TIME SERIES ANALYSIS AND FORECASTING ELECTRICITY PRICES IN TURKEY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

SEYED AMIR HAMED ZAKERI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

DECEMBER 2015

Approval of the thesis:

TIME SERIES ANALYSIS AND FORECASTING ELECTRICITY PRICES IN TURKEY

submitted by **SEYED AMIR HAMED ZAKERI** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** __________

Prof. Dr. Ayşen Dener Akkaya
Head of Department, **Statistics** __________

Assist. Prof. Dr. Ceylan Yozgatlıgil
Supervisor, **Statistics Department, METU** __________

Assoc. Prof. Dr. Ömür Uğur
Co-supervisor, **Inst. of Applied Mathematics, METU** __________

Examining Committee Members:

Prof. Dr. İnci Batmaz
Statistics Department, METU __________

Assist. Prof. Dr. Ceylan Yozgatlıgil
Statistics Department, METU __________

Assoc. Prof. Dr. Ömür Uğur
Inst. of Applied Mathematics, METU __________

Assoc. Prof. Dr. Cem Iyigün
Endustrial Engineering Department, METU __________

Assoc. Prof. Dr. Seher Nur Sülkü
Econometrics Department, Gazi University __________

Date: __________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: SEYED AMIR HAMED ZAKERI

Signature :

ABSTRACT

TIME SERIES ANALYSIS AND FORECASTING ELECTRICITY PRICES IN TURKEY

Zakeri, Seyed Amir Hamed
M.S., Department of Statistics
Supervisor : Assist. Prof. Dr. Ceylan Yozgatlıgil
Co-Supervisor : Assoc. Prof. Dr. Ömür Uğur

December 2015, 83 pages

Due to the liberalization of the electricity market, prices are now determined based on contracts on regulated markets and their behavior is mainly driven by constant supply and demand forces. Power producers and consumers need accurate price forecasting tools in a competitive market. Price forecasts give important information for producers and consumers to plan bidding strategies to maximize their benefits and utilities. Analysis of hourly electricity prices in Turkey is challenging due to the existence of multiple seasonality. In this study, we construct a time series model and obtain short-term forecasts of hourly electricity prices using multiple regression method. We used lagged price values, demand as the exogenous variable and dummy variebles for Saturdays and Sundays to capture the seasonality in the price

Keywords: Electricity Price Forecasting, Time Series, ARMAX, GARCH

ÖZ

ZAMAN SERISI ANALIZI VE TÜRKIYE'DE ELEKTRIK FIYAT TAHMINI

Zakeri, Seyed Amir Hamed
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi : Yrd. Doç. Dr. Ceylan Yozgatlıgil
Ortak Tez Yöneticisi : Doç. Dr. Ömür Uğur

Aralık 2015 , 83 sayfa

Elektrik piyasalarındaki liberalleşme sonucunda, elektrik fiyatları normal piyasa fiyatları üstündeki sözleşmelerle belirlenmektedir ve piyasa hareketleri arz ve taleplerine göre hareket etmektedir. Rekabetçi piyasalar, elektrik üreticileri ve kullanıcılarını elektrik fiyatlarının doğru öngörmeye itmektedir. Fiyat öngörüleri üretici ve tüketicilere elektrik kullanımı ve açık arttırmalarda önemli bilgiler sunmaktadır. Türkiyedeki saatlik elektrik fiyatlarının belirlenmesi, serinin birden fazla mevsimellellik içermesi nedeniyle zorlayıcıdır. Bu çalışmada, saatlik elektrik fiyatlarının modellenmesi ve kısa dönemli öngörülerin elde edilmesi planlanmaktadır.

Anahtar Kelimeler: Elektrik Fiyat Tahmini, Zaman Serileri, ARMAX, GARCH

To my soul mate

Bahar

and to my parents

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Professor Ceylan Yozgatligil, who has been a constant support for me throughout my time at ODTU. Her kindness, eagerness to help and her knowledge made the bumpy path to graduation into a smooth and silken road. I would also like to thank my co-advisor, Professor Ömür Uğur, from whom I was able to hone my skill and make the bridge between statistics and finance. I would also like to thank Doctor Erkan Kalayci for providing me the data do to my study. Many thanks to the committee for guiding me to become a better researcher. And finally I would like to thank my parents for their support. They have been the greatest source of inspiration for me in my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

EPF	Electricity Price Forecasting
ARMA	Auto Regressive Moving Average
ARMAX	Auto Regressive Moving Average with Exogenous variable
GARCH	Generalized Auto Regressive Conditional Heteroscedasticity
i.i.d	independent and identically distributed
ME	Mean Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
MPE	Mean Percentage Error
MAPE	Mean Absolute Percentage Error
COR	Correlation
ACF	AutoCorrelation Function
PACF	Partial AutoCorrelation Function
EXAA	Energy Exchange Austria

CHAPTER 1

INTRODUCTION

1.1 Liberalization of Electricity Market

Since 1980, reforms have spread to different facets of the economy in the world, from air transportation and banking to ports, railroads and even food services and communication. The aim of all these efforts is to reduce and replace the government control with the free market. Many people believe that liberalization brings about significant benefits to the consumers; however there is no consensus on that.

By early 1990, most of the major economies had plans to initiate such a reform in the electricity market. Reforms were based on the three principle of the "standard textbook model". The first principle advocates separation of generation, transmission, distribution and marketing functions of electricity. That is, these activities do not have a monopolistic nature and instead of the government, can be performed by firms on a competitive basis. The second principle states that those firms could be privatized. It was believed that private companies are more efficient and have better management systems. The third principle advocates establishment of authoritative institutions to oversee the operation of market players and protect public interest. On 1982, Chile first applied the model on its electricity market and then some European countries such as the UK, Norway, Sweden and Finland started the reform.

In a liberalized market electricity prices is determined by contracts on regular markets. In this market, there is no possibility for arbitrage. The fluctuation of supply depends on the demand. That is, when the demand increases or decrease, supply will also increase or decrease.

Electricity as a commodity has a unique feature. It is inelastic, that is, it cannot be stored. However, demand shows considerable variability and great weather and business cycle dependence. Some accidents such as power plant outages and transmission grid unreliability add to the complexity and reduce the predictability. Therefore, the resulting spot prices exhibit strong seasonality at the annual, weekly and daily levels, as well as mean reversion, very high volatility and abrupt, short-lived and generally unanticipated extreme price changes known as spikes or jumps. These characteristics make modeling and forecasting electricity price very challenging and academically interesting.

The literature on electricity price forecasting have different aims and use different methodologies based on the temporal horizon of the study. For the profitability analysis and power planning, long run horizons are studied, whereas, to get a forecast distribution for the price, medium run is studies are carried out. The evaluation of derivatives is based on the spot prices determined by the market.

There are a number of classifications of the methods for electricity price modeling. Weron (2014) classified these methods into five categories as follows:

- Multi-agent models: These models, simulate the system and its players (companies and agents) and built the price process through matching supply and demand.
- Fundamental models: These models explain the price behavior by modeling the effect of economical and physical variables on the price.
- Reduced-form models: These models describe the statistical properties of electricity prices over time, in order to evaluate the derivatives and for risk management.
- Statistical models: These models are used in load forecasting or implementations of econometric models in the power market.
- Computational intelligence models: These models combine elements of

learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems.

Many studies in the literature use hybrid methods for modeling and price forecasting by combining techniques from two or more of the groups listed above. Figure 1.1 from [118] illustrates these models with their sub-branches.

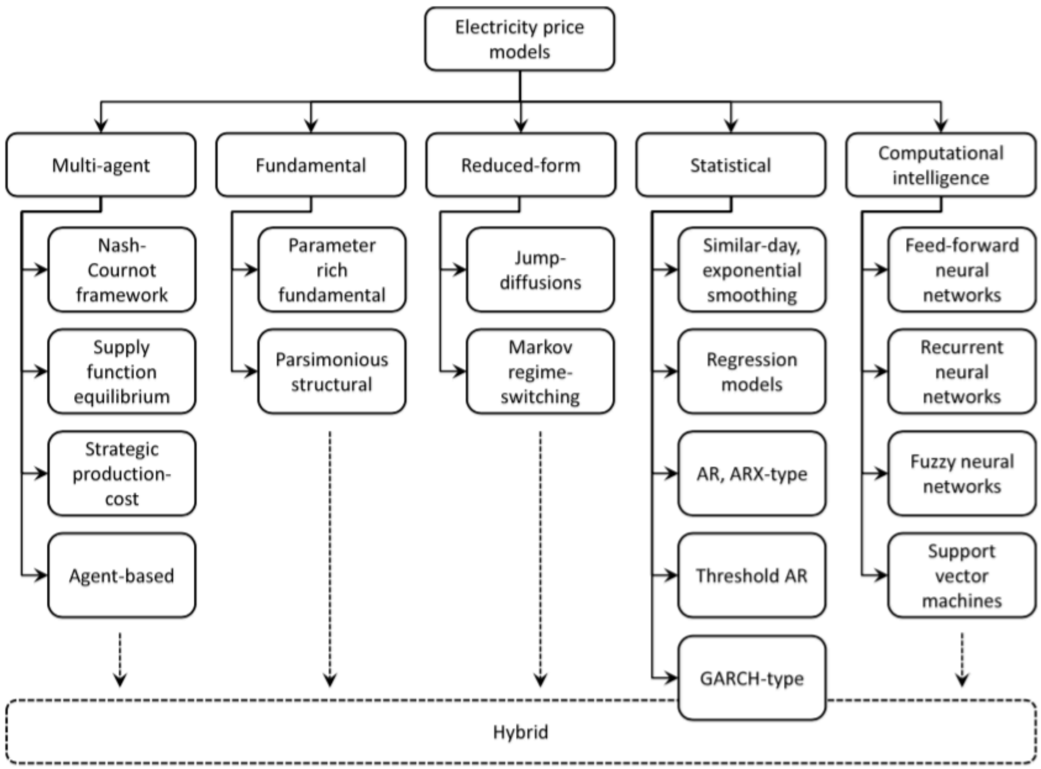

Figure 1.1: *Taxonomy of electricity price modeling approaches.*

It is almost globally accepted that privatization will increase efficiency and decrease the cost of electricity for end users. Although in the only published study, [51] showed that the privatization of electricity distribution companies in Turkey has not yielded the expected results within the first four years of implementation. However, it does not invalidate the liberalization and its benefits for consumers. As the process is going on, the trading companies involved are increasingly in need of accurate price forecasts.

In Turkey, there have not been extensive studies on electricity price. Most of the previous studies were focused on modeling the demand, and there is a lot of room for improvements. Therefore, to further explore the topic, we decided to study electricity prices in Turkey using more advanced methods.

For this purpose, we used ARX and GARCH statistical methods to model electricity prices. We used 2-year hourly historical data from 2012 to 2014 for electricity spot price and demand. We used an autoregressive model with demand as the exogenous variable, and dummy variables for weekends to handle seasonality.

Before getting into the the details of our modeling, we review the current state of Turkish Electricity Market.

1.2 Turkish Electricity Market

After World War I, Turkey started rebuilding its damaged country. Since the financial resources were scarce, it was not possible to invest in the electricity sector and most of electricity was supplied by the foreign private companies such as German, Italian, Hungarian, and Belgian companies. On 1963, the Ministry of Energy and Natural Resources (MENR) of Turkey was founded and the Turkish Electricity Authority (TEK) was established on 1970. TEK took charge of all the electricity activities and responsibilities except the distribution in Turkey. Since 1980, when an export oriented strategy was adopted, various restructuring models were introduced to attract the investors. On 1984, the monopoly power of TEK was removed, so that other private entities can invest and engage in all the electricity activities, including generation, transmission, and distribution.

Since 1984, several financial models were tested in order to attract investment in electricity sector, but they all failed until 2001, when Electricity Market Law (EML) was ratified. The aim of the law was to establish a transparent and competitive electricity market which financially benefits the end users. On 1993 TEK was split into two separate entities, Turkish Electricity Generation Transmission Company (TEAS) which was in charge of generation and transmission activities, and the Turkish Electricity Distribution Company (TEDAS) responsible for distribution and retail sale activities. As a consequence of the

EML, TEDAS was further split into three entities, namely; EUAS, TEIAS, and TETAS which were responsible for generation, transmission and trading activities respectively. Energy Market Regulatory Authority was also established to supervise the activities of market participant as a consequence of EML.

The reform brought significant changes to the monopolistic electricity system of Turkey. The new market is based on bilateral contracts and a complementary residual balancing mechanism. All the electricity generated should be traded either directly or indirectly in the market. TEIAS is responsible for the balancing and settlement market and all the electricity generators above 20MW should submit bids and offers to TEIAS. The balancing mechanism has two phases; the first phase is Day-Ahead scheduling which is done by TEIAS who also sets the hourly prices for the next day. The second phase is the within the day and real-time bid and offer acceptances by TEIAS to meet the fluctuations in supply and demand. When there is a bid, TEIAS settles the trade by using system marginal prices, offer/bid prices and system imbalance price. System imbalance price is the weighted average of hourly system marginal prices within the particular settlement period. There are three settlement periods, namely day, peak and night periods. For each period calculations are carried out monthly and market participants are charged based on their trading and imbalance positions. [9].

Currently, electricity is either traded by bilateral contracts or in the balancing and settlement market. Generators can be categorized into five groups as illustrated in Figure 1.2 from [19]. The first group is EUAS, the biggest state owned electricity generator, and it subsidiaries, affiliates, partnerships, and portfolio generation groups. On aggregate, this group produces 44.4% of total electricity in Turkish market as of 26 June 2012. The second group includes private generator who have build-operate (BO), build-operate-transfer (BOT), or transfer of operational rights (TOOR) contracts with government. This group which do not compete in the market, generate 17% of total electricity in Turkey and sell it to TETAS. The third group is the independent power generators with 32.8% share of the total. The forth group includes autoproducers or self-generators with a 5.8% share of the total. The last group includes small, unlicensed renewable or micro-cogeneration generators who are able to trade in the market through distribution companies. However, it is expected that these small companies soon become part of the Turkish electricity industry [19].

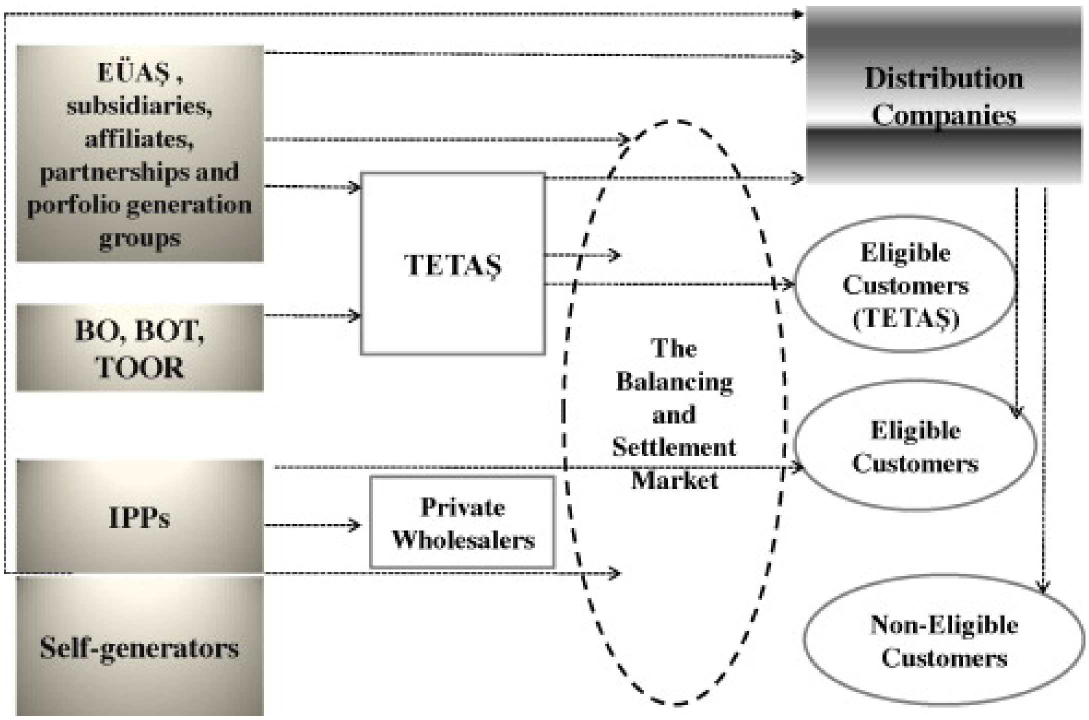

Figure 1.2: *Breakdown of Turkish Electricity Market.*

For a detailed review of the electricity market in Turkey and assessment of its current state, one can refer to [19], [18], [24], [4], [9], and [48].

