this strategy can significantly improve bidding strategies and maximize traders profit.

Abstract (Italian)

Nell'ambito dei mercati elettrici liberalizzati, negli ultimi anni l'interesse verso una buona modellazione e un'accurata previsione di variabili da essi provenienti, ad es. domanda, prezzi, produzione etc., è andato via via crescendo. Ciò anche perché in molti mercati elettrici, i prezzi e i volumi giornalieri vengono determinati mediante un sistema di aste (semi-)orarie che ha luogo il giorno precedente a quello della consegna fisica; una previsione accurata permette quindi un'efficiente gestione del sistema elettrico.

La modellazione e la previsione di queste variabili, tuttavia, è resa difficile dal fatto che le serie storiche di domanda e prezzi, sono caratterizzate dalla presenza di vari tipi di periodicità, annuale, settimanale e giornaliera, da una media e una varianza che non sono costanti nel tempo, da picchi improvvisi e dalla dipendenza da diversi effetti di calendario.

Questa tesi si occupa proprio di questo difficile compito e lo fa seguendo due approcci principali. Nel primo approccio vengono modellate e previste, in un contesto out-of-sample, le serie storiche della domanda e dei prezzi ufficialmente riportati dal Gestore dei Mercati Energetici. A tal fine, viene considerato un modello a componenti additive che include una parte deterministica ed una componente residua stocastica. La parte deterministica, in particolare, contiene varie componenti che descrivono la dinamica di lungo periodo, quella periodica annuale e settimanale e gli effetti di calendario. Le prime tre componenti vengono stimate utilizzando delle splines del tempo mentre gli effetti di calendario vengono modellati mediante variabili dummy. La componente residuale, invece, viene trattata in maniera stocastica mediante vari modelli, univariati e multivariati, con diversi livelli di complessità. Sia nel caso univariato che in quello multivariato sono stati considerati modelli parametrici e non parametrici, nonché modelli basati sull'approccio funzionale.

La classe dei modelli univariati comprende modelli lineari autoregressivi (AR), modelli (auto)regressivi non parametrici e non lineari basati su splines (NPAR) e modelli funzionali a risposta scalare. Questi ultimi, a loro volta, possono essere formulati secondo una specificazione parametrica (FAR) o non parametrica (NPFAR). Relativamente alla classe dei modelli multivariati, invece, sono stati considerati modelli vettoriali autoregressivi (VAR) e modelli funzionali a risposta funzionale, sia nella versione parametrica (FFAR) che in quella non parametrica (NPFFAR). Tutti questi modelli sono stati stimati e confrontati in termini di capacità previsiva nell'ambito della previsione a 1 giorno e out-of-sample. Per verificare le performance dei modelli sono stati considerati i dati provenienti da 5 tra i principali mercati

elettrici: il mercato inglese (APX Power UK), il mercato del Nord Pool (NP), quello italiano (IPEX), quello di Pennsylvania-New Jersey-Maryland electricity market (PJM) ed, infine, quello portoghese (OMIE(Po)). Il periodo analizzato va dal 2009 al 2014. I primi cinque anni sono stati utilizzati per la stima dei modelli mentre l'intero 2014 è stato lasciato per le previsioni out-of-sample. La performance predittiva è stata valutata prima mediante indici descrittivi e poi mediante un test statistico per attestare la significatività delle differenze.

I risultati suggeriscono che, in generale, l'approccio multivariato produce previsioni più accurate dell'approccio univariato e che, nell'ambito dei modelli multivariati, i modelli basati sull'approccio funzionale risultano i migliori, anche se il VAR è comunque competitivo in diverse situazioni. Questi risultati possono essere letti anche come un segnale della presenza o meno di non linearità nei vari processi generatori dei dati. Anche se il confronto con altri lavori non è mai del tutto omogeneo, gli errori di previsione ottenuti sono tendenzialmente più piccoli di quelli riportati in letteratura.

Nella seconda parte della tesi il tema della previsione dei prezzi dell'elettricità è stato riconsiderato seguendo un percorso completamente diverso. L'idea di fondo di questo nuovo approccio è quella di modellare non le serie dei prezzi di mercato, ma le curve di domanda e di offerta giornaliere mediante modelli funzionali, di prevederle un giorno in avanti, e di trovare l'intersezione tra le due curve previste. Questa intersezione fornisce la previsione della quantità e del prezzo di equilibrio (market clearing price and volume). Questa metodologia richiede di aggregare, secondo uno specifico ordine, tutte le offerte di vendita e le richieste di acquisto presentate ogni (mezz')ora. Ciò produce delle spezzate lineari a tratti che vengono trasformate dall'approccio funzionale in curve lisce (smooth functions). Per questo fine, sono state considerati modelli funzionali parametrici (FFAR) e nonparametrici (NPPFAR). Come benchmark è stato stimato un modello ARIMA scalare alle serie storiche dei prezzi di equilibrio (clearing prices) ottenuti dall'incrocio tra le curve di domanda e di offerta. L'applicazione di questo metodo è stata fatta limitatamente al caso del mercato italiano. Come precedentemente, i risultati suggeriscono una migliore abilità previsiva dell'approccio funzionale rispetto al modello ARIMA. Tra i vari modelli considerati, quello funzionale non parametrico ha fornito i risultati migliori.

Va sottolineato poi che un aspetto rilevante, che va oltre il miglioramento nell'accuratezza previsiva, è che l'approccio basato sulla previsione delle curve di offerta e di domanda può essere utilizzato per ottimizzare le strategie di offerta/acquisto da parte degli operatori e, di conseguenza, per massimizzare il profitto dei traders.

Acknowledgements

All praise to the Almighty, the Lord of the universes, the most beneficent and the most merciful who empowered me and granted me the wisdom, health and strength to undertake this research task and enabled me to its completion.

I would like to thank my supervisor, Professor Francesco Lisi for his invaluable advice, guidance and support throughout the process of this research. I must to acknowledge his professionalism, supervision and good humor that help me to complete my research.

Special thanks to the academic committee of the PhD program and respective course instructors for their valued wisdom and knowledge that gave me the strength and capability for the successful completion of my studies here in Padova. I am deeply indebted to the Department of Statistical Sciences of the University of Padova, for having provided me this wonderful opportunity and for having offered a dynamic, friendly, and thought-provoking environment.

I appreciate my classmates of XXVIII PhD cycle and other researchers of the department who make my moments enjoyable and wiped away my loneliness with their active presence and support. Their company provided me awesome and unforgettable moments. I would like to thank the technical and administrative staff, especially, Mrs. Patrizia Piacentini for her co-operation and fruitful assistance during these three years.

A very special thanks to Dr. Enrico Edoli (Phinergy s.r.l.) for providing supply and demand curves data set.

I cannot evaluate, but feel the love and affections of my parents. Their sacrifices, overall supports and voices gave me energy and inspiration every time and never let me to fall in my entire life. I wish to express my deepest sense of love, respect and gratefulness to them. I am

also thankful to my brothers: Haroon ur Rasheed, Saif Ullah, Abid Ullah and Ikram Ullah for their unconditional love and continuous moral support. I want to give special thanks to my wife for her moral and emotional support and to my sons: Talha Shah and Saad Shah for their cute smiles full of energy. Last, but not least, I want to thanks my whole family and friends for their ongoing encouragement and support.

Finally, I would like to dedicate this dissertation to my beloved father, Haji Jan Gul.

Padova, 31/01/2016.

Ismail Shah

Contents

Abstract	ii
Acknowledgement	vii
Contents	ix
List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Overview	1
1.2 Main contributions of the thesis	4
2 Electricity Sector, Liberalization Process and Specific Features	9
2.1 Introduction	9
2.2 Electricity Markets Liberalization	12
2.2.1 The British Electricity Market	14
2.2.2 The Nordic Electricity Market	15
2.2.3 The PJM Electricity Market	16
2.2.4 The Italian Electricity Market	16
2.2.5 The OMEI(Po) Electricity Market	17
2.2.6 Other Electricity Markets	18
2.3 Electricity Time Series Features	18
2.3.1 Seasonality and Calendar Effects	19
2.3.2 Volatility, Outliers and Jumps	22
2.3.3 Non-normality and Non-stationarity	24
2.3.4 Mean Reversion and Other Features	26
3 Literature Review for Electricity Demand and Prices	27
3.1 Statistical Models and Methods	29
3.2 Artificial Intelligence Techniques	31

4	Predictive Models	33
4.1	Introduction	33
4.2	AutoRegressive Models	34
4.3	Nonparametric AutoRegressive Models	36
4.4	Vector AutoRegressive Models	38
4.5	Functional Data Analysis	39
4.5.1	Basis Functions	40
4.5.1.1	Fourier Basis	40
4.5.1.2	B-spline Basis	41
4.5.2	Functional AutoRegressive Models	42
4.5.3	Nonparametric Functional AutoRegressive Models	43
4.5.4	Functional-Functional AutoRegressive Models	46
4.5.5	Nonparametric Functional-Functional AutoRegressive Models	47
5	Modeling and Forecasting Electricity Demand and Price Time Series	49
5.1	Introduction	49
5.2	General Modelling Framework	52
5.3	Modeling the Stochastic Component	55
5.3.1	Univariate Models	56
5.3.2	Multivariate Modeling	57
5.4	Out-of-Sample Forecasting	58
5.4.1	Demand Forecasting	59
5.4.2	Price Forecasting	66
5.5	Conclusion	72
6	Modeling and Forecasting Supply and Demand Curves	73
6.1	Introduction	73
6.2	Price Formation Process in IPEX	75
6.3	Prices Prediction with Supply and Demand Curves	78
6.3.1	Application to GME Data	79
6.4	Optimizing Bidding Strategy	86
6.5	Conclusion	88
7	Conclusion and Further Research	89
	Bibliography	91
	Curriculum Vitae	103

List of Figures

2.1	The electricity value chain	10
2.2	One and two side auction	13
2.3	APX: (left) Annual seasonality for the period 01/01/2009 - 31/12/2010. (right) NP: Daily and weekly periodicity for demand data in the period 24/04/2010 - 07/05/2010.	20
2.4	APX: Periodogram of half-hourly electricity demand for the period 01/01/2013 to 31/12/2014	20
2.5	Average daily curves for the period 01/01/2014 to 31/12/2014 for (right) NP (left) PJM . . .	21
2.6	IPEX: Daily demand curves for the period 1/4/2011 - 30/4/2011. Solid lines: weekdays; dashed lines: Saturdays; dotted lines: Sundays. Solid line at the bottom: bank holiday (25th April).	21
2.7	(left): Temperature Vs electricity demand (source: Parker (2003)).(right) IPEX: Average daily electricity demand in each season for 2014.	22
2.8	(left) IPEX: Box plots for hourly demand for the period 01/01/2009 - 31/12/2014. (right) PJM: Box plots for hourly prices for the period 01/01/2009 - 31/12/2014	23
2.9	(left) PJM: Hourly electricity spot prices for the period 01/01/2013 - 31/12/2014. (right) A schematic supply stack with superimposed two potential demand curves (source Weron et al. (2004b))	24
2.10	PJM: hourly electricity spot prices for the period 01/01/2009 to 31/12/2010. (left) Normalized histogram with superimposed nonparametric density in red (right) quantile-quantile plot	25
2.11	Daily electricity demand for (right) APX, for the period 01/01/2006 - 31/12/2014 and (left) PJM, for the period 01/01/2001 - 31/12/2014 with superimposed linear (red) and a nonlinear (green) trend.	25
2.12	(left) APX: Half-hourly electricity prices (right) Hourly electricity prices for European Energy Exchange (source Erni (2012))	26
4.1	An example of Fourier (left) and B-spline (right) expansion with $k = 5$ basis functions.	41
4.2	Common symmetrical kernels	45

5.1	APX: Half-hourly price time series for the period 12/01/2014 - 18/01/2014.	50
5.2	APX: Daily price curves for the period 01/01/2014 - 31/12/2014.	51
5.3	(left) IPEX: Hourly demand cross correlation structure (right) PJM: Hourly prices cross correlation structure	51
5.4	APX: Load period 9. $\log(D_{t,j})$, $\hat{f}_1(T_{t,j})$, $\hat{f}_2(Y_{t,j})$, and $\hat{f}_3(W_{t,j})$ ($\hat{f}_3(W_{t,j})$ is only for ten weeks) components.	54
5.5	Electricity Demand: Hourly MAPE values for (top left) PJM (top right) NP (middle) UKPX (bottom left) IPEX (bottom right) OMIE (Po). lines are (solid) VAR (dashed) FFAR (dotted) NPFFAR.	65
5.6	Electricity Price: Hourly MAPE values for (top left) NP (top right) APX (bottom left) PJM (bottom right) IPEX. lines are (solid) VAR (dashed) FFAR (dotted) NPFFAR.	71
6.1	Supply (blue) and two demand (red) hypothetical curves	74
6.2	IPEX: Supply (red) and Demand (blue) curves (left) and their zoomed version (right)	77
6.3	Supply and demand curves original (black) and smoothed (dotted red)	81
6.4	IPEX load period 9: market clearing price (black) and equilibrium price (red) for 485 day.	82
6.5	IPEX: supply and demand curves in original (red) and forecasted (black)	85
6.6	IPEX: what-if simulations: (left) Predicted supply and demand curves (dashed) with modified supply curves (solid) (right) and zoom on a neighbourhood of the intersection	87
6.7	IPEX: what-if simulations: (left) original supply and demand curves (dashed) with modified supply curves (solid) (right) and zoom on a neighbourhood of the intersection	87

List of Tables

5.1	Electricity Demand: Prediction accuracy statistics: AutoRegressive (AR), Nonparametric AutoRegressive (NPAR), Functional AutoRegressive (FAR), Nonparametric Functional AutoRegressive (NPFAR), Vector AutoRegressive (VAR), Functional Functional AutoRegressive (FFAR), Nonparametric Functional Functional AutoRegressive (NPFFAR). . . .	61
5.2	Electricity Demand: P-values for the DM test for equal prediction accuracy versus the alternative hypothesis that model in the row is more accurate than model in the column (squared loss function used)	63
5.3	Electricity Demand: Hourly DS-MAPE	64
5.4	Electricity Price: Prediction accuracy statistics: AutoRegressive (AR), Nonparametric AutoRegressive (NPAR), Functional AutoRegressive (FAR), Nonparametric Functional AutoRegressive (NPFAR), Vector AutoRegressive (VAR), Functional Functional AutoRegressive (FFAR), Nonparametric Functional Functional AutoRegressive (NPFFAR).	68
5.5	Electricity Price: P-values for the DM test for equal prediction accuracy versus the alternative hypothesis that model in the row is more accurate than model in the column (squared loss function used)	69
5.6	Electricity Price: Hourly DS-MAPE	70
6.1	IPEX: Supply and demand bids	79
6.2	IPEX: Prediction accuracy statistics: Nonparametric Functional Functional AutoRegressive (NPFFAR), Functional Functional AutoRegressive (FFAR), AutoRegressive Integrated Moving Average (ARIMA)	83
6.3	IPEX: P-values for the DM test for equal prediction accuracy versus the alternative hypothesis that model in the row is more accurate than model in the column (squared loss function used)	84

Chapter 1

Introduction

1.1 Overview

Before the liberalization, electricity sector was fully controlled by state-owned companies. In this monopolistic structure, the variation in the electricity prices was minimal and the main attention was paid to demand forecasting and long-term planning and investment in this sector. The electricity sector undergone through drastic reforms in the late 80's when the state owned monopolistic structure was reorganized into liberalized and competitive power markets. The main idea behind restructuring was to promote competition among generators, retailers and consumers by encouraging private investments in production, supply and retail sectors. The first electricity reforms were introduced in Chile in 1982, and in the following years the phenomenon spread throughout the world particularly in Europe. The British electricity sector started its liberalization in 1990 followed by Norway in 1992 and so on. Currently, many EU countries, including Italy, have their own liberalized electricity market as well as Australia, America, Canada, New Zealand, Japan and many other developed countries. The number of liberalized electricity markets is steadily growing worldwide, but the trend is most visible in Europe.

The liberalization not only brought important benefits to consumers such as low prices, more choices, reliable and secure electric supply but it also introduced a new field of research. The accurate modeling and forecasting of different variables related to the markets e.g. prices,

demand, production etc. became more crucial due to market structure. In most countries, the electricity market consists of different markets including a day-ahead market where prices and demand are determined the day before the delivery by means of (semi-) hourly concurrent auctions for the next day. For each auction, producers/buyers submit their offers bids willing to sell/buy a certain amount of electricity at a given price. These bids are aggregated by an independent system operator in order to construct the aggregated supply and demand curve which determines the market clearing price and quantity. Since electricity is a flow commodity in the sense that it cannot be stored in large amount, over or under-estimation of electric load can cause serious problem to electric utility providers, energy suppliers, system operators and other market participants. For example, in case of underestimation, agents rely on highly responsive but expensive generating plants since low cost generating plants need a long time to start-up and so are not useful for serving short-duration peaks. On the other hand overestimation of electricity demand leads to unnecessary production or excessive purchases of energy which can cause substantial financial losses. Adequate forecasting, instead, leads to less expensive, reliable and secure power operation and planning and allows the cash flow analysis, least cost planning, integrated resource planning, financial procurement, regulatory rule-making and demand side management etc (Bunn, 2004a). However forecasting electricity markets are not straight forward due to the specific features these markets exhibit. There exist a large variability in end-user demand throughout the year due to seasonal variation resulting in multiple periodicities, non-constant mean and variance, spikes or sudden jumps etc. in the price and load series. Calendar effects are evident as the daily load and prices profiles are different for different days of the week and the behavior deviates from the typical behavior on bank holidays, bridging holidays etc. Technical problems such as plant outages and grid line unreliability add more variability to the system. The load series usually contain few outliers however; the price series show high volatility and unexpected jumps, also called spikes. In fact, the volatility is by far stronger for electricity prices compared to any other financial commodity (Weron, 2007).

In the literature, different methods have been discussed to account for these specific features effects before modeling the demand/price series in order to achieve stationarity and mini-

mizing distorting effects on forecasting. These effects are either modeled in a deterministic or stochastic way. In the deterministic approach, piecewise constant functions or dummies are widely used to model the multiple periodicities and the specific calendar conditions such as bank holidays, bridging effect etc. (Escribano et al., 2011; Fanone et al., 2013; Fleten et al., 2011; Gianfreda and Grossi, 2012; Lisi and Nan, 2014; Lucia and Schwartz, 2002). In some cases, components are modeled using sum of sinusoidal functions of different frequencies, sometime, equipped also with linear trend for the long term dynamics (Bierbrauer et al., 2007; Erlwein et al., 2010; Nan et al., 2014). Other authors considered polynomials, splines, wavelet decomposition, moving averages and in some cases state space models with linear trend to model different components (De Livera et al., 2011; Dordonnat et al., 2010; Janczura and Weron, 2010; Schlueter, 2010). In the second case, components are viewed as stochastic processes. Some authors suggest modeling of long term dynamics by a random walk or Brownian motion with the assumption of unit root while other also treated the seasonal components as stochastic (Bosco et al., 2010, 2007; Koopman et al., 2007). The stochastic approach is widely used for the case of spikes/jumps and is modeled by diffusion models with Poisson jumps or by Markov-switching models (Borovkova and Permana, 2006; Hellström et al., 2012; Pirino and Renò, 2010; Weron et al., 2004a,b). Lastly, it is worth mentioning that in both cases, deterministic or stochastic, the authors who modeled the specific calendar effects e.g. bank holidays, bridging effects etc. considered dummies. Once these components are estimated, the residuals (stochastic) component is obtained by subtracting them from original (unadjusted) demand/price time series, whose dynamics is modeled using different models with increasing level of complexity

For the modeling of residual part, two approaches can be considered, univariate and multivariate. Since an individual auction is held for each load period and the load pattern is quite different across the different days of the week, the first approach treats each load period separately, consequently, (48)24 (half-)hourly models, reflecting the incorporation of the daily total series. However, the load profile suggests the presence of correlation among different load periods within a day that can be used when modeling the series and thus leads to a multivariate approach. For both approaches, various techniques have been pro-

posed in the literature, see for example (Weron, 2014, and references therein). Different parametric models, such as regression models (e.g. multiple regression), time series models (e.g. ARIMA and its extensions) and models based on exponential smoothing techniques (e.g. Holt-winters and its extensions) that account for multiple seasonalities are extensively used (Bianco et al., 2009; Charlton and Singleton, 2014; De Livera et al., 2011; Ediger and Akar, 2007; Hong et al., 2010; Taylor, 2012). Semi-parametric and state space models are also employed to forecast short-term electric load and prices (Dordonnat et al., 2008; Fan and Hyndman, 2012). On the other hand, nonparametric techniques are always attractive for researchers due to their flexibility to functional form specifications, non-linearity and detection of structures that are usually undetected by traditional parametric methods. These techniques under dependence are useful for forecasting in time series and are frequently used (Härdle and Vieu, 1992; Hart, 1991; Shang et al., 2010). Artificial neural network (ANN) are extensively used for load forecasting due to their nonlinear and nonparametric features (Hippert et al., 2001; Zhang et al., 1998). Prediction problems are also addressed with other computational intelligence based methods such as fuzzy logic, support vector machines etc. (Mohandes, 2002; Pandian et al., 2006). Although mathematical structure and complexity of all the models differ, it is difficult to find a single model that outperforms all others in every situation. In general, each model has its own advantages and disadvantages when it comes to practice.

1.2 Main contributions of the thesis

The main goal of this thesis is to model and forecast variables related to electricity markets such as, prices, demand etc. To this end, different approaches are considered and applied to electricity market data. This work considers the deterministic approach for the component estimation, and analyzes several ways of modeling the residual component. Both for demand and prices, different classes of models are estimated and compared in terms of forecasting ability with respect to the original (unadjusted) time series. In particular, different univariate as well as multivariate models, parametric and nonparametric, have been consid-

ered for five electricity markets, namely, British electricity market (APX Power UK), Nord Pool electricity market (NP), Italian electricity market (IPEX), Pennsylvania-New Jersey-Maryland electricity market (PJM) and Portuguese electricity market (OMIE(Po)). These markets substantially differ in generation modes, market maturity, size and policies implemented, geographical location and land electricity demand and have been widely considered in the literature. Our data set consists of 24 (or 48) observations for each day, corresponding to the number of daily auctions. The class of univariate models includes parametric autoregressive models (AR), nonparametric and nonlinear regression models based on splines (NPAR) and scalar-response functional models, that in turns can be formulated parametrically (FAR) or non parametrically (NPFAR). The multivariate models are vector autoregressive models (VAR) and functional-response, parametric (FFAR) and nonparametric (NPFAR), models. Linear AR(p) models are well-known and widely used (Brockwell and Davis, 2006). They describe the daily dynamics of load/price taking into account a linear combination of the last p values. In the nonparametric nonlinear (NPAR) case, the relation between current load/price and its lagged values has not a specific parametric form allowing, potentially, any kind of nonlinearity. Vector autoregressive (VAR) models are well-known multivariate models able to account for linear relationships among different time series. In this approach each variable (in our case the demand/price at each load period) is a linear function of past lags of itself and of the other variables. On the other hand, functional models consider the demand/price daily profile as a single functional object. Generally, statistical models combine information either across or within sample units to make inference about the population, functional data analysis (FDA) considers both. Although functional data analysis has been extensively used in other fields, limited literature is available for time series prediction and the books (Ferraty and Vieu, 2006; Ramsay et al., 2009) are comprehensive references for parametric and nonparametric functional data analysis. Its main advantage with respect to vector autoregressions (VAR) is that VAR are multivariate finite dimensional models, while functional models, being infinite dimensional, bypass the problem of the number of variables and allow to use additional information (e.g. smoothness, derivatives) contained in the functional structure of the data. The use of the functional

approach is one of the main contributions of this thesis. In fact, although it is not completely new, the use of the functional approach in the energy markets is not still widespread. In the following, the contents of the thesis have been divided in two points, corresponding to two different kinds of problems that have been considered.

1) The first part addresses the issue of modeling and out-of-sample forecasting electricity demand and price time series. To this end, I referred to the additive component model suggested by Lisi and Nan (2014) that assumes some deterministic components and a stochastic residual component. The deterministic components include a long-term dynamics, annual and weekly periodicities and calendar effects. Different possibilities for the estimation of these components were considered and the final selection was made based on the minimum prediction error. The first three components were estimated using splines while the calendar effects were modeled using dummy variables. In case of demand, data for indicated margin was available for APX and hence included as an extra covariate to the model. The demand structure for OMIE(Po) changed dramatically in the start of 2012 and therefore a dummy variable accounting for this level shift has been included to the model. For the prices, forecasted demand used as an extra covariate in the model. All these extra covariates were found highly significant. For the residual component, different univariate and multivariate models have been considered with increasing level of complexity. Within both classes, linear parametric and nonlinear nonparametric models as well as functional based models have been estimated and compared in a one day-ahead out-of-sample forecast framework. Data from 2009 to 2014 were used for all five electricity markets included in our study. The first five years were used for models estimation while the year 2014 was left for one day ahead out-of-sample forecast. Thus, globally, we have $365 \times 24 \times 48 = 8760 \times 48 = 420480$ one-day-ahead predictions allowing for a thorough analysis of the forecasting results. To compare the forecasting performance, global mean absolute percentage error (MAPE), daily specific mean absolute percentage error (DS-MAPE) and mean square percentage error (MSPE) were computed for each model. To assess the significance of the differences among different summary statistics, Diebold and Mariano (DM) (Diebold and Mariano, 1995) test for equal predictive accuracy was used.

The results suggest, as expected, the multivariate approach leads to better results than the univariate one. Within univariate models, the results clearly showed superior performance of scalar-response functional models compared to others. The significance of the results were evaluated and confirmed by DM test. In case of multivariate models, the functional models perform generally better with VAR being a competitive model in some cases. The results also lead to another important finding correspond to the performance of parametric and nonparametric approach that showed strong linkage with underlying process. For IPEX and OMIE (Po), the nonparametric and nonlinear approach performs better, suggesting possible nonlinearities in the underlying process. For the other three markets, the parametric approach produces better results. Lastly, the obtained results were compared with other works in the literature. Although different works refer to different time periods, we compare the results with the authors who used the same prediction accuracy statistics. The comparison suggests that our forecasting errors are smaller compared with the state-of-art prediction techniques used in the literature.

2) In the second part of this thesis the issue of electricity price forecasting is revisited and a completely and, at my best knowledge, new approach is used. It is based on the idea of modeling the daily supply and demand curves, predicting them and finding the intersection of the predicted curves in order to find the predicted market clearing price and volume. For this task the functional approach is quite suitable because for each given day, the number of bids data, corresponding to the number of producers/buyers in the market, is very large. Thus, finite dimensional (both univariate and multivariate) forecasting techniques cannot be used due to the large number of variables. On the contrary, functional models consider a single day as a single functional object and the bids, points on this functional object.