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Weron in a recent publication [118] reviewed most of major literature on electricity price forecasting (EPF) from 1989 to 2013. He performed both a bibliometrics analysis and a critical review of the publications. Therefore, for the sake of simplicity and to avoid redundancy, we will review the highlights of his work and focus on documents published after 2013 till October 2015.

2.2 Bibliometrics Analysis

A bibliometrics analysis is a quantitative analysis of the literature. However, a critical review is a qualitative analysis. For the bibliometrics analysis, we used Scopus database that is a very popular and well-structured database. Weron used WoS and Scopus databases, but as WoS is a subset of Scopus, and Scopus has a much more user-friendly interface, we limit our review to publications only found by Scopus.

Figure 2.1 from [118] indicates the number of articles and conference papers published from 1989 till 2013. As there were few publications before 2000, the cumulative sum is used for illustrations. The number of publications was increasing until 2009/2010 but decreased as the conference papers declined.

Most of the articles, as classified by Scopus, were categorized in decreasing order

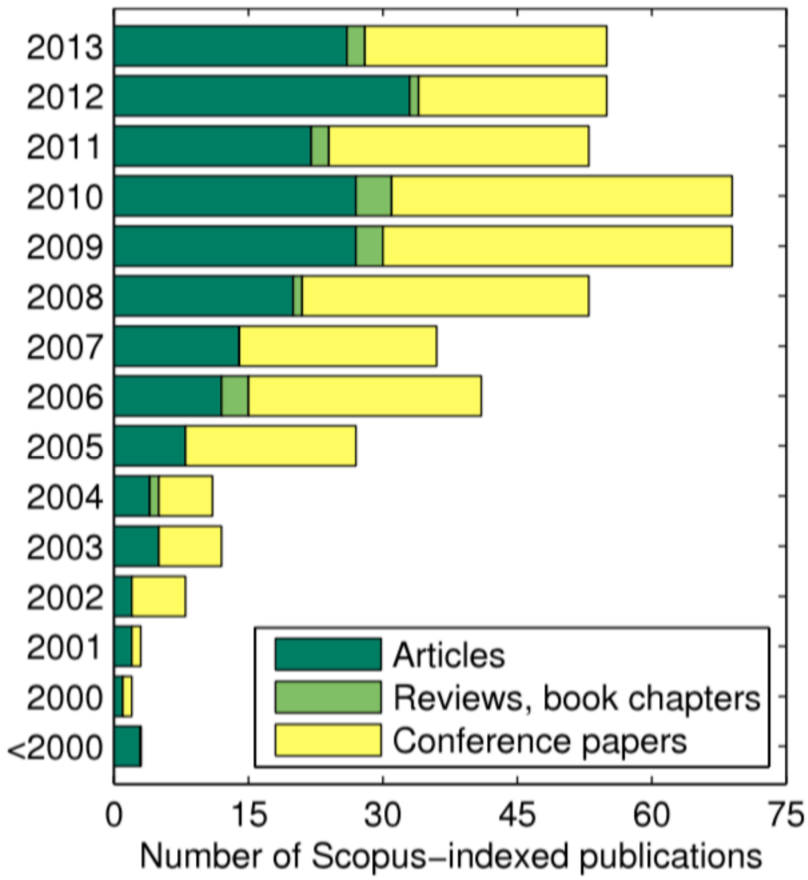

Figure 2.1: *Number of publications by year.*

as engineering or energy, computer science, mathematics, business, management & accounting and economics, econometrics & finance. As we can see in Figure 2.2 from [118], of the ten most popular journals, IEEE Transactions on Power Systems is the most popular one, which has almost equal number of publication on EPF using Neural Networks, Time Series, and other methods.

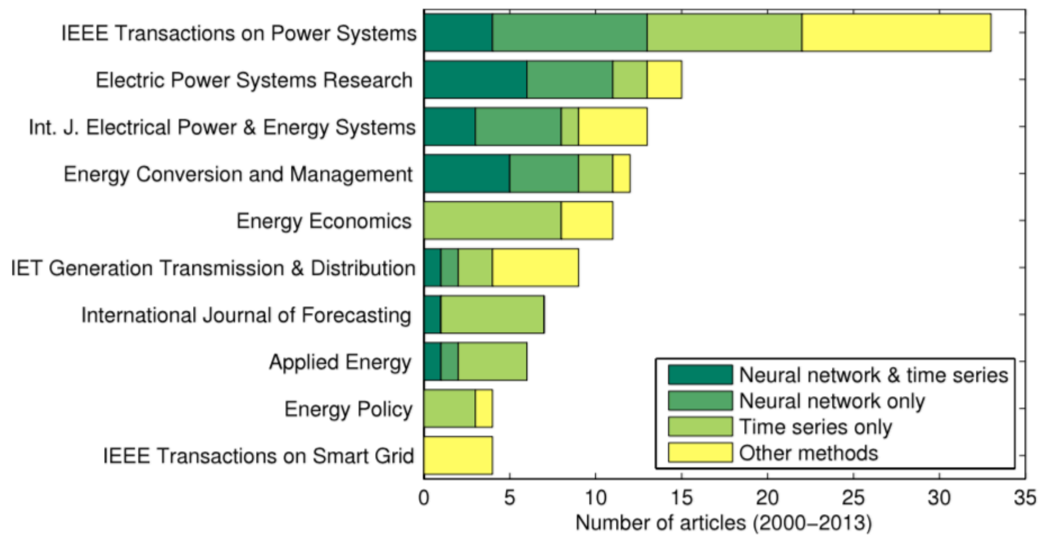

Figure 2.2: *Classification by subjects in major journals.*

Our bibliometric review reveals that there have been 64 and 44 publications on EPF in 2014 and 2015 respectively. We used the built-in keyword TITLE-ABS-KEY to query Scopus database and limited the results by publication year. This query will search for “electricity price forecasting” in the titles, abstracts and keywords of all publications issued since January 2014 to November 2015. Then, the results were refined by inspecting each entry and filtering irrelevant ones.

Among all the publication released in this period, 62% were journal articles, 27.8% were conference papers, 7.4% were articles in the press and 2.8% were reviews. Moreover, based on Scopus, most of the publications were categorized in energy, engineering and computer science fields. Not surprisingly, the Journal of Energy Economics had the most number of papers, followed by International Journal of Electrical Power and Energy Systems and the Journal of Energy Conversion and Management. See Figure 2.3 and 2.4.

Furthermore, China and India, with 17 and 15 publications respectively, had the

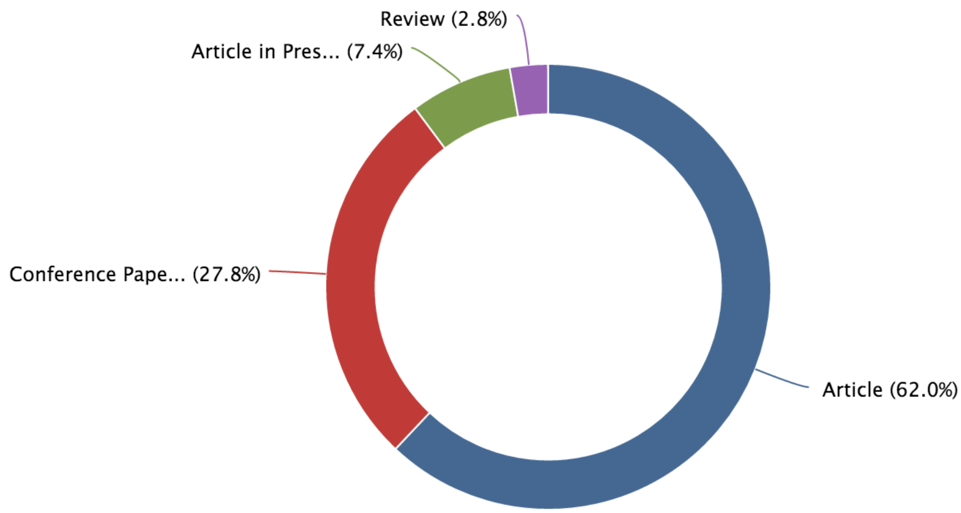

Figure 2.3: *Type of publications since 2014.*

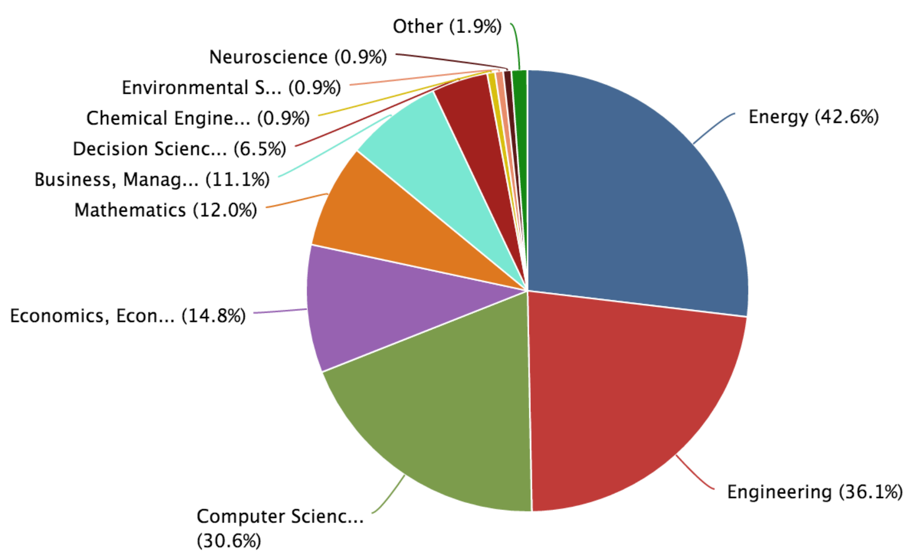

Figure 2.4: *Classification of publications by sector.*

most contribution to the field. Meanwhile, Turkey with four publications is the 12th in the ranking, see figure 2.5. Finally, Weron with seven publications and the most number of citations was the most active researcher in the field.

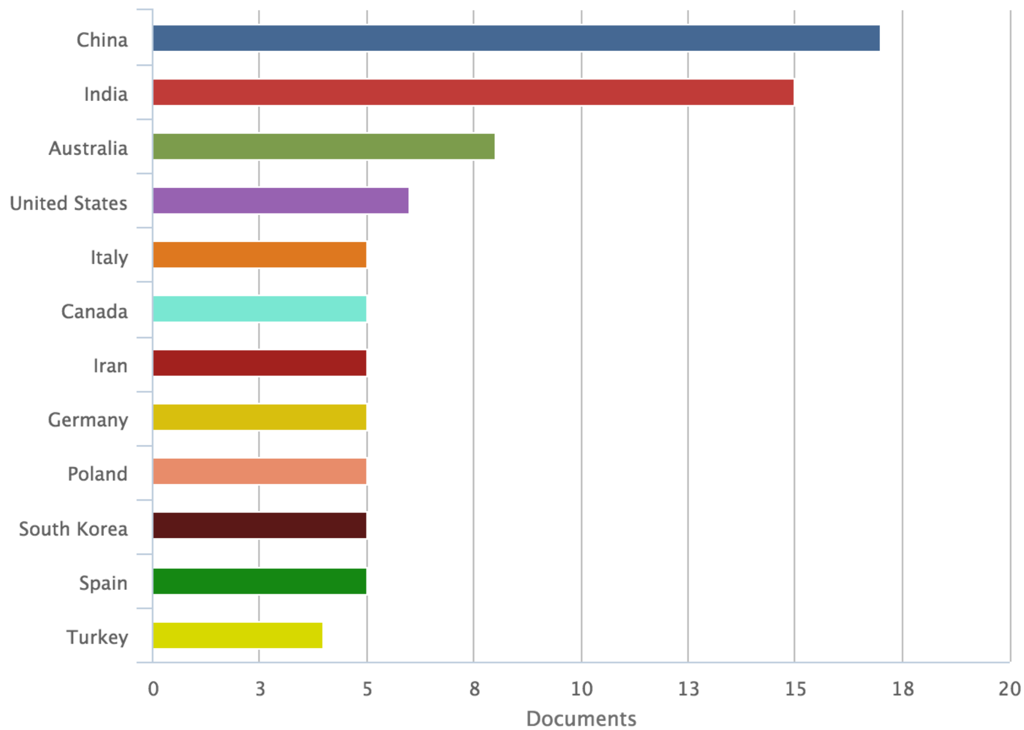

Figure 2.5: *Number of publications by country.*

2.3 Critical Analysis

2.3.1 Studies on EPF in the World

The first major article on EPF appears to be [16]. The author reviewed some of the main issues and techniques related to forecasting of daily loads and prices in competitive markets. He concluded that "forecasting of loads and prices are mutually intertwined activities and that game theory and the economic perspective cannot be an accurate basis for daily forecast". He suggests methods that use separate models for each load period (variable segmentation), neural network techniques for modeling nonlinear behavior, and forecast combinations.

In a discussion article [7] author explains the need for short-term price forecasts, reviews problems related to EPF, and puts forward proposals to such predictions. They argue that "time series techniques (AR, ARIMA, GARCH) are usually successful only in the areas where the frequency of the data is low, such as weekly patterns". They advocate the use of artificial intelligence and hybrid approaches that are "capable of tracking the hard nonlinear behaviors of hourly load and especially price signals".

One of the best studies on EPF may be considered to be [119]. In this study, twelve time series models for forecasting short-term spot price in auction type electricity market were compared. They used, AR model with its extensions, spike preprocessed, threshold model and semiparametric autoregressions as well as mean reverting jump diffusion models. They used the hourly spot price and system-wide loads for California and a series of hourly spot prices and air temperature for Nordic market. They found out that models which use system load as the exogenous variable usually have better performance compared to pure price models. However, when temperature is used as the exogenous variable, price model does not outperform the pure price model. They also found out that in general, semiparametric models have better point and interval forecasts than other model. Specifically, these model can perform well under diverse market conditions.

A survey article [2] reviews 47 time series and neural network papers published between 1997 and 2006. It concludes that "there is no systematic evidence of out-performance of one model over the other models on a consistent basis, which may be attributed to the substantial differences in price developments in different power markets". In a more recent article [3], the same authors classify EPF models as three categories, namely heuristics, simulations, and statistical model. The latter includes time series and artificial intelligence models.

A recent survey article [28] reviews neural networks, support vector machines, three class of time series models namely ARMA, ARMAX, and GARCH, and functional principal component analysis (FPCA) models for EPF. The authors suggest using multivariate factor models and especially robust FPCA, which is proved to have a better performance than both the standard FPCA and an AR model whose mean value varies by time in a limited forecasting study.

In one of the chapters of Wiley Encyclopedia of Electrical and Electronics Engineering, Martos and Conejo review time series model for short-term and medium-term electricity price forecasting. For day-ahead predictions, they focus on ARIMA and seasonal ARIMA models, and for medium-term horizons they use vector ARIMA and unobserved component models.

In [20], authors examine the structural approach for electricity modeling, and emphasize its advantages compared to traditional reduced-form models. They study several recent articles and recommend a structural framework for spot prices, which encompasses demand, capacity and fuel prices.

Authors of [43] suggest the use of Realized GARCH models for estimating the volatility of daily price in the EPEX power markets. They argue that "the model specifications extract the volatility-related information from realized measures, which improves the in-sample fit of the data". Furthermore, "the evidence on the out-of-sample predictability reinforces the value of the specifications as the forecast quality is improved over the benchmark EGARCH model under eight conventional criteria".

In a comparative study [31], authors compared the prediction performances of three models; a discrete-time univariate econometric model (ARMA-GARCH) and two computational intelligence models, namely Neural Networks and Support Vector Machines. They found out that the Support Vector Machine methodology gives a better forecasting accuracy for price time series, closely followed by the econometric technique.

In a study on Italian Power Exchange market [23], authors propose an econometric model for short-term forecasting of the daily single national price of electricity. They use constants, regressors, moving averages, weekly and seasonal dummies, autoregressive and heteroskedastic variables. Their results show a significant decrease in error of short-term forecasts in comparison with linear least squares method which was conventionally used in the literature.

Authors in [34] studied the use of univariate time series models for modeling electricity prices in Leipzig Power Exchange. The found out that when each hour of the day is modeled separately, the performance of the model forecast is better compared to models for the whole time series data. Also they found out that, modeling spikes using a simple probabilistic process leads to better forecasting abilities of the models.

In [75], authors apply a method similar to [119] on the Nord Pool hourly day-ahead price. They used Nordic demand and Danish wind power as the exogenous variables. Also, they modeled the price across all hours in their analysis period rather than a single hour of a day.

In [83] authors set up different time series and assessed their short-term forecasting power in the electricity spot market. They used AR, ARX, ARX-GARCH, TARX, and Markov regime-switching models to model electricity spot price in the California Power Exchange. They also found the point and interval forecast for their model. They concluded that (i) nonlinear models outperform the linear model and (ii) additional GARCH component generally decreases the efficiency of point forecasting.

Authors of [15] studied the use of ARIMA, ARIMA-EGARCH and ARIMA-EGARCH-M models for modeling hourly electricity prices in Midwest Independent System Operator (MISO). They concluded that "no model outperform the others in terms of in-sample forecasting performance. However, ARIMA-EGARCH-M model outperforms the other models in terms of out-of-sample forecasting performance".

In [10], authors compared an ARMAX model with Gradient Boosting Regression which is a new technique. They showed that a multi-model approach has better performance in terms of error metrics. They also argued that "Gradient Boosting can deal with seasonality and autocorrelation out-of-the-box and achieve lower rate of normalized mean absolute error on real-world data".

In [46], authors propose another approach for modeling electricity.Their model combined several univariate and multivariate time series methods which represent the energy produced with clean energies, such as wind and hydro. They finally argued that their model is the optimal model compared to other studies.

In a similar study [130], authors introduce an econometric model for hourly electricity prices of the European Power Exchange and incorporate features such as renewable energies. They call it a VAR-TARCH model with wind power, solar power and load as influences on time series. They used an efficient iteratively reweighed lasso approach for estimation and claim that their model outperformed several existing models.

2.3.2 Studies on EPF in Turkey

Although, the literature on EPF is very diverse and active, there are few studies on this topic in Turkey. So far, most of the studies on Turkish market has been focused on modeling the demand. Furthermore, they mostly used Neural Networks for modeling and statistical methods are not studied in depth. The following paragraphs explain some of studies on EPF in Turkey.

In [109], authors modeled electricity price in Turkish electricity market using ARIMA model and feed forward neural network model. They compared the performance of both models and concluded that it is possible to forecast weekly electricity price with and average error rate of 8.5%. In another study [12], the same authors examined the effect of historical prices and loads, calendar data, weather conditions and currencies on short-term EPF in Turkish electricity market. They tested the combinations of feature subsets on the feed forward neural network forecast model, and find out that "the best feature subset combination is calendar data, historical prices and load prediction".

In another study [70], authors studied the use of artificial neural networks and proper artificial neural network configurations for price modeling. They examine various sets of parameters and network topologies to find the best suitable configuration. They finally compare their model with a time series model.

The overall analysis of the publications reveals that the literature on EPF is almost saturated by statistical modeling techniques, although to the best of our knowledge, there is not enough studies on Turkish electricity market. Moreover, there is an increasing interest in using artificial intelligence methods (Neural Networks, Support Vector Machines, Wavelet Transforms, Machine Learning Techniques, etc.) and hybrid methods. Considering the importance of the accurate modeling and forecasting for market participants and to extent the studies on EPF in Turkey, we decided to study application of statistical methods for EPF in Turkish electricity market.

CHAPTER 3

METHODOLOGY

3.1 Introduction

Electricity is an inelastic commodity, that is, it cannot be stored. This unique feature makes price series have some unique characteristics. These series exhibit daily, weekly and annual seasonality. They also have extreme values, called spikes. These characteristics make modeling such series rather difficult.

The statistical models to forecast the current electricity price usually use a combination of previous price together with the current and/or previous values of some exogenous factors. These factors are usually consumption and productions or weather variables. There have been numerous studies on the most important factors. But, it is not possible to generate a universal formula for all markets, since each market has its own characteristics and behaves differently. There are two general categories for these models, namely the additive and multiplicative models. The additive model use summation of the variables to generate forecasts, while in the multiplicative model, the final result is obtained by multiplication of variables.

Statistical models are attractive, because they have some physical interpretation so that the end user can easily understand them. However, they also have some limitations in modeling nonlinear behavior of the variables.

3.2 Pre-processing the data

Before starting the modeling, it is usually useful to process the data. The aim of the pre-processing is usually to make the data easier to model. In the case of electricity prices, this process includes finding the outliers and dealing with them using standard techniques. Hawkins (1980) defined outlier as " an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism".

Since identification of the spikes has a significant effect on model estimation, they should carefully be treated before modeling. In the literature spikes are defined as values that surpass a specific threshold for a short period of time. But there is no consensus on either the threshold value or the time period. There are a number of methods introduced in the literature for outlier detection. Authors in [60] studied various techniques for treatment of spikes in the electricity price series and compared their effect on the estimation of model parameters. Some of the methods they used are:

- *Fixed price threshold:* In this method, when the price exceeds some predefined threshold it is considered as a spike.

- *Variable price threshold:* In this method, a certain percentage of the extreme values (highest or lowest) are considered as outliers.

- *Fixed Price Change Threshold* In this method, when the change in the price exceed some specific threshold, the value is considered as spike.

- *Variable price change threshold:* In this method, when the price exceeds three times the standard deviation, it is considered as spike and is removed from the data. This procedure continue recursively until all the spikes are identified and removed from the data.

For a complete review of these methods refer to [60]. There are also a number of R packages to deal with outliers, namely; {outlier}, {extremevalues} and etc.

3.3 Statistical Model for EPF

As explained in the first Chapter, statistical methods to model electricity prices, as discussed by [118], are categorized into 5 types:

- Similar day and exponential smoothing methods,
- Regression models,
- AR-type time series models,
- ARX-type time series models,
- Heteroskedasticity and GARCH-type models.

In the following, each method is briefly explained and some studies on their application are introduced.

3.3.1 Similar day and exponential smoothing methods

In this rather simple and popular method, a forecast is based on the similar days and hours in the past. That is we have to look for the similar days, which have the same characteristics in the historical data and take them as the forecast for future. The similar characteristics include similar day of the week, day of the year, holiday type, and weather or consumption figures. It is also possible to use a linear combination or regression of several similar days instead of a single day.

In [85] authors proposed a naive test to examine the accuracy of the models based on the similarities of the days. They argued that a Monday, Saturday and Sunday are similar to the respective days of the previous weeks. A Tuesday is similar to the previous Monday and the same rule applies for Wednesdays, Thursdays and Fridays. They argued that models that are not well calibrated would fail to pass this test.

Exponential Smoothing is another simple method for forecasting. It is mostly used for load forecasting, however there are some studies using this method for EPF as well [108]. In this method the prediction is constructed from an exponentially weighted average of past observations:

$$\hat{X}_t = S_t = \alpha \cdot X_t + (1-\alpha) \cdot S_{(t-1)} \tag{3.1}$$

In this formula the smoothed value S_t is the weighted average of the previous observations, where the weights decrease exponentially depending on the value of parameter $\alpha \in (0,1)$. It is also possible to add a seasonal and trend component to the formula. [37, 47, 58] can be referred to for the application of this method on EPF.

3.3.2 Regression Models

In regression analysis we study the relationship between variable. It includes many modeling and analyzing techniques which try to estimate the relationship between a dependent variable and some independent variables called regressors. They help us understand the behavior of the dependent variable when only one of the independent variables changes and the others are held fixed. Multiple regression is based on the least squares method. In this method, the best fit to the data is obtained by minimizing the sum-of-squares of the difference between the observed and the predicted values. The relationship between variables is usually assumed to be linear in multiple regression. Whereas, in a nonlinear regression this relationship is nonlinear. A linear regression can be formulated as:

$$P_t = \mathbf{B}_t\mathbf{X}_t + \epsilon_t = b_1 X_t^{(1)} + \cdots + b_k X_t^{(k)} + \epsilon_t, \tag{3.2}$$

where $\mathbf{B}$ is a $1 \times k$ vector of constant coefficients, X_t is the $k \times 1$ vector of regressors, and ϵ_t is the error term. The regressors are chosen from the variables, which are believed to be correlated to the electricity price P_t. In this case, we use Maximum Likelihood method for estimation.

In the case that price driver effects evolve continuously, we have a *time-varying regression (TVR)* as:

$$P_t = \mathbf{B_t}\mathbf{X_t} + \epsilon_t = b_{1,t} X_t^{(1)} + \cdots + b_{k,t} X_t^{(k)} + \epsilon_t, \tag{3.3}$$

where $\mathbf{B_t}$ is now a $1 \times k$ vector of time-varying coefficients. TVR model parameters can be estimated using state space methods and Kalman filter, as discussed by [40].

Despite many other alternatives, regression methods are still one the most popular methods in EPF. However, they are usually combined with other methods to get a better forecast. [33, 63, 66, 71] are some examples for application of this approach in EPF.

3.3.3 AR-type and ARX-type time series models

AR-type time series models

AutoRegressive Moving Average or ARMA(p, q) model is the standard time series model that takes into account the random nature and time correlations of the phenomenon under study. The current value of X_t is expressed by an autoregressive (AR) part that includes p past values of X_t and a moving average (MA) part that consist of q past values of the noise.

$$\phi(B)X_t = \theta(B)\epsilon_t. \tag{3.4}$$

Here B is the backward shift operator, that is, $B^h X_t = X_{t-h}$ and $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, similarly, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ where $\phi_1, \cdots, \phi_p$ and $\theta_1, \cdots, \theta_q$ are coefficients of AR and MA polynomials respectively. Finally, ϵ_t is a white noise (i.i.d noise) with zero mean and finite variance, which is often denoted as WN$(0, \sigma^2)$. If $q = 0$ the model is autoregressive AR(p) and if $p = 0$ it is Moving Average MA(q).

In the ARMA modeling, we assume that the data is weakly stationary, and if it is not, transformation is needed. A simple way to make a non-stationary series to stationary one is differencing. The resulting model is called Autoregressive Integrated Moving Average or ARIMA(p, d, q). This model assumes d times differencing before estimating p and q, and can be written as:

$$\phi(B)\nabla^d X_t = \theta(B)\epsilon_t, \tag{3.5}$$

where $\nabla X_t = (1 - B)X_t$ is a lag-1 differencing operator, and h-lag differencing can be defined as $\nabla_h X_t = (1-B^h)X_t = X_t - X_{t-h}$. Sometimes the simple differencing is not enough to make the series stationary and a differencing with longer lags is required. These models are called Seasonal ARIMA or SARIMA. The general notation for such models is ARIMA$(p, d, q) \times (P, D, Q)_s$. The (p, d, q) and $(P, D, Q)_s$ represent the order of the nonseasonal and seasonal parts respectively, and s is the lag of the seasonal part. The mathematical notation for such a models is:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D X_t = \theta(B)\Theta(B^s)\epsilon_t, \tag{3.6}$$

By taking $\tilde{X}_t = \nabla^d\nabla_s^D X_t$, we can simply convert a SARIMA model to ARMA. Therefore, the estimation process for both ARIMA and SARIMA models are analogous to that of an ARMA process. That is, firstly we have to do model identification to find the order of the model and then model estimation, which is estimating the coefficients using standard methods such as least squares or maximum likelihood estimation. After finding the right ARMA-type model, forecasting can be carried out using Durbin-Levinson algorithm. The details of this procedure is explained in classical time series analysis textbooks.

There are many studies that used ARMA and its variants for EPF. [33, 34, 83] are some examples of such studies.

ARX-type time series models

The ARMA process that was formulated by (3.4) only used the previous values of price and the error to forecast the future. However, there are some external factors that influence electricity price as well. These external factors are called exogenous variables and they are usually load and consumption or weather data.

The models that incorporate these variables are often called time series with exogenous variable. Therefore, ARX, ARMAX, ARIMAX, SARIMAX are generalized versions of AR, ARMA, ARIMA, and SARIMA models respectively. The ARX-type models are called regression models in the literature as well.

The mechanism for incorporating the exogenous variables into ARMA-type model is straightforward. For example, in the ARMAX model, the current value of the spot price X_t is expressed in terms of its past values and the previous value of the noise, together with the present and past values of the exogenous variable(s). The *AutoRegressive Moving Average model with exogenous variables* $V^1, \ldots, V^{(k)}$ or ARMAX$(p, q, r_1, \ldots, r_k)$, is formulated as:

$$\phi(B)X_t = \theta(B)\epsilon_t + \sum_{i=1}^{k} \psi^i(B)V_t^{(i)}, \tag{3.7}$$

where r_i are the orders of the exogenous factors, where $i = 1, \ldots, k$ and $\psi^i(B) = \psi_0^i + \psi_1^i B + \cdots + \psi_{r_i}^i B^{r_i}$ with ψ_j^i as coefficients.

The *transfer function* of the ARMAX model is therefore:

$$X_t = \frac{\theta(B)}{\phi(B)}\epsilon_t + \sum_{i=1}^{k} \tilde{\psi}^i(B)V_t^{(i)} \tag{3.8}$$

We can use the least squares method for estimation of ARX models. In this method the difference of the square of the right-hand side and the left-hand side of the equation (3.7) is minimized. For calibration of the model, we can use *maximum likelihood* technique. In this method, we try to minimize the difference between the model outputs and the observed values by selecting the right model parameters. In addition to these methods, there are other methods what can be used for estimation and calibration.

Time series models with exogenous variables have been extensively used in EPF. For example [33, 75, 83, 85, 119] are some of these studies.