In this approach, the raw bids data for demand and supply corresponding to each (half-) hour is first aggregated in a specific order. The functional approach converts the resulted piece wise curves into smooth functions using B-spline approximation. To consider the weekly periodicity, data are divided into seven groups representing a single day of week. Thus, e.g., for the prediction of Monday, the historical data from all available previous Mondays were used. The application of this approach is limited to the Italian market because it requires a

lot of data that are not always simple to obtain. Note that these data are available only with a eight-day-lag and thus, in a real context, eight-days-ahead forecasting is required. For this issue, parametric functional model (FFAR) and the nonlinear nonparametric counterpart (NPFFAR) were considered. As benchmark, an ARIMA model was fitted to the scalar time series corresponding to the clearing prices obtained from the crossing points of supply and demand curves. In this case we obtained one-day-ahead predictions and compared to the results obtained with our functional approach. We consider data for the period January 2014 to April 2015. The whole year 2014 is used for model estimation while the last four months are used for out-of-sample forecasts. Mean absolute error (MAE), root mean square error (RMSE) and MAPE were used to summaries the results. The results showed superior forecasting performance of our functional approach. In general, the MAE were significantly lower ranging from 5% to 20% for different load periods. The MAPE values showed the difference between 1% to 4% in favor of functional models. The significance of the differences was also confirmed by DM test. Among different models, the nonparametric functional model produces better results compared to parametric models.

Apart from the improvement in forecasting accuracy, it is important to stress that forecasting the entire demand/supply curves can substantially improve the supplier/buyer bidding strategy resulting in a significant financial gain. Despite their good forecasting abilities for electricity price/demand, an important drawback related to the classical time series models is the fact that they do not provide insight to the supply and demand mechanism consequently to the price/demand formation process. With the current approach, if the forecasted curves are available, a trader who requires a moderate quantity to sell/buy can rise/lower the price by submitting an extra non-standard offer for an extra small quantity. As forecasting the whole curves gives deep insight into the market, our analysis showed that this strategy can significantly improve bidding strategies and maximize traders profit.

Chapter 2

Electricity Sector, Liberalization Process and Specific Features

2.1 Introduction

Electricity is a unique commodity that is essential for the development of any society or country. It helps to utilize human abilities and capabilities to produce goods and services efficiently, communicate more easily and to trade all around the world. Humans poverty, health, education, income etc. are strongly linked with the availability of this commodity. According to world health organization (WHO), around three billion people lack access to modern fuels for cooking and heating and use traditional stoves burning biomass (wood, animal dung and crop waste) and coal resulting four million premature deaths every year. The impact of electricity on human life is very strong and therefore, extensive studies have been made in different directions related to this sector.

Electricity is itself not the primary source of energy but the energy released by other sources and converted by mankind for the use of end-user. These resources are broadly divided into two categories: renewable and nonrenewable. Renewable resources such as hydro, solar, wind etc. are replenished naturally and over relatively short periods of time. On the other hand, nonrenewable energy resources e.g. coal, nuclear, oil, natural gas etc. are available in limited supplies and usually take long period to replenished. Both the categories are mainly

made up of the following energy resources:

- *Chemical energy* is obtained through chemical reactions or absorbed in the creation of chemical compounds such as oil, coal, natural gas, biomass etc.
- *Nuclear energy* is obtained through the radioactive decay of some unstable nuclide's such as plutonium, uranium etc.
- *Potential energy* is obtained through the forces of gravity pulling something towards earth. The most common is the one that stored in the water.
- *Kinetic energy* is obtained through the motion of an object. The most common form is that obtained through windmill that converts the energy of moving air into electricity.
- *Solar energy* is obtained through conversion of sunlight into electricity, either directly using photovoltaic (PV), or indirectly using concentrated solar power (CSP).

The marginal cost of producing electricity is different for different resources. Electricity generated from nuclear, hydro and wind have low generation cost compare to generated by other fuels such as coal, gas, diesel, etc. As the demand increases, more expensive generation units are used for generation that result increase in electricity prices. Before liberal-

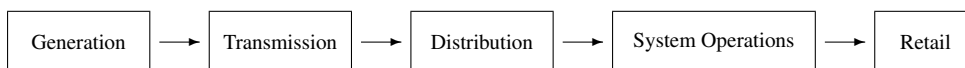


Figure 2.1 The electricity value chain

ization, electricity firms were vertically integrated in five major components also known as “electricity value chain” given in Figure 2.1. They comprised of generation, transmission, distribution, system operations and retail.

- *Generation* refers to the process of installing a power plant and converting primary energy resource to electricity.

- *Transmission* refers to the transportation/transmission of the generated electricity. Power plants are often installed far from the population and therefore high voltage transmission lines are installed which stepped up (transformed) the voltage to travel fast and cover long distances.
- *Distribution* refers to providing low voltage electricity to homes and industries. Substations receive high voltage electricity and step down the voltage for the delivery and use of end-users.
- *System operator* refers to the process of monitoring the system continuously and balancing supply and demand to avoid electric grid blackouts. As demand fluctuates throughout the day, system operator monitors and balance the system throughout the day so that production and demand match perfectly and continuously.
- *Retail* refers to the process of delivery service for sale to retail customers. The retail companies directly sell electricity to end-users and responsible for providing billing, customer services etc. facilities.

The economy of a country is heavily dependent on availability and efficient management of electricity. Any mismanagement or shortages results significant crises for the economy and for this reason, until late eighties, electricity sector was fully controlled by state owned companies and was highly regulated. In this monopolistic structure, the variations in the electricity prices were minimal and the main attention was paid to demand forecasting and long-term planning and investment in this sector. Inspired from the successful liberalization of various sectors of the economy, electricity sector undergone through drastic reforms in late eighties that reorganized the state owned monopolistic structure into liberalized and competitive power markets. The main aim behind liberalization was to rely on competitive forces to encourage investment and efficiency that benefits all the participants of the market and consequently the economy.

2.2 Electricity Markets Liberalization

The liberalization process started first in Chile in 1982 by introducing reforms, the 1982 Electricity act, to electricity sector that dissolve the state owned monopolistic structure by commercialization and part privatization followed by large scale privatization in 1986. The main idea behind liberalization was to increase industry efficiency, price stability, heightened competition, and enhanced security of supply. Soon after deregulation, many (macro-) economic indicators show considerable improvements that encouraged this phenomenon to spread throughout the world. In Europe, the British electricity sector was the first that started its liberalization in 1990 followed by Norway in 1992 and so on. Currently, many EU countries, including Italy, have their own liberalized electricity market as well as Australia, America, Canada, New Zealand, Japan and many other developed countries.

Electricity market reorganization unbundled the vertically integrated utilities that historically managed generation, transportation and supply of electricity and introduce competition mainly in generation and retail sector as all the competitors need non-discriminatory access to the other three components; transmission, distribution and system operations. Liberalization resulted mainly into two wholesale electricity markets; power pools and power exchange. The difference between these two is not trivial as they share many characteristics however they can be distinguish by two criteria: initiative and participation (Boisseleau, 2004). The power pools are the result of public initiative and the participation is mandatory i.e. no trading is allowed out side the pool while the power exchange is launched on private initiative and the participation is voluntary. Power pools are further divided into two types namely technical and economic pools. In technical pools, the power production cost and the network capacity is the main factor for dispatch. The power plants are ranked on merit order by their production cost and the electric utilities optimize their power generation with respect to cost minimization and optimal technical dispatch. Economic pools have been initiated with the idea of competition among generators. This pool is one sided auction market where the participants are only generators and the participation is mandatory. In this market, the producers bid based on the prices for which they willing to run their power plants. These bids are aggregated to obtain supply curve by independent system operators. Finally,

the market clearing price (MCP) and volume (MCV) are obtained through the intersection point of supply curve and estimated demand.

On the other hand, power exchange are two side auction markets where the market participants are generators, distributors, large consumers and traders. The main idea behind the establishment of power exchange was to facilitate the trade of electricity in a short term with the promotion of competition and liquidity. The market clearing prices (also called spot prices) and volumes are determined through two sided auctions in a day-ahead market where trading terminates typically the day before the delivery. Generally the auctions con-

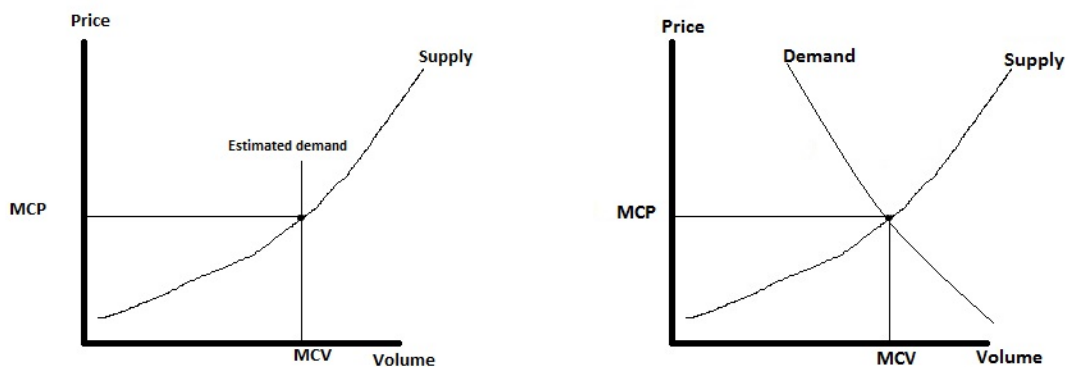


Figure 2.2 One and two side auction

ducted once per day where producers and buyers submit their offers bids willing to sell/buy a certain amount of electricity and its corresponding minimum price for each load period. These bids are aggregated by an independent system operator in order to construct the aggregated supply and demand curves which determine the market clearing price and quantity. The buyers who bid above or equal to market clearing price pay the price and the suppliers who bids below or equal are paid the same price. This pricing scheme is also called uniform pricing (non-discriminatory) in contrast to pay-as-bid (discriminatory) where a supplier is paid the amount for his transacted quantity based on his marginal cost.

Liberalized electricity markets are nowadays situated all around the world. These markets share many characteristics but also differ substantially in generation modes, market ma-

turity, size and policies implemented, geographical location and land electricity demand. From last two decades, extensive studies have been made on these markets in different directions. In the following, some of the markets that are considered in this thesis for empirical analysis are illustrated.

2.2.1 The British Electricity Market

The liberalization of UK electricity sector is due to structural changes and regulatory reforms introduced in late 80's in order to dissolve the state owned monopolistic structure and to introduce a competitive electricity wholesale market. Since transmission and distribution are natural monopolies, the main objective of the reforms was to privatise the generation and supply sector. Hence in 1990, the UK electricity market is reorganized into England and Wales electricity pool and the state owned monopoly is divided into three companies, namely, National Power, Powergen and Nuclear Electric. The pool was compulsory day-ahead one sided market where the trading was carried out on half-hourly basis. National Power and Powergen had 50% and 30% shares respectively due to which market power in generation was a significant problem as Nuclear Electric was providing the based load nuclear power and essentially was a price taker. Market manipulation by these two companies resulted in a less competitive environment and hence the average price remain 24£/MWh in the years 1994-96 (Bunn, 2004b).

With the introduction of New Electricity Trading Arrangements (NETA) in 2001 (from 2005, NETA is called British Electricity Trading Transmission Arrangements, BETTA), the pool was replaced by fully liberalized bilateral contracting and voluntary spot trading market resulted in a balanced market share for electric utilities both in generation and retail sectors. These reforms resulted three independent power exchanges namely, UK Power exchange (UKPX), UK Automated Power Exchange (APX UK) and International Exchange (IE, formerly named International Petroleum Exchange (IPE)). In 2004, APX and UKPX merged into APX Group and a year later, Scotland was included to the UK electricity market. Currently this market is fully competitive and one of the mature market in the world that exhibits strong linkage between market price and market fundamentals Karakatsani and

Bunn (2008).

2.2.2 The Nordic Electricity Market

Soon after the liberalization of British electricity sector, the phenomenon spreads throughout the world, particularly in Europe. The Nordic electricity market was established in 1992 as a consequence of the reforms introduced in Norwegian energy act 1991. The deriving factor for the reforms was the increasing dissatisfaction of electric sector performance in terms of economic efficiency in resource utilization. In the beginning, this market consisted of Norway only however in the preceding years Sweden (1996), Finland (1998) and Denmark (2000) were also included and was called Nord pool. Nord pool was the first international power exchange and currently over 380 companies from 20 countries actively trade on this market.

Nord pool electricity market is comprised of different markets. The day-ahead market (spot market) where power trading for physical delivery is carried out is called Elspot. The participation in this market is voluntary and the minimum contract size is 0.1 MWh. The prices are determined in a two-sided concurrent auction system (demand and supply) for each hour of the day. The resulted price is used as the reference price for settling financial power contracts and a benchmark for bilateral transactions. The adjustment market, Elbas is a short-term physical delivery market that allows players to modify the injection/withdrawal schedules that they have defined in Elspot. Eltermin and Eloption are financial markets that offer different kind of financial products.

The Nord pool electricity market is quite unique since the larger portion of electricity is generated from hydro source. The electric production from hydro in Norway, Sweden and Iceland are over 99, 85 and 76 percent respectively. The peculiar price dynamics are originated by this feature as prices are very sensitive to atmospheric conditions. Electric production varies from season to season depending on rainfall and snow conditions. The water shortage in 2002-2003 resulted in substantial price increase and put the electricity market under tremendous pressure. Consequently, all the Nordic power plants were used for production and significant increase in imports and decrease in demand was observed in that period.

2.2.3 The PJM Electricity Market

The PJM (Pennsylvania-New Jersey-Maryland) Interconnection is a regional transmission organization (RTO) that manages all the movement of wholesale electricity in all or parts of thirteen states and the District of Columbia in United States of America. It is an independent and neutral party that operates a competitive wholesale electricity market and also responsible to manage and ensure the reliability of high-voltage electricity grid that provide electricity to over 61 million people. The PJM Interconnection started the transition to an liberalized, independent and neutral organization in 1993 when the PJM Interconnection Association was formed to administer the power pool. In 1997, PJM became a fully independent organization and opened its first bid-based energy market. The PJM announced as the nation's first fully functioning independent system operator (ISO) in late 1997 and approved as a regional transmission organization (RTO) in 2001 to operate the transmission system in multi-state areas. From 2001, The PJM interconnection is growing continuously and a number of utilities transmission systems are integrated to PJM resulted in the world largest competitive market where continuous buying, selling and delivery of wholesale electricity is carried out. The growing number of utilities integrated to PJM enhances its abilities to meet consumer demand for electricity with diversified resources availability.

The PJM consist of two generating capacity markets, two energy markets, an ancillary service market and a financial transmission entitlements market. The energy markets consist of two different market: (1) a day-ahead market where most of the quantity is traded through hourly auctions for the next operating day based on generation offers, demand bids, and bilateral transaction schedules submitted into the day-ahead market and (2) a real time (five minutes) market that is a real time balancing market where locational marginal prices are calculated at five minute intervals based on actual grid operating conditions.

2.2.4 The Italian Electricity Market

In Italy, the liberalization process of the electricity sector started in 1999 when the state-controlled entity, ENEL, was replaced with a market mechanisms more suited to the new

framework for power trade. The main objective of liberalization was to regularize the market for electricity power production and trading under principles of neutrality, transparency, objectivity and competition among producers. The first regulated wholesale electricity market in Italy started to operate as a pool in April 2004. In 2005, the pool was replaced by an exchange namely Italian Power Exchange (IPEX) by liberalizing the demand side bidding. Since July 2007, when the market became fully liberalized, it has gone through different phases of market maturity. According to TERNA, the Italian transmission network operator, the total electricity demand for the year 2014 was 309 TWh, a 3% decline from 2013, third consecutive annual decrease and is 10% lower than observed in 2007 when it reached to its maximum. Compared to 2013, net production decreased around 4%, and was 267.6 TWh resulted an increase in import by 3.7% (up to 43.7 TWh) that covers 14% of Italian demand. The price for base load showed a decline of 10 Euro/MWh compared to year 2013 and reached to 52.1 Euro/MWh, the lowest level in the last decade.

The Italian Spot Power Exchange is split in several markets namely the Day-ahead market (MGP), Adjustment/intra-daily market (MA) and the Ancillary services market (MSD). The Gestore dei Mercati Energetici S.p.A (GME) is responsible for scheduling injections and withdrawals of electricity for the next day in 24 concurrent auctions one each for a hour in a day. The adjustment market (MA) allows the utilities to modify their injection/withdrawal schedules defined in MGP. TERNA use the MSD to retrieve the resources required for controlling the power system, i.e the creation of an energy reserve and real-time balancing. The Italian electricity market is divided into different zones and the prices across the zones can differ due to transmission limits and supplier's behaviour. With the introduction of Italian derivatives energy exchange, producers and consumers are allowed to hedge positions, to have bilateral contracts for the prices far in the future and to physically trade energy.

2.2.5 The OMEI(Po) Electricity Market

Spain was the first Iberian country that liberalized its electricity sector. In 1997, Operadore del Mercado Espanol de Electricidad (OMEL) was created to manage and run wholesale electricity market. In November 2001, the formation of an Iberian electricity market (MI-

BEL) is initiated by Portuguese and Spanish governments however the start of the joint organized market took longer due to many political and technical reasons. In July 2006, MIBEL start its operations in both, Portugal and Spain. Like any other market, the Iberian electricity market has a day-ahead and intra-daily markets. OMEL has changed its name to Operador Del Mercadeo Iberico De Energia (OMIE, operator of the Iberian Market) and is in charge of managing the MIBEL day-ahead market where electricity is traded on hourly basis for the twenty-four hours of the following day. In the intra-day markets, buyers and sellers are allowed to readjust their commitment made in day-ahead market for purchasing and selling up to four hours ahead of real time. In 2014, OMIE covers 80% trading of electricity amounting for 11 billion Euros carried out in Spain and Portugal. With over 800 participants, this market is one of Europe's more liquid ones and the prices are generally below the average compare to Europe's major markets.

2.2.6 Other Electricity Markets

Nowadays liberalized electricity markets are situated all around the world. In Europe, including above mentioned countries, Austria, Belgium, Czech Republic, Estonia, France, Finland, Germany, Hungary, Latvia, Lithuania, Luxemburg, the Netherlands, Poland, Romania, Slovakia and Slovenia liberalized their electricity sector. Many developed countries e.g. Australia, Japan, Turkey, America, Canada etc. have their own liberalized electricity market. The main motive behind liberalizing and restructuring electricity market in all countries is to ensure security of supply and efficient production by introducing competition in different sectors.

2.3 Electricity Time Series Features

Liberalization not only brought important benefits to consumers such as low prices, more choices, reliable and secure electric supply but it also introduced a new field of research. The accurate modeling and forecasting of different variables related to these markets e.g. prices, demand, production etc. became more crucial due to market structure. Electric utilities,

generators, system operators and other participants are highly interested in forecasting these variables at different horizons, that are referred in the literature as short-, medium- and long-term. Daily and weekly predictions are very common and are used for real time operations, control of power system and generator maintenance whereas monthly and yearly maximum and minimum loads are important for planning purposes and risk management studies. In the literature, short-term prediction received higher attention because in many electricity markets daily demand and prices are determined the day before the delivery by mean of (semi-)hourly auctions for the following day. Therefore efficient modeling and forecasting for these variables is an important issue in competitive electricity markets. However this task is challenging due to the specific features these markets exhibit that are substantially different from other financial commodities markets and hence require dedicated modeling techniques. A physical constraint to electricity is that it cannot be economically stored and must be delivered to end-user at the time when it is produced. Due to this characteristic, minor fluctuations in electricity demand can cause serious problems for electric utilities, generators, distributors and other market participants. The main peculiarity referring to price volatility is also attributed to this characteristic. On the other hand, the end user treats electricity as a facility resulting the demand to vary throughout the day, week and year. These and other attributes generate different features found in both, demand and prices time series that are commonly shared across the electricity markets. In the following, some of the main features found in demand and prices series are described.

2.3.1 Seasonality and Calendar Effects

Perhaps the most common and prominent features that the electricity demand exhibit are seasonality and calendar effects. In most markets, electricity demand shows three regular cycles: daily, weekly and annual. An example of the three periodicities is given in Figure 2.3. Annual cycle can be clearly seen from the graph as the variation in the demand level throughout the year has similar pattern in both years. Daily cycle is evident from the similarity of the demand from one day to the next whereas an weekly cycle can be seen by comparing the demand level on same days of different weeks. Electric consumption is

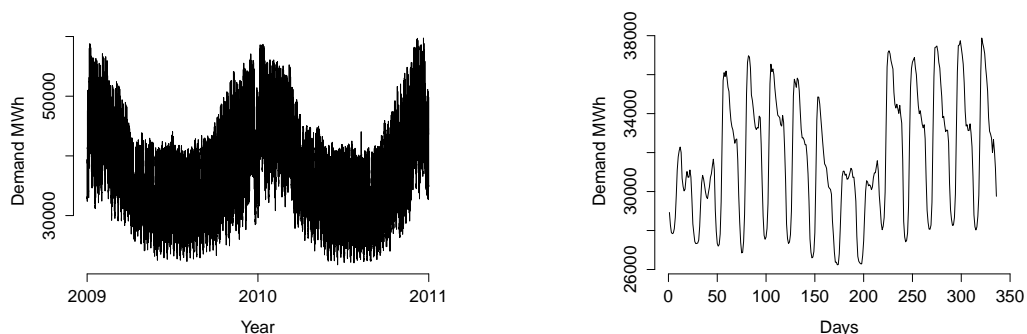


Figure 2.3 APX: (left) Annual seasonality for the period 01/01/2009 - 31/12/2010. (right) NP: Daily and weekly periodicity for demand data in the period 24/04/2010 - 07/05/2010.

closely related to human behaviour, daily life and economical activities. The daily cycle is responsible for the highest part of variability in the data (see Figure 2.4). The consumption varies throughout the day and night following the working habit of population that can be seen in Figures 2.5. The demand is considerably lower at night and starts increasing in the morning when people start their daily life. Depending on the market, the daily cycle can have one or more peaks during the day corresponding to the living habits of the population and atmospheric conditions of the country. Due to this variation, many researchers argued

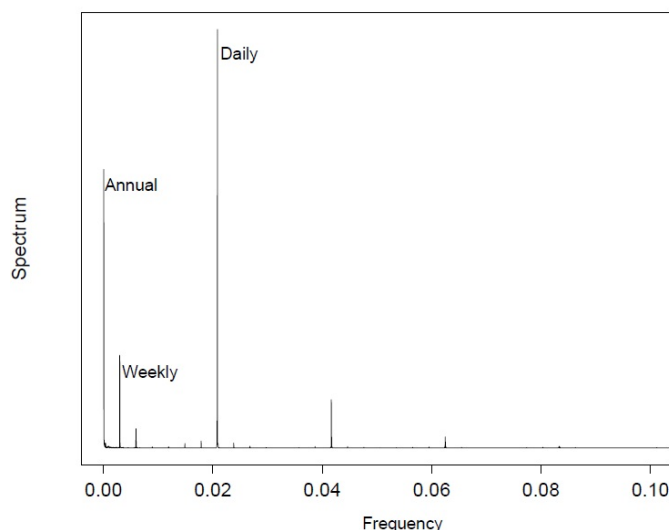


Figure 2.4 APX: Periodogram of half-hourly electricity demand for the period 01/01/2013 to 31/12/2014

and considered the daily load pattern as (48)24 (half-)hourly different time series for modeling purposes. The weekly cycle encompasses for relatively lower variation in the data. As

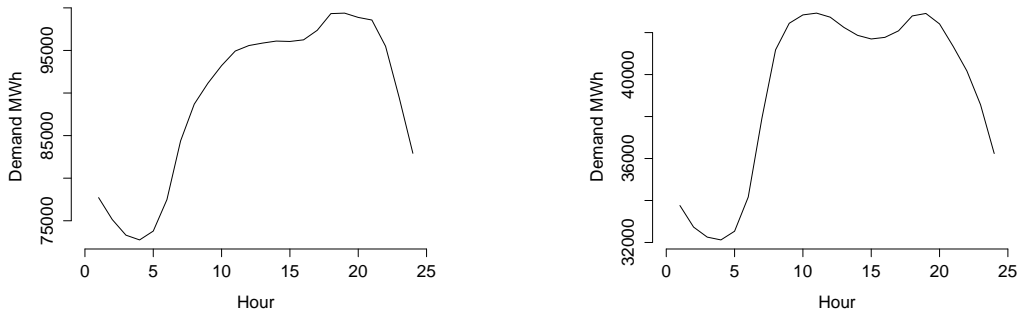


Figure 2.5 Average daily curves for the period 01/01/2014 to 31/12/2014 for (right) NP (left) PJM

can be seen in Figure 2.6, load profiles are relatively different on different days of the week. Saturday and Sunday load profiles are considerably lower than those of other weekdays. In general, demand is lower during national holidays, bridge holidays (a day between two

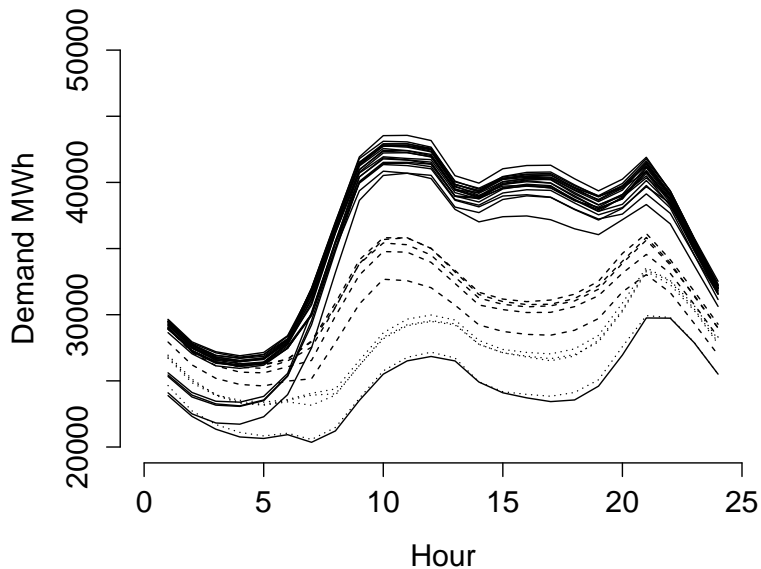


Figure 2.6 IPEX: Daily demand curves for the period 1/4/2011 - 30/4/2011. Solid lines: weekdays; dashed lines: Saturdays; dotted lines: Sundays. Solid line at the bottom: bank holiday (25th April).

not-working day) and any other not-working day. A steep decline on late Friday and a steep increasing on Monday morning are commonly observed in load data. This feature is also called “weekend effect” and is also present during other not-working days. This characteristic often affect electricity prices as the prices on Saturdays, Sundays and other not-working days are relatively unstable compare to other days. Annual periodicity is often referred to the seasonal fluctuations caused by variation in temperature and length of day. As can be seen from Figure 2.7, the electricity consumption is higher during summer and winter due to the growing use of air conditioning and heating, respectively, and is lower in autumn and spring. In fact, atmospheric conditions such as wind velocity, cloud cover, humidity, pre-

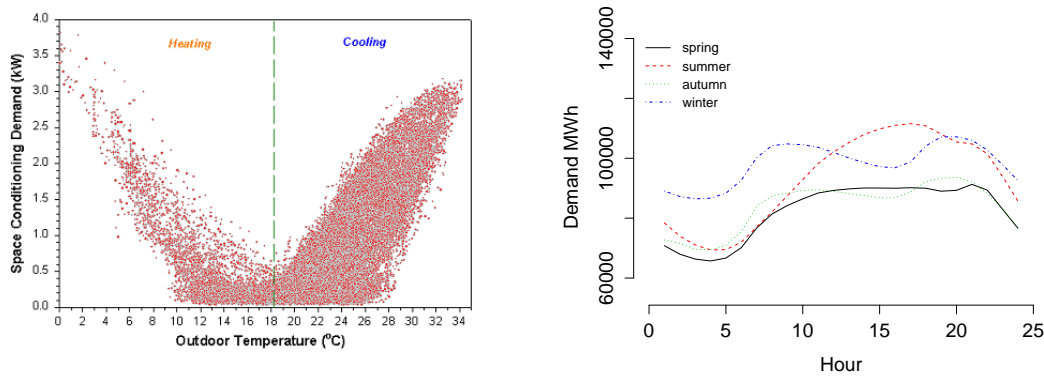


Figure 2.7 (left): Temperature Vs electricity demand (source: Parker (2003)).(right) IPEX: Average daily electricity demand in each season for 2014.

cipitation, rainfall and snowfall not only originate the yearly cycle but also explain the short term variation in electricity demand. In general, electricity demand and atmospheric temperature hold strong nonlinear relationship as can be seen from the Figure 2.7. In addition, the prolong use of artificial lights also assert to the demand increase in winter.

2.3.2 Volatility, Outliers and Jumps

Electricity demand often contain few outliers however; the price series show high volatility and unexpected jumps (also called spikes) due to meteorological, economical, technical and other influential factors. Figure 2.8 shows an example of outliers and volatility in demand

and prices data respectively. Price volatility is known as one of the most pronounced features and a direct consequence of electricity market liberalization. Electricity spot prices are highly volatile and the price can vary extremely within a short period of time. In fact, the

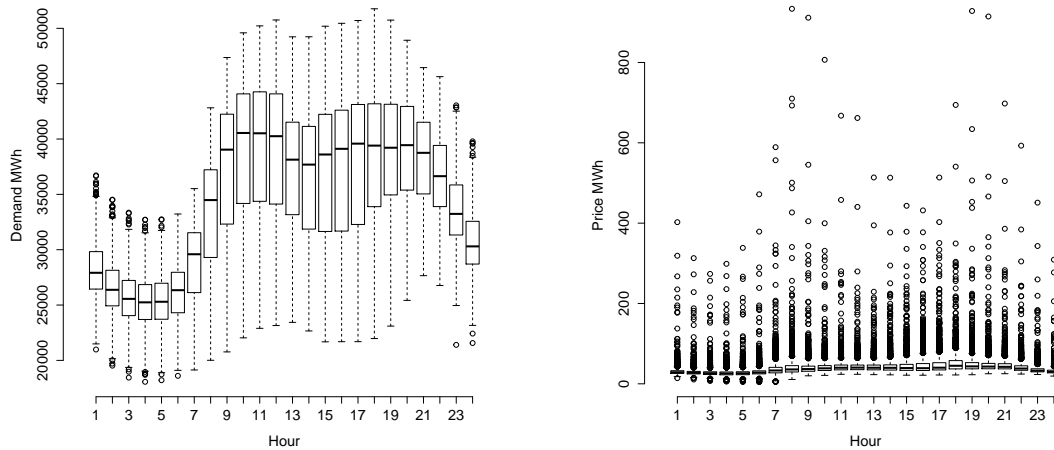


Figure 2.8 (left) IPEX: Box plots for hourly demand for the period 01/01/2009 - 31/12/2014. (right) PJM: Box plots for hourly prices for the period 01/01/2009 - 31/12/2014

volatility is by far stronger for electricity prices compared to any other financial commodity. Price spikes or jumps that are known as short lived, abrupt and generally unanticipated extreme price changes are commonly observe in electricity price series. In Figure 2.9 (left), an example of this feature is given when the spot prices increases substantially to many folds of its normal value and then drops back to the previous level soon after. Generally, these price spikes are short lived and much more extreme in magnitude. To understand well the reasons of these spikes, one should remember that electricity markets have distribution and transmission constraints that make them different from other commodity markets. Electricity cannot be economically stored and it has capacity and transmission constraint as well as the system must be balanced in real times. Any temporary imbalance in supply and demand due to any influential factor or technical reasons can cause price spikes. An important market structure element that plays vital role in market price determination is the diversity of generation plants and their corresponding marginal costs per unit of production. A schematic supply stack corresponding to different sources of energy with two potential demand curves

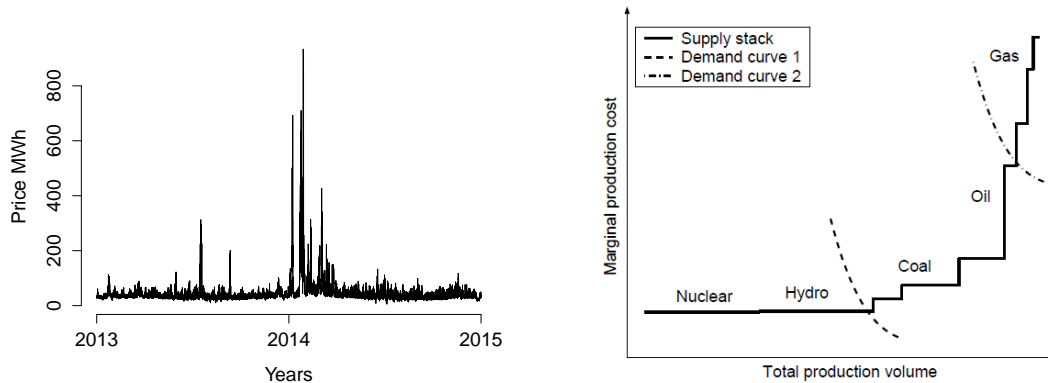


Figure 2.9 (left) PJM: Hourly electricity spot prices for the period 01/01/2013 - 31/12/2014. (right) A schematic supply stack with superimposed two potential demand curves (source Weron et al. (2004b))

superimposed is given in Figure 2.9 (right). As can be seen from the graph when the demand is low, electricity is produced and supplied from low marginal cost sources. As soon as the demand increases, the marginal production cost increases since the more expensive fuels plants start operations. Even a small increase in electricity demand can force prices to increase substantially. Once the cause of spike goes away, the prices fall back to their average level. Price spikes are non constant and are highly variable with respect to time scale. In general, they occur during peak load hours when the electric consumption is high.

2.3.3 Non-normality and Non-stationarity

In most electricity markets, the distributional properties of the spot electricity price series appear non-normal and highly positively skewed. For instance, Figure 2.10 shows these features for the PJM market for the period 01/01/2009 to 31/12/2010. The histogram shows positive skewness suggesting the greater likelihood of large price increases than price falls. Some authors suggest that the leptokurtic or heavy tailed feature indicates of inverse leverage effect. This means that positive jumps in prices amplify the conditional variance of the underlying process more than negative ones. On the other hand, extensive literature argued about the possible non stationarity of the demand series. In general, it has been widely observed and described that electricity demand series are non-stationary. Apart from other

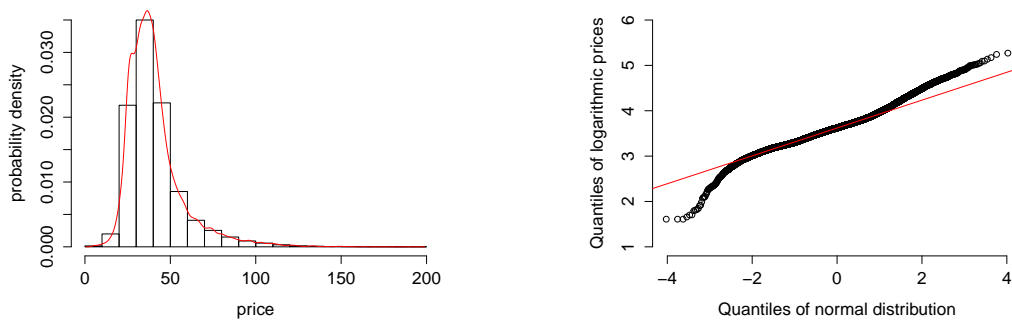


Figure 2.10 PJM: hourly electricity spot prices for the period 01/01/2009 to 31/12/2010. (left) Normalized histogram with superimposed nonparametric density in red (right) quantile-quantile plot

features it exhibit, electricity demand shows an overall trend due to the country economic situation, atmospheric changes, technological advancement and other related factors. For example, demand data for APX and PJM markets are plotted in Figure 2.11 that shows an overall trend. Generally, the trend can be increasing/decreasing and linear or nonlinear. In the case of APX, one can see that data exhibit a linear trend where for PJM, a nonlinear trend is more appropriate. In some cases, structural breaks or level shifts (see for example Figure 2.11) are also observed in demand series that are generally resulted from the expansion of the market or by the introduction of new regulatory laws.

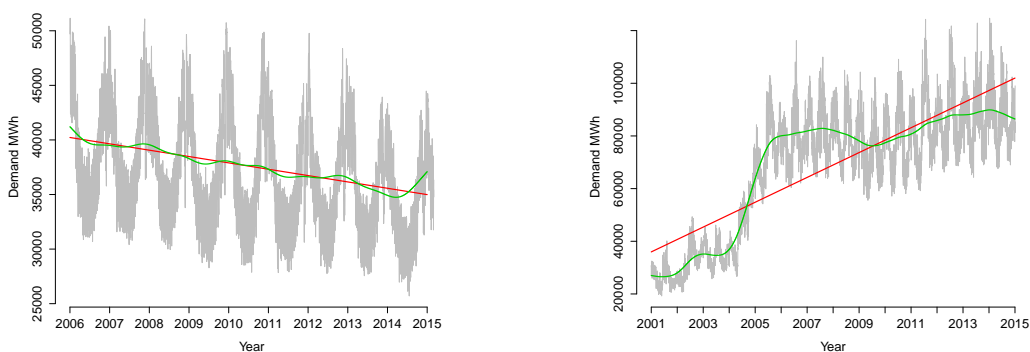


Figure 2.11 Daily electricity demand for (right) APX, for the period 01/01/2006 - 31/12/2014 and (left) PJM, for the period 01/01/2001 - 31/12/2014 with superimposed linear (red) and a nonlinear (green) trend.

2.3.4 Mean Reversion and Other Features

In general, electricity prices are regarded to be mean-reverting. Mean reversion is a process refers to a stochastic process that displays a tendency to remain near or to revert to its historical mean value. In other words, this process suggests that prices or returns eventually move back towards the overall mean of underlying commodity. As explained in section 2.3.2, in electricity market any temporary imbalance in supply and demand can cause a price spike. However, once the cause of spike goes away, the prices fall back to their average level suggesting strong mean reversion characteristics in price series (see for example Figure 2.12). On the other hand, in some markets (e.g. the French and German/Austrian day-ahead mar-

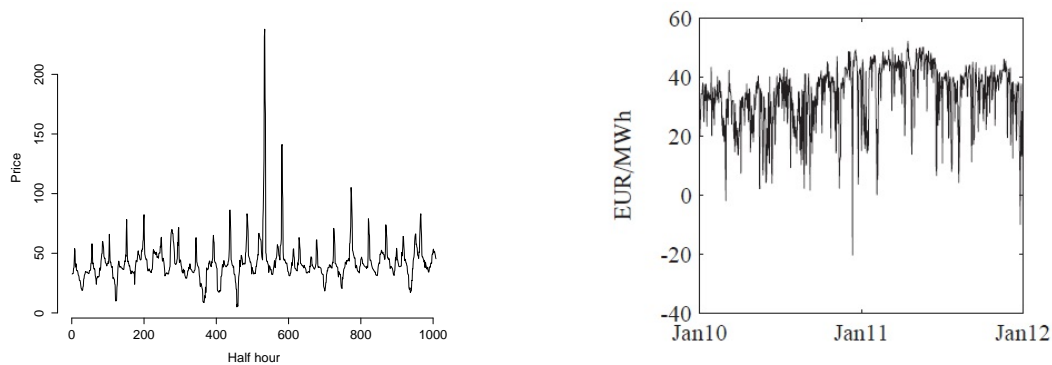


Figure 2.12 (left) APX: Half-hourly electricity prices (right) Hourly electricity prices for European Energy Exchange (source Erni (2012))

ket) electricity prices can turn negative when a high inflexible generation hit a low demand. Inflexible power resources (e.g. Nuclear) cannot be shut down and restarted in a fast and cost efficient manner. In case of low demand, prices fall signalling generators to reduce production in order to avoid overloading of the grid. In this case, often generators accept negative prices as it is less expensive to keep power plant online than to shut down.

Chapter 3

Literature Review for Electricity

Demand and Prices

Accurate modeling and forecasting electricity demand and prices are very important issues for decision making in deregulated electricity markets. Before the liberalization, electric utilities and researchers were mainly interested in demand forecast as variation in prices were minimal and considered after regular intervals. Different techniques were developed to describe and forecast the dynamics of electricity load, see for instance (Bunn and Farmer, 1985; Gross and Galiana, 1987; Moghram and Rahman, 1989, and references therein). On the other hand, restructuring of electric sector has increased the need of efficient and accurate forecasting for both demand and prices due to market structure. As electricity markets are not equally developed across countries, extensive studies have been made to model and forecast different variables related to these markets (Bunn, 2004a; Hahn et al., 2009; Weron, 2007).

In general, different forecasting horizons are described in the literature that are used for different purposes. For the efficient management of day-to-day operations of the power system, short term forecasts are important (Kyriakides and Polycarpou, 2007; Studiorum, 2007; Yang, 2006) whereas for generator maintenance, acquiring energy resources, balance sheet calculations, new structural investment based on the anticipation of future demand or for long term bilateral contracts to ensure low financial risk, medium and long term fore-

casts are very important (Boisseleau, 2004; Feinberg and Genethliou, 2005; Gellings, 1996; Gonzalez-Romera et al., 2006; Wang et al., 2012). This chapter will mainly focus on short term forecast as it received higher attention in the literature and also for congruency with the general scope of this thesis.

Before describing different techniques used for modeling and forecasting electricity demand and prices, a short description of the methods used to describe different components (see section 2.3) corresponding to market specific features are given below. Indeed, short term forecast proved to be very challenging task due to these specific features. In the literature, different methods have been discussed to account for these specific features effects before modeling the demand/price series in order to achieve stationarity and minimizing distorting effects on forecasting (Lisi and Nan, 2014). These effects are either modeled in a deterministic or in a stochastic way. In the first approach, the multiple periodicities in the data are often described by polynomial functions (Bordignon et al., 2013; Veraart and Veraart, 2012), piecewise constant functions or dummies (Espinoza et al., 2005; Fan and Hyndman, 2012; Raviv et al., 2013) or sinusoidal regression (Fanone et al., 2013; Weron et al., 2004b), sometime, equipped also with linear trend for the long term dynamics (Escribano et al., 2011; Fleten et al., 2011; Gianfreda and Grossi, 2012; Lucia and Schwartz, 2002; Soares and Medeiros, 2008). In some cases, splines, wavelet decomposition, moving averages and state space models are used to model different components (Cho et al., 2013; De Livera et al., 2011; Dordonnat et al., 2010; Goude et al., 2014; Janczura and Weron, 2010; Nedellec et al., 2014; Schlueter, 2010; Trueck et al., 2007). In the stochastic approach, components are viewed as stochastic processes. Some authors suggest modeling the long term dynamics by a random walk or Brownian motion with the assumption of unit root while others also treated the seasonal components as stochastic (Bosco et al., 2010, 2007; Koopman et al., 2007; Pappas et al., 2008; Soares and Souza, 2006). The stochastic approach is widely used in the case of price prediction and especially for the case of spikes/jumps that are often modeled by diffusion processes with Poisson jumps or by Markov-switching models (Borovkova and Permana, 2006; Hellström et al., 2012; Pirino and Renò, 2010; Weron et al., 2004a,b). Lastly, it is worth mentioning that in both cases, deterministic and stochastic, the authors

who modeled the specific calendar effects e.g. bank holidays, bridging effects etc. considered dummies (Bisaglia et al., 2010; Sahay and Tripathi, 2014). Once these components are estimated, the residuals (stochastic) component is obtained by subtracting them from original (unadjusted) demand/price time series, whose dynamics is modeled using different models with increasing level of complexity.

For modeling of residual part, two approaches are generally considered: univariate and multivariate. Since an individual auction is held for each load period and the demand/price pattern is quite different across the different days of the week, the first approach treats each load period separately, consequently, (48)24 (half-)hourly models, reflecting the incorporation of the daily total serie (Ramanathan et al., 1997). However, the daily demand/price profile suggests the presence of correlation among different load periods within a day that can be used when modeling the series and thus leads to a multivariate approach. For both approaches, various techniques have been proposed in the literature, see for example (Weron, 2014, and references therein). These methods can be broadly classified into two group: statistical models and methods, that follow more classical approaches, and techniques from the field of artificial and computational intelligence that are more flexible in handling complexities and non-linearities (Hahn et al., 2009).

3.1 Statistical Models and Methods

Statistical models such as regression models, exponential smoothing and time series models are widely used for electricity price and demand forecasting problems. Generally statistical models require an explicit mathematical expression that defines the relationship between demand/price and several input factors. These models generally perform better in case of short term forecasting and are extensively studied (Kim et al., 2002; Kyriakides and Polycarpou, 2007; Taylor and McSharry, 2007; Weron and Misiorek, 2005).

Regression models e.g. multiple regressions are used to learn the relationship between the variable of interest and other independent variables. In our specific context, they model the relationship between the current demand/price level with other external factors e.g. tem-

perature, calendar conditions, fuel prices etc. (Bianco et al., 2009; Charlton and Singleton, 2014; Hong et al., 2010; Hor et al., 2005; Karakatsani and Bunn, 2008; Nan, 2009). In general, regression models are easy to implement and the interpretation is easy to comprehend. However some authors argued about the inherent problems in identification of the correct model that often resulted from complex nonlinear relationship between price/demand and other influential factors (Kyriakides and Polycarpou, 2007).

On the other hand, modeling techniques based on exponential smoothing are very famous in time series as they are widely used to model multiple periodicities in the data. In this approach, the variable of interest is predicted as an exponentially weighted average of the past observations. This method uses a weighting factor known as the smoothing constant (parameter) that reflects the weight given to the most recent data values. Smoothing constant value lies between 0 and 1 and it determines the degree of smoothing consequently how responsive the model is to the fluctuations in the data. In electricity context, these techniques have been widely used to model and predict electricity demand and prices (Carpio et al., 2014; De Livera et al., 2011; Taylor, 2010, 2012; Taylor et al., 2006). Exponential smoothing techniques are generally easy to perform and with relatively low computational cost. In the short term forecast they generally produce better results as they give greater weights to the more recent observations. However the forecast is relatively sensitive to the specification of smoothing constant.

Time series approaches have been extensively used in forecasting for decades (Box and Jenkins, 1970). These methods are among the oldest methods applied to both, price and demand forecasting (Contreras et al., 2003; Hagan and Behr, 1987; Hahn et al., 2009). In general, these methods are capable to account for both, the time correlation and random nature existing in the underlying phenomenon. Autoregressive moving average (ARMA) model and its different extensions like seasonal ARIMA, ARIMA with exogenous variables (ARIMAX), autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) models are often used classical time series methods in demand and price forecasting problems (Amjady, 2001; Conejo et al., 2005; Espinoza et al., 2005; Garcia et al., 2005; Hao, 2007; Knittel and Roberts, 2005; Kristiansen,

2012). To account for the possible correlation among different load period, vector autoregressive (VAR) models are also used to forecast demand and price series (Raviv et al., 2013; Studiorum, 2007). Although time series models are considered very flexible at handling different time series patterns however they generally have no structural interpretation.

3.2 Artificial Intelligence Techniques

Artificial intelligence methods are considered as nonparametric techniques and are always attractive for researchers due to their flexibility to functional form specifications, non-linearity and detection of structures that are usually undetected by traditional parametric methods. These methods generally map the input output relationship without exploring the underlying process. These techniques mainly comprised of artificial neural network (ANN), fuzzy systems, support vector machines (SVM) and evolutionary computation.

Artificial neural networks (ANN) are one of the most popular computational intelligence techniques that are widely used for electric load and price forecasting. These models are made up of various architectures of highly interconnected processing elements called neurons. Neural networks learn by example and relate a set of input variables to a set of output variables in such a manner that it achieve complicated input-output mappings without explicit programming and extract relationships both, linear and nonlinear between data sets presented during a learning process. In practice, choices concerning the number of architectures e.g., Boltzmann machine, Hopfield, back-propagation, connectivity of layers, and uni- or bi-directional links among inputs and outputs are made before apply ANN models to electricity data. These techniques are frequently used in the case of load forecasting (Dai and Wang, 2007; Hayati and Shirvany, 2007; Hippert et al., 2001; Kandil et al., 2006; Zhang et al., 1998) however many applications can be found in case of price forecasting as well (Pao, 2007; Singhal and Swarup, 2011). Generally statistical models assume linearity and/or normality of the underlying process whereas ANN do not assume any assumptions and are more flexible however they required high processing time for large neural networks.

on the other hand, Fuzzy system (FS) and support vector machines (SVM) are extensively used for electricity demand and price time series forecasting problems (Amjady, 2006; Chen et al., 2004; Niu et al., 2007; Pai and Hong, 2005; Ranaweera et al., 1996). A fuzzy system is based on fuzzy logic that analyzes analog input values on continuous values between 0 and 1. SVM maps the data into a high dimensional space using nonlinear mapping and then perform simple linear functions that create linear boundaries in the new space. Both these methods are flexible in adopting any complex nonlinear structure in the data.

Beside these methods, a relatively new modeling technique based on functional data analysis, where the ideal unit is a function defined on bounded interval, is also used for demand and price prediction (Aneiros-Pérez et al., 2011; Jonas Andersson, 2010; Vilar et al., 2012). In this approach, the daily demand profile is used as a single functional object allowing to use additional information (e.g. smoothness, derivatives) contained in the functional structure of the data. Functional data approach is very popular in different scientific fields, but little explored in time series context and (Ferraty and Vieu, 2006; Ramsay et al., 2009) are comprehensive references for nonparametric and parametric functional data analysis.

Chapter 4

Predictive Models

4.1 Introduction

In regression analysis, the main interest lies in finding how a response variable y depends on some explanatory variable $\mathbf{x} = (x_1, \dots, x_k)$. The problem herein is to find a measurable function $f(\mathbf{x})$ that, in some sense, is the best approximation of y . To understand well, consider a random sample $(y_i, \mathbf{x}_i)_{i=1}^n$ where \mathbf{x}_i is possibly a k -vector, i.e. $\mathbf{x}_i = (x_{1,i}, \dots, x_{k,i})$ and let the following holds:

$$\begin{aligned} y_i &= f(x_{1,i}, \dots, x_{k,i}) + \varepsilon_i, \quad i = 1, \dots, n \\ &= f(\mathbf{x}_i) + \varepsilon_i, \end{aligned} \tag{4.1}$$

where $f(\cdot)$ is a real valued function called regression function, $\varepsilon_i, i = 1, \dots, n$ is an error term satisfying $E[\varepsilon_i | \mathbf{x}_i] = 0$ and $\text{Var}(\varepsilon_i) = \sigma(\mathbf{x}_i)$ is a positive real valued function. The variable y_i is the response variable whereas \mathbf{x}_i is the explanatory variable. The main interest lies in finding the regression function $f(\cdot)$ such that this function minimizes the mean squared error. i.e.

$$E[y_i - f(\mathbf{x}_i)]^2 = \min_f E[y_i - \hat{f}(\mathbf{x}_i)]^2.$$

In general, the estimation of regression function can be done by two different approaches namely parametric and nonparametric. In the case of parametric modeling, the structure

of the regression function is known (e.g. linear) up to a finite set of parameters. In general, parametric models are estimated fast (by low computational cost) and they provide good results even for small sample sizes in case of well chosen model. On the other hand, nonparametric modeling approach does not assume any specific structure for the regression function allowing the data to “speak for itself”. In this case the regression function cannot be described by finite number of parameters. Since nonparametric models require fewer assumptions about the data, they prove better in situations where the true distribution is unknown. In the following, different classes (univariate, multivariate and functional) of parametric and nonparametric models are presented that will also be used in the next chapters. Before going in details, For regression purpose some notations used are described below.

- y : generic real random variable.
- \mathbf{x} : generic multivariate random variable (finite dimensional)
- \mathcal{X} : generic functional random variable (infinite dimensional)
- χ : generic (non-random) functional element of \mathcal{X}

4.2 AutoRegressive Models

Linear AutoRegressive (AR) models are well-known parametric models that are commonly used for modeling univariate time series (Brockwell and Davis, 2006). In AR modeling, the variable of interest is modeled by its own p lagged values using linear regression techniques. The value p referred to the order of the model that determines the number of steps into the past needed to predict the current value. Thus an autoregressive model of order p can be written as

$$y_i = c + \sum_{j=1}^p \phi_j y_{i-j} + \varepsilon_i, \quad (4.2)$$

where y_i is a stationary series, c and ϕ_j ($j = 1, \dots, p$) are constant parameters and $\varepsilon_i \sim D(0, \sigma_\varepsilon^2)$ is a white noise process. We refer to model 4.2 as an AR(p) model. Depending on

the distributional assumptions, parameters can be estimated by ordinary least squares (OLS) or by maximum likelihood (ML) methods.