3.3.4 Threshold Autoregressive Models

There are generally two class of regime-switching models. One is the model that the regime is determined by an observable variable, and the other is the model that the regime is determined by an unobservable variable. In the former, the regimes have already occurred in the past, however, in the latter, we can never be sure whether the regime change has occurred or not.

Threshold Autoregressive (TAR) models are the most important models of the first class. They were first introduced by Tong and Lim [52] and assume specific values for the observable variable v_t relative to a threshold value T. The formulation of these models is as follows:

$$\phi_1(B)X_tI_{(v_t \leq T)} + \phi_2(B)X_tI_{(v_t > T)} = \epsilon_t, \tag{3.9}$$

where
$\phi_i(B) = 1 - \phi_{i,1}B - \cdots - \phi_{i,p}B^p$,
$i = 1, 2$,
B is the backward shift operator,
$I_{(.)}$ denotes the indicator function, and
X_t is the spot electricity price.

It is possible to have more than two regimes, the formulation of such a model is as follows:

$$\begin{aligned} &\phi_1(B)X_tI_{(v_t \leq T_1)} + \phi_2(B)X_tI_{(T_1 < v_t \leq T_2)} + \cdots \\ &+ \phi_n(B)X_tI_{(T_{n-1} < v_t \leq T_n)} + \phi_{n+1}(B)X_tI_{(T_n < v_t)} = \epsilon_t, \end{aligned} \tag{3.10}$$

and
$i = 1, 2, \ldots, n+1$,
Furthermore, it is possible to add exogenous variable(s) to the TAR model to make a TARX model.

When the threshold variable is taken as the lagged value of the price, the resulting model is called *Self Exciting TAR* (SETAR) model. This model can be modified to allow gradual transition between regimes, and resulting in a *Smooth Transition AR*(STAR) model. Logistic function as a popular choice for the transition function is expressed as:

$$G(X_{t-d}; \gamma, T) = [1 + exp\{-\gamma(X_{t-d} - T)\}]^{-1}, \tag{3.11}$$

where d is the lag and γ determines the smoothness of the transition. This model is known as the *Logistic STAR*(LSTAR) model.

For a review of studies on EPF using TAR models, one can refer to [30, 55, 83, 94, 95].

3.3.5 Heteroskedasticity and GARCH-type models

In conventional econometric models, the variance of the disturbance term is assumed to be constant. However, in many cases, such as electricity prices, the time series exhibit periods of unusually large volatility, followed by period of relative tranquility. In such circumstances, the assumption of a constant variance (homoskedasticity) is inappropriate.

The *Autoregressive Conditional Heteroskedastic* (ARCH) model, introduced by Engle (1982), was the first model to address this issue. In this model, the conditional variance of the time series $\{X_t\}$ is represented by an autoregressive process, which is a weighted sum of squared preceding observations:

$$\hat{\epsilon_t}^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \alpha_2 \hat{\epsilon}_{t-2}^2 + \cdots + \alpha_q \hat{\epsilon}_{t-q}^2 + v_t, \tag{3.12}$$

where v_t is a white-noise process.

There are many possible application of ARCH models since the residuals in (3.12) can come from an autoregression, an ARMA model, or a standard regression model.

Since the conditional variance are best estimated simultaneously using maximum likelihood techniques, and it is better to specify v_t as a multiplicative disturbance, we can reformulate (3.12) to:

$$\epsilon_t = v_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \tag{3.13}$$

where v_t is a white-noise process such that $\sigma_v^2 = 1$, v_t and ϵ_{t-1} are independent of each other, and α_0 and α_1 are constants such that $\alpha_0 > 0$ and $0 \leq \alpha_1 \leq 1$.

Bollerslev (1986) extended Engle's original work by developing a technique that allows the conditional variance to be an ARMA process. If we let the process be such that:

$$\epsilon_t = v_t \sqrt{h_t},$$

where $\sigma_v^2 = 1$,

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}^2, \tag{3.14}$$

Since $\{v_t\}$ is a white-noise process, the conditional and unconditional means of ϵ_t are equal to zero. The conditional variance of ϵ_t is the ARMA process given by the expression h_t in (3.14).

The **generalized** ARCH(p, q) model, called **GARCH**(p, q), allows for both autoregressive and moving average components in the heteroskedastic variance. The benefit of a GARCH model is that, an ARCH model may have a more parsimonious GARCH(p, q) representation that is much easier to identify and estimate. This is particularly true since all the coefficients in the (3.14) must be positive. Moreover, to ensure that the variance is finite, all roots of the characteristic functions in (3.14) must lie inside the unit circle. Clearly, the more parsimonious model will entail fewer coefficient restrictions.

The key feature of GARCH models is that the conditional variance of the *disturbances* of the $\{X_t\}$ sequence constitutes an ARMA process. Hence, it is expected that the *squared residuals* from a fitted ARMA model should display this characteristic pattern.

That is, if there is conditional heteroskedasticity, the correlogram of the squared residuals should be suggestive of such a process. The algorithm to construct the correlogram of the squared residuals is as follows:

Step 1: Estimate the $\{Y_t\}$ sequencing using the "best fitting" ARMA model (or regression model) and obtain the squares of the fitted errors $\{\epsilon_t^2\}$. Also calculate the sample variance of the residuals $\hat{\sigma}^2$ defined as

$$\hat{\sigma}^2 = \sum_{t=1}^{T} \hat{\epsilon}_t^2 / T,$$

where T = number of residuals

Step 2: Calculate and plot the sample autocorrelations of the squared residuals as:

$$\rho_i = \frac{\sum_{t=i+1}^{T} (\hat{\epsilon}_t^2 - \hat{\sigma}^2)(\hat{\epsilon}_{t-i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^{T} (\hat{\epsilon}_t^2 - \hat{\sigma}^2)},$$

Step 3: In large samples, the standard deviation of ρ_i can be approximated by

$T^{-0.5}$. Individual values of ρ_i that are significantly different from zero are indicative of GARCH errors. Ljung-Box Q-Statistics can be used to test for groups of significant coefficients. The statistic

$$Q = T(T+2)\sum_{i=1}^{n} \rho_i^2/(T-i),$$

has an asymptotic χ^2 distribution with n degrees of freedom if the $\{\epsilon_t^2\}$ sequence is serially uncorrelated. Rejecting the null hypothesis that the $\{\epsilon_t^2\}$ are serially uncorrelated is equivalent to rejecting the null hypothesis of no ARCH or GARCH errors. In practice we consider values of n up to $T/4$.

In fact, the identification and estimation of GARCH models are analogous to that of (S)AR(IMA) models; Maximum Likelihood (ML) is the preferred algorithm. By itself, the GARCH model is not attractive for short-term EPF; however, when they are coupled with an AR-type model, it presents an interesting alternative; the (S)AR(IMA)-GARCH model, where the residuals of the regression part are modeled further with a GARCH process.

Although electricity prices exhibit heteroskedasticity, the GARCH models are not always as successful as expected. Some studies on application of such models in EPF include [38, 64, 69].

3.4 Model Selection Criteria

There is not always a single model that can describe the relative behavior of the data. So, we need to choose the best model. As George Box said, “All models are wrong, but some are useful”. With that in mind, we try to find the model that best approximates the reality. That is the model that minimizes the loss of information. Kullback and Leibler (1951) developed a measure called Kullback-Leibler information to address this issue. On 1973 Akaike recommended to use Kulback and Leibler information for model selection. He established an information criteria based on maximum likelihood to estimate the Kulback-Liebler information. His measure called the Akaike information criteria is defined as:

$$AIC = -2\log(L) + 2K, \tag{3.15}$$

Where K is the number of estimated parameters in the model, and L is the likelihood of the model. Schwarz (1978) derived the Bayesian information criteria as:

$$BIC = -2\log(L) + K\log(n), \tag{3.16}$$

where n is the number of observations or the sample size, K is the number of parameter to estimate. As AIC and BIC measure the loss of information in the model, for selecting the best model, one should find the value of AIC or BIC for all the models and select the model with the minimum value of AIC or BIC respectively.

In this study, we only use the AIC and BIC for selecting the best model. However, there are other criteria and methods which can be used for model selection. For a detailed explanation of the model selection process and the other criteria and methods refer to [17] and [131].

3.5 Model Performance Measures

When we find the best model, we can forecast the future. If we denote y_t as the observation at time t and f_t as the forecast value at time t, then $e_t = f_t - y_t$ is the forecast error at time t. Also, we define $\text{Mean}(y_t) = \frac{1}{n}\sum_{i=1}^{n} y_t$, for simplicity of the formulas.

There are a number of accuracy measures to assess the forecast. Some are scaled dependent measures. These measures are useful when comparing different models applied to the same data set. Some examples of such measures are:

- Mean Squared Error (MSE) = $\text{Mean}(e_t^2)$
- Root Mean Squared Error (RMSE) = $\sqrt{\text{MSE}}$
- Mean Absolute Error (MAE) = $\text{Mean}(|e_t|)$,

Some other measures are based on percentage errors. The percentage error is calculated as $p_t = 100e_t/y_t$. The advantage of these measures is that they are independent of the scale. So they can be used across different data sets. Some examples of these measures are:

- Mean Absolute Percentage Error (MPE) = Mean(p_t),
- Mean Absolute Percentage Error (MAPE) = Mean($|p_t|$),
- Root Mean Square Percentage Error (RMSPE) = $\sqrt{\text{Mean}({p_t}^2)}$,

There are other types of forecast accuracy measures as well. For a more complete explanation and discussion on these measures one can refer to [57].

CHAPTER 4

ANALYSIS

4.1 Data Description

In this study, we used the hourly electricity demand (in MWh) and price (in TL/MWh) of Turkey with permission from the authorities at Enerji Piyasaları İşletme A.Ş.(EPİAŞ)[1]. We used the data from 1st January 2012 to 11 June 2014. This time period is equivalent to 21,432 hours. We divided the data into a training set and a test set. The training set constitutes the data from 1 January 2012 to 1 January 2014, which is about 81.86% of the total data and the remaining part was used the test set. We built our models in the training set and then tested them on the test set. Tables 4.1 provides some descriptive statistics for the whole price data.

Table4.1: *Summary Statistics for Price (TL/MWh)*

Min	Max	Range	Median	Mean	Var	Std.Dev
0	2,000	2,000	152	151.048	2,433.101	49.32

There are some interesting facts about the data. Firstly, the minimum price of electricity is zero, which suggest that at some point, it was traded for free. This unusual phenomenon is another characteristic of electricity markets. Since electricity can not be stored, when the supply exceeds the demand, it should be sold in the market for free. Secondly, the variance of the price is very high compared to its mean and the median. This is another characteristic of electricity prices which is called high volatility.

[1] https://www.epias.com.tr/index.php

4.2 Exploratory Data Analysis

Figure 4.1 shows the hourly plot of price from 1 January 2012 to 11 June 2014. As we can see there are some very sharp spikes. The huge spike in the beginning of the data is the price on 13 February 2012, when the price reached the maximum of 2,000 TL/MWh at noon. We could not find any explanation for this incident, but it was most probably due to an unexpected cutoff in electricity supply at that time.

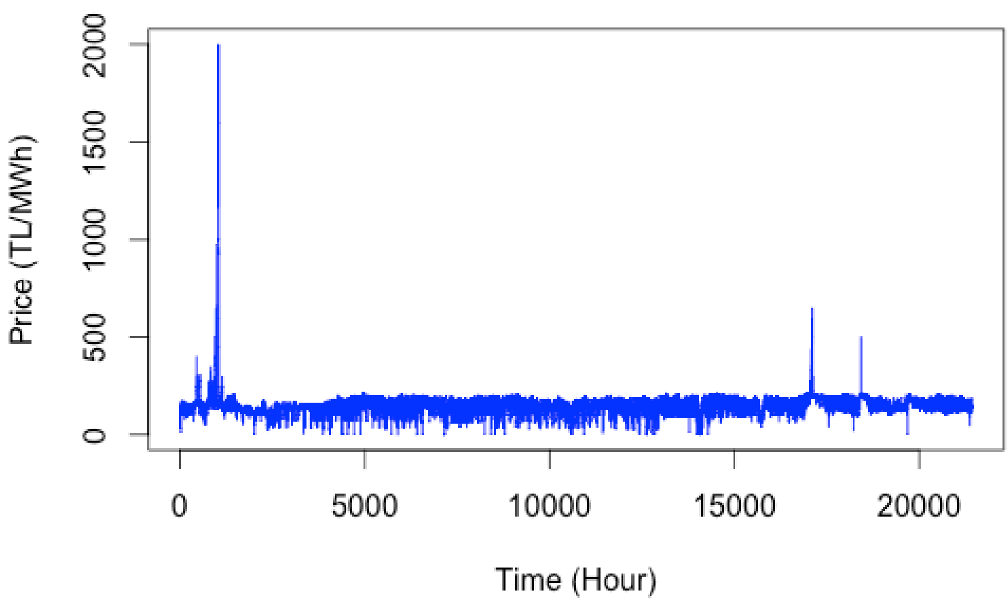

Figure 4.1: *Hourly Price (1 Jan 2012 - 11 Jun 2014).*

The spikes in the data, bring some difficulties in modeling. Some researchers simply filter out the spikes and use the rest of the data for modeling. Some others pre-process the data so that the outliers or the spike would not be so sharp. It is also possible to use different models for the spiky region and the normal region (regime switching models). For an study on application of various techniques for pre-processing the spikes in electricity prices, one can refer to [60]. In our study, we first used the original data (without any pre-processing) and built the model based on that data. Then we processed the data and applied the model on the new dataset and compared the forecast accuracy of the two approaches.

4.3 Pre-processing spikes

In order to detect the outliers in the price, we took the 99.7% confidence interval for the price as the acceptable limit and assumed the values outside this limit as outlier. If we assume the mean value of price as M and the standard deviation of the price as S, then the 99.7% confidence interval is equivalent to $M \pm 3 \cdot S$. In practice, starting with the first observation in the data, we check whether the price is out of the confidence interval or not. If so, it is considered as an outlier and subsequently replaced with the price value at the boundary of the interval. Note that, the standard deviation should be recalculated each time an outlier is detected. Since after replacement of outliers with smaller values, the standard deviation will decrease.

Figure 4.2 illustrates outliers in the price. There are some spikes located in the middle of the price series, other than those huge ones previously shown in Figure 4.1.

After replacing the outliers with corresponding boundary value according to the procedure outlined above, we get the spike processed price series which is plotted in Figure 4.3. It may look that the price series has become more spiky after the outlier treatment. However, this is not true, as in Figure 4.1 the spikes were only invisible due to the range of Y-axis. The summary statistics for the price after spike treatment are given in Table 4.2.

Table4.2: *Summary statistics for spike processed price (TL/MWh)*

Min	**Max**	**Range**	**Median**	**Mean**	**Var**	**Std.Dev**
10	299	289	152	150.3	1436.66	37.90

The exogenous variables for modeling price is different for each country. Some countries use temperature, some others use demand and a few countries use other economic measures for this purpose. According to a study that was done on

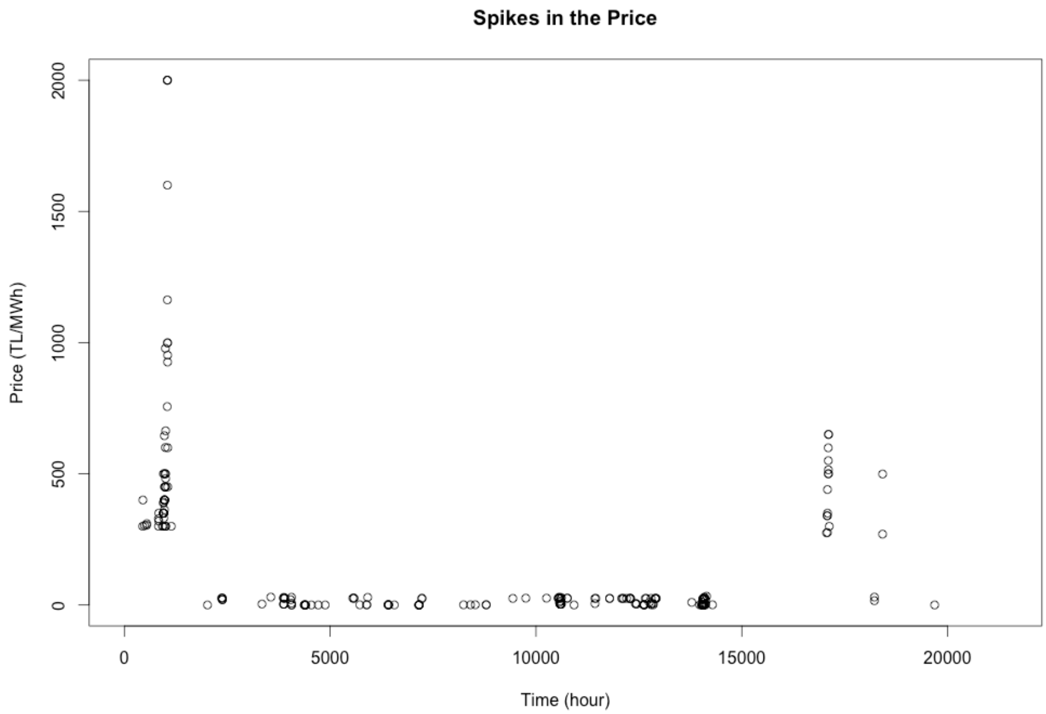

Figure 4.2: *Spikes in the price series*

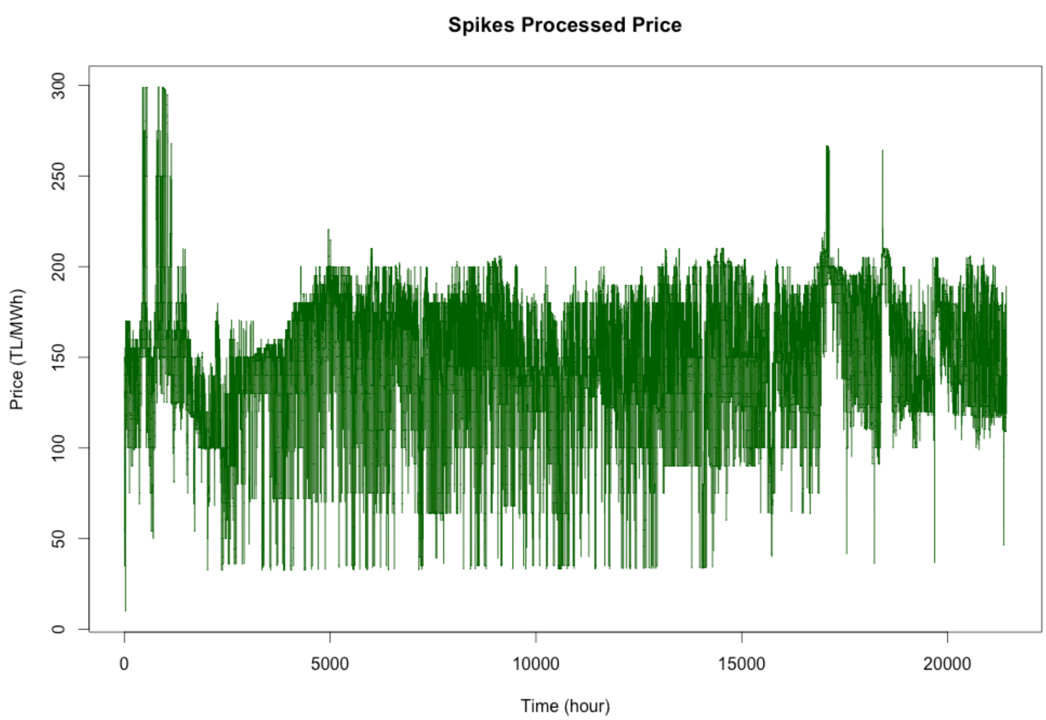

Figure 4.3: *Spike processed price plot*

2012 by Kamil Demirberk in his Masters thesis, the relation between electricity price and temperature is not significant and therefore temperature should not be considered as a regressor in modeling of electricity prices in Turkey. As we will show in the following section, demand has a significant effect on electricity prices in Turkey and should be considered as exogenous variable in regression.

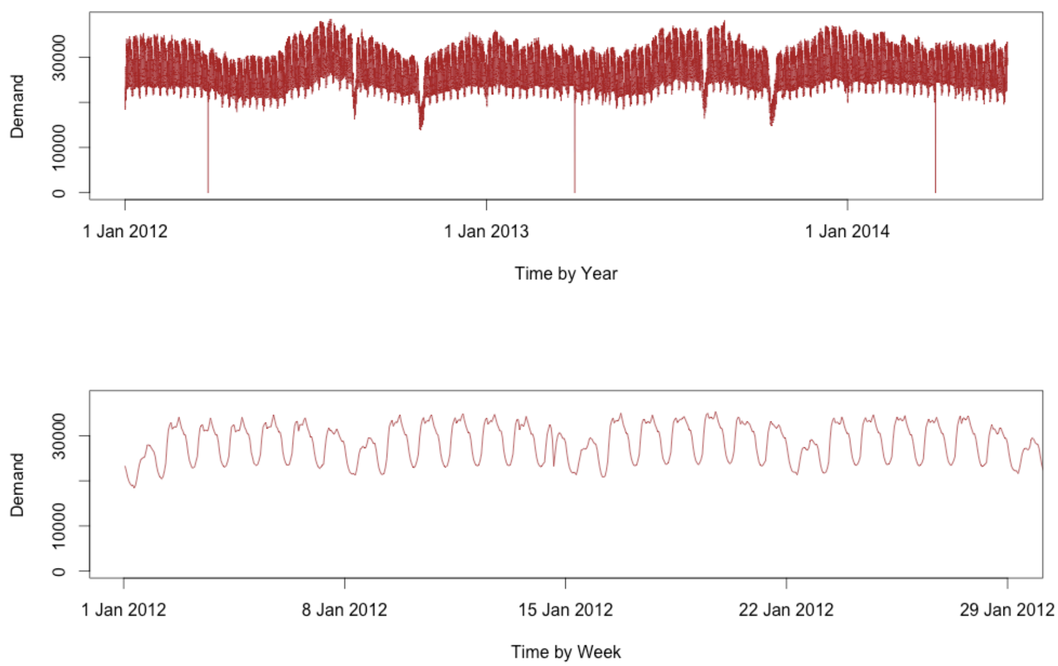

Figure 4.4: *Hourly plot of demand.*

Figure 4.4 shows the plot of demand for the whole time period and for a 4-week time window. There are three apparent seasonal patterns in the plot, namely daily, weekly and yearly seasonality.

4.4 Modeling

We briefly explained the statistical models used in the literature for EPF in Chapter 3. In this study, we modeled the price using the regression method. In this model, we used the lagged values of price as the regressors, demand as the exogenous variable, and dummy variables for Saturdays and Sundays to handle seasonality.

In general, the steps in building a time series model can be summarized as:

1. Perform exploratory data analysis to look for abnormalities in the data,
2. Pre-process or transform the data where necessary,
3. If the data is non-stationary, make it stationary by taking the difference of the data,
4. Examine the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the data for model identification,
5. Check the model recursively by plotting the ACF plot of the residuals and do portmanteau tests on the residuals for model estimation. (Use Information Criteria to find the best model)
6. If the residuals are not white noise, remodel the data or if there is heteroskedasticity, use GARCH models,
7. Perform model diagnostics tests by:
    - Jarque-Bera Normality Test for checking normality of errors,
    - Durbin-Watson and Ljung-Box tests for checking serial correlation,
    - Breusch-Pagan test for checking heteroskedasticity,
8. If the tests pass, calculate the forecasts.
9. Check the forecast accuracy

4.4.1 An ARX-type model for Price

We used the original price data, that is data without any treatment on the spikes, to build our models. Then we processed the spikes and applied the initial model on the new price and compared the accuracy of the forecast.

As we saw in Figure 4.1, price has a non-stationary series, since the variance of the price is not constant over time. This implies that we may need to difference the series to make it stationary. The statistical tests for regular and seasonal unit roots also suggest the existence of one regular unit root for both the price and demand series. However, since differencing results in loss of information, and also it leads to dealing with cointegration in the series and vector correction modeling, we preferred not to use differencing and leave it for future studies.

To build an autoregressive model we need to find the significant lags of price as the regressors. For this purpose we used the ACF and PACF plots of the price series. As we can see in Figure 4.5, both ACF and PACF have a decaying oscillating behavior that suggest an ARMA(p,q) model for the price. However, as there is multiple seasonality in the price, namely daily, weekly and yearly seasonality, a simple ARMA(p,q) model is not a good model for price.

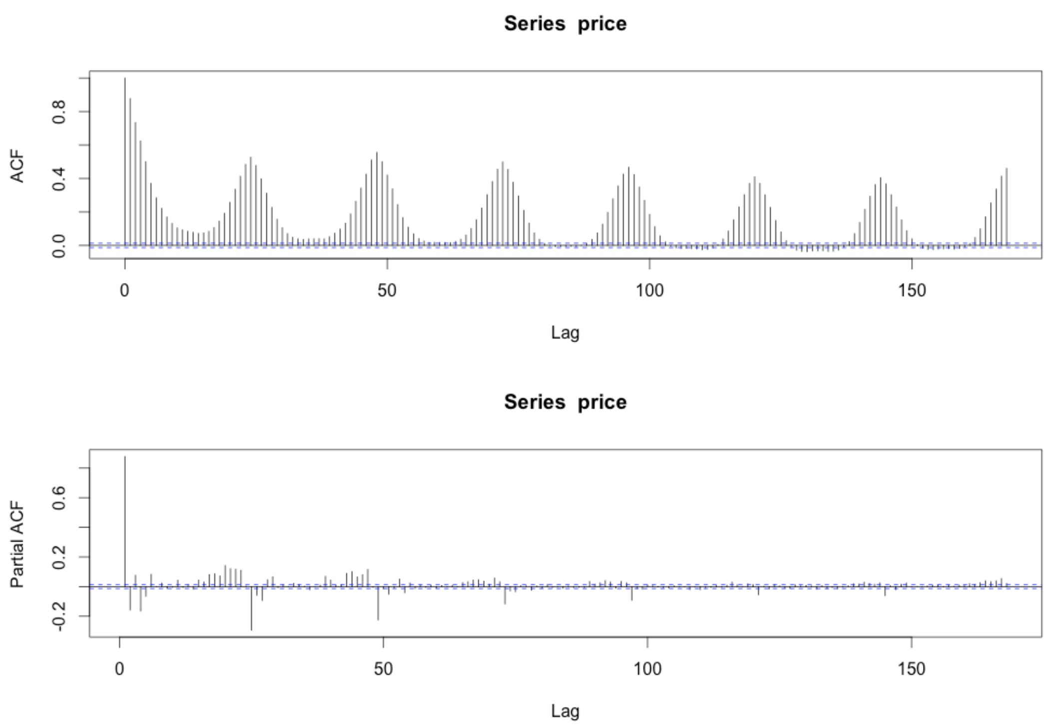

Figure 4.5: *ACF and PACF plot of price.*

After numerous modeling attempts, we finally came up with (4.1) as the best regression model for price.

$$
\begin{aligned}
P_t = \alpha_0 + \alpha_1 P_{t-1} + \alpha_2 P_{t-2} + \alpha_3 P_{t-3} + \alpha_4 P_{t-4} + \alpha_5 P_{t-5} + \alpha_{18} P_{t-18} + \\
\alpha_{24} P_{t-24} + \alpha_{36} P_{t-36} + \alpha_{48} P_{t-48} + \alpha_{72} P_{t-72} + \alpha_{168} P_{t-168} \\
+ X_t + D_{sat} + D_{sun} + u_t \qquad (4.1)
\end{aligned}
$$

where:

P_t : is the current value of price,

P_{t-i} : is the i-th lagged price value,

X_t : is the demand as the exogenous variable of the model,

D_{sat} and D_{sun} are the dummy variables for Saturday and Sunday, and

u_t : is the error term.

In order to choose the right set of dummies to handle seasonality we first put dummy variables for all days of a week in the model. Then by trial and error, we removed the insignificant days and ended up with only Saturdays and Sundays to handle seasonality. We also tested using multiple sine and cosine functions with various periods as regressors, however, neither of those terms had a significant contribution to our model and subsequently were removed.

The estimated values for model parameters and the significant level of each parameter is given in Table 4.4. Table 4.3 shows the descriptive statistics for the residuals of the model. The maximum value of the residual is 1,250.86 which indicates that the model was not able to capture the spikes in the price. It should also be noted that the value of Adjusted R-squared is 0.9795, which is an indicator for the good performance of the model.

Table4.3: *Descriptive statistic for residuals*

Min	1Q	Median	3Q	Max
-428.85	-7.68	-0.50	6.28	1,250.86

The best model is the model which has the least number of parameters and can finely predict the real data. The latter phrase means that the best model is the one that has white noise residuals. However, as mentioned in Table 4.3 the residuals of the model have high variability and does not look like a white noise. So, we tried to model the residuals using regression and came up with (4.2) as the best model for residuals.

$$\begin{aligned} u_t = \beta_0 + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \beta_3 u_{t-3} + \beta_6 u_{t-6} + \beta_{24} u_{t-24} \\ + \beta_{48} u_{t-48} + \beta_{72} u_{t-72} + \beta_{96} u_{t-96} + \beta_{168} u_{t-168} + v_t \end{aligned} \tag{4.2}$$

Table4.4: *Model Parameter Estimation*

Regressors	Estimate	Std.Error	t-value	Pr(>|t|)	Sig. Level
P_{t-1}	9.062e-01	7.687e-03	117.886	$<$ 2e-16	***
P_{t-2}	-2.724e-01	1.027e-02	-26.523	$<$ 2e-16	***
P_{t-3}	2.335e-01	1.032e-02	22.631	$<$ 2e-16	***
P_{t-4}	-8.510e-02	1.027e-02	-8.284	$<$ 2e-16	***
P_{t-5}	-5.978e-02	7.030e-03	-8.503	$<$ 2e-16	***
P_{t-18}	2.719e-02	3.234e-03	8.407	$<$ 2e-16	***
P_{t-24}	4.842e-02	4.460e-03	10.857	$<$ 2e-16	***
P_{t-48}	8.415e-02	4.311e-03	19.519	$<$ 2e-16	***
P_{t-72}	4.176e-02	4.370e-03	9.556	$<$ 2e-16	***
P_{t-168}	4.124e-02	4.115e-03	10.021	$<$ 2e-16	***
X_t	2.356e-04	3.465e-05	6.798	1.10e-11	***
D_{sat}	-2.196e+00	5.027e-01	-4.368	1.26e-05	***
D_{sun}	-6.049e+00	5.117e-01	-11.822	$<$ 2e-16	***

To be a good model, the residuals (v_t) of (4.2) should look like a white noise. In other words, the ACF plot of the residual should not have any significant lags and all the values of autocorrelation should lie inside the white noise bounds. Figure 4.6 shows the ACF and PACF plots for the residuals. It is clear from the graphs that the model is not sufficient.