An equivalent representation of the above model using backshift operator B is given by

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_j B^j) y_i = c + \varepsilon_i ,$$

where the back shift operator B^p is defined as

$$B^p y_i = y_{i-p} .$$

In case of stationarity, the statistical properties such as mean, variance, autocorrelation of y_i are all constant over time. If y_i is not stationary, a commonly used approach called “differencing” that can help stabilize the mean of y_i by removing changes in the level of y_i , and so eliminating trend and seasonality, is applied to y_i . The d time differenced series y_i^d is defined as

$$y_i^d = y_i - y_{i-d} .$$

Finally, the generic AR(p) model for the d times differenced series is commonly written as

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y_i = \phi_0 + \varepsilon_i . \quad (4.3)$$

In general, AR models are considered very flexible at handling different time series patterns. Changes in the values of parameters $\phi_0, \phi_1, \dots, \phi_j$ results in different time series pattern whereas the variance of ε_j only affect the scale of time series. Note that model 4.2 can be expressed in terms of past forecast errors in a regression-like model by repeated substitution of y_i . For example, for AR(1) process the repeated substitution yields

$$y_i = \phi y_{i-1} + \varepsilon_i = \dots = \phi^q y_{i-q} + \sum_{k=0}^{q-1} \phi^k \varepsilon_{i-k}$$

Provided $|\phi| < 1$ and continuing to iterate backward, we can write an AR(1) model as a linear process given by

$$y_i = \sum_{k=0}^{\infty} \phi^k \varepsilon_{i-k} \quad (4.4)$$

Model 4.4 is called moving average model and denoted by $MA(\infty)$ as the order of the model is ∞ . Generally, the order of MA model is fixed in prior, say q , and is written as:

$$y_i = \sum_{k=0}^q \phi_k \varepsilon_{i-k} \quad (4.5)$$

If we combine an AR and a MA model, we are end up with autoregressive moving average (ARMA) model. Further, if the differentiation is done before applying the ARMA model, the model become ARIMA where “I” is an acronym for integration. The full model is written as ARIMA(p,d,q) and given in the following form:

$$y_i^d = \sum_{j=1}^p \phi_j y_{i-p}^d + \sum_{k=1}^q \theta_k \varepsilon_{i-k} + \varepsilon_i \quad (4.6)$$

where y_i^d is the d times differenced series, ϕ_j ($j = 1, \dots, p$) and θ_k ($k = 1, \dots, q$) are AR and MA parameters respectively. In general, the order of p and q is selected by using autocorrelation function (ACF) and partial autocorrelation function (PACF) or by different information criteria.

4.3 Nonparametric AutoRegressive Models

Model 4.2 can be generalized relaxing the linearity by considering NonParametric AutoRegressive (NPAR) models. In this case the relation between the variable of interest and covariates has not a specific parametric form allowing, potentially, any kind of nonlinearity. To avoid the so-called curse of dimensionality, that refers to the exponentially decay of precision by increasing the dimension of regressor (Wasserman, 2006), usually an additive form

is considered that assumes no interactions among the explanatory variables. i.e.

$$\begin{aligned} y_i &= f(\mathbf{x}_i) + \varepsilon_i \\ &= f_1(x_{1,i}) + \cdots + f_j(x_{j,i}) + \varepsilon_i, \end{aligned} \quad (4.7)$$

where $f_j(j = 1, \dots, k)$ are arbitrary smoothing functions describing the existing relations between the response and explanatory variables and ε_i is an error term. Here it is worth mentioning that covariates x_j 's can refer to both an exogenous and a lagged variable in (4.7).

Generally, when more explanatory variables are added, the interpretation of a nonparametric model becomes impossible in addition with the curse of dimensionality issue. This leads to additive regression models that essentially apply local regression to low dimensional projections of the data. As each component is estimated separately, they are the analogues of the coefficients in linear regression.

Functions f_j in (4.7) can be described and estimated in several ways referring to different methods (Hastie and Tibshirani, 1990). In this work, they are described by penalized cubic smoothing spline functions and estimated by back-fitting (iterative procedure) algorithm. Splines generally provide a flexible way of estimating an underlying regression function. They are piecewise polynomial functions of a fixed degree, say k , that are continuous and has continuous derivatives of orders $1, \dots, k - 1$ at their knots points (knots are a vector of interior points specified on the variable, see also section 4.5.1.2). In this modeling approach, the choice for the degree of polynomial function is important as well as number and the placement of knots (Rice and Rosenblatt, 1983). A common choice for the degree is 3 refers to cubic splines. For the other two issues, penalized cubic smoothing splines are commonly used as they circumvent the problem of knots selection (as they place knots at all input points) simultaneously controlling for over fitting by shrinking the coefficients of the estimated function. More precisely, we directly solve for the function $f(\cdot)$ by choosing

the coefficients that minimizes the following objective function:

$$L(m, \lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - f(\mathbf{x}_i)\}^2 + \lambda \int \{f''(\mathbf{x})\}^2 d\mathbf{x}. \quad (4.8)$$

Here the first term refers to the goodness of fit while the second penalizes roughness. The term f'' refers to the second derivative of f with respect to \mathbf{x} . The term λ is a smoothing parameter controlling the degree of penalization and often selected by cross validation techniques. The solution to this minimization problem is given by

$$\hat{f}_\lambda = \underset{f}{\operatorname{argmin}} L(f, \lambda). \quad (4.9)$$

4.4 Vector Autoregressive Models

In multiple time series context, vector autoregressive (VAR) models are perhaps the most popular and widely used models able to account for linear relationships among different time series. Unlike univariate, VAR is a multivariate modeling technique that considers multiple equation system or a multiple time series generalisation of AR models.

In VAR models, each variable is a linear function of past lags of itself and of the other variables taking into account the interdependence among variables included in the model. The vector autoregressive model of order p , denoted as VAR(p), is as follows:

$$\mathbf{x}_i = \Phi_1 \mathbf{x}_{i-1} + \dots + \Phi_p \mathbf{x}_{i-p} + \boldsymbol{\varepsilon}_i \quad (4.10)$$

where \mathbf{x}_i is a multivariate random variable, Φ_j ($j = 1, \dots, p$) are coefficient matrices and $\boldsymbol{\varepsilon}_i \sim N(0, \Sigma_\varepsilon)$ is a vector of innovations that may be contemporaneously correlated. As in case of AR, parameters can be estimated by OLS or ML methods.

In case of stationary series, VAR is fitted directly to the data otherwise differentiation are made before fitting a VAR model. In general, two choices to be made in prior using a VAR model to forecast. The first one corresponds to the number of variables, say j , whereas the second is the number of lags, say p , to be included in the system. Thus, the total number of

coefficients to be estimated in a VAR model is equal to $j + j^2 p$. In practice, small number of j is preferred and often correlated variables included only to the model as they are useful in forecasting each other. Generally, cross validation techniques and different information criteria are commonly used for the selection of number of lags. Apart from the fact that VAR models provide a systematic way to capture rich dynamics of the given multiple time series, they become difficult to estimate when the number of variables get higher. To overcome this issue, a relatively new technique that depends on functional data analysis is described in the following.

4.5 Functional Data Analysis

Due to technological advances in recent years, the data collection process became increasingly sophisticated that leads to increasingly complex, high dimensional and structured data. Nowadays in many research areas data are often densely sampled over time, space or any other continua. A large fraction of these data are characterized as functional data as they usually reflect the influence of the underlying smooth function that generates the observations. This phenomenon lead to the rapid development of the relatively new area called functional data analysis (FDA).

Generally, statistical models combine information either across or within sample units to make inference about the population, FDA considers both. The main idea behind FDA is to consider each function as a single structured object instead of a collection of data points. This in turn bypass the problem of the number of variables and, in addition, allow to use additional information (e.g. smoothness, derivatives) contained in the functional structure of the data.

Functional data analysis (FDA) considers data that are curves, surfaces or any other mathematical object varying over a continuum. Functional data are observed at discrete sampling values which may or may not be equally spaced. These data are converted to functional objects using basis functions in such a way that it resemble original data with some smoothing that eliminates the most obvious noise. Once the data are converted to functional form,

it can be evaluated at all values over some interval. In the following, basis functions are described in details as they are considered the main building blocks for FDA.

4.5.1 Basis Functions

The key building blocks for functional data analysis are basis functions. A basis is a standard set of functions such that their linear combination defines the actual functional observations. i.e. A functional observation χ is constructed as a linear combination of basis functions as

$$\chi(j) = \sum_{k=1}^K \lambda_k \zeta_k(j), \quad \forall j \in T \quad (4.11)$$

where λ_k are constant parameters and $\zeta_k(j)$ are basis functions. Note that the use of finite number of basis coefficients also allows to yields estimates and inference in an infinite dimensional (functional) space. The commonly used basis functions include Fourier, B-spline, polynomial, exponential and principle components. The choice of an appropriate basis function depends on the characteristics of given data. In the following, Fourier and B-spline basis are explained in details as they are considered the most important and widely used in case of periodic and non-periodic data, respectively, and also have been used in this thesis.

4.5.1.1 Fourier Basis

Fourier basis are commonly used for periodic functions defined on some interval, say T . These basis are a linear combination of sin and cosines function of increasing frequency where λ_k 's take the following form:

$$\begin{aligned} \lambda_0(j) &= 1/\sqrt{|T|} \\ \lambda_{2l-1}(j) &= \frac{\sin(l\omega j)}{\sqrt{|T|/2}} \\ \lambda_{2l}(j) &= \frac{\cos(l\omega j)}{\sqrt{|T|/2}}, \end{aligned}$$

where $l = 1, \dots, (k-1)/2$ and k is an odd number representing number of basis functions. The frequency ω determines the length and period of the interval $|T| = 2\pi/\omega$. Note that the number of basis is critical as it trade-off between bias and sampling variance. Small number of basis functions unable to capture interesting features of the curves resulting large bias but small sampling variance. In contrast, large number of basis functions over-fits the data resulting in small bias but large sampling variance. Cross validation techniques are often used to choose an appropriate number of basis. An advantage of using Fourier basis is its fast computation by Fast Fourier Transformation algorithm. Another important features is related to the existence of continuous derivatives as the basis involve sin and cosine functions.

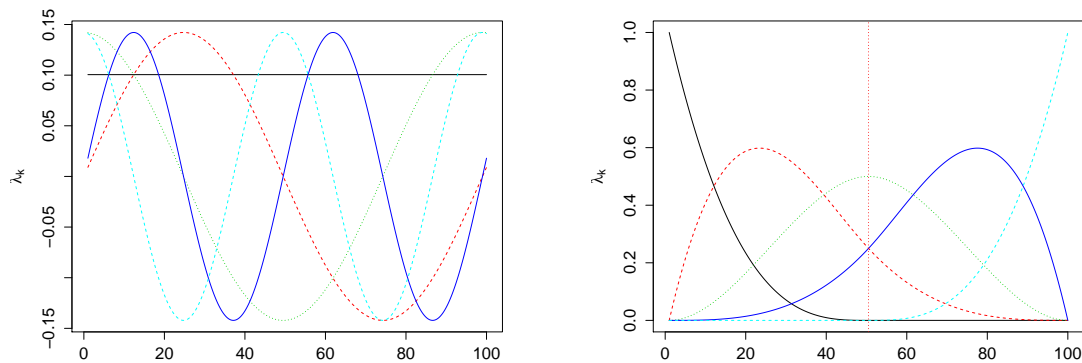


Figure 4.1 An example of Fourier (left) and B-spline (right) expansion with $k = 5$ basis functions.

4.5.1.2 B-spline Basis

B-splines or basis-splines are generally used for non-periodic data. They are defined on a sequence of knots in the interval T and are constructed by piece-wise from polynomial functions of order z that are smoothly connected at each knots. The degree of the B-spline polynomial is z minus one. More formally, for degree d and a non decreasing knots sequence

q_0, q_1, \dots, q_m , the basis functions are defined as:

$$\zeta_\tau(j) = N_{\tau,z}(j), \quad \tau = 1, \dots, m+l-1.$$

where $N_{\tau,z}(j)$ is the τ -th B-spline of order z build by following recursive scheme:

$$N_{\tau,0}(j) = \begin{cases} 1, & \text{if } q_\tau < j < q_{\tau+1} \\ 0, & \text{otherwise} \end{cases}$$

$$N_{\tau,l}(j) = \frac{j - q_\tau}{q_{\tau+l-1} - q_\tau} N_{\tau,l-1}(j) + \frac{q_{\tau+l} - j}{q_{\tau+l} - q_{\tau+1}} N_{\tau+1,l-1}(j).$$

for $\tau = 1, \dots, m+l-1$ where $l = 0, \dots, k$. The B-spline order and number of knots determine the number of basis functions. The B-spline function of degree d is a parametric curve computed as a linear combination of B-splines basis given by

$$\chi(j) = \sum_{\tau=1}^k \lambda_\tau N_{\tau,z}(j) \quad \forall \quad j \in T$$

B-spline basis are widely used in many interdisciplinary areas due to their greater flexibility, relatively fast evaluation of the basis functions and the existence of their derivatives (Lowther and Shene, 2003).

Once the observed discrete data is converted to smooth functions with the help of basis functions, different class of functional models can be used for regression purposes. In the following, different functional models are explained in detail that are also used for electricity demand and price prediction in the following chapters.

4.5.2 Functional AutoRegressive Models

The functional version of the classical linear parametric regression model was first introduced by Ramsay and Dalzell (1991). It will be referred as Functional AutoRegressive (FAR) model. With this approach, a scalar response y_i is predicted by functional covariates

\mathcal{X}_i using a linear relationship given by

$$\hat{y}_i = \int \beta_j \mathcal{X}_i \partial j + \varepsilon_i, \quad i = 1, \dots, n \quad \text{and} \quad j \in T \quad (4.12)$$

where \mathcal{X}_i is the functional predictor, β is a functional coefficient evaluated at j points and ε_i is random errors with mean zero, finite variance σ^2 and $E[\mathcal{X}_i \varepsilon_i] = 0$. A common approach to estimate the model is to represent both functional parameter and functional predictor in basis functions. i.e.

$$\begin{aligned} \mathcal{X}_i &= \sum_{k=1}^{K_1} \lambda_{ik} \zeta_{ik}(j) = \lambda' \zeta(j) \\ \beta_j &= \sum_{k=1}^{K_2} b_{ik} \gamma_{ik}(j) = B' \gamma(j). \end{aligned}$$

Then, the functional linear model is estimated by the expression:

$$\hat{y}_i = \langle \mathcal{X}_i, \beta_j \rangle = \lambda' \zeta(j) \gamma'(j) \hat{B} = \tilde{X} \hat{B}$$

where

$$\tilde{X} = \lambda' \zeta(j) \gamma'(j) \quad \text{and} \quad \hat{B} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' y_i$$

and

$$\hat{y}_i = \tilde{X} (\tilde{X}' \tilde{X})^{-1} \tilde{X}' y_i = H y_i$$

where H is the hat matrix with degrees of freedom equal to $tr(H)$. Other details about this model can be found in Ramsay et al. (2009).

4.5.3 Nonparametric Functional AutoRegressive Models

The previous model can be generalized by considering a nonparametric specification that allows to model the (scalar) response as a nonlinear function of functional covariates, provided that both satisfy some strong mixing conditions (Rosenblatt, 1956). Denoting by nonparametric FAR (NPFAR), the following functional based on a Nadaraya–Watson-type

kernel estimator can be used to predict the scalar response y_i is by functional predictor \mathcal{X}_i

$$\hat{f}(\chi) = \hat{y}_i = \frac{\sum_{i=1}^n y_i K(b^{-1}\xi(\chi, \mathcal{X}_i))}{\sum_{i=1}^n K(b^{-1}\xi(\chi, \mathcal{X}_i))} \quad (4.13)$$

where $K(\cdot)$ is a priori real chosen function on R^n called the kernel satisfying

$$\int K^2(u) \partial u < \infty \quad \text{and} \quad \int K(u) \partial u = 1$$

and y_i is the scalar response, \mathcal{X}_i is a functional random variable valued in some semi-metric space (E, ξ) and χ is a fixed element of ξ . b is a smoothing parameters (bandwidth) corresponding to the optimal number of neighbours selected by a cross validation procedure:

$$R_{-i}(\chi) = \frac{\sum_{g=1, g \neq i}^n y_g K(b^{-1}\xi(\chi, \mathcal{X}_g))}{\sum_{g=1, g \neq i}^n K(b^{-1}\xi(\chi, \mathcal{X}_g))}$$

and

$$\begin{aligned} \text{CV}(b) &= \sum_{i=1}^n (y_i - R_{-i}(\chi)) \\ b_{opt} &= \underset{b}{\text{argmin}} \text{CV}(b) . \end{aligned}$$

Model 4.13 can also be written as

$$\hat{f}(\chi) = \sum_{i=1}^n y_i \omega_{i,b} ,$$

where

$$\omega_{i,b} = \frac{K(b^{-1}\xi(\chi, \mathcal{X}_i))}{\sum_{i=1}^n K(b^{-1}\xi(\chi, \mathcal{X}_i))} ,$$

which is a weighted average as

$$\sum_{i=1}^n \omega_{i,b} = 1 .$$

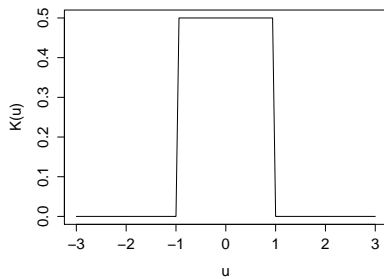
This kind of estimator is widely used in the finite dimensional case. Here the main difference comes from the semi-metric ξ which computes the proximity between curves, permitting to

avoid the “curse of dimensionality” problem. The proximity between the q -order derivatives of two curves χ_1 and χ_2 is computed as:

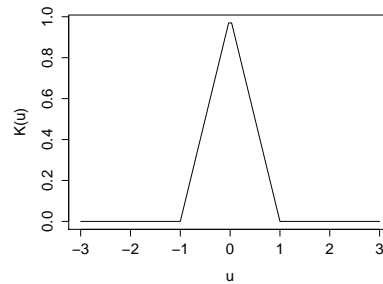
$$\xi_q(\chi_1, \chi_2) = \sqrt{\int (\chi_1^q(j) - \chi_2^q(j))^2 \partial j},$$

where χ_i^q denotes the q -th derivative of the curve χ_i . In general, the choice of semi-metric and bandwidth is more crucial than kernel. Some of the most commonly used kernel function are:

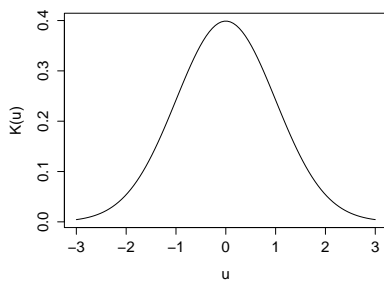
- Uniform kernel $K(u) = \frac{1}{2} \mathbb{1}_{[-1,1]}(u)$
- Triangle kernel $K(u) = (1 - |u|) \mathbb{1}_{[-1,1]}(u)$



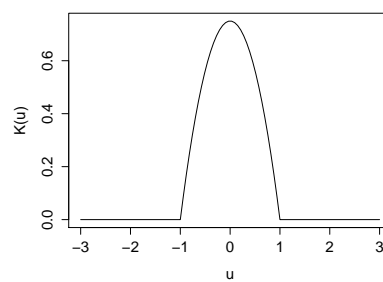
Uniform kernel



Triangle kernel



Gaussian kernel



Epanechnikov kernel

Figure 4.2 Common symmetrical kernels

- Gaussian kernel $K(u) = (2\pi)^{-1/2} \exp(-\frac{u^2}{2})$
- Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{[-1,1]}(|u|)$

Note that, in functional regression analysis, only an asymmetrical kernel are considered allowing only a positive input to the kernel function. This modification is based on the assumption that asymmetrical kernel function has its maximum (peak) at zero and decrease monotonically as the input increase. This assumption ensures that if the function of interest χ is close to \mathcal{X}_i , the response value y_i plays a more important role in the estimate of $\hat{f}(\chi)$ than when its far. Further note that as soon as $\xi(\chi, \mathcal{X}_i) > b$, we have $\omega_{i,b}(\chi) = 0$, hence in the estimation of $\hat{f}(\chi)$ is only taken into account among y_i 's those for which the corresponding χ_i 's are distant from χ of at most b . Further details can be found in Ferraty and Vieu (2006) along with many asymptotic properties.

4.5.4 Functional-Functional AutoRegressive Models

The section 4.5.2 reviewed a regression model with a scalar response but a functional explanatory variable. This section will generalize the previous one by allowing also the response variable to be functional; hence a completely functional regression model that has both response and explanatory variables functional, and will be discussed in time series context. The resulting autoregressive model of order one (FFAR(1)) in its general form can be written as:

$$\mathcal{X}_i = \rho(\mathcal{X}_{i-1}) + \eta_i, \quad (4.14)$$

where ρ is a linear parametric operator and η_i is a functional random error taking values in an infinite dimensional (functional) space. More specifically, FFAR(1) can be written as

$$\mathcal{X}_i = \int_s \Phi(s,j) \mathcal{X}_{i-1} \partial s + \eta_i(j), \quad j \in T \quad (4.15)$$

where $\Phi(s,j)$ is the functional parameter (regression function) that creates dependence similar to coefficient matrix in a first order VAR model. To estimate the model, we used the approach described in Bosq (2000) that deal with the covariance and cross covariance oper-

ators estimation of a Hilbert space H valued autoregressive process. This approach assume that the mean of the process is zero hence in practice, require demeaning the data. For simplicity, let \mathcal{X} be a zero-mean H -valued random variable such that $E\|\mathcal{X}^2\| < \infty$. Then the autocorrelation operator ρ can be written as:

$$\rho(\chi) = (DC^{-1})(\chi) \quad \chi \in H ,$$

where the covariance operator C is

$$C(\chi) = C_{\mathcal{X}_i, \mathcal{X}_i}(\chi) = E(\langle \mathcal{X}_i, \chi \rangle \mathcal{X}_i) ,$$

and D is the one-step cross covariance operator defined as

$$D(\chi) = C_{\mathcal{X}_i, \mathcal{X}_{i+1}}(\chi) = E(\langle \mathcal{X}_i, \chi \rangle \mathcal{X}_{i+1}) .$$

A natural estimator for C and D is the empirical covariance and cross covariance operator respectively defined as

$$C_n(\chi) = \frac{1}{n} \sum_{i=1}^n (\langle \mathcal{X}_i, \chi \rangle \mathcal{X}_i)$$

$$D_n(\chi) = \frac{1}{n-1} \sum_{i=1}^{n-1} (\langle \mathcal{X}_i, \chi \rangle \mathcal{X}_{i+1}) .$$

Since $C_n(\chi)$ behave erratically because of the unboundedness of C^{-1} in functional context, the data is projected to a suitable finite dimensional subspace. Generally the space spanned by the largest eigenvectors of C is used and the number of eigenvalues are determined by cross validation.

4.5.5 Nonparametric Functional-Functional AutoRegressive Models

A nonlinear version of the previous model is the NonParametric Functional response with Functional covariates AutoRegressive model (NPFFAR) that is based on a functional kernel

regression. This model can be written as

$$\mathcal{Y}_i = \varphi(\mathcal{X}_i) + \eta_i, \quad (4.16)$$

where φ is a possible nonlinear operator, \mathcal{Y}_i and \mathcal{X}_i both are functional variable and, again, the residual η_i is a functional random error. Note that \mathcal{X}_i can be a lagged variable or any other exogenous functional variable. The estimation of φ is based on kernel methods, more specifically

$$\hat{\mathcal{Y}}_i = \frac{\sum_{i=1}^n \mathcal{Y}_i K(b^{-1}\xi(\mathcal{X}, \mathcal{X}_i))}{\sum_{i=1}^n K(b^{-1}\xi(\mathcal{X}, \mathcal{X}_i))}, \quad (4.17)$$

where K is an asymmetrical kernel function and b is the bandwidth selected by cross-validation as

$$R_{-i}(\mathcal{X}) = \frac{\sum_{g=1, g \neq i}^n \mathcal{Y}_g K(b^{-1}\xi(\mathcal{X}, \mathcal{X}_g))}{\sum_{g=1, g \neq i}^n K(b^{-1}\xi(\mathcal{X}, \mathcal{X}_g))}$$

and

$$\text{CV}(b) = \sum_{i=1}^n (\mathcal{Y}_i - R_{-i}(\mathcal{X}))$$

$$b_{opt} = \underset{b}{\text{argmin}} \text{CV}(b).$$

The stated model satisfies many asymptotic properties that can be seen in Ferraty et al. (2012).

Chapter 5

Modeling and Forecasting Electricity

Demand and Price Time Series

5.1 Introduction

In competitive electricity markets, accurate modeling and forecasting electricity demand and prices are crucial for effective planning and operations of the power system. As electricity is a flow commodity, in the sense that it cannot be stored economically, over- or under-estimation of electricity demand can cause serious problems to electric utility providers, energy suppliers, system operators and other market participants. For example, overestimation can lead to excessive energy purchase or unnecessary production, resulting in a substantial financial loss. On the other hand, underestimation requires to balance the system at any cost and this can originate financial distress. Unlike for other stock commodities, in electricity market any shocks to supply or demand cannot be smoothed using pre stored inventory. Consequently electricity prices exhibit high volatility that leads to significant price risk for market participants. Price prediction is also important for cash flow analysis, financial procurement, least cost planning and optimal bidding strategies. For all these reasons, efficient modeling and forecasting of electricity loads and prices gained an increasing attention in last two decades (Weron, 2007).

The market participants are interested in electricity demand (and price) forecasting at differ-

ent horizons, that are commonly referred in the literature as short-, medium- and long-term. Although, literature describe no common time intervals for these forecasting horizons however in general, short-term forecasts refers to the forecast interval ranging from one hour to one week and are important to ensure the reliability and efficient operations of power grid, load management, scheduling and dispatching of generating units.

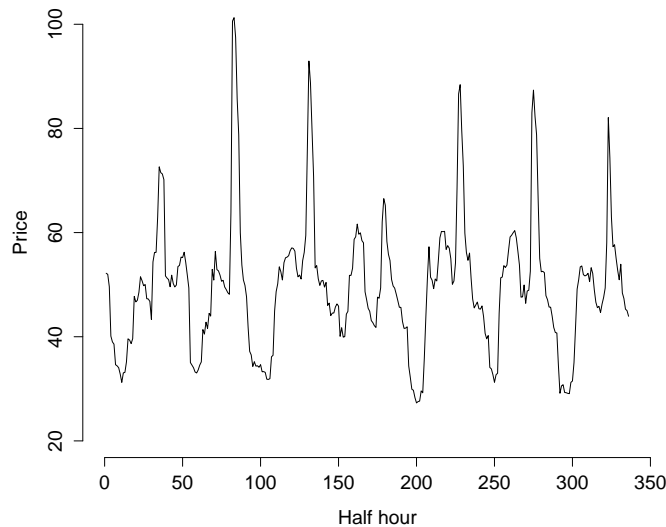


Figure 5.1 APX: Half-hourly price time series for the period 12/01/2014 - 18/01/2014.