The normality tests on the residual (p-value<2.2e-16) also confirms the inadequacy of the model. Figure 4.7 shows the Q-Q plot of the residuals (v_t). As we can see, the plot has heavy tails, which is another indicator for non-normality of the errors.

Furthermore, Breusch–Godfrey test for serial correlation (p-value<2.2e-16) also rejects the null hypothesis H_0 of uncorrelated errors. And Breusch-Pagan test for heteroskedasticity (p-value<2.2e-16), rejects the null hypothesis of homoskedasticity and proves the existence of heteroskedasticity in the data. These findings can also be proved by looking at the significant lags in the ACF and PACF plot of residuals.

To model the variability in the data we need to use a GARCH model. We built various GARCH models according to the significant lags in ACF and PACF plots of the residuals. The best model was chosen based on the AIC and BIC criteria

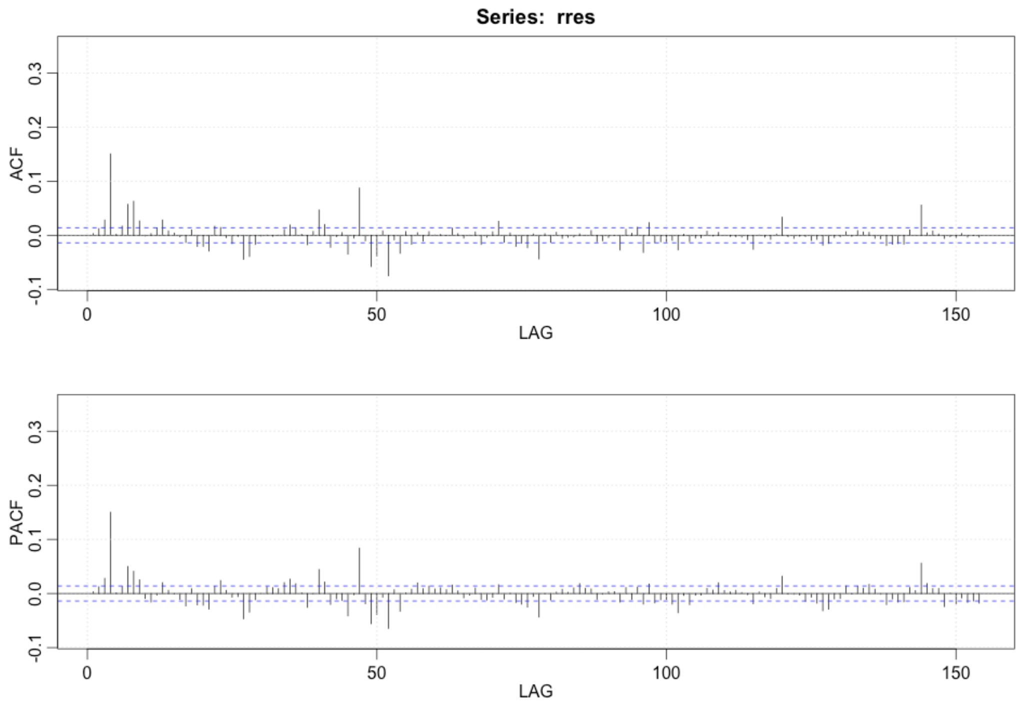

Figure 4.6: *ACF and PACF of v_t*

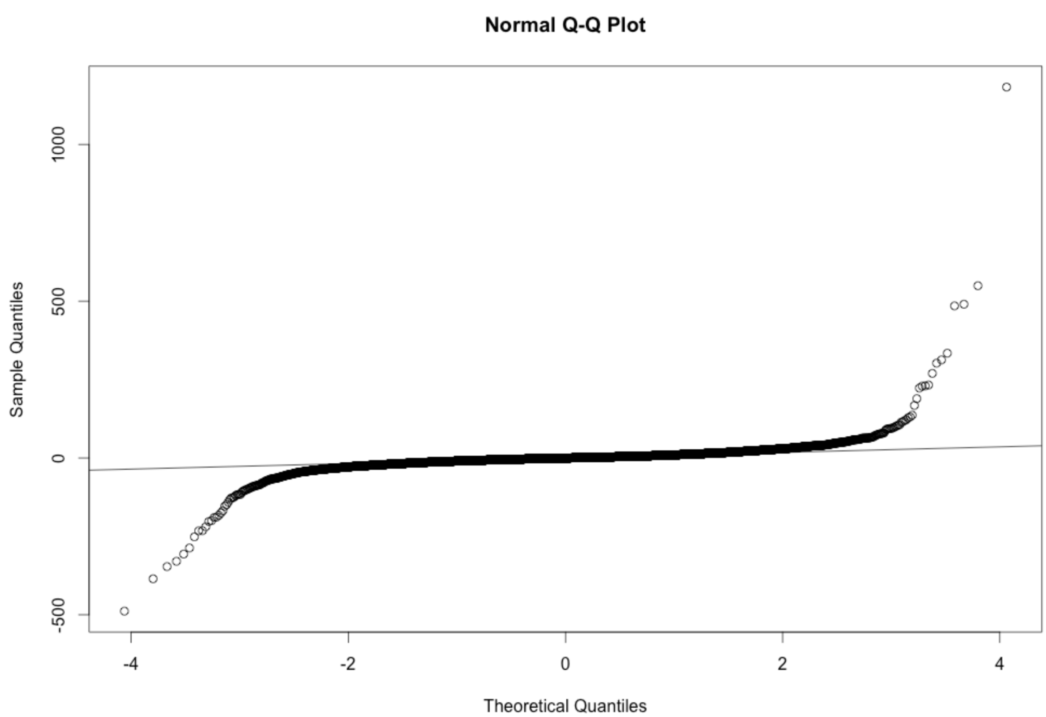

Figure 4.7: *Q-Q plot of the v_t.*

and also the ACF plot of the squared standardized residuals. As explained in Section 3.4, model selection criteria, the best model is the one which has the minimum AIC and BIC values and also has the least number of parameters. The ACF plot of the squared standardized residuals should look like a white noise when the model fits the data. In this study we built the GARCH models presented in Table 4.5 for the residuals (u_t). After examining each model, **ARMA(8,8)GARCH(24,1)** was chosen as the best model. It should be noted that choosing the model only based on AIC and BIC would lead us to choose ARMA(24,24)GARCH(24,1), but this model has too many parameters that were not making a big difference. Also, the model did not have a white noise ACF plot for squared standardized residuals.

Table4.5: *Information Criteria for models*

Model	**AIC**	**BIC**
ARMA(24,24)GARCH(24,1)	7.463061	7.492131
ARMA(24,4)GARCH(24,1)	7.563173	7.584593
ARMA(5,4)GARCH(24,1)	7.745489	7.759642
ARMA(24,24)GARCH(1,1)	7.554887	7.575159
ARMA(8,8)GARCH(24,1)	7.721317	7.738147
ARMA(7,7)GARCH(24,1)	7.728746	7.744811

The ACF plot of the square standardized residuals for the selected model behave normally as shown in Figure 4.8. In this figure the plots in the first row are for the residuals and those in the second row are for the standardized residuals. In each row, the first column is the plot of the residuals, the second is the ACF plot and the third is the ACF for the squared valued of residuals.

After identifying the GARCH model we need to incorporate the GARCH effect of the residuals into the original model for the price (4.1). For this purpose we simply refit (4.1) to the data and put the weights of the regression equal to the inverse of the conditional variance of the selected GARCH model. Note that, since the residuals of the model were not normally distributed, as shown in Figure 4.7, we assumed a Student t-distribution for the conditional distribution in the regression. For a more detailed explanation of this approach refer to [96]. When we refit the model using this approach, the Adjusted R-squared value of the refitted model becomes 0.9964.

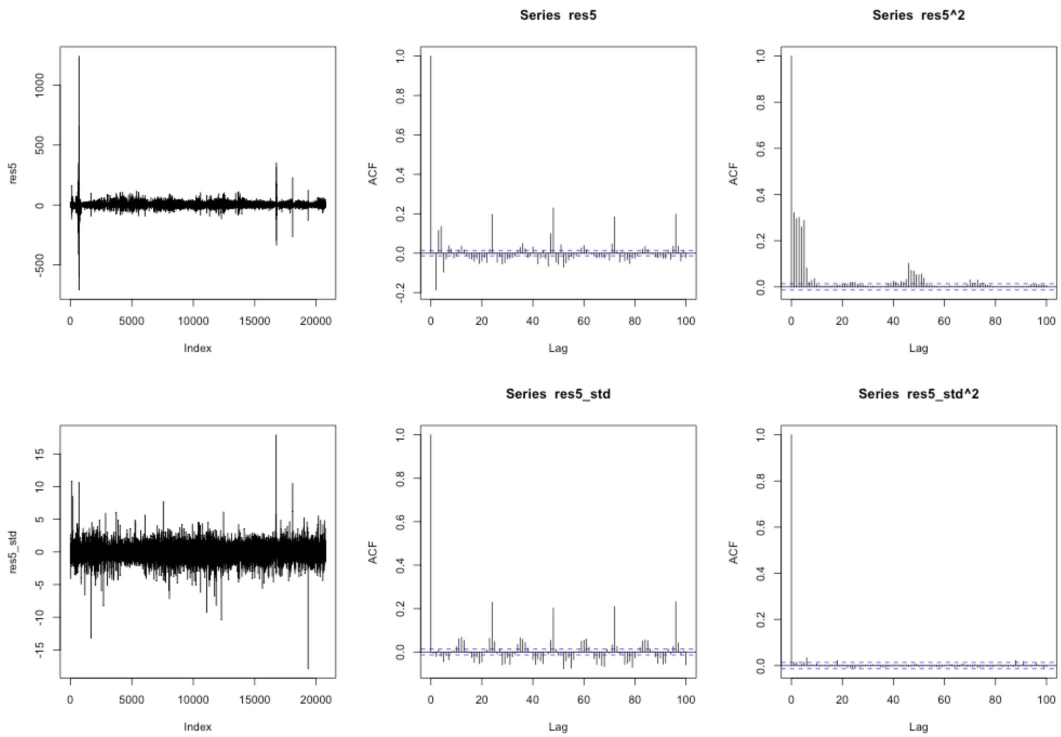

Figure 4.8: *ACF plots of squared standardized residuals*

4.5 Forecasting

For forecasting the price, we first need to forecast the demand. Since demand is one of the components in the price model. Although in this study we have the values of demand for the training and the test set, however in reality we do not know the exact values for demand.

4.5.1 Forecasting Demand

We applied two methods to forecast the demand, namely Double Seasonal Holt-Winters modeling, and ARIMA modeling. For the first model we used a procedure called *dshw*() from package {Forecast} in R and the second model was built using the *auto.arima*() procedure in the same package. For a detailed explanation of these procedure an interested reader can refer to [56]. In the following subsections we explain the outcome of forecasting demand using each of these methods.

4.5.1.1 Forecasting using dshw()

Since there is multi-seasonality in demand series, we decided to use a model considering this behavior. Double Seasonal Holt-Winter method, one of the exponential smoothing methods, covers this characteristics of the data. Taylor on 2003 developed Double Seasonal Holt-Winters method. This method uses two seasonal cycles where the shorter cycles repeats itself inside the longer one. If we let m_1 and m_2 as the periods of the short and the long cycles in this method, then:

$$y_t = l_{t-1} + b_{t-1} + s^{(1)}_{t-m_1} + s^{(2)}_{t-m_2} + \epsilon_t, \tag{4.3}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t, \tag{4.4}$$

$$b_t = b_{t-1} + \beta\epsilon_t, \tag{4.5}$$

$$s^{(1)}_t = s^{(1)}_{t-m_1} + \gamma_1\epsilon_t, \tag{4.6}$$

$$s^{(2)}_t = s^{(2)}_{t-m_2} + \gamma_2\epsilon_t, \tag{4.7}$$

where ϵ_t~NID$(0, \sigma^2)$ and γ_1 and γ_2 are the smoothing parts for seasonal components. Such a model is called a DS(m_1,m_2) model with the seasonal cycles as:

$$c^{(1)}_t = (s^{(1)}_t, s^{(1)}_{t-1}, \ldots, s^{(1)}_{t-m_1+1})', \tag{4.8}$$

$$c^{(2)}_t = (s^{(2)}_t, s^{(2)}_{t-1}, \ldots, s^{(2)}_{t-m_2+1})'. \tag{4.9}$$

In this model, we set two seasonality periods. The first period was set to 24 for daily and the second one to 168 for weekly seasonality. The procedure automatically estimates the parameter of the model using least squares method and does the forecast for h steps ahead.

Figure 4.9 illustrates the graph for demand. In this graph, both the actual values of demand (in dash line), and the forecast values are plotted for two weeks ahead. As we can see, the forecast are only good for the first day and after that the forecast values are always less that the actual ones. However, it should be noted that, the seasonal pattern in the forecast is similar to the pattern in demand.

The goodness of fit for this model is presented in Table 4.6 and the forecast accuracy measures are given in Table 4.7.

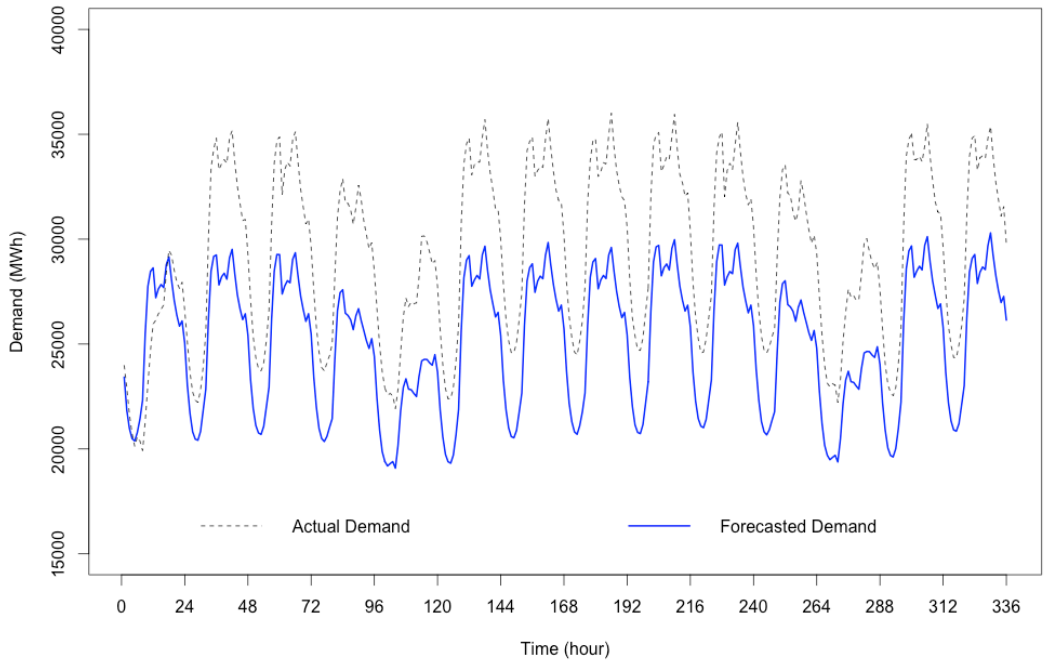

Figure 4.9: *Plot of demand forecast using dshw() procedure*

Table4.6: *Accuracy measure for demand model*

ME	RMSE	MAE	MPE	MAPE	MASE
-8.065	340.885	239.682	-0.038	0.876	0.246

Table4.7: *Demand forecast accuracy measures using dshw()*

ME	RMSE	MAE	MPE	MAPE
-483.622	1931.417	1506.25	-2.34	6.188

4.5.1.2 Forecasting using auto.arima()

It is also possible to model a series using *auto.arima*() procedure in package {Forecast}. Using this method for demand data, we found an ARIMA(24,1,1) model. The accuracy measure for this model are shown in Table 4.8.

Table4.8: *Demand accuracy measures for predicted values vs. actual value for ARIMA(24,1,1) model*

ME	RMSE	MAE	MPE	MAPE	MASE
0.23	636.89	336.30	-Inf	Inf	0.34

Table 4.9 shows the accuracy measures of demand forecast for one day ahead.

Table4.9: *Demand forecast accuracy measures for ARIMA(24,1,1) model*

ME	RMSE	MAE	MPE	MAPE	MASE
-1,725.23	3,389.89	2,701.64	-7.35	11.02	11.023457

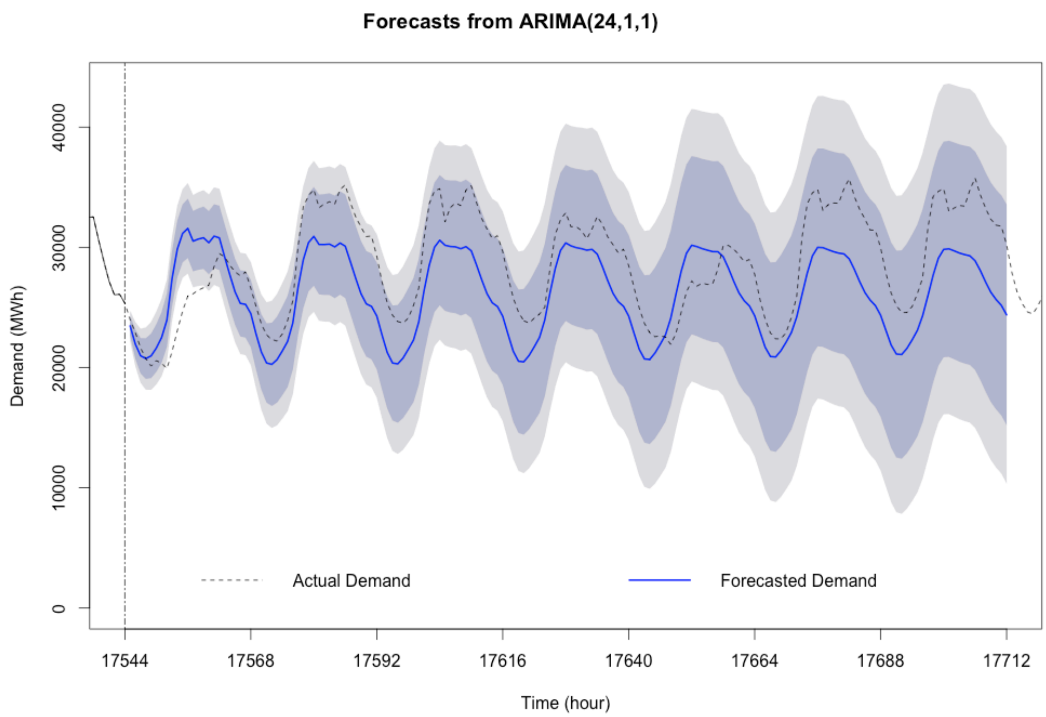

Figure 4.10: *Demand forecast for ARIMA(24,1,1) model*

Comparing the accuracy of demand forecast using the two methods, we realize that the demand forecast using *dshw*() method is slightly better than the forecast using *auto.arima*() method. Therefore, for calculation of price forecast we used the result of *dshw*(). This finding can also be visually inspected by comparing 4.10 and 4.9.

4.5.2 Forecasting Price

The predicted values of the price and the actual prices have a correlation of 0.965. This can be seen in the plot of the actual price values (Dashed line) and the predicted ones in Figures 4.11, 4.12 and 4.13. It can be said that the model behavior is acceptable where there are no high jumps or spikes in the price.

However the model fails to fully capture the spikes. Table 4.10 indicates the accuracy measures of the model.

Table4.10: *Model Accuracy Measures*

Measure	Description	Value
ME	Mean Error	0.60
RMSE	Root Mean Squared Error	22.56
MAE	Mean Absolute Error	10.59
MPE	Mean Percentage Error	-488.49
MAPE	Mean Absolute Percentage Error	495.47
COR	Correlation between Price and Prediction	0.965

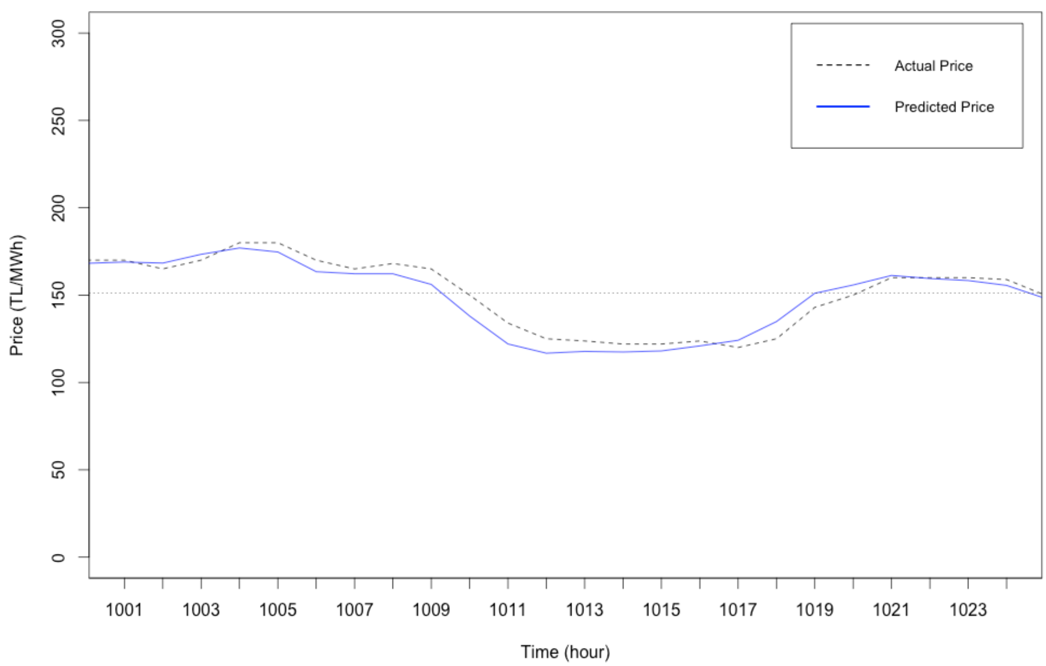

Figure 4.11: *Daily comparison of actual versus predicted price values*

To assess the behavior of a model in future we need to forecast the data and check the forecast accuracy. After forecasting the exogenous variable in the model (demand), we can forecast the price. For this purpose we used the *predict*() function in package {stats} in R, and set h to 24, for one day ahead prediction, and then plotted the forecast values.

We used both of our demand forecasts and compared the resulting price forecast together. The accuracy measure given in Table 4.11 indicate that the price

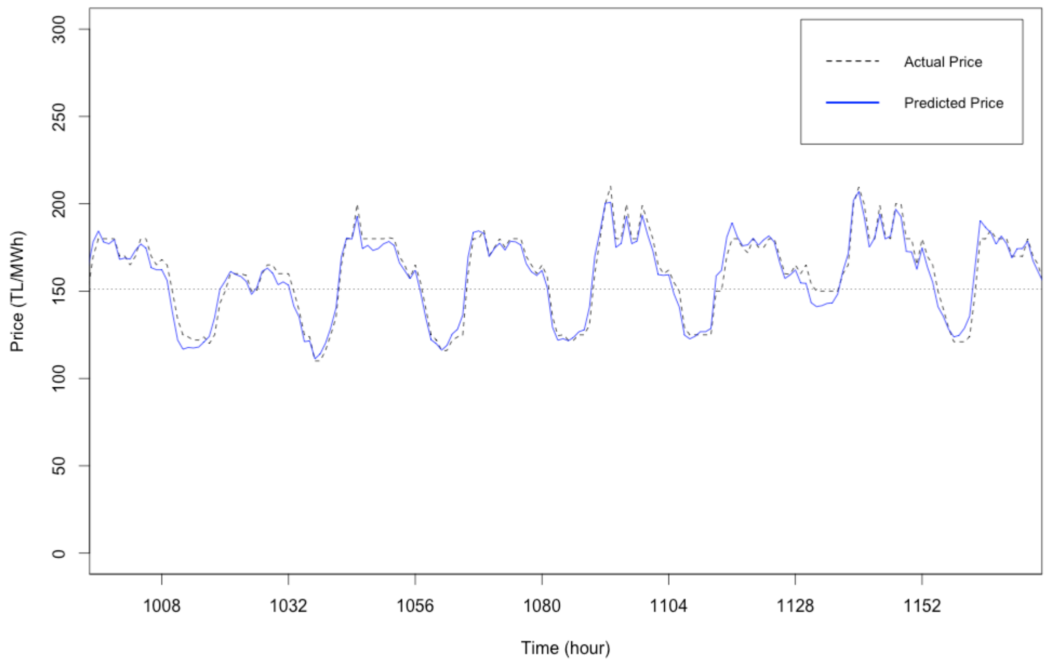

Figure 4.12: *Weekly comparison of actual versus predicted price values*

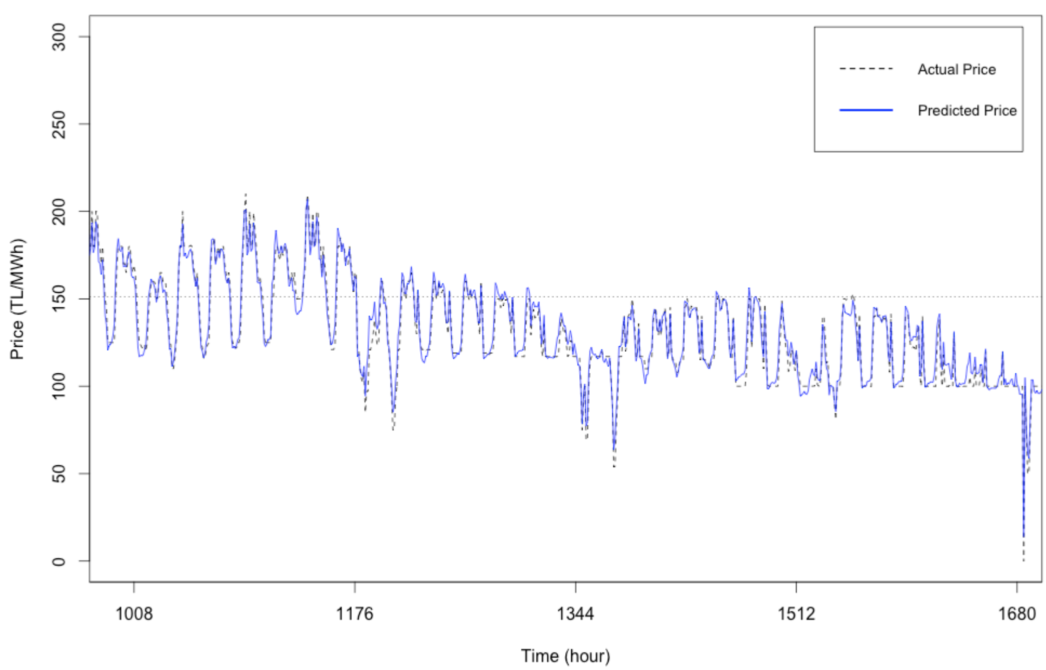

Figure 4.13: *Monthly comparison of actual versus predicted price values*

forecast does not change significantly when we use different forecast for demand. This is due to the small estimate value for demand parameter in the model. So we chose the forecast obtained using double seasonal Holt-Winters model as our main forecast.

Table4.11: *Accuracy measures for the two price forecast*

ME	RMSE	MAE	MPE	MAPE
-0.197	0.282	0.224	-0.111	0.126

Figure 4.14 shows the plot of predicted price values versus the actual values for the first day in the test set. As we can see, the predicted values are close to the actual values for this particular day where there is no spikes. However, as we mentioned in Table 4.10, due to the existence of spikes, the accuracy measures for the model are bigger than these values.

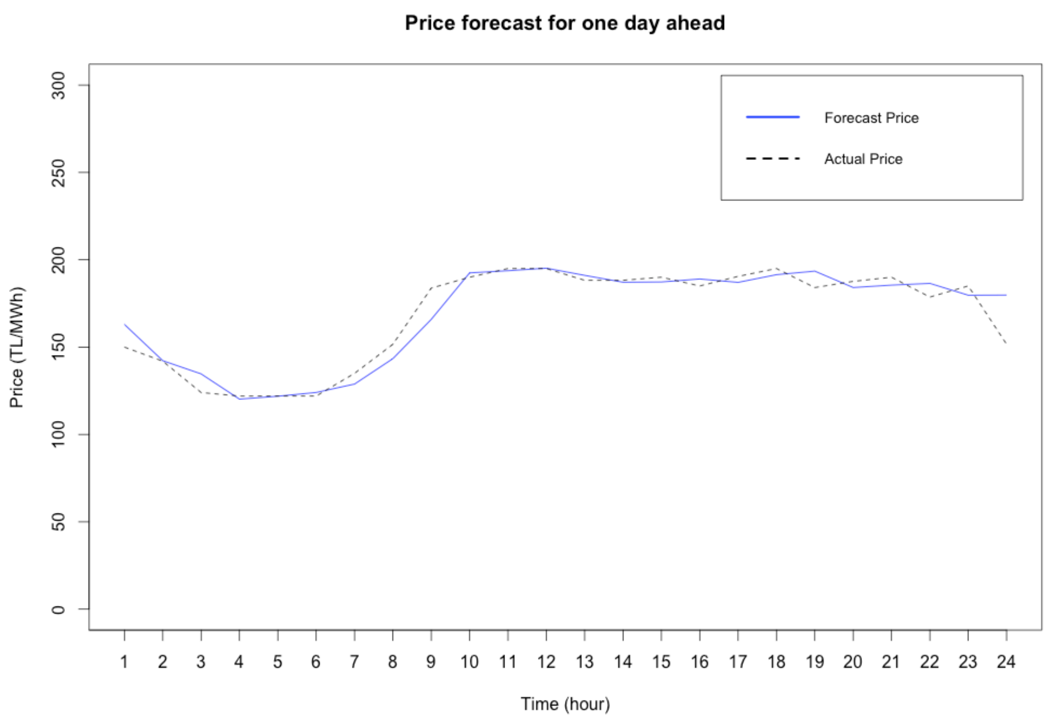

Figure 4.14: *Price forecast for one day ahead (The first day in the test set)*

Table4.12: *Price forecast Accuracy Measures for one day ahead*

ME	**RMSE**	**MAE**	**MPE**	**MAPE**
-0.8872706	8.6182022	5.8491320	-0.7326048	3.6207281

Finally, we ran the analysis again using the spike processed data and followed all the modeling steps. But, we could not get better forecast. This is due to the fact that, spikes have a significant effect on model parameter estimation. Therefore, when we process the spikes, a new model should be constructed for the new data. Obviously the new price data will have less volatility and therefore it would be easier to fit a model to it. It may also be possible to use standard ARIMA models or other conventional modeling approaches. However, doing so is beyond the scope of this study and is left for future studies.