However, forecasting electricity markets are not straight forward due to specific features these markets exhibit (see Chapter 2 for details). Medium-term forecast are generally ranging from a week to few months and are commonly used for scheduling unit maintenance, planning fuel procurement and revenue assessment. On the other hand, long-term forecast generally involved the prediction of monthly and yearly maximum/minimum and are often used for planning and risk management purposes. In the literature, short-term prediction received higher attention because in many electricity markets daily demand and prices are determined the day before the physical delivery by mean of (semi-)hourly auctions for the following day. Therefore, adequate forecasting results in less expensive, reliable and secure power operations and planning (Bunn, 2004a).

Demand time series generally show strong daily, weekly and annual cycles. Daily and

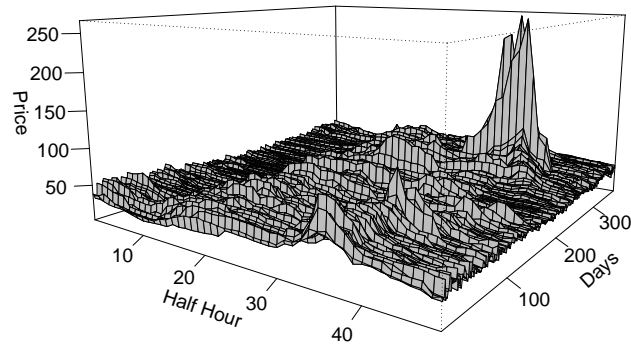


Figure 5.2 APX: Daily price curves for the period 01/01/2014 - 31/12/2014.

weekly cycles are also evident, in general, in price time series (see Figure 5.1) along with other attributes like non-constant mean and variance, jumps etc. As electricity consumption is strongly linked with human activities, the daily load profile is different for different days of the week. This effect is also present in the price series. Calendar effects are evident from load series as load behaviour deviates from the typical behaviour on bank holidays, bridging holidays etc. leading to highly unstable prices on such days. Variation in daily load profile

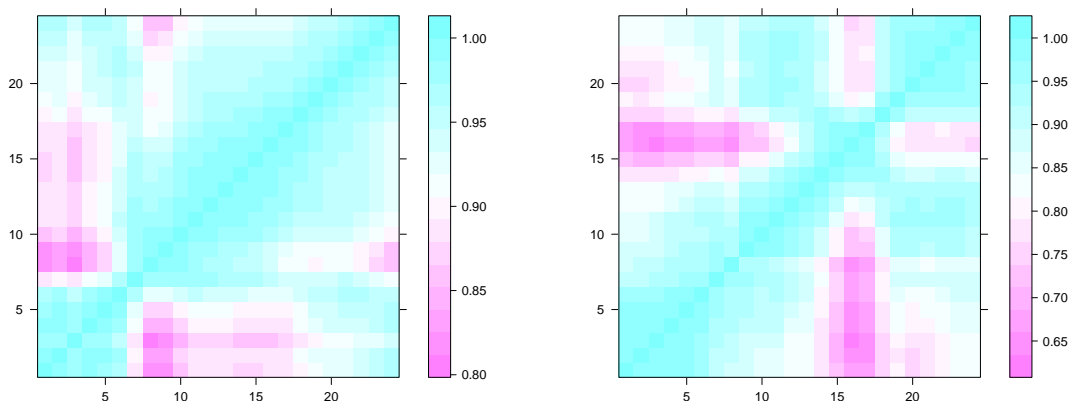


Figure 5.3 (left) IPEX: Hourly demand cross correlation structure (right) PJM: Hourly prices cross correlation structure

in winter and summer is high due to meteorological effects. Demand series may contain outliers however the price series are usually highly volatile and contain large number of extreme prices also known as spikes or jumps, see e.g. Figure 5.2, where daily price curves for APX are plotted for the year 2014. Generally, price and demand series have hourly or half-hourly frequency and they are highly correlated. In Figure 5.3, cross correlation structure is plotted for both price and demand data. In case of demand the overall correlation ranges between 0.8 and 0.99 and between 0.65 and 0.99 for prices.

5.2 General Modelling Framework

Our main goal is to forecast next-day electricity demand and prices using different modeling techniques. For this purpose, data from five different electricity markets, namely the British electricity market (APX Power UK), the Nord Pool electricity market (NP), the Italian electricity market (IPEX), the Pennsylvania-New Jersey-Maryland electricity market (PJM) and the Portuguese electricity market (OMIE(Po)) will be used to study thoroughly the performance of the different forecasting methods described in Chapter 4. These markets substantially differ in generation modes, market maturity, size and policies implemented, geographical location and land electricity demand. For British electricity market, the daily demand/prices are determined the day before the delivery by means of half-hourly auctions while for all other markets by hourly auctions for the following day. Thus, our data set consists of 24 or 48 observations for each day, corresponding to the number of load periods. We denote $D_{t,j}$ and $P_{t,j}$, the demand and price respectively for day t ($t = 1, \dots, n$) and load period j ($j = 1, \dots, 24(48)$). In this work, Data from January 1st, 2009 to December 31st, 2014 is considered for all the above markets. The first five years are used for models estimation while post-sample accuracy was evaluated for the final year by expanding window technique. In the following, models will be estimated separately for each load period and, if necessary, predictions will be put together to re-obtain the hourly time series. This allows us to avoid of modeling the daily periodic component of the time series.

We will explain our modeling approach by referring to demand series however everything

will be analogous in case of price unless stated otherwise. To account for the features stated in section 5.1, we assume that the dynamics of the (log) Demand, $\log(D_{t,j})$, can be adequately described by the following model.

$$\log(D_{t,j}) = f_1(T_{t,j}) + f_2(Y_{t,j}) + f_3(W_{t,j}) + f_4(C_{t,j}) + d_{t,j} . \quad (5.1)$$

Here each f_i is a smooth function of the corresponding component where $T_{t,j}$ represents long-term dynamics, $Y_{t,j}$ and $W_{t,j}$ annual and weekly periodicities respectively. The component $C_{t,j}$ describes several calendar effects, i.e. the bank holidays effect, the bridge effect etc., of course, depending on the understudy market features. Finally, $d_{t,j}$ is the residual component describing the short-term dynamics. The inclusion of a separate yearly component $Y_{t,j}$ in the model is important since mostly demand/price series show regular annual cycles whereas the long-term component $T_{t,j}$ is basically a smooth curve presenting the trend structure in the series. Long-run and intra-annual components can also be modeled jointly by means of a unique component. However in our analysis, the separate estimation results improves prediction accuracy. For homogeneity, only bank holidays (national holidays) are used as calendar effect in model 5.1 although for some markets other features are available that can be significant.

In this work, except for the residual component $d_{t,j}$, all other components included in model 5.1 are regarded as deterministic. Cubic smoothing splines of time and of an yearly and of a weekly cycles are used to describe and estimate the long-term component $T_{t,j}$, the yearly component $Y_{t,j}$ and the weekly component $W_{t,j}$, respectively. The calendar effect component, $C_{t,j}$, is described by a dummy variable that takes value one for national holiday and zero for all other days, thus $C_{t,j} = \sum_{i=1}^2 \alpha_i I_{t,j}$, with $I_{t,j} = 1$ if t refers to a bank holiday and 0 otherwise. Coefficients α 's are estimated by ordinary least squares (OLS). Figure 6.5 shows an example of estimated components in case of electricity demand for a specific load period. The four panels starting from top left contains the actual log demand, the estimated trend, and the estimated yearly and weekly components. As can be seen from the graphs, the yearly component shows regular cycles whereas the long-run component highlights the

trend structure in the data. Once these components are estimated, the residual or stochastic component, is obtained by

$$d_{t,j} = \log(D_{t,j}) - (\hat{f}_1(T_{t,j}) + \hat{f}_2(Y_{t,j}) + \hat{f}_3(W_{t,j}) + \hat{f}_4(C_{t,j})), \quad (5.2)$$

whose dynamics is modeled using different models with increasing level of complexity. As the deterministic components are functions of time or of calendar conditions that are known at any time, out-of-sample predictions are straightforward once the model has been estimated. Since both $T_{t,j}$ and $Y_{t,j}$ represent long-term dynamics with respect to the prediction horizon, we used $\hat{f}_1(T_{t+1,j}) = \hat{f}_1(T_{t,j})$ and $\hat{f}_2(Y_{t+1,j}) = \hat{f}_2(Y_{t,j})$. Finally, the forecasted day-ahead demand is obtained by

$$\hat{D}_{t+1,j} = \exp\left\{\hat{f}_1(T_{t+1,j}) + \hat{f}_2(Y_{t+1,j}) + \hat{f}_3(W_{t+1,j}) + \hat{f}_4(C_{t+1,j}) + \hat{d}_{t+1,j}\right\} \quad (5.3)$$

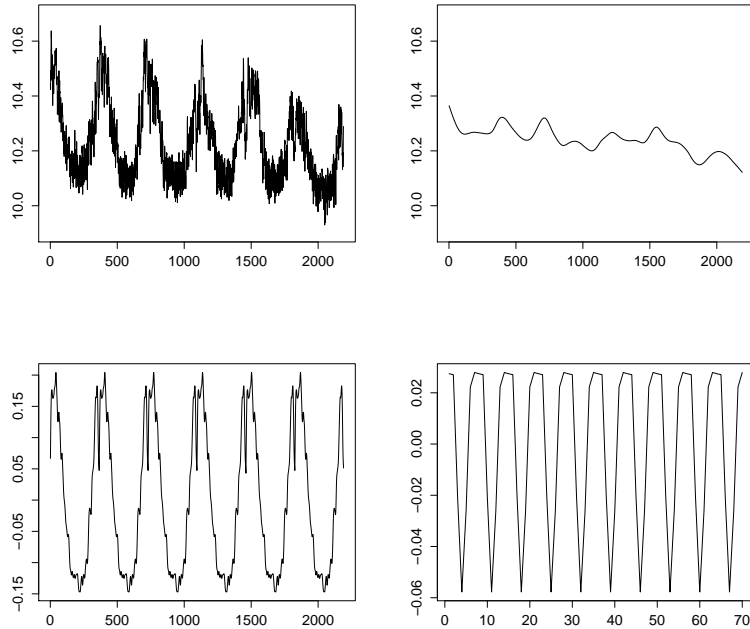


Figure 5.4 APX: Load period 9. $\log(D_{t,j})$, $\hat{f}_1(T_{t,j})$, $\hat{f}_2(Y_{t,j})$, and $\hat{f}_3(W_{t,j})$ ($\hat{f}_3(W_{t,j})$ is only for ten weeks) components.

5.3 Modeling the Stochastic Component

This section describes how to model and estimate the residual stochastic component $d_{t,j}$. To this end, the models and techniques described in Chapter 4 will be used. In particular, we will estimate univariate as well as multivariate models. For each of these two cases, three different ways to specify the models will be considered following (i) a (linear) parametric, (ii) a (nonlinear) non-parametric and (iii) a functional approach.

Univariate models describe the serial dependence between present and past demand/price levels for each given load period, without accounting for possible relations among the levels in different load periods and their lags.

Multivariate models are thought to do this job. In particular, the classical vector autoregressive model we will consider, describes the dynamics of the vector \mathbf{d}_t composed by the intra-daily observations representing the daily profile, namely

$$\mathbf{d}_t = \{(d_{t,1}, \dots, d_{t,24(48)})\}, \quad t = 1, \dots, n. \quad (5.4)$$

A different way to consider the daily profile is referring to the functional data analysis. This is a relatively new approach that looks at the daily demand/price profile as a functional random variable and, thus, treats it as a single object. This, in turn, bypasses two kinds of problems arising in the traditional statistical approach. The first is the number of variables (the demand/price level in each load period), up to 48 in our case. The second one is connected with the possible existence of strong correlation between variables (see Figure 5.3) that can be problematic in the context of multivariate linear models. To model the whole daily profile, the functional approach first transforms the data into a smooth continuous curve by means of some basis functions. In the following we will denote the functional object corresponding to the demand/price daily profile by:

$$\delta_t = \delta(d_{t,1}, \dots, d_{t,24(48)}) = \sum_{g=1}^K \lambda_g \zeta_g(j) \quad (5.5)$$

where λ_g are constant parameters and $\zeta_g(j)$ are basis functions. In this work, the latter are represented by Fourier basis functions. The use of smooth functional objects allows to

exploit additional information contained in the structure of the data, for example the curve derivatives.

5.3.1 Univariate Models

We think of a univariate model as a model whose response is scalar. Within the class of univariate models we consider a linear parametric AutoRegressive (AR) model, its additive nonparametric counter-part and functional models using functional covariates with scalar response.

Linear AR(p) models describe the daily dynamics of $d_{t,j}$ taking into account a linear combination of the last p values of $d_{t,j}$ and is given by

$$d_{t,j} = \phi_0 + \sum_{i=1}^p \phi_i d_{t-i,j} + \varepsilon_{t,j} \quad (5.6)$$

where $\phi_0, \phi_1, \dots, \phi_p$ are constant parameters and $\varepsilon_{t,j} \sim D(0, \sigma_\varepsilon^2)$ is a white noise process. After an inspection of the correlograms, we fit an AR(2) to the residual series $d_{t,j}$.

The additive nonparametric counter-part of previous model is NPAR where the relation between $d_{t,j}$ and its lagged values has not a specific parametric form allowing, potentially, any kind of nonlinearity and is given by

$$d_{t,j} = f_1(d_{t-1,j}) + \dots + f_p(d_{t-p,j}) + \varepsilon_{t,j} \quad (5.7)$$

where f_i are smoothing functions describing the existing relation between each lagged variable and $d_{t,j}$. Functions f_i are described by cubic smoothing spline functions and estimated by the back-fitting algorithm. As in case of AR, order two ($p=2$) is used in the above model. The functional version of the classical linear parametric regression model where a scalar response is predicted by functional covariates using a linear relationship is called Functional AutoRegressive (FAR) and is given by

$$\hat{d}_{t,j} = \int \beta_j \delta_{t-1} \partial j + \varepsilon_t, \quad t = 1, \dots, n-1 \quad (5.8)$$

where δ_{t-1} is the functional predictor (in our current context, it is the functional (smoothed) version of the daily demand profile for day $t - 1$), β is functional coefficient evaluated at j points and ε_t is the error component.

The nonparametric counterpart of previous model is called NPFAR that allows to model the (scalar) response as a nonlinear function of functional covariates. In this case, the following functional based on a Nadaraya–Watson-type kernel estimator is used to predict the scalar response

$$\hat{d}_{t,j} = \frac{\sum_{t=1}^{n-1} d_{t-1,j} K(b^{-1}\xi(\tilde{\delta}_{t-1}, \delta_{t-1}))}{\sum_{t=1}^{n-1} K(b^{-1}\xi(\tilde{\delta}_{t-1}, \delta_{t-1}))} \quad (5.9)$$

where $d_{t-1,j}$ is the demand value for a fixed load period, K represents an asymmetric kernel function, δ_{t-1} is a functional random variable valued in some semi–metric space (E, ξ) and $\tilde{\delta}_{t-1}$ is a fixed element of ξ . Finally, b is a smoothing parameter (bandwidth) corresponding to the optimal number of neighbours selected by cross validation.

5.3.2 Multivariate Modeling

In the multivariate approach both the response variable and the covariates are not scalar but contain information on the whole daily demand profile of a given day. This allows to account for possible interdependences among demand in different load periods. We consider two classes of multivariate models: vector autoregressive models and functional models.

Vector AutoRegressive (VAR) models are well-known multivariate models able to account for linear relationships among different time series. In this approach, each variable (in our case the vector of the demand) is a linear function of p past lags of itself and of the other variables

$$\mathbf{d}_t = \Phi_1 \mathbf{d}_{t-1} + \dots + \Phi_p \mathbf{d}_{t-p} + \varepsilon_t \quad (5.10)$$

where \mathbf{d}_t is the vector of the demand at the 24(48) load periods of day t , Φ_i ($i = 1, \dots, p$) are coefficient matrices and $\varepsilon_t = (\varepsilon_{t,1}, \dots, \varepsilon_{t,24(48)}) \sim N(0, \Sigma_\varepsilon)$. Parameters are estimated by ML methods.

Our second approach for multivariate models refers to functional modeling. In particular, we generalize the univariate functional approach by allowing also to the response variable

to be functional. As for the univariate case, this approach also considers two functional models, a parametric and a nonparametric one.

Functional response with Functional covariates AutoRegressive of order one (FFAR(1)) is given by

$$\delta_t = \int_{s \in (0, 24(48))} \delta_{t-1}(j) \Phi(s, j) \partial s + \eta_t(j), \quad j \in (0, 24(48)) \quad (5.11)$$

where $\Phi(s, j)$ is the functional parameter (regression function), δ_t is the functional version of daily demand profile for day t and η_t is a functional random error.

A nonlinear version of the previous model is the NonParametric Functional response with Functional covariates AutoRegressive model (NPFAR) that is based on a functional kernel regression. This model can be written as

$$\delta_t = \varphi(\delta_{t-1}) + \eta_t \quad (5.12)$$

where φ is a (possibly) nonlinear operator and, again, the residual η_t is a functional random error. The estimation of φ is based on kernel methods, more specifically

$$\hat{\delta}_t = \hat{\varphi}(\delta_{t-1}) = \frac{\sum_{i=1}^{n-1} \delta_{i-1} K(b^{-1} \xi(\tilde{\delta}_{i-1}, \delta_{i-1}))}{\sum_{i=1}^{n-1} K(b^{-1} \xi(\tilde{\delta}_{i-1}, \delta_{i-1}))} \quad (5.13)$$

where K is a kernel function, in this case an Epanechnikov kernel, and b is the bandwidth, selected by cross-validation.

5.4 Out-of-Sample Forecasting

In this section the modeling strategy presented previously is implemented to different electricity markets and one-day ahead forecasting results are obtained for both demand and price time series. For each case, demand and price, this section will in turn examine the forecasting ability of different models previously defined.

5.4.1 Demand Forecasting

For the current issue, data from 2009 to 2014 are used for all five electricity markets included in our study. Few missing values in the demand series were interpolated by averaging the neighboring values. For APX, indicated margin was included as an extra covariate to the model. The demand structure for OMIE(Po) changed dramatically in the start of 2012 and therefore a dummy variable accounting for this level shift has been included to the model. For all the markets, model 5.1 is used to estimate and forecast day-ahead demand for the complete year 2014 by expanding window techniques where at each estimation step, the calibration sample is enlarged by one day. Thus, globally, we have $365 \times 24(48) = 8760(17520)$ one-day-ahead predictions allowing for a more thorough analysis of the forecasting results. To compare the forecasting performance, global mean absolute percentage error (MAPE), daily specific mean absolute percentage error (DS-MAPE) and mean square percentage error (MSPE) are computed for each model defined as:

$$\text{MAPE} = \text{mean}(|e_{t,j}/D_{t,j}|) \times 100$$

$$\text{MSPE} = \text{mean}(|e_{t,j}/D_{t,j}| \times 100)^2$$

$$\text{DS-MAPE}_{\text{day}} = \text{mean}(|e_{t \in \text{day},j}/D_{t,j}|) \times 100$$

$$\text{MAE} = \text{mean}(|e_{t,j}|)$$

where $e_{t,j} = (D_{t,j} - \hat{D}_{t,j})$, with $D_{t,j}$ and $\hat{D}_{t,j}$ the observed and the forecasted demand values for day t ($t = 1, \dots, 365$), load period j ($j = 1, \dots, 24(48)$) and $t \in \text{day}$ denotes the days Monday, ..., Sunday. The median and Q(0.99) quantiles of the absolute percentage errors are also calculated. To assess the significance of the differences between the summary statistics, the Diebold and Mariano (DM) test (Diebold and Mariano, 1995) for equal predictive accuracy has been performed for each couple of indicators. DM test measure the accuracy of each forecast by a particular loss function. To understand well, let a pair of h -steps ahead forecasts from two different models produce $e_{t,1}$ and $e_{t,2}$ errors where $\{e_{t,i}\}_{i=1}^2$ is the difference between observed and forecasted series. Two popular loss functions are

$$\begin{aligned} \text{absolute error loss: } & g(e_{t,i}) = |e_{t,i}|, \\ \text{squared error loss: } & g(e_{t,i}) = (e_{t,i})^2. \end{aligned}$$

The DM test is based on the loss differential given by

$$d_t = g(e_{t,1}) - g(e_{t,2}).$$

The null of equal predictive accuracy is then

$$H_0: E(d_t) = 0 \quad \forall t,$$

The DM test statistic is defined as

$$\text{DM} = \frac{\bar{d}}{\sqrt{\hat{L}(\bar{d})}}, \quad (5.14)$$

where

$$\begin{aligned} \bar{d} &= \frac{1}{n} \sum_{t=1}^n d_t, \\ \hat{L}(\bar{d}) &= \text{var}(\bar{d}_t) + 2 \sum_{k=1}^{h-1} \hat{\gamma}_k, \end{aligned}$$

with

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

Finally, Diebold and Mariano show that under the null of equal predictive accuracy, the statistic DM has an asymptotic standard normal distribution.

Results concerning the out-of-sample prediction for all the markets are listed in Table 5.1. They suggest, as expected, multivariate models perform better than univariate models. Although the differences among the multivariate models are not huge, the results for APX, NP and PJM show the better performance of FFAR and VAR compared to others. For OMIE(Po) and IPEX, nonparametric functional models outperforms the others. Note that the relative performance of parametric and nonparametric approach may depend on the underlying pro-

Market	Models	MAPE	MSPE	Median	Q(0.99)
APX	AR	1.98	7.05	1.53	8.10
	NPAR	2.17	8.72	1.66	9.48
	FAR	1.69	5.14	1.32	6.98
	NPFAR	1.76	5.64	1.38	7.28
	VAR	1.67	5.02	1.30	6.93
	FFAR	1.67	5.01	1.30	6.92
	NPFFAR	1.76	5.63	1.37	7.31
NP	AR	2.53	11.49	1.98	10.70
	NPAR	3.01	16.47	2.27	12.28
	FAR	2.28	9.19	1.78	9.44
	NPFAR	2.27	9.12	1.75	9.12
	VAR	2.10	7.85	1.61	8.84
	FFAR	2.10	7.83	1.63	8.81
	NPFFAR	2.27	9.09	1.75	9.09
PJM	AR	3.44	21.85	2.56	14.31
	NPAR	5.13	45.92	4.07	19.44
	FAR	2.91	14.61	2.30	11.23
	NPFAR	2.86	14.97	2.19	11.45
	VAR	2.43	10.99	1.79	10.53
	FFAR	2.41	10.91	1.77	10.50
	NPFFAR	2.84	14.84	2.14	11.44
OMIE (PO)	AR	3.29	22.07	2.29	15.04
	NPAR	3.47	25.16	2.37	17.44
	FAR	3.80	25.47	3.00	14.31
	NPFAR	3.15	20.27	2.15	14.19
	VAR	3.48	24.21	2.44	15.87
	FFAR	3.44	22.14	2.52	14.21
	NPFFAR	3.15	20.26	2.14	14.29
IPEX	AR	1.92	9.25	1.30	11.21
	NPAR	2.21	12.44	1.45	12.87
	FAR	1.89	8.75	1.32	10.93
	NPFAR	1.85	8.23	1.28	10.14
	VAR	1.95	8.90	1.36	10.40
	FFAR	1.88	8.69	1.31	10.80
	NPFFAR	1.83	8.18	1.26	10.04

Table 5.1 Electricity Demand: Prediction accuracy statistics: AutoRegressive (AR), Nonparametric AutoRegressive (NPAR), Functional AutoRegressive (FAR), Nonparametric Functional AutoRegressive (NPFAR), Vector AutoRegressive (VAR), Functional Functional AutoRegressive (FFAR), Nonparametric Functional Functional AutoRegressive (NPFFAR).

cess as can be seen in our experiment. The linear parametric approach produces better results for the first three markets while the nonparametric for the last two indicating underlying nonlinearities in the process. These nonlinearities can correspond to the economic situation of the country, some specific calendar effects (e.g. for IPEX, the August effect, end of year effect, bridging effect)¹ and other related phenomena. Since filtering removes the maximum nonlinearities by removing different components from the data resulting in more linear process in the residuals, we do not expect large differences between FFAR and VAR. Notice that for all markets, the minimum MAPE values can be considered satisfactory compared with the state-of-art demand prediction techniques. Although a comparison of results from the literature is difficult due to the different periods considered, some reference works are listed here. Using six different models based on exponential smoothing techniques and forecasting the whole year 2009 (public holidays are excluded) for APX, (Taylor, 2012) results a MAPE of higher than 2% for all the models. For the same out-of-sample period used in this paper for IPEX, (Bernardi and Petrella, 2015) reported day-ahead minimum MAPE value of 2.90. Another study(Göb et al., 2013) reports a MAPE value of 2.54 for IPEX. Both these papers used a verity of model however their results are significantly higher compared to ours. For PJM and OMIE(Po), the day-ahead MAPE values reported in (Sahay and Tripathi, 2014) and (Cuambe, 2013) are relatively higher compared to the results listed in Table 5.1. To assess the significance of the differences between the summary statistics listed in table 5.1 , the Diebold and Mariano (DM) test (Diebold and Mariano, 1995) for equal predictive accuracy has been performed for each couple of indicators. The results concerning DM test are listed in table 5.2. Each element of the table is the p-value of a hypothesis system where the null hypothesis assumes no difference in the accuracy of the forecasters in the row/column against the alternative hypothesis that the predictor in the row is more accurate than predictor in the column. The p-values in table 5.2 confirm the significance of multivariate model compared to univariate and hence the results obtained from table 5.1. Again, for the first three markets, parametric multivariate model, FFAR and VAR, perform relatively better than others while for the last two markets the nonparametric

¹For homogeneity, these are not included to the model

Market	Models	AR	NPAR	FAR	NPFAR	VAR	FFAR	NPFFAR
APX	AR	-	< 0.01	0.99	0.99	0.99	0.99	0.99
	NPAR	0.99	-	0.99	0.99	0.99	0.99	0.99
	FAR	< 0.01	< 0.01	-	0.16	0.87	0.99	0.16
	NPFAR	< 0.01	< 0.01	0.84	-	0.92	0.96	0.35
	VAR	< 0.01	< 0.01	0.13	0.08	-	0.82	0.08
	FFAR	< 0.01	< 0.01	0.01	0.04	0.18	-	0.04
	NPFFAR	< 0.01	< 0.01	0.84	0.65	0.92	0.96	-
NP	AR	-	< 0.01	0.99	0.99	0.99	0.99	0.99
	NPAR	0.99	-	0.99	0.99	0.99	0.99	0.99
	FAR	< 0.01	< 0.01	-	0.37	0.99	0.99	0.41
	NPFAR	< 0.01	< 0.01	0.63	-	0.99	0.99	0.94
	VAR	< 0.01	< 0.01	< 0.01	< 0.01	-	0.56	< 0.01
	FFAR	< 0.01	< 0.01	< 0.01	< 0.01	0.44	-	< 0.01
	NPFFAR	< 0.01	< 0.01	0.59	0.06	0.99	0.99	-
PJM	AR	-	< 0.01	0.99	0.99	0.99	0.99	0.99
	NPAR	0.99	-	0.99	0.99	0.99	0.99	0.99
	FAR	< 0.01	< 0.01	-	0.08	0.99	0.99	0.12
	NPFAR	< 0.01	< 0.01	0.92	-	0.99	0.99	0.85
	VAR	< 0.01	< 0.01	< 0.01	< 0.01	-	0.64	< 0.01
	FFAR	< 0.01	< 0.01	< 0.01	< 0.01	0.36	-	< 0.01
	NPFFAR	< 0.01	< 0.01	0.88	0.15	0.99	0.99	-
OMIE (Po)	AR	-	0.03	< 0.01	0.97	0.01	0.21	0.96
	NPAR	0.97	-	0.33	0.99	0.60	0.89	0.99
	FAR	0.99	0.67	-	0.99	0.88	0.99	0.99
	NPFAR	0.03	0.01	< 0.01	-	< 0.01	< 0.01	0.37
	VAR	0.99	0.40	0.12	0.99	-	0.98	0.99
	FFAR	0.79	0.11	< 0.01	0.99	0.02	-	0.99
	NPFFAR	0.04	0.01	< 0.01	0.63	< 0.01	< 0.01	-
IPEX	AR	-	0.01	0.82	0.99	0.52	0.89	0.99
	NPAR	0.99	-	0.99	0.99	0.99	0.99	0.99
	FAR	0.18	< 0.01	-	0.94	0.07	0.64	0.98
	NPFAR	0.01	< 0.01	0.06	-	0.01	0.14	0.99
	VAR	0.48	0.01	0.93	0.99	-	0.97	0.99
	FFAR	0.11	< 0.01	0.36	0.86	0.03	-	0.92
	NPFFAR	< 0.01	< 0.01	< 0.01	0.01	< 0.01	0.08	-

Table 5.2 Electricity Demand: P-values for the DM test for equal prediction accuracy versus the alternative hypothesis that model in the row is more accurate than model in the column (squared loss function used)

functional models have superior forecasting performance. In case of IPEX and OMIE(Po), although the summary statistics in table 5.1 shows the better performance of AR compare

Market	Models	Day of the week						
		M	T	W	T	F	S	S
APX	AR	2.26	1.69	1.80	1.62	1.80	2.23	2.44
	NPAR	2.22	2.01	2.02	1.85	1.91	2.44	2.67
	FAR	2.05	1.43	1.54	1.42	1.55	1.91	1.89
	NPFAR	2.08	1.42	1.64	1.45	1.58	2.05	2.14
	VAR	1.99	1.40	1.51	1.42	1.55	1.86	1.93
	FFAR	1.99	1.39	1.54	1.40	1.52	1.91	1.93
	NPFFAR	2.11	1.40	1.61	1.44	1.58	2.04	2.15
NP	AR	2.74	2.17	2.20	2.33	2.47	2.80	2.84
	NPAR	3.37	2.51	2.48	2.74	2.71	3.25	3.64
	FAR	2.12	1.93	1.98	2.09	2.21	2.45	3.02
	NPFAR	2.36	1.90	1.94	2.01	2.09	2.64	2.79
	VAR	2.02	1.74	1.78	1.83	2.14	2.32	2.70
	FFAR	2.05	1.75	1.82	1.82	2.12	2.30	2.71
	NPFFAR	2.34	1.89	1.93	2.01	2.10	2.65	2.79
PJM	AR	3.47	3.99	3.24	3.60	3.49	3.09	3.21
	NPAR	5.22	5.40	4.90	5.16	5.31	5.03	4.84
	FAR	3.49	2.77	2.84	2.89	2.76	2.79	2.87
	NPFAR	3.45	3.10	2.91	2.83	2.66	2.52	2.62
	VAR	2.99	2.36	2.27	2.29	2.35	2.27	2.49
	FFAR	2.96	2.37	2.21	2.29	2.32	2.27	2.49
	NPFFAR	3.36	3.14	2.89	2.81	2.64	2.50	2.59
OMIE (Po)	AR	4.09	3.01	3.36	2.80	2.41	3.23	4.04
	NPAR	3.50	3.35	3.53	2.69	2.55	3.84	4.77
	FAR	4.41	3.49	3.86	3.40	3.11	3.76	4.52
	NPFAR	4.00	2.76	3.09	2.56	2.41	3.41	3.84
	VAR	4.01	3.00	3.76	2.86	2.68	3.27	4.70
	FFAR	4.08	3.05	3.54	2.80	2.65	3.48	4.41
	NPFFAR	4.02	2.76	3.05	2.54	2.42	3.41	3.82
IPEX	AR	1.76	1.77	1.76	1.55	1.85	2.29	2.46
	NPAR	2.10	2.00	1.92	1.82	2.01	2.57	2.93
	FAR	1.88	1.72	1.70	1.54	1.84	2.18	2.37
	NPFAR	1.70	1.61	1.71	1.61	1.82	2.21	2.25
	VAR	1.99	1.76	1.76	1.58	1.88	2.27	2.34
	FFAR	1.83	1.62	1.68	1.57	1.83	2.23	2.37
	NPFFAR	1.68	1.60	1.68	1.59	1.81	2.22	2.23

Table 5.3 Electricity Demand: Hourly DS-MAPE

to some multivariate models, p-values in table 5.2 indicate of no statistical significant difference among these models.

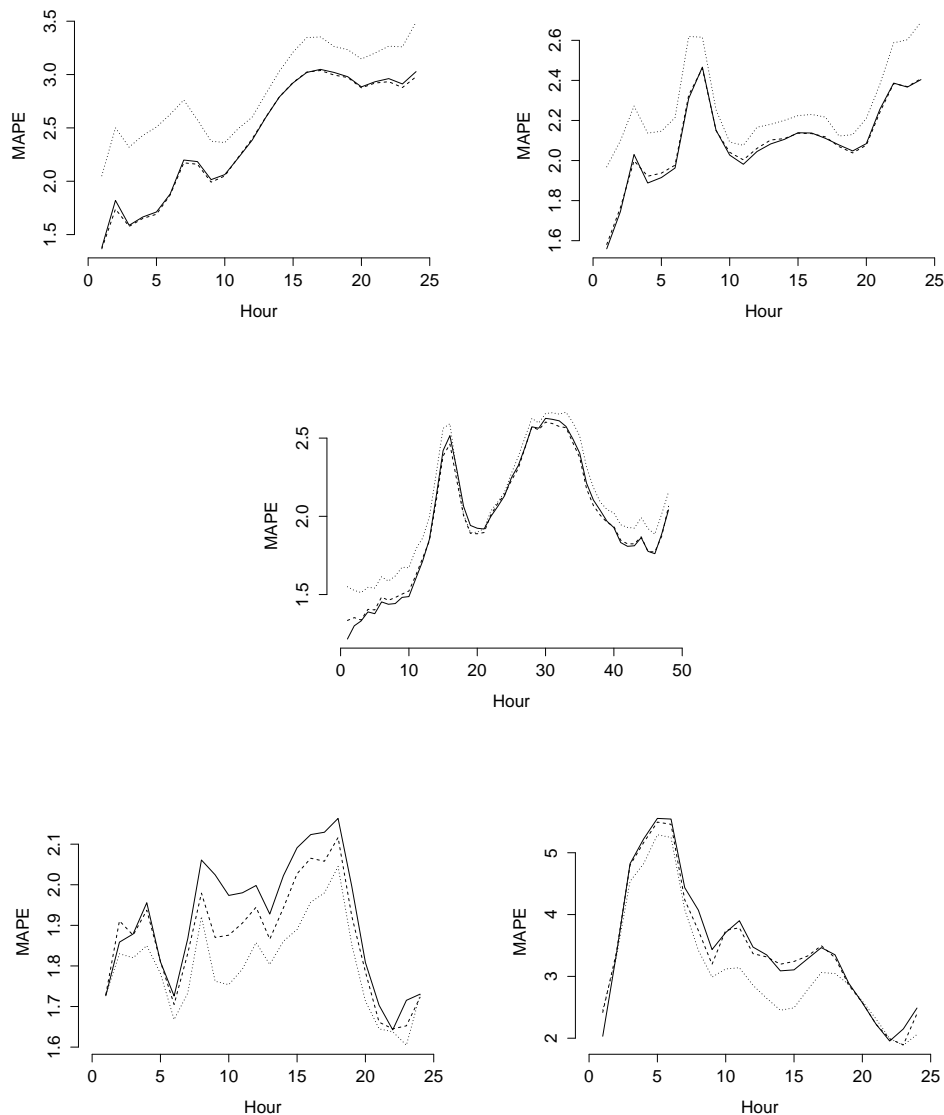


Figure 5.5 Electricity Demand: Hourly MAPE values for (top left) PJM (top right) NP (middle) UKPX (bottom left) IPEX (bottom right) OMIE (Po). lines are (solid) VAR (dashed) FFAR (dotted) NPPFFAR.

The day specific mean absolute percentage errors are given in Table 5.3. From this table, we can see that except for PJM, DSMAPE is relatively higher on Monday, Saturday and Sunday and smaller on other weekdays for all other markets. For PJM, Monday, Tuesday and Wednesday accounts for higher prediction error than other days of the week.

Finally, the hourly mean absolute percentage error (H-MAPE) for multivariate models (only for VAR, FFAR and NPFFAR) are plotted in Fig. 5.5. Again the good forecasting ability

of parametric (linear) models for NP, PJM and UKPX is evident from the graph while non-parametric (nonlinear) model for the other two markets. The H-MAPE for NP and PJM are lower at starting load period and higher before midnight. The H-MAPE for UKPX follows the usual trend seen for daily demand with two peaks; one around 09:00 a.m. and another around 16:00 p.m. and considerably low errors for all other periods. OMIE (Po) shows high variation all the day and it reaches to higher in the morning at around 5 a.m. On the other hand, the H-MAPE is considerably lower for IPEX compare to other markets at all load periods. For example, The H-MAPE values for NPPFAR exceeds 2% at 5 and 6 p.m only and is lower at all other load periods.

5.4.2 Price Forecasting

In case of price prediction, data from 2009 to 2014 are used for four electricity markets namely, NP, APX, PJM and IPEX. Data for OMIE(Po) was not available for the same period and therefore excluded from the analysis. The few missing values in the price series is treated as in the case of demand in section 5.4.1. Forecasted demand was included as an extra covariate to the model. For all the markets, model 5.1 is used to estimate and forecast one-day-ahead prices for the complete year 2014 by expanding window techniques. The forecasting ability of each model was evaluated by different prediction accuracy statistics stated in the previous section.

A critical issue concerning the treatment of price spikes (also called jumps, outlier) is also considered. Generally, authors do not show agreement on whether these observations have to be excluded or included in the price series before estimating the model. As it is commonly known and widely agreed that price spikes account for greater part of the total variation in the data, yet there is no common definition for the price spikes. For this issue, four different approaches are considered to identify price spikes.

- Fixed price threshold: In this approach all the values exceeding from a subjectively choose price are classified as price spikes. In our case, we choose a fixed price for every markets and this price varies across the markets.

- Variable price threshold: In this approach, a fixed percentage of the highest (or/and lowest) prices are treated as price spike. The fixed percentage in our case is upper and lower 1% of prices.
- Fixed price change thresholds: in this approach, price increments exceeding some subjectively chosen (fixed) threshold (we choose 25%) are classified as price spikes.
- Variable price change thresholds: In this approach, any price corresponding to the price increments exceed three standard deviation of all prices is classified as price spike.

In general, different techniques are used to replace the price spikes. For the current issue, they are replaced by four neighboring prices avoiding the possibility of two consecutive outliers. After the data treated for price spikes and results obtained (not reported here), they do not provide evidence of statistically significant improvement compare to the results obtained from untreated data in the context of out-of-sample forecasting. Therefore the analysis is done without replacing any price spike.

The result concerning out-of-sample day-ahead prices forecasting are listed in Table 5.4. The reported results indicate that multivariate models produce lower prediction error compared to univariate models in general. However, the differences are small compared to error sizes. Within the univariate approach, functional models perform relatively better with respect to their non functional counterparts. For NP and APX markets, the parametric approach produces better results and FAR, VAR and FFAR models produce lower forecasting errors. For the other two markets, PJM and IPEX, nonparametric models NPFAR and NPFFAR produce better results indicating the existence of possible nonlinearities in the underlying process. The forecasting errors are also high for PJM and IPEX compare to other two markets. In general, NP and APX are considered more mature and stable markets than PJM and IPEX since they are still going through different maturity phases and yet facing many challenges.

To assess the significance of the differences between the summary statistics listed in Table 5.4, the DM test for equal predictive accuracy has been performed for each couple of indicators and results are reported in Table 5.5. Each element of the table is the p-value of

Market	Models	MAPE	MAE	Median
NP	AR	6.08	13.60	3.67
	NPAR	6.87	15.49	3.86
	FAR	5.24	12.01	3.11
	NPFAR	5.96	13.27	3.78
	VAR	5.24	12.06	3.23
	FFAR	5.47	11.99	3.20
	NPFFAR	5.94	13.30	3.72
APX	AR	9.95	4.17	7.18
	NPAR	10.23	4.32	7.45
	FAR	9.51	4.01	6.73
	NPFAR	9.87	4.18	7.10
	VAR	9.54	4.03	6.68
	FFAR	9.52	4.02	6.82
	NPFFAR	9.78	4.12	7.05
PJM	AR	13.12	10.49	8.00
	NPAR	20.42	19.19	8.35
	FAR	12.84	9.91	7.81
	NPFAR	12.45	9.90	7.70
	VAR	12.80	9.79	7.72
	FFAR	12.46	10.26	7.64
	NPFFAR	12.37	9.97	7.62
IPEX	AR	13.92	6.03	8.80
	NPAR	15.05	6.52	9.61
	FAR	13.95	6.12	9.02
	NPFAR	13.49	5.95	8.58
	VAR	13.46	5.94	8.38
	FFAR	13.55	5.99	8.61
	NPFFAR	13.47	5.88	8.54

Table 5.4 Electricity Price: Prediction accuracy statistics: AutoRegressive (AR), Nonparametric AutoRegressive (NPAR), Functional AutoRegressive (FAR), Nonparametric Functional AutoRegressive (NPFAR), Vector AutoRegressive (VAR), Functional Functional AutoRegressive (FFAR), Nonparametric Functional Functional AutoRegressive (NPFFAR).

a hypothesis system as in case of demand. As in case of demand, The p-values in Table 5.5 confirm the overall significance of multivariate model compared to univariate however in many cases univariate models shows equal predictive. Except the poor performance of NPAR, in general the differences found in summary statistics in Table 5.4 are not statistically significant and in particular in multivariate case.

As in case of demand, the obtained values of different descriptive statistics for different models are compared with the state-of-art price prediction techniques from the literature.

Market	Models	AR	NPAR	FAR	NPFAR	VAR	FFAR	NPFFAR
NP	AR	-	0.02	0.92	0.95	0.99	0.99	0.96
	NPAR	0.98	-	0.99	0.99	0.99	0.99	0.99
	FAR	0.08	0.01	-	0.38	0.78	0.75	0.34
	NPFAR	0.04	0.01	0.62	-	0.92	0.98	0.55
	VAR	< 0.01	0.01	0.22	0.08	-	0.60	0.07
	FFAR	< 0.01	0.01	0.25	0.02	0.40	-	0.03
	NPFFAR	0.04	0.01	0.66	0.45	0.93	0.97	-
APX	AR	-	0.01	0.91	0.78	0.50	0.94	0.80
	NPAR	0.99	-	0.99	0.99	0.98	0.99	0.99
	FAR	0.09	< 0.01	0.50	0.36	0.03	0.64	0.41
	NPFAR	0.22	0.01	0.64	-	0.21	0.65	0.55
	VAR	0.50	0.02	0.97	0.79	-	0.99	0.80
	FFAR	0.06	0.00	0.36	0.35	0.01	-	0.35
	NPFFAR	0.20	0.01	0.59	0.45	0.20	0.65	-
PJM	AR	-	0.11	0.83	0.91	0.87	0.70	0.89
	NPAR	0.89	-	0.89	0.87	0.89	0.89	0.89
	FAR	0.17	0.11	-	0.33	0.79	0.23	0.34
	NPFAR	0.09	0.13	0.67	-	0.79	0.32	0.47
	VAR	0.13	0.11	0.21	0.21	-	0.12	0.24
	FFAR	0.30	0.11	0.77	0.68	0.88	-	0.68
	NPFFAR	0.11	0.11	0.66	0.43	0.76	0.32	-
IPEX	AR	-	< 0.01	0.06	0.90	0.37	0.28	0.93
	NPAR	0.99	-	0.99	0.99	0.99	0.99	0.99
	FAR	0.94	< 0.01	-	0.98	0.77	0.79	0.99
	NPFAR	0.10	< 0.01	0.02	-	0.11	0.09	0.45
	VAR	0.63	< 0.01	0.23	0.89	-	0.44	0.89
	FFAR	0.72	< 0.01	0.21	0.91	0.56	-	0.93
	NPFFAR	0.07	< 0.01	0.01	0.55	0.11	0.07	-

Table 5.5 Electricity Price: P-values for the DM test for equal prediction accuracy versus the alternative hypothesis that model in the row is more accurate than model in the column (squared loss function used)

Although a completely fair comparison of the results are not possible due to the fact that different authors considered different time periods and different prediction accuracy measures, however some reference works are listed here. Using multivariate exponential smoothing and dynamic factor model ideas, Carpio et al. (2014) results the median of absolute percentage error 4.2% for NP that is relatively higher compare to our. In case of IPEX, Petrella and Sapio (2009) reported a root mean square error (RMSE) of 11.58 obtained by ARX-EGARCH model that is significantly higher of our 8.40 (RMSE are not reported here) ob-

Market	Models	Day of the week						
		M	T	W	T	F	S	S
NP	AR	8.91	5.92	4.67	5.20	4.95	6.37	6.50
	NPAR	9.26	6.39	5.39	6.05	5.14	7.94	7.76
	FAR	6.86	5.40	4.38	4.79	3.95	5.41	5.90
	NPFAR	7.66	5.94	4.82	5.55	4.52	6.04	7.13
	VAR	6.50	5.78	4.40	4.66	4.06	5.47	5.79
	FFAR	7.04	6.69	4.45	4.52	4.03	5.44	6.07
	NPFFAR	7.71	5.85	4.81	5.48	4.56	6.10	7.03
APX	AR	10.46	10.48	9.87	8.79	8.93	11.76	9.42
	NPAR	11.09	10.31	9.81	9.07	9.38	12.23	9.80
	FAR	9.86	10.18	9.62	8.57	8.34	11.04	9.05
	NPFAR	10.22	10.21	9.88	8.77	8.99	11.47	9.31
	VAR	9.79	10.35	9.67	8.34	8.65	10.88	9.21
	FFAR	9.92	10.13	9.74	8.54	8.51	10.85	9.03
	NPFFAR	10.20	10.17	9.81	8.79	8.95	11.37	9.25
PJM	AR	13.40	15.33	14.38	14.36	13.20	11.39	9.91
	NPAR	16.74	18.01	37.82	23.66	19.13	15.60	12.31
	FAR	13.78	13.32	15.55	14.06	12.50	10.67	10.05
	NPFAR	12.51	14.20	12.87	13.89	12.77	10.72	9.83
	VAR	14.41	13.48	15.14	13.21	11.73	11.27	10.53
	FFAR	12.72	14.22	13.26	12.44	12.60	11.48	10.65
	NPFFAR	12.47	14.28	13.01	13.85	12.69	10.63	9.80
IPEX	AR	14.17	10.60	9.53	10.43	8.96	13.91	30.00
	NPAR	14.75	12.77	11.60	10.64	10.18	15.21	30.36
	FAR	15.03	10.62	9.63	10.66	9.30	13.88	28.71
	NPFAR	15.25	9.98	9.22	9.87	8.05	13.77	28.93
	VAR	15.21	10.20	8.72	10.45	9.15	13.14	27.56
	FFAR	15.40	10.07	9.04	10.28	9.23	13.25	27.75
	NPFFAR	15.15	9.96	9.15	9.80	8.11	13.65	28.67

Table 5.6 Electricity Price: Hourly DS-MAPE

tained with NPFFAR.

The day specific mean absolute percentage errors (DS-MAPE) are given in Table 5.6. As in case of demand, from this table we can see that except for PJM, DS-MAPE are relatively higher on Monday, Saturday and Sunday and smaller on other weekdays for all other markets. For PJM, Monday, Tuesday and Wednesday accounts for higher prediction error than other days of the week. The effect of parametric and nonparametric approach is also evident in this table as for NP and APX, parametric approach produces lower errors while

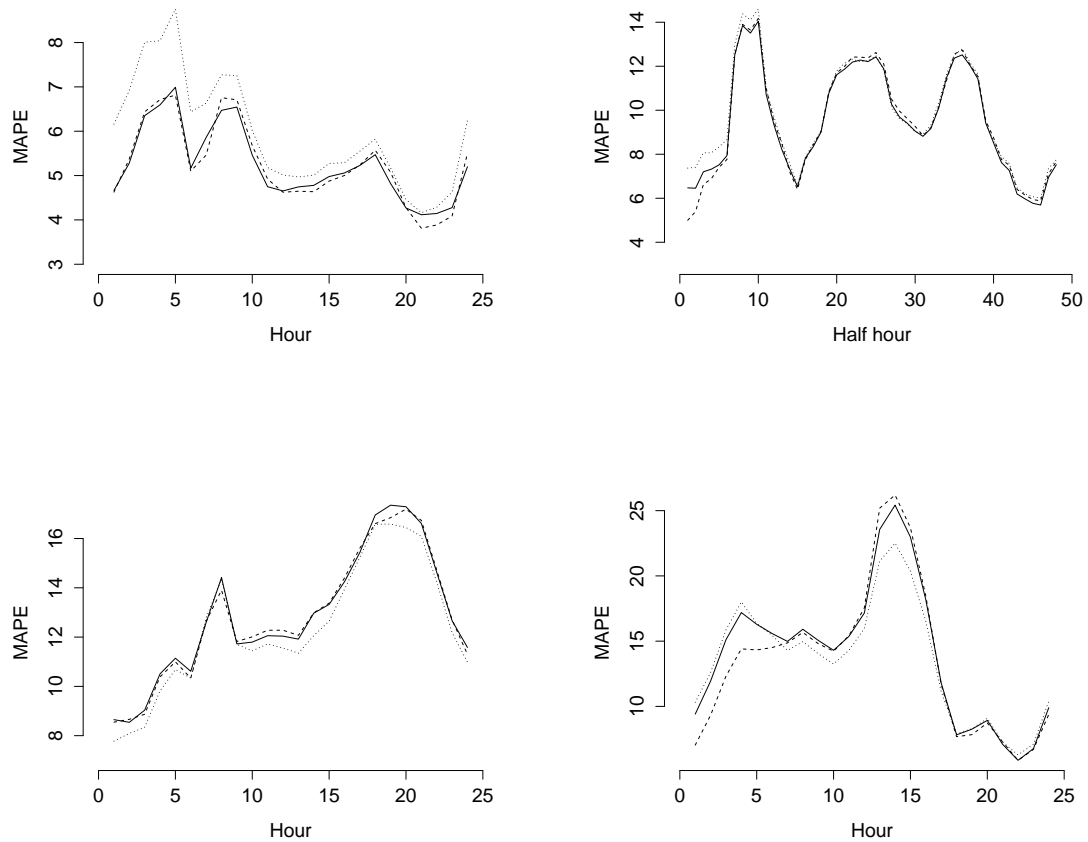


Figure 5.6 Electricity Price: Hourly MAPE values for (top left) NP (top right) APX (bottom left) PJM (bottom right) IPEX. lines are (solid) VAR (dashed) FFAR (dotted) NPFFAR.

nonparametric approach for the other two markets. Finally, the hourly MAPE values for all market can be visualized from Figure 5.6. The results are plotted only for VAR, FFAR and NPFFAR as they perform relatively better than others. All the four graphs are showing multiple bumps indicating high MAPE values for corresponding hours. In general, these hours accounts for greater variation in the data and receives high number of spikes (outliers). For example in case of IPEX, it is well known that hours from 14:00 to 16:00 are highly volatile and not easy to predict.

5.5 Conclusion

Since the liberalization of electricity sector, modeling and forecasting electricity demand and prices gained an increasing attention in competitive electricity markets. This chapter considered these issues by using data from five different electricity markets. For both issues, an additive component model was considered that includes some deterministic and a stochastic residual components. The deterministic components include a long-term dynamics, annual and weekly periodicities and calendar effects. For the residual component, different univariate and multivariate models have been considered with increasing level of complexity. In both cases, linear parametric and nonlinear nonparametric models, as well as functional based models, have been estimated and compared in a one day-ahead out-of-sample forecast framework. The analyses suggest that the multivariate approach leads to better results than the univariate one and that, within the multivariate framework, functional models are the most accurate, with VAR being a competitive model in some cases. The results also lead to another important finding concerning to the performance of parametric and nonparametric approach that showed strong linkage with underlying process. Finally the obtained results were compared with other works in the literature that suggest our forecasting errors are smaller compared with the state-of-art prediction techniques used in the literature.

Chapter 6

Modeling and Forecasting Supply and Demand Curves

6.1 Introduction

In deregulated electricity markets, the study of price prediction is equally important for producers, buyers, investors and other load serving bodies for various reasons. These includes, among others, the cash flow analysis, least cost planning, integrated resource planning, financial procurement, optimal bidding strategies, regulatory rule-making and demand side management. As can be seen from Chapter 5, although electricity loads forecasting has reached to a comfortable state of performance, achieving the same results for prices seem still a long way to go. Indeed, this task has been proved to be quite challenging for both practitioners and academics due to the peculiar characteristics of electricity prices.

Before liberalization of the electric sector, when the market was highly regulated and controlled (usually) by state owned companies, the electric utilities were mainly interested in efficient forecasting of electric load as the variation in the electricity prices was minimal and changes in prices were considered after regular time intervals. The reorganization of the electricity sector completely changed the nature of price series. Perhaps one of the most evident and prominent consequence of liberalization is reflected in electricity prices that are highly volatile. In general, literature described many factors affecting the demand and sup-

ply balance consequently results in high volatility of prices. Among others, one of the main factors corresponds to meteorological events such as temperature variation, precipitations, snow and rainfalls that are essential elements for hydro production, sun light and wind flow that affects the electricity production from solar and windmill respectively. Other factors include changes in fuel prices, new regulatory conditions and unexpected events. Moreover,

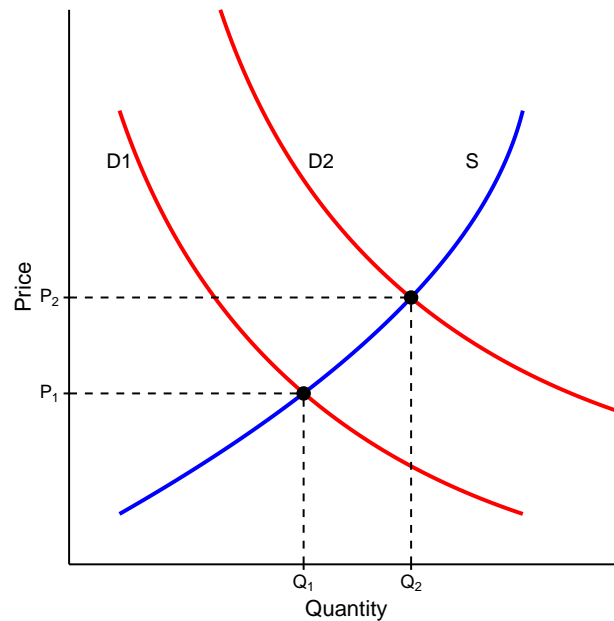


Figure 6.1 Supply (blue) and two demand (red) hypothetical curves

electricity markets exhibit many limiting factors. Electricity cannot be economically stored and it has a production and transmission lines capacity. Technically it is impossible to produce more electricity than the aggregated amount of production capacity and it should not exceed the transmission line capacity in order to ensure the reliability of electricity grid. These restrict market flexibility in short term as increase in generation or transmission capacity is expensive and time consuming.

Beside these factors, less importance is given to the price formation process that, in fact, is essential for the study of price prediction. In the electricity markets, the process of price formation follows the basic law of demand and supply frequently used in finance and macroeconomics. According to supply and demand theory, in a competitive market the price of the

underlying commodity should reflect the relative scarcity of the supply for a given demand level. In competitive electricity market, the offers from suppliers are rejected with higher incremental costs if the demand level is low and hence suppliers with the lowest incremental costs remain in the competition resulting in a relatively low equilibrium prices. As the demand increases, the suppliers with lower incremental cost use up their production capacity first followed by increasingly expensive suppliers that eventually raise the equilibrium price.

In the electricity markets, different suppliers can directly get affected by different factors stated above. For example, a producer whose production mainly depends on fossil fuels or gas is directly get affected by the variation in prices of these resources. Any new regulations imposed on energy from nuclear resource directly affect the corresponding suppliers. Different participants use different bidding strategies in order to optimize their profits. For example, a widely used strategy is to “bid zero” where participants submit a bid containing their desired quantity corresponding to a zero price. In real, a “zero price” bid means that the quantity submitted in the bid is required at any price or simply at market clearing price since demand is always larger than any electric utility production capacity however in this case the equilibrium price is settled by other bidders. These and other factors necessitate for looking the problem of price prediction in more detail.

In the following, we describe a different approach for forecasting electricity prices based on modeling and forecasting daily supply and demand curves. In addition, this method has a remarkable advantage as it provides deep insights to the price formation process and, hence, is useful in optimizing bidding strategies. In this thesis, the application of this approach is limited to the Italian market because it requires a lot of data that are not always simple to obtain. Before going in detail, we describe first the prices formation process in a competitive electricity market with specific reference to IPEX.

6.2 Price Formation Process in IPEX

The Italian electricity market (IPEX) consists of different markets, including a day-ahead market (see section 2.2.4). The day-ahead market is managed by Gestore del Mercato Elet-

trico (GME) where prices and demand are determined the day before the delivery by means of hourly concurrent auctions. The market sitting starts at 8 a.m. of the ninth day before the day of physical delivery and closes at mid-day (12 p.m.) of the day before of delivery. The market participants, producers and demander's, submit offers/bids where they specify the quantities and the minimum/maximum price at which they are willing to sell/purchase. These bids typically consist of a set of energy blocks and their corresponding prices with other relevant information for every hour of the next day. These bids are submitted through an online web portal called "market participant interface" that is also used to manage and display invoicing data and payables/receivables resulting from transactions that are already concluded in the previous days. These bids are then aggregated by an independent system operator (ISO) in order to construct the supply and demand curves. Only one agent is responsible for this task and his role is very important for many reasons including reliability, independence, non discrimination, unbundling and efficiency. ISO ensures reliability of power grid by coordinating short term operations, independence by not allowing any entity to control the criteria or operating procedures and non-discriminatory access for all market participants without distinction as to customer identity or affiliation. Services unbundling for utilization by the market participants and efficient operating procedures and pricing of services are also responsibility of an ISO.

Once the offers and bids are received by the ISO, supply and demand curves are established by summing up individual supply and demand schedules. In the case of demand, the first step is to replace "zero prices" bids by the market maximum price (for IPEX, the market maximum price is 3000 Euro) without changing the corresponding quantities. After this replacement, the bids are sorted from the highest to the lowest with respect to prices. The corresponding value of the quantities is obtained by cumulating each single demand bid. For supply curve, in contrast, the offers are sorted from the lowest to the highest with respect to prices and the corresponding value of the quantities is obtained by cumulating each single supply offer. Both the curves are monotonically increasing/decreasing as they result from accumulation of each single offer/bid. The supply curve has an upward slope and describes the relationship between the unit price and total quantity offered by suppliers. The

increasing nature of the slope highlights the fact that lower electricity demand will be satisfied by most efficient plants available. With the increase in demand, suppliers will produce electricity with less efficient plants and with high production costs. On the other hand, the demand curve has a downward slope and describes the relation between the unit price and the total quantity required by consumers (Buzoianu et al., 2005). The market equilibrium is the point where both curves intersect each other and the price balances supply and demand schedules. This point determines the market clearing price and the quantity. Accepted offers

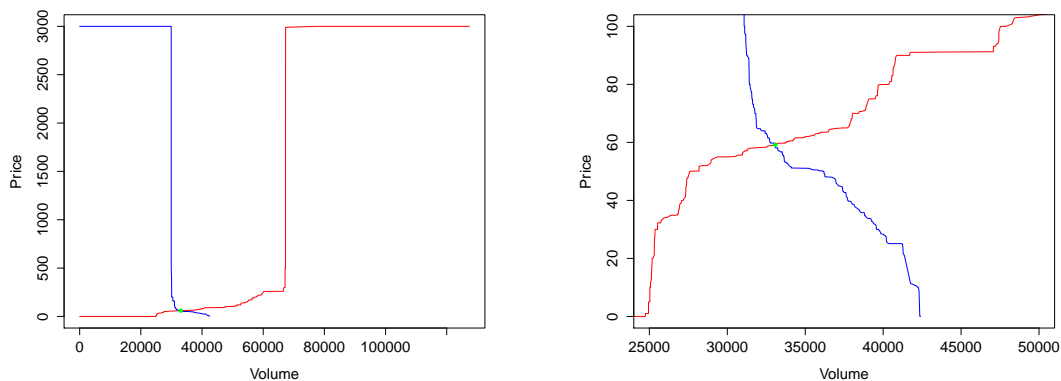


Figure 6.2 IPEX: Supply (red) and Demand (blue) curves (left) and their zoomed version (right)

and bids are those that fall to the left of the intersection of the two curves and all of them are exchanged at the resulted price. An example of the two curves, of the intersection point and their zoomed version is plotted in Figure 6.2.

However, at GME the equilibrium price is different from the market clearing price as the latter accounts for other transactions, e.g. transmission capacity limits between zones, electricity imports from other countries etc. All demand bids and supply offers pertaining to both, pumping unit and consuming units, belonging to foreign virtual zones, that are accepted, are valued at the marginal clearing price of the zone to which they belong. The accepted demand bids pertaining to consuming units belonging to Italian geographical zones are valued at the “Prezzo Unico Nazionale” (national single price, PUN); this price is equal to the average of the prices of geographical zones, weighted for the quantities purchased in

those zones (more information on the GME website¹). The results (market clearing prices and quantities for each hour for the following day) of the day-ahead market (MGP) are made available within 12.55 p.m. of the day before that of delivery.

6.3 Prices Prediction with Supply and Demand Curves

Despite the good forecasting abilities for electricity price/demand of the classical time series models, an important drawback related to them is the fact that they do not provide insight to the supply and demand mechanism consequently to the price/demand formation process. The classical approaches do not explain the relationships between market clearing price (uniform price) and different influential factors, i.e. fuel prices, bidding strategies etc. that can be essential in the problem of price prediction. On the other hand, although auctions are of great use for creating competitive environment, however the information provided by each participant in these auctions are secrete. A main concern to the market players is how to handle incomplete information about other competitors. If the information about other participants are available, it can certainly help the agent to improve his bidding strategies and ends up with significant profits.

This motivates us to look at the problem of price prediction in a completely different and new direction. In the following, the issue of electricity price forecasting is revisited in a context where beside the prediction itself, other remarkable advantages will also emerge. Our approach is quite new in the literature and, at my best knowledge, is completely new for the Italian electricity market. Although, a similar kind of approach has been used by Canale (2014) for the Italian natural gas balancing platform, it is important to know that their work do not consider the limitations related to the data as the data are available with eight-days-lag. Secondly, they performed in-sample prediction and obtained one-day-ahead predictions whereas, in real context, eight-day-ahead predictions are required, thus, leading to a kind of “dummy” experiment. Lastly, in our case both parametric and nonparametric models are considered whereas they consider only a parametric model.

¹<http://www.mercatoelettrico.org>

This approach is based on the idea of modeling the daily supply and demand curves, predicting them and finding the intersection of the predicted curves in order to find the predicted market clearing price and volume. Predicting the whole curves will also help in the optimization of bidding strategies. For this task the functional approach is quite suitable because for each given day, the number of offers/bids data, submitted by producers/buyers in the market, is very large. Thus, finite dimensional (both univariate and multivariate) forecasting techniques could be unsuitable due to the large number of variables. On the contrary, functional models consider a single day as a single functional object and the bids/offers, points on this functional object.

6.3.1 Application to GME Data

To evaluate the performance of our new approach for price prediction, the data from the Italian electricity market is used. In general, the offers and bids data are not easily available and difficult to find. In addition, the size of these data is often very large due the number of bids/offers for each load period and hence not always easy to handle. Note that, for the Italian market these data are available only with a eight-day-lag and thus, in a real forecasting context, eight-days-ahead forecasts are required. For each hour of the day, the

S#	Date	Hour	Buy/Sell	Quantity	Price	Status	Operator	...
1	01/01/2014	15	B	143.83	0.00	ACC	Bilateralista	...
2	01/01/2014	15	B	230.00	0.00	ACC	Bilateralista	...
3	01/01/2014	15	B	17.10	0.00	ACC	IM SPA	...
4	01/01/2014	15	B	1.72	0.00	ACC	Bilateralista	...
5	01/01/2014	15	B	133.33	60.00	ACC	REI S.P.A.	...
6	01/01/2014	15	B	763.51	0.00	ACC	Bilateralista	...

Table 6.1 IPEX: Supply and demand bids

original data (published by GME) consist of different information corresponding to a single buyer or supplier and reported in a XML table format, where every row represents a single bid/offer with its own date, trader name, type(sell or buy), awarded price, awarded quantity

and other related information. An example of the data set is given in Table 6.1.

This work considers data from 1st January 2014 to 30 April 2015 (485 days) where the whole year 2014 is used for model estimation while the last four months are kept for out-of-sample (eight-days-ahead) forecasting. In the first step, the raw bids and offers data for each load period are aggregated, as described in section 6.2, in order to construct the supply and demand curves. We denote $b_{t,j}$ and $s_{t,j}$, the demand curve and supply curve respectively for day t ($t = 1, \dots, n$) and load period j ($j = 1, \dots, 24$). As can be seen from Figure 6.2 (right), the upper flat peaks in demand and supply curves occur due to few constant volumes corresponding to the market maximum price. To recover more functional shape of the curves, these few constants are deleted from the data. The resulted piecewise curves are converted to smooth functions using B-spline basis given by

$$\psi_t^b(j) = \psi(b_{t,j}) = \sum_{g=1}^K \lambda_{tg} \zeta_{tg}(j) \quad \forall j \in J$$

and

$$\psi_t^s(j) = \psi(s_{t,j}) = \sum_{g=1}^K \alpha_{tg} \zeta_{tg}(j) \quad \forall j \in J$$

where λ_{tg} and α_{tg} are constant parameters and $\zeta_{tg}(j)$ are basis functions. An example of smoothed curves along with the original for a particular day are plotted in Figure 6.3. To economize on notation, we will use $\psi_t(j)$ from this point on for both demand and supply curves.

For this issue, a parametric functional model (FFAR) and its nonlinear nonparametric counterpart (NPFAR) were considered. In case of FFAR, a functional response (in our case supply/demand curve) is predicted by its lagged value and is given by

$$\hat{\psi}_t(j) = \int_s \Phi(s,j) \psi_{t-1}(j) \partial s + \eta_t(j), \quad j \in J \quad (6.1)$$

where $\Phi(s, j)$ is the functional parameter (regression function) and $\eta_t(j)$ is the functional error. In the nonparametric case, the relationship is nonlinear and is given by

$$\hat{\psi}_t(j) = \frac{\sum_{t=1}^{n-1} \psi_{t-1}(j) K(b^{-1} \xi(\tilde{\psi}_{t-1}(j), \psi_{t-1}(j)))}{\sum_{t=1}^{n-1} K(b^{-1} \xi(\tilde{\psi}_{t-1}(j), \psi_{t-1}(j)))} \quad (6.2)$$

where K is an Epanechnikov kernel function and b is the bandwidth corresponding to optimal number of neighbors selected by cross-validation. Three different summary measures

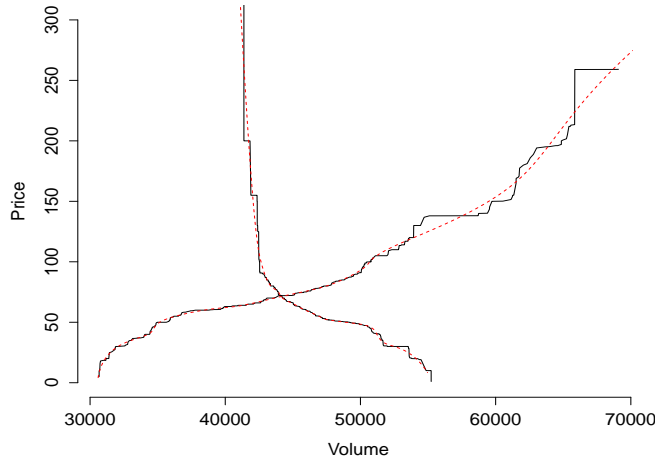


Figure 6.3 Supply and demand curves original (black) and smoothed (dotted red)

namely, Mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the out-of-sample forecasting performance. As benchmark an ARIMA(1,1,1) model was fitted to the scalar time series corresponding to the market clearing prices (PUN) for the same period. In this case one-day-ahead predictions are obtained and compared to the results obtained with our functional approach. Here it is important to know that, as explained in section 6.2, the equilibrium prices are different from market clearing prices in GME as the latter account for other transactions that are not available in our case. Hence, they are different data sets (see for instance Figure 6.4). The forecasting horizons used for both data sets are also different as in case of market clearing price, one-day ahead prediction are required whereas for equilibrium prices (our approach) eight-days ahead. Therefore the results are not directly comparable, however the reader can

get an idea of the performance of both approaches by looking to different error summary measures and particularly to MAPE that summarize relative percentage errors. Finally before applying the models, data are divided into seven groups, each representing a single day of week, in order to account for weekly periodicity. Thus, e.g., for the prediction of Monday, the historical data from all available previous Mondays were used.

For all the stated models, results are listed in Table 6.2. Comparing the parametric and non-parametric functional models, the results in Table 6.2 suggest the superior performance of nonparametric over its parametric counterpart in general. This suggests that the price formation mechanism, and consequently the supply and demand curves process, may contain some nonlinearities. While different error indicators fluctuate all over the day, it can be seen that for both functional models, they are high soon after midday and are comparatively low at the ending hours of the day. For the first 12 hours of the day, the percentage errors remain between 7 to 12 percent while for the second part, 6 to 15 percent indicating more variation

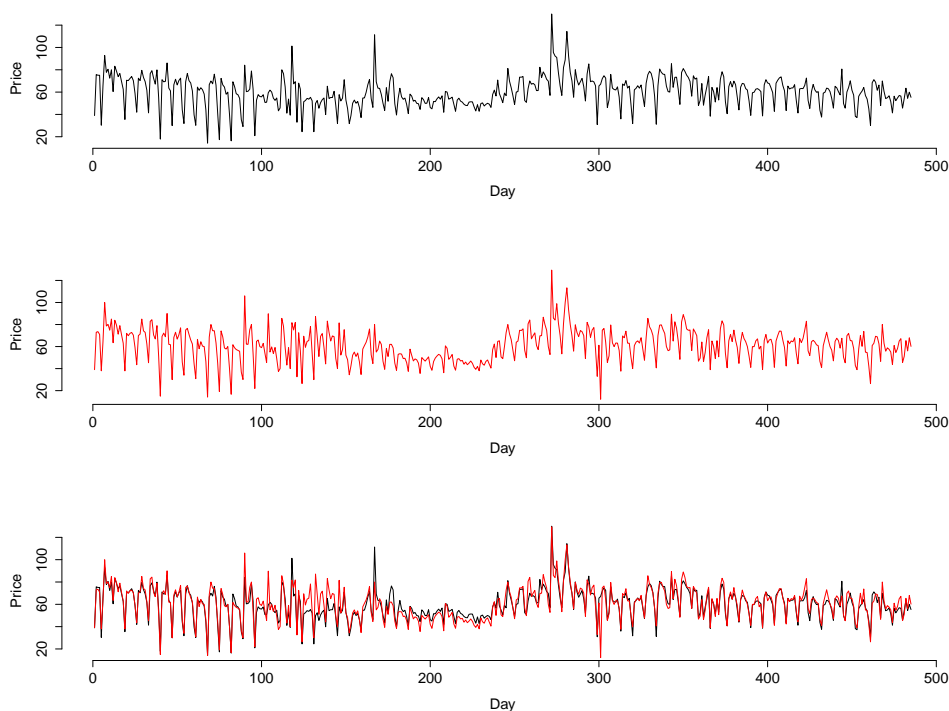


Figure 6.4 IPEX load period 9: market clearing price (black) and equilibrium price (red) for 485 day.

Model	Hour	MAE	MAPE	RMSE	Hour	MAE	MAPE	RMSE
NPFFAR	01	4.29	8.79	6.06	13	5.08	12.24	7.10
FFAR		4.53	9.18	6.15		5.50	12.35	7.19
ARIMA		4.10	8.82	5.86		4.59	10.48	6.25
NPFFAR	02	4.21	9.81	5.64	14	5.68	13.38	7.76
FFAR		4.25	9.70	5.68		6.89	15.64	8.58
ARIMA		3.98	9.95	5.76		5.32	13.11	7.02
NPFFAR	03	4.13	10.49	5.38	15	5.73	12.44	8.07
FFAR		4.59	11.45	5.84		6.72	14.34	8.61
ARIMA		4.27	12.25	6.13		6.05	14.35	8.15
NPFFAR	04	4.47	12.03	6.07	16	4.76	9.91	6.49
FFAR		4.94	12.79	6.30		6.75	13.37	9.32
ARIMA		4.78	15.65	6.70		5.37	11.69	7.34
NPFFAR	05	4.55	12.49	6.03	17	4.8	9.42	6.58
FFAR		4.70	12.49	6.09		5.48	10.51	6.98
ARIMA		5.10	16.67	7.07		4.93	9.83	6.59
NPFFAR	06	4.17	10.26	5.70	18	4.77	8.81	6.33
FFAR		4.69	11.40	6.30		5.86	10.46	7.37
ARIMA		4.79	13.61	6.77		4.22	7.57	5.66
NPFFAR	07	4.06	8.15	5.20	19	5.78	9.73	8.16
FFAR		4.62	9.43	5.81		7.62	12.31	9.44
ARIMA		5.09	11.31	6.51		4.57	7.28	6.03
NPFFAR	08	3.99	7.27	5.09	20	5.49	8.13	7.67
FFAR		5.27	9.51	6.60		6.94	10.33	9.42
ARIMA		5.40	11.04	7.14		4.03	6.23	5.61
NPFFAR	09	4.68	8.07	6.25	21	5.63	8.28	7.31
FFAR		6.88	11.71	9.09		6.91	10.25	10.09
ARIMA		7.20	13.65	9.54		4.16	6.52	5.95
NPFFAR	10	5.01	8.85	6.39	22	3.92	6.99	5.25
FFAR		6.21	10.69	8.13		4.11	7.20	5.39
ARIMA		5.95	10.94	8.25		3.21	5.72	4.45
NPFFAR	11	5.42	9.83	7.24	23	3.45	6.53	4.43
FFAR		7.11	13.11	9.11		3.86	7.01	4.93
ARIMA		5.66	10.60	8.23		3.06	5.89	4.06
NPFFAR	12	5.29	10.13	7.09	24	3.22	6.73	4.34
FFAR		7.05	13.70	9.15		4.27	8.35	5.25
ARIMA		5.66	11.39	7.78		2.94	6.08	3.77

Table 6.2 IPEX: Prediction accuracy statistics: Nonparametric Functional Functional AutoRegressive (NPFFAR), Functional Functional AutoRegressive (FFAR), AutoRegressive Integrated Moving Average (ARIMA)

Model	Hour	NPFFAR	FFAR	ARIMA	Hour	NPFFAR	FFAR	ARIMA
NPFFAR	01	-	0.61	0.64	13	-	0.29	0.80
FFAR		0.39	-	0.85		0.71	-	0.92
ARIMA		0.36	0.15	-		0.20	0.08	-
NPFFAR	02	-	0.71	0.47	14	-	0.01	0.52
FFAR		0.29	-	0.39		0.99	-	0.98
ARIMA		0.53	0.61	-		0.48	0.02	-
NPFFAR	03	-	0.23	0.08	15	-	0.02	0.05
FFAR		0.77	-	0.25		0.98	-	0.54
ARIMA		0.92	0.75	-		0.95	0.46	-
NPFFAR	04	-	0.35	0.12	16	-	<0.01	0.01
FFAR		0.65	-	0.29		0.99	-	0.98
ARIMA		0.88	0.71	-		0.99	0.02	-
NPFFAR	05	-	0.56	0.04	17	-	0.15	0.45
FFAR		0.44	-	0.10		0.85	-	0.73
ARIMA		0.96	0.90	-		0.55	0.27	-
NPFFAR	06	-	0.12	0.06	18	-	0.02	0.64
FFAR		0.88	-	0.21		0.98	-	0.95
ARIMA		0.94	0.79	-		0.36	0.05	-
NPFFAR	07	-	0.02	<0.01	19	-	0.02	0.96
FFAR		0.98	-	0.05		0.98	-	0.98
ARIMA		0.99	0.95	-		0.04	0.02	-
NPFFAR	08	-	<0.01	<0.01	20	-	0.02	0.88
FFAR		0.99	-	0.10		0.98	-	0.99
ARIMA		0.99	0.90	-		0.12	<0.01	-
NPFFAR	09	-	<0.01	<0.01	21	-	0.42	0.84
FFAR		0.99	-	0.05		0.58	-	0.97
ARIMA		0.99	0.95	-		0.14	0.03	-
NPFFAR	10	-	<0.01	<0.01	22	-	0.39	0.88
FFAR		0.99	-	0.34		0.61	-	0.99
ARIMA		0.99	0.66	-		0.12	<0.01	-
NPFFAR	11	-	0.02	0.04	23	-	0.15	0.82
FFAR		0.98	-	0.86		0.85	-	0.92
ARIMA		0.96	0.14	-		0.18	0.08	-
NPFFAR	12	-	<0.01	0.07	24	-	<0.01	0.90
FFAR		0.99	-	0.96		0.99	-	0.99
ARIMA		0.93	0.04	-		0.10	<0.01	-

Table 6.3 IPEX: P-values for the DM test for equal prediction accuracy versus the alternative hypothesis that model in the row is more accurate than model in the column (squared loss function used)

in the second part of the day. Comparing the eight-days-ahead forecasting results obtained with our functional approach with those obtained with one-day-ahead with ARIMA, one can notice that the functional models produce better results in general. The differences are relatively higher among MAPE values as this indicator is more sensitive to outliers. For the first 12 hours of the day, the superior performance of functional models compare to ARIMA is evident however for some hours in the second part of the day, ARIMA produced better results. Finally, among all the three models, the performance of NPFFAR is superior in general. An example of the predicted and original supply and demand curves are illustrated in Figure 6.5. The observed and forecasted price (volume) for this reference day is 51.14 and 50.81 (32976 and 32716) respectively.

To assess the significance of the differences between the summary statistics listed in Table 6.2, the Diebold and Mariano (DM) test for equal predictive accuracy has been performed for each couple of indicators. The results concerning DM test are listed in Table 6.3 and confirm the significance of overall prediction accuracy of NPFFAR compare to both FFAR and ARIMA. On most load periods in the first half of the day, NPFFAR is highly

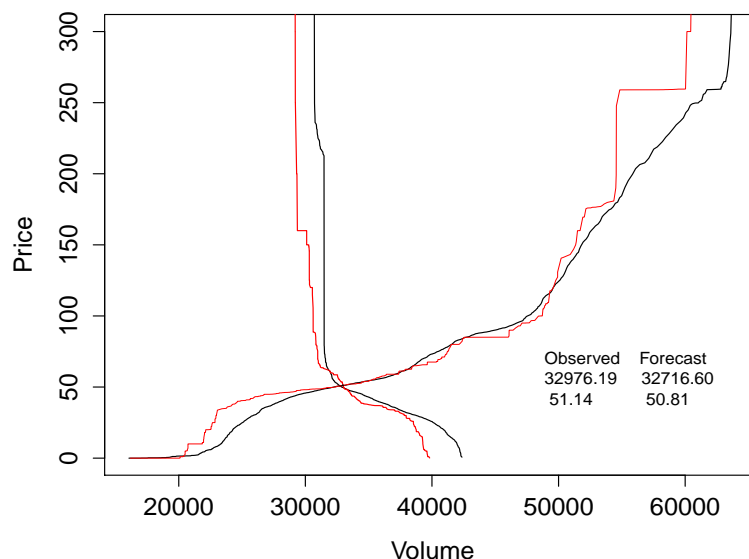


Figure 6.5 IPEX: supply and demand curves in original (red) and forecasted (black)

significant whereas in the second half, no model shows superiority in general. Further it can be seen that, on five percent significance level, ARIMA perform statistically better only on one load period whereas NPPFAR shows statistically superior results on eight load periods.

6.4 Optimizing Bidding Strategy

Apart from the improvement in forecasting accuracy, it is important to stress that forecasting the entire demand/supply curves can substantially improve the supplier/buyer bidding strategy resulting in a significant financial gain. As explained in section 6.3, a main concern to the market players is how to handle incomplete information about other competitors as the biddings are done secretly. Since this approach considers the forecasting of entire curve (supply/demand), the above issue can be addressed with this technique. Once the forecasted curves are available, a trader can see his position with different bidding strategies that can rise/lower the price.

To clearly understand, let suppose a trader requires a moderate/large quantity to buy tomorrow. To buy his desired quantity, one strategy he can use is to bid zero which will lead him to pay market clearing price but in that case the market clearing price will be determined by other participants. Alternatively, the trader can predict supply and demand curves with the current approach and can apply different bidding strategies to both curves and see the outcomes of his experiments. To be more precise, let a buyer want to buy a large quantity tomorrow. He obtained the predicted curves given in Figure 6.5 with the current approach. To lower the price, the buyer can use different strategies that will results in different gains. For example, one strategy involves the buyer to submit an extra non-standard supply offer for a small quantity with a low price. In the predicted supply curve, the buyer submitted an artificial quantity size of 100 MWh with 1 Euro price. The modified supply and demand curves show decline in price of $50.81 - 50.58 = 0.23$ Euro and result in a profit of $\text{quantity} \times 0.23$ Euro for the buyer. The modified curve is represented by a solid line in Figure 6.6 along with a zoom in its neighborhood of the intersection.

Now, an important question is to know how much it will translate to the original curves.

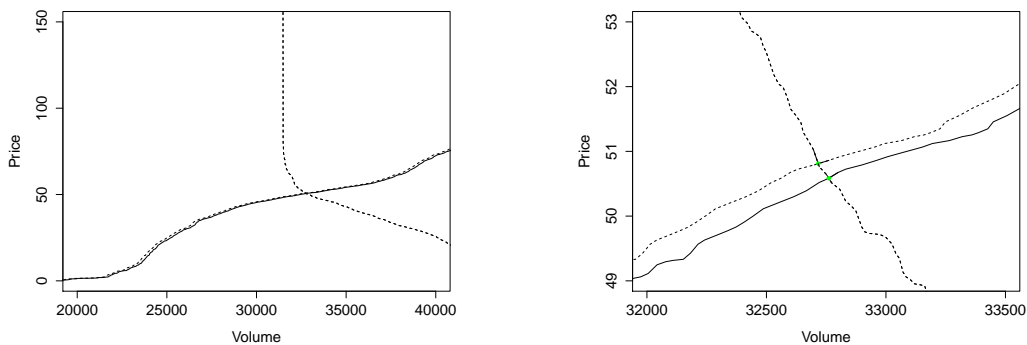


Figure 6.6 IPEX: what-if simulations: (left) Predicted supply and demand curves (dashed) with modified supply curves (solid) (right) and zoom on a neighbourhood of the intersection

The same non-extra bid (with the same quantity and price) is added to the original curves and the modified curves are plotted in Figure 6.7. In this case the original price lower from 51.14 to 50.71 leading the trader to save quantity $\times 0.43$ Euro. With the increase of the quantity offered with low price, the shape of the supply curve will lower from original resulting in low prices. In general, with this strategy to move the intersection point, the lower the offered price, the higher need to be the offered quantity. Evidently this strategy will lower the prices with no consequences on market clearing volume from a practical viewpoint. In general, with the current approach suppliers and other participant can also optimize their bidding strategies.

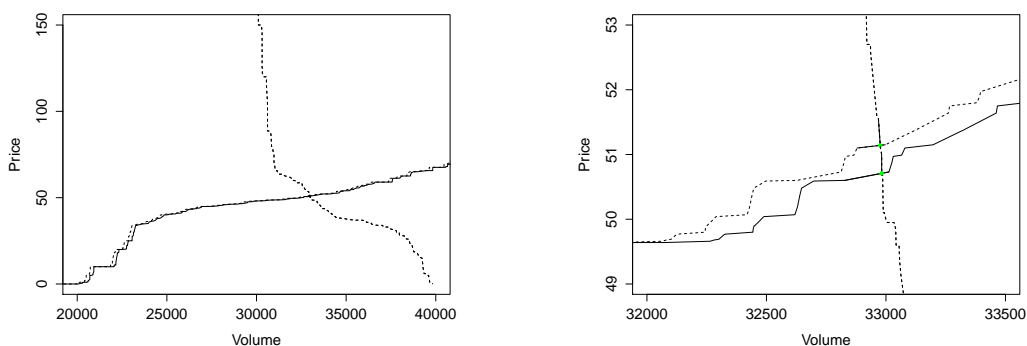


Figure 6.7 IPEX: what-if simulations: (left) original supply and demand curves (dashed) with modified supply curves (solid) (right) and zoom on a neighbourhood of the intersection

6.5 Conclusion

This chapter revisited the issue of price prediction using a completely different approach. Our new approach was based on the idea of modeling and forecasting supply and demand curves and finding the intersection of the predicted curves in order to obtain market clearing price. For this issue, functional modeling techniques were used as they were best suited for this task. Indeed, our current approach exploit the advantages of functional modeling in more detail. The application of this approach was limited to the Italian electricity market where the data are available with a eight-day-lag and thus, eight-days-ahead forecasts were required. As a benchmark, an ARIMA was applied directly to market clearing price time series that is available with one-day-lag and thus, one-day-ahead forecasts were obtained. Results concerning out-of-sample forecasts suggest the usefulness and superior performance of our new approach compare to the classical ARIMA model.

Apart from the improvement in forecasting accuracy, it is important to stress that our new approach can be used for optimizing bidding strategies. As forecasting the whole curves gives deep insight into the market, our analysis showed that this strategy can significantly improve bidding strategies and maximize traders profit.

Chapter 7

Conclusion and Further Research

The liberalization of electricity sector not only brought important benefits to consumers such as low prices, more choices, reliable and secure electric supply but it also introduced a new field of research. Accurate modeling and forecasting of different variables related to the market e.g. prices, demand, production etc. became more crucial due to market structure. In deregulated electricity markets, the demand and prices are determined the day before the delivery by means of (semi-) hourly concurrent auctions for the next day. Thus, accurate forecasting is very important issue for an efficient management of power grid. In the past, various techniques have been developed both for price and demand prediction with different levels of complexity and final performance. However, it is difficult to find a single model that outperforms all others in every situation. This thesis addressed the issue of modeling and forecasting electricity demand and prices following two different approaches.

In the first approach, the issue of out-of-sample forecasting for both electricity price and demand was addressed using an additive component model that includes some deterministic and a stochastic residual components. The deterministic components include a long-term dynamics, annual and weekly periodicities and calendar effects, whereas the stochastic component describes the intra-daily and the short-term dynamics. Deterministic components were estimated using splines and dummy variables. For the stochastic component, different univariate and multivariate models were used with varying level of complexity. In particular, different functional models were used and compared with more traditional models that

have been extensively used for price and demand prediction. For a detailed study, five different electricity markets were used and the out-of-sample forecasting performance of the different models was summarized by several descriptive indicators. The analyses suggest that the multivariate approach leads to better results than the univariate one and that, within the multivariate framework, functional models are, in general, the most accurate. This is confirmed also by the Diebold and Mariano test. The obtained results were compared with other works in the literature suggesting our forecasting errors are smaller compared with the state-of-art prediction techniques.

The second approach mainly focused on price prediction in a context where beside the prediction itself, other remarkable advantages will also emerge. For this issue a completely different approach was used. It is based on the idea of modeling supply and demand curves, predicting them and finding the intersection point of the predicted curves in order to obtain the market clearing price and volume. For this issue, only data from the Italian electricity market were used. Different functional models were used for this task as they were best suited for this issue. As benchmark, a classical time series model (ARIMA) was fitted to the scalar time series corresponding to the market clearing prices obtained from the crossing points of supply and demand curves. The results show superior forecasting performance of our functional approach compared to ARIMA. Beside the improvement in forecasting accuracy, this approach leads to a tremendous advantage compared to classical approaches. As forecasting the whole curves gives deep insight into the market, our analysis showed that this strategy can significantly improve bidding strategies and maximize traders profit.

In the light of our findings, functional modeling approach perform relatively better than the classical approaches and is effective for both price and demand prediction.

In our further research, the functional techniques will be used for classification purposes. Clearly, this is another advantage and shows the usefulness of this approach. In our research, classification will be used to classify the daily/hourly loads and prices profile. This can be useful for both to reduce the number of models to be estimated and as a prediction technique itself.

Bibliography

- Amjady, N. (2001). Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *Power Systems, IEEE Transactions on*, 16(3):498–505.
- Amjady, N. (2006). Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *Power Systems, IEEE Transactions on*, 21(2):887–896.
- Aneiros-Pérez, G., Cao, R., and Vilar-Fernández, J. M. (2011). Functional methods for time series prediction: a nonparametric approach. *Journal of Forecasting*, 30(4):377–392.
- Bernardi, M. and Petrella, L. (2015). Multiple seasonal cycles forecasting model: the italian electricity demand. *Statistical Methods & Applications*, pages 1–25.
- Bianco, V., Manca, O., and Nardini, S. (2009). Electricity consumption forecasting in italy using linear regression models. *Energy*, 34(9):1413–1421.
- Bierbrauer, M., Menn, C., Rachev, S. T., and Trück, S. (2007). Spot and derivative pricing in the eex power market. *Journal of Banking & Finance*, 31(11):3462–3485.
- Bisaglia, L., Bordignon, S., and Marzovilli, M. (2010). Modelling and forecasting hourly spot electricity prices: some preliminary results.
- Boisseleau, F. (2004). *The role of power exchanges for the creation of a single European electricity market. Market design and market regulation*. TU Delft, Delft University of Technology.

- Bordignon, S., Bunn, D. W., Lisi, F., and Nan, F. (2013). Combining day-ahead forecasts for british electricity prices. *Energy Economics*, 35:88–103.
- Borovkova, S. and Permana, F. J. (2006). Modelling electricity prices by the potential jump-diffusion. In *Stochastic Finance*, pages 239–263. Springer.
- Bosco, B., Parisio, L., Pelagatti, M., and Baldi, F. (2010). Long-run relations in european electricity prices. *Journal of applied econometrics*, 25(5):805–832.
- Bosco, B. P., Parisio, L. P., and Pelagatti, M. M. (2007). Deregulated wholesale electricity prices in italy: an empirical analysis. *International Advances in Economic Research*, 13(4):415–432.
- Bosq, D. (2000). *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media.
- Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. Oakland, CA: Holden-Day.
- Brockwell, P. J. and Davis, R. A. (2006). *Introduction to time series and forecasting*. Springer Science & Business Media.
- Bunn, D. and Farmer, E. D. (1985). Comparative models for electrical load forecasting.
- Bunn, D. W. (2004a). *Modelling prices in competitive electricity markets*. J. Wiley.
- Bunn, D. W. (2004b). Structural and behavioural foundations of competitive electricity prices. *Power*, 70(80):90.
- Buzoianu, M., Brockwell, A., and Seppi, D. J. (2005). A dynamic supply-demand model for electricity prices.
- Canale, A, V. S. (2014). Constrained functional time series: an application to demand and supply curves in the italian natural gas balancing platform, mox–report no. 42/2014.

- Carpio, J., Juan, J., and López, D. (2014). Multivariate exponential smoothing and dynamic factor model applied to hourly electricity price analysis. *Technometrics*, 56(4):494–503.
- Charlton, N. and Singleton, C. (2014). A refined parametric model for short term load forecasting. *International Journal of Forecasting*, 30(2):364–368.
- Chen, B.-J., Chang, M.-W., and Lin, C.-J. (2004). Load forecasting using support vector machines: A study on eunite competition 2001. *Power Systems, IEEE Transactions on*, 19(4):1821–1830.
- Cho, H., Goude, Y., Brossat, X., and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, 108(501):7–21.
- Conejo, A. J., Contreras, J., Espínola, R., and Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, 21(3):435–462.
- Contreras, J., Espinola, R., Nogales, F. J., and Conejo, A. J. (2003). Arima models to predict next-day electricity prices. *Power Systems, IEEE Transactions on*, 18(3):1014–1020.
- Cuambe, I. D. F. (2013). Electricity load demand forecasting in portugal using least-squares support vector machines.
- Dai, W. and Wang, P. (2007). Application of pattern recognition and artificial neural network to load forecasting in electric power system. In *Natural Computation, 2007. ICNC 2007. Third International Conference on*, volume 1, pages 381–385. IEEE.
- De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

- Dordonnat, V., Koopman, S. J., and Ooms, M. (2010). Intra-daily smoothing splines for time-varying regression models of hourly electricity load. *Journal of Energy Markets*, 3:17–52.
- Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., and Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24(4):566–587.
- Ediger, V. Ş. and Akar, S. (2007). Arima forecasting of primary energy demand by fuel in turkey. *Energy Policy*, 35(3):1701–1708.
- Erlwein, C., Benth, F. E., and Mamon, R. (2010). Hmm filtering and parameter estimation of an electricity spot price model. *Energy Economics*, 32(5):1034–1043.
- Erni, D. (2012). *Day-Ahead Electricity Spot Prices-Fundamental Modelling and the Role of Expected Wind Electricity Infeed at the European Energy Exchange*. PhD thesis, University of St. Gallen.
- Escribano, A., Ignacio Peña, J., and Villaplana, P. (2011). Modelling electricity prices: International evidence*. *Oxford Bulletin of Economics and Statistics*, 73(5):622–650.
- Espinoza, M., Joye, C., Belmans, R., and Moor, B. D. (2005). Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. *Power Systems, IEEE Transactions on*, 20(3):1622–1630.
- Fan, S. and Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *Power Systems, IEEE Transactions on*, 27(1):134–141.
- Fanone, E., Gamba, A., and Prokopczuk, M. (2013). The case of negative day-ahead electricity prices. *Energy Economics*, 35:22–34.
- Feinberg, E. A. and Genethliou, D. (2005). Load forecasting. In *Applied mathematics for restructured electric power systems*, pages 269–285. Springer.

- Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109:10–28.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer.
- Fleten, S.-E., Heggedal, A. M., and Siddiqui, A. (2011). Transmission capacity between norway and germany: a real options analysis. *Journal of Energy Markets*, 4(1):121–147.
- Garcia, R. C., Contreras, J., Van Akkeren, M., and Garcia, J. B. C. (2005). A garch forecasting model to predict day-ahead electricity prices. *Power Systems, IEEE Transactions on*, 20(2):867–874.
- Gellings, C. W. (1996). Demand forecasting for electric utilities. *Lilburn, GA (US); Fairmont Press, Inc.*
- Gianfreda, A. and Grossi, L. (2012). Forecasting italian electricity zonal prices with exogenous variables. *Energy Economics*, 34(6):2228–2239.
- Göb, R., Lurz, K., and Pievatolo, A. (2013). Electrical load forecasting by exponential smoothing with covariates. *Applied Stochastic Models in Business and Industry*, 29(6):629–645.
- Gonzalez-Romera, E., Jaramillo-Moran, M. A., and Carmona-Fernandez, D. (2006). Monthly electric energy demand forecasting based on trend extraction. *Power Systems, IEEE Transactions on*, 21(4):1946–1953.
- Goude, Y., Nedellec, R., and Kong, N. (2014). Local short and middle term electricity load forecasting with semi-parametric additive models. *Smart Grid, IEEE Transactions on*, 5(1):440–446.
- Gross, G. and Galiana, F. D. (1987). Short-term load forecasting. *Proceedings of the IEEE*, 75(12):1558–1573.

- Hagan, M. T. and Behr, S. M. (1987). The time series approach to short term load forecasting. *Power Systems, IEEE Transactions on*, 2(3):785–791.
- Hahn, H., Meyer-Nieberg, S., and Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 199(3):902–907.
- Hao, C. (2007). A new method of load forecasting based on generalized autoregressive conditional heteroscedasticity model [j]. *Automation of Electric Power Systems*, 15:012.
- Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series. *Journal of Time Series Analysis*, 13(3):209–232.
- Hart, J. D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 173–187.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC Press.
- Hayati, M. and Shirvany, Y. (2007). Artificial neural network approach for short term load forecasting for illam region. *World Academy of Science, Engineering and Technology*, 28:280–284.
- Hellström, J., Lundgren, J., and Yu, H. (2012). Why do electricity prices jump? empirical evidence from the nordic electricity market. *Energy Economics*, 34(6):1774–1781.
- Hippert, H. S., Pedreira, C. E., and Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *Power Systems, IEEE Transactions on*, 16(1):44–55.
- Hong, T., Gui, M., Baran, M., and Willis, H. (2010). Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–8.

- Hor, C.-L., Watson, S. J., and Majithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. *Power Systems, IEEE Transactions on*, 20(4):2078–2085.
- Janczura, J. and Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy economics*, 32(5):1059–1073.
- Jonas Andersson, J. L. (2010). Modeling and forecasting electricity consumption by functional data analysis. *Journal of Energy Markets*, 3(1):3–15.
- Kandil, N., Wamkeue, R., Saad, M., and Georges, S. (2006). An efficient approach for short term load forecasting using artificial neural networks. *International Journal of Electrical Power & Energy Systems*, 28(8):525–530.
- Karakatsani, N. V. and Bunn, D. W. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4):764–785.
- Kim, C.-i., Yu, I.-K., and Song, Y. (2002). Prediction of system marginal price of electricity using wavelet transform analysis. *Energy Conversion and Management*, 43(14):1839–1851.
- Knittel, C. R. and Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, 27(5):791–817.
- Koopman, S. J., Ooms, M., and Carnero, M. A. (2007). Periodic seasonal reg-arfima-garch models for daily electricity spot prices. *Journal of the American Statistical Association*, 102(477):16–27.
- Kristiansen, T. (2012). Forecasting nord pool day-ahead prices with an autoregressive model. *Energy Policy*, 49:328–332.
- Kyriakides, E. and Polycarpou, M. (2007). Short term electric load forecasting: A tutorial. In *Trends in Neural Computation*, pages 391–418. Springer.

- Lisi, F. and Nan, F. (2014). Component estimation for electricity prices: Procedures and comparisons. *Energy Economics*, 44:143–159.
- Lowther, J. and Shene, C.-K. (2003). Teaching b-splines is not difficult! In *ACM SIGCSE Bulletin*, volume 35, pages 381–385. ACM.
- Lucia, J. J. and Schwartz, E. S. (2002). Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of derivatives research*, 5(1):5–50.
- Moghram, I. and Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. *Power Systems, IEEE Transactions on*, 4(4):1484–1491.
- Mohandes, M. (2002). Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research*, 26(4):335–345.
- Nan, F. (2009). Forecasting next-day electricity prices: from different models to combination. *University of Padua*.
- Nan, F., Bordignon, S., Bunn, D. W., and Lisi, F. (2014). The forecasting accuracy of electricity price formation models. *International Journal of Energy and Statistics*, 02(01):1–26.
- Nedellec, R., Cugliari, J., and Goude, Y. (2014). Gefcom2012: Electric load forecasting and backcasting with semi-parametric models. *International Journal of forecasting*, 30(2):375–381.
- Niu, D., Li, J., and Li, J. (2007). Middle–long electric power load forecasting based on co-integration and support vector machine. In *Natural Computation, 2007. ICNC 2007. Third International Conference on*, volume 1, pages 596–600. IEEE.
- Pai, P.-F. and Hong, W.-C. (2005). Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management*, 46(17):2669–2688.

- Pandian, S. C., Duraiswamy, K., Rajan, C. C. A., and Kanagaraj, N. (2006). Fuzzy approach for short term load forecasting. *Electric Power Systems Research*, 76(6):541–548.
- Pao, H.-T. (2007). Forecasting electricity market pricing using artificial neural networks. *Energy Conversion and Management*, 48(3):907–912.
- Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G., Katsikas, S., and Liatsis, P. (2008). Electricity demand loads modeling using autoregressive moving average (arma) models. *Energy*, 33(9):1353–1360.
- Parker, D. S. (2003). Research highlights from a large scale residential monitoring study in a hot climate. *Energy and Buildings*, 35(9):863–876.
- Petrella, A. and Sapio, S. (2009). A time series analysis of day-ahead prices on the italian power exchange. In *Energy Market, 2009. EEM 2009. 6th International Conference on the European*, pages 1–6. IEEE.
- Pirino, D. and Renò, R. (2010). Electricity prices: a nonparametric approach. *International Journal of Theoretical and Applied Finance*, 13(02):285–299.
- Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., and Brace, C. (1997). Shorter-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, 13(2):161–174.
- Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*, volume 7. Springer.
- Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 539–572.
- Ranaweera, D., Hubele, N., and Karady, G. (1996). Fuzzy logic for short term load forecasting. *International journal of electrical power & energy systems*, 18(4):215–222.
- Raviv, E., Bouwman, K. E., and Van Dijk, D. J. (2013). Forecasting day-ahead electricity prices: Utilizing hourly prices.

- Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *The annals of Statistics*, pages 141–156.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43.
- Sahay, K. B. and Tripathi, M. (2014). Day ahead hourly load forecast of pjm electricity market and iso new england market by using artificial neural network. In *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*, pages 1–5. IEEE.
- Schlueter, S. (2010). A long-term/short-term model for daily electricity prices with dynamic volatility. *Energy Economics*, 32(5):1074–1081.
- Shang, H. L. et al. (2010). Nonparametric modeling and forecasting electricity demand: an empirical study. *Monash Econometrics and Business Statistics Working Papers*, (19/10).
- Singhal, D. and Swarup, K. (2011). Electricity price forecasting using artificial neural networks. *International Journal of Electrical Power & Energy Systems*, 33(3):550–555.
- Soares, L. J. and Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to brazilian data. *International Journal of Forecasting*, 24(4):630–644.
- Soares, L. J. and Souza, L. R. (2006). Forecasting electricity demand using generalized long memory. *International Journal of Forecasting*, 22(1):17–28.
- Studiorum, A. M. (2007). Econometric models for the analysis of electricity markets.
- Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152.
- Taylor, J. W. (2012). Short-term load forecasting with exponentially weighted methods. *Power Systems, IEEE Transactions on*, 27(1):458–464.

- Taylor, J. W., de Menezes, L. M., and McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1):1–16.
- Taylor, J. W. and McSharry, P. E. (2007). Short-term load forecasting methods: An evaluation based on european data. *Power Systems, IEEE Transactions on*, 22(4):2213–2219.
- Trueck, S., Weron, R., and Wolff, R. (2007). Outlier treatment and robust approaches for modeling electricity spot prices.
- Veraart, A. E. and Veraart, L. A. (2012). Modelling electricity day-ahead prices by multivariate lévy semistationary processes. *CREATES Research Papers 2012-2013, Aarhus university*.
- Vilar, J. M., Cao, R., and Aneiros, G. (2012). Forecasting next-day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*, 39(1):48–55.
- Wang, C.-h., Grozev, G., and Seo, S. (2012). Decomposition and statistical analysis for regional electricity demand forecasting. *Energy*, 41(1):313–325.
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.
- Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*, volume 403. John Wiley & Sons.
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030 – 1081.
- Weron, R., Bierbrauer, M., and Trück, S. (2004a). Modeling electricity prices: jump diffusion and regime switching. *Physica A: Statistical Mechanics and its Applications*, 336(1):39–48.

-
- Weron, R. and Misiorek, A. (2005). Forecasting spot electricity prices with time series models. In *Proceedings of the European Electricity Market EEM-05 Conference*, pages 133–141.
- Weron, R., Simonsen, I., and Wilman, P. (2004b). Modeling highly volatile and seasonal markets: evidence from the nord pool electricity market. In *The application of econophysics*, pages 182–191. Springer.
- Yang, J. (2006). *Power system short-term load*. PhD thesis, TU Darmstadt.
- Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62.

Ismail Shah

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.
Tel. +39 3425124098
e-mail: shah@stat.unipd.it

Current Position

Since January 2013; (expected completion: March 2016)

PhD Student in Statistical Sciences, University of Padova.

Thesis title: Modeling and Forecasting Electricity Market Variables

Supervisor: Prof. Francesco Lisi

Research interests

- Time series analysis
- Functional data analysis
- Nonparametric smoothing techniques
- Applied statistics

Computer skills

- R
- STATA
- SPSS
- MatLab

Education

September 2009 – August 2010

Master of Science, Statistics.

Lund University, Lund University School of Economics and Management

Title of dissertation: "Human Development Index and Human Poverty Index for Indian states, 2005: multivariate statistical analysis of basic indicators."

Supervisor: Prof. Björn Holmquist

Final mark: 60ECTS

September 2005 – August 2007

Master of Statistics.

University of Peshawar, Pakistan

Final mark: 817/1200

Work experience

December 2012 – Till date

University of Swabi, Swabi, Pakistan.

Lecturer (Statistics).

May 2012 – November 2012

Abdul Wali Khan University Mardan, Mardan, Pakistan.

Lecturer (Statistics),

Awards and Scholarship

2013-2015

PhD Scholarship, Università Degli Studi Di Padova, Italy.

Language skills

Pushto: native; Urdu: native; English: fluent; Italian: basic.

Publications

Shah, I. and Lisi, F., 2015. "Day-ahead electricity demand forecasting with nonparametric functional models". In European Energy Market (EEM), 2015 12th International Conference on the (pp. 1-5). IEEE.

Durante, D., Shah, I. and Torelli, N., (2014). "Bayesian nonparametric modeling of contraceptive use in India", arXiv:1405.7555

Shah, I. and Waleed, M., 2010. "The Fisher Equation, Belgium before and after Euro currency." Lund University Sweden.

References

Prof. Francesco Lisi

Department of Statistical Sciences
University of Padua
Via Cesare Battisti 241
35121 Padua, Italy
Phone: +39 049 8274182
e-mail: francesco.lisi@unipd.it

Prof. Björn Holmquist

Department of Statistics
Lund University
Tycho Brahes väg 1
Box 743, 22007 Lund, Sweden
Phone: +46 46 222 8926
e-mail: Bjorn.Holmquist@stat.lu.se