CHAPTER 5

CONCLUSION AND FUTURE RESEARCH

The liberalization of electricity market in Turkey started on 2001 and the market became fully liberalized on 2013. In such a market, an accurate price forecast is fundamental to market participants. However, since the Turkish liberalized electricity market is young, there are not enough studies on forecasting the electricity prices when compared to the industrial countries. Therefore, we decided to work on using statistical models for modeling and forecasting electricity prices in Turkey

In Chapter one, we introduced the idea of liberalization in electricity market and it benefits for consumers, then we discussed the state of electricity market in Turkey briefly. In Chapter two, we reviewed the major literature on EPF in the World and in Turkey. This helped us figure out the current trends in modeling and find the gap in Turkey. To the best of our knowledge, there have not been enough studies on EPF in Turkey in general, and studies using statistical modeling techniques are specifically rare. Subsequently, we decided to work on using statistical modeling techniques in EPF in Turkey. In Chapter three we briefly introduced the main statistical methods used for EPF in the literature and explained the criteria for selecting the best model and forecast accuracy measures.

In Chapter four we focused on modeling the price. We first split the data set into two separate sets, namely the training set and the test set. The models were build based on the training set and were tested on the test set. Before starting to model the price, we pre-processed the spikes. Our approach was to replace all the values outside 99.7% confidence limit by the boundary value of interval. For this purpose, we calculated mean plus three times standard of the price and the upper limit and mean minus three times standard deviation of the price as the lower limit. Starting with the first entry, we calculated the limit for all the entries in price series and replaced the value where necessary. We first built our models based on the original series, and then used that model on the spike processed series.

For modeling, we used an ARX type model for the price. Since there was multiple seasonality in the price series, we decided to use a regression model instead of SARIMA models. In this model, lagged values of the price were used as regressors to make the autoregressive part. The lags were chosen according the the ACF and PACF plots of the price. We used the demand as exogenous variable, and dummy variables for the weekends (Saturdays and Sundays) to handle the seasonality. It should be noted that, we did not do differencing on the price. Doing so, leads us to deal with cointegration and vector correction modeling. The diagnostic tests on the model proved existence of heteroskedasticity in the price. So we modeled the residuals of the model using a ARMA(8,8)GARCH(24,1) model and then incorporated this effect on the initial model. By adding the GARCH part we were able to simulate price movements mostly in the steady state, and to some extent in the spiky regions.

There are always room for improvements. In the follow-up studies, one can build the model on the spikes processed series, and even use differencing to make the series stationary. There are many variation of GARCH modeling, such as EGARCH or TGARCH models. The use of these models can also be studied in the context of electricity price modeling. However, as concluded in Chapter two, the main trend in EPF in the world is to use hybrids methods. In these methods, different modeling approaches are used to model price. Artificial intelligence techniques, such as neural networks and support vector machines can be combined with statistical techniques to simulate various characteristics of electricity prices. To the best of our knowledge, there have not been studies using hybrid methods on Turkey's data. So in future researchers may consider these approaches as well.

REFERENCES

[1] S. Aggarwal, L. Saini, and A. Kumar. Day-ahead price forecasting in ontario electricity market using variable-segmented support vector machine-based model. *Electric Power Components and Systems*, 37(5):495–516, 2009. cited By 19.

[2] S. Aggarwal, L. Saini, and A. Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power and Energy Systems*, 31(1):13–22, 2009. cited By 132.

[3] S. Aggarwal, L. Saini, and A. Kumar. Short term price forecasting in deregulated electricity markets: A review of statistical models and key issues. *International Journal of Energy Sector Management*, 3(4):333–358, 2009. cited By 2.

[4] K. A. Akkemik and F. Oĝuz. Regulation, efficiency and equilibrium: A general equilibrium analysis of liberalization in the Turkish electricity market. *Energy*, 36(5):3282–3292, 2011.

[5] E. Akyildirim, A. Göncü, and A. Altarovici. Modeling long-term seasonality and spikes of the spot electricity prices in Turkey. *Bogazici Journal*, 28(2):1–18, 2014.

[6] N. Amjady. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems*, 21(2):887–896, 2006. cited By 165.

[7] N. Amjady and M. Hemmati. Energy price forecasting: Problems and proposals for such predictions. *IEEE Power and Energy Magazine*, 4(2):20–29, 2006. cited By 63.

[8] C. N. Babu and B. E. Reddy. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing*, 23:27–38, oct 2014.

[9] N. Bagdadioglu and N. Odyakmaz. Turkish electricity reform. *Utilities Policy*, 17(1):144–152, mar 2009.

[10] G. Barta, G. B. G. Nagy, S. Kazi, and T. Henk. *Intelligent Decision Technologies*, volume 39 of *Smart Innovation, Systems and Technologies*. Springer International Publishing, Cham, 2015.

[11] M. Bessec, J. Fouquau, and S. Meritet. Forecasting electricity spot prices using time-series models with a double temporal segmentation. *Applied Economics*, pages 1–18, sep 2015.

[12] G. Biricik, O. O. Bozkurt, and Z. C. Taysi. Analysis of features used in short-term electricity price forecasting for deregulated markets. In *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, pages 600–603. IEEE, may 2015.

[13] S. Bordignon, D. W. Bunn, F. Lisi, and F. Nan. Combining day-ahead forecasts for British electricity prices. *Energy Economics*, 35:88–103, 2013.

[14] A. Botterud, T. Kristiansen, and M. D. Ilic. The relationship between spot and futures prices in the Nord Pool electricity market. *Energy Economics*, 32(5):967–978, 2010.

[15] N. Bowden and J. E. Payne. Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models. *Energy Economics*, 30(6):3186–3197, nov 2008.

[16] D. Bunn. Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, 88(2):163–169, 2000. cited By 243.

[17] K. P. Burnham. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304, nov 2004.

[18] E. Camadan. Why 2012 Will Be So Important for the Restructured Turkish Electricity Market. *The Electricity Journal*, 24(10):70–78, dec 2011.

[19] E. Camadan and F. Kölmek. A Critical Evaluation of Turkish Electricity Reform. *The Electricity Journal*, 26(1):59–70, jan 2013.

[20] R. Carmona and M. Coulon. *Quantitative Energy Finance*. Springer New York, New York, NY, 2014.

[21] J. Carpio, J. Juan, and D. López. Multivariate Exponential Smoothing and Dynamic Factor Model Applied to Hourly Electricity Price Analysis. *Technometrics*, 56(4):494–503, dec 2014.

[22] M. Cerjan, M. Matijaš, and M. Delimar. Dynamic Hybrid Model for Short-Term Electricity Price Forecasting. *Energies*, 7(5):3304–3318, may 2014.

[23] A. Cervone, E. Santini, S. Teodori, and D. Z. Romito. Electricity price forecast: A comparison of different models to evaluate the single national price in the Italian energy exchange market. *International Journal of Energy Economics and Policy*, 4(4):744–758, 2014.

[24] T. Çetin. Institutional change in the turkish energy markets. In *Understanding the Process of Economic Change in Turkey: An Institutional Approach*, pages 275–318. Nova Science Publishers, Inc., 2013.

[25] N. Chaâbane. A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction. *Journal of Applied Statistics*, 41(3):635–651, oct 2013.

[26] N. Chaâbane. A hybrid ARFIMA and neural network model for electricity price prediction. *International Journal of Electrical Power & Energy Systems*, 55:187–194, feb 2014.

[27] J. S. Chan, S. B. Choy, and C. P. Lam. Modeling Electricity Price Using A Threshold Conditional Autoregressive Geometric Process Jump Model. *Communications in Statistics - Theory and Methods*, 43(10-12):2505–2515, may 2014.

[28] S. Chan, K. Tsui, H. Wu, Y. Hou, Y.-C. Wu, and F. Wu. Load/price forecasting and managing demand response for smart grids: Methodologies and challenges. *IEEE Signal Processing Magazine*, 29(5):68–85, 2012. cited By 23.

[29] D. Chen and D. Bunn. The Forecasting Performance of a Finite Mixture Regime-Switching Model for Daily Electricity Prices. *Journal of Forecasting*, 33(5):364–375, aug 2014.

[30] D. Chen and D. W. Bunn. Analysis of the Nonlinear Response of Electricity Prices to Fundamental and Strategic Factors. *IEEE Transactions on Power Systems*, 25(2):595–606, may 2010.

[31] S. Cincotti, G. Gallo, L. Ponta, and M. Raberto. Modeling and forecasting of electricity spot-prices: Computational intelligence vs classical econometrics. *AI Communications*, 27(3):301–314, 2014.

[32] A. Clements, R. Herrera, and A. Hurn. Modelling interregional links in electricity price spikes. *Energy Economics*, 51:383–393, sep 2015.

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

[34] J. Crespo Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77(1):87–106, 2004.

[35] E. Crisostomi, C. Gallicchio, A. Micheli, M. Raugi, and M. Tucci. Prediction of the Italian electricity price for smart grid applications. *Neurocomputing*, 170:286–295, dec 2015.

[36] S. K. Dash, R. Bisoi, and P. K. Dash. A hybrid functional link dynamic neural network and evolutionary unscented Kalman filter for short-term electricity price forecasting. *Neural Computing and Applications*, aug 2015.

[37] A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, dec 2011.

[38] A. K. Diongue, D. Guégan, and B. Vignal. Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied Energy*, 86(4):505–510, apr 2009.

[39] G. Du, L. Wu, L. Liu, and J. Wang. Electricity price forecasting by a combined method. In *WIT Transactions on Information and Communication Technologies*, volume 51, pages 1191–1198. WITPress, mar 2014.

[40] J. Durbin. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616, aug 2002.

[41] O. Efimova and A. Serletis. Energy markets volatility modelling using GARCH. *Energy Economics*, 43:264–273, may 2014.

[42] J. C. R. Filho, C. d. M. Affonso, and R. C. de Oliveira. Energy price prediction multi-step ahead using hybrid model in the Brazilian market. *Electric Power Systems Research*, 117:115–122, dec 2014.

[43] M. Frömmel, X. Han, and S. Kratochvil. Modeling the daily electricity price volatility with realized measures. *Energy Economics*, 44:492–502, jul 2014.

[44] K. Gajowniczek and T. Ząbkowski. Short Term Electricity Forecasting Using Individual Smart Meter Data. *Procedia Computer Science*, 35(C):589–597, 2014.

[45] R. C. Garcia, J. Contreras, S. Member, and M. V. Akkeren. A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices. 20(2):867–874, 2005.

[46] C. García-Martos, E. Caro, and M. Jesús Sánchez. Electricity price forecasting accounting for renewable energies: optimal combined forecasts. *Journal of the Operational Research Society*, 66(5):871–884, mar 2014.

[47] E. S. Gardner. Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22(4):637–666, oct 2006.

[48] A. Göksu. Turkish energy market law and governance: How to mobilize investments. *Insight Turkey*, 12(3):129–145, 2010.

[49] P. G. Gould, A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman, and F. Vahid-Araghi. Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191(1):207–222, nov 2008.

[50] L. Grossi and F. Nan. Robust Self Exciting Threshold AutoRegressive models for electricity prices. In *11th International Conference on the European Energy Market (EEM14)*, pages 1–6. IEEE, may 2014.

[51] K. H. and T. M. The effect of power distribution privatization on electricity prices in turkey: Has liberalization served the purpose? *Energy Policy*, 63:614–621, 2013. cited By 0.

[52] K. S. L. H. Tong. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3):245–292, 1980.

[53] A. Hayfavi and I. Talasli. Stochastic multifactor modeling of spot electricity prices. *Journal of Computational and Applied Mathematics*, 259(PART B):434–442, mar 2014.

[54] R. Herrera and N. González. The modeling and forecasting of extreme events in electricity spot markets. *International Journal of Forecasting*, 30(3):477–490, jul 2014.

[55] O. Hili. Estimation of a multiple-threshold AR(p) model. *Statistical Methodology*, 5(2):177–186, mar 2008.

[56] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1–22, 2008.

[57] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, oct 2006.

[58] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, jul 2002.

[59] S. Islyaev and P. Date. Electricity futures price models: Calibration and forecasting. *European Journal of Operational Research*, 247(1):144–154, may 2015.

[60] J. Janczura, S. Trück, R. Weron, and R. C. Wolff. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, 38:96–110, jul 2013.

[61] T. Jónsson, P. Pinson, H. Nielsen, and H. Madsen. Exponential Smoothing Approaches for Prediction in Real-Time Electricity Markets. *Energies*, 7(6):3710–3732, jun 2014.

[62] I. Juárez, J. Mira-McWilliams, and C. González. Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests. *IET Generation, Transmission & Distribution*, 9(11):1120–1128, aug 2015.

[63] N. V. Karakatsani and D. W. Bunn. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4):764–785, oct 2008.

[64] N. V. Karakatsani and D. W. Bunn. Fundamental and behavioural drivers of electricity price volatility. *Studies in Nonlinear Dynamics and Econometrics*, 14(4), 2010.

[65] V. Kekatos, Y. Zhang, and G. B. Giannakis. Electricity Market Forecasting via Low-Rank Multi-Kernel Learning. *IEEE Journal of Selected Topics in Signal Processing*, 8(6):1182–1193, dec 2014.

[66] C.-i. Kim, I.-K. Yu, and Y. Song. Prediction of system marginal price of electricity using wavelet transform analysis. *Energy Conversion and Management*, 43(14):1839–1851, sep 2002.

[67] M. K. Kim. A new approach to short-term price forecast strategy with an artificial neural network approach: Application to the nord pool. *Journal of Electrical Engineering and Technology*, 10(4):1481–1492, 2015.

[68] M. K. Kim. Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithms. *IET Generation, Transmission & Distribution*, 9(13):1553–1563, oct 2015.

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

[70] M. A. Kölmek and . Navruz. Forecasting the day-ahead price in electricity balancing and settlement market of Turkey by using artificial neural networks. *Turkish Journal of Electrical Engineering & Computer Sciences*, 23(3):841–852, 2015.

[71] S. J. Koopman, M. Ooms, and M. A. Carnero. Periodic Seasonal Reg-ARFIMA–GARCH Models for Daily Electricity Spot Prices. *Journal of the American Statistical Association*, 102(477):16–27, mar 2007.

[72] V. V. Kossov. Medium-term forecasting of demand prices on example of electricity prices for industry. *Studies on Russian Economic Development*, 25(5):456–466, sep 2014.

[73] P. Kou, D. Liang, L. Gao, and J. Lou. Probabilistic electricity price forecasting with variational heteroscedastic Gaussian process and active learning. *Energy Conversion and Management*, 89:298–308, jan 2015.

[74] R. M. Kovacevic and D. Wozabal. A semiparametric model for electricity spot prices. *IIE Transactions*, 46(4):344–356, mar 2014.

[75] T. Kristiansen. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, 49:328–332, oct 2012.

[76] B. Q. Li, X. Liang, D. S. Yang, B. Hu, and D. Li. The Establishment of Electricity Price Model Based on User Satisfaction and Power Supply Enterprise Return. *Applied Mechanics and Materials*, 488-489:1348–1353, jan 2014.

[77] N. Ludwig, S. Feuerriegel, and D. Neumann. Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems*, 24(1):19–36, jan 2015.

[78] K. Maciejowska, J. Nowotarski, and R. Weron. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *International Journal of Forecasting*, apr 2015.

[79] K. Maciejowska and R. Weron. Forecasting of daily electricity prices with factor models: utilizing intra-day and inter-zone relationships. *Computational Statistics*, 30(3):805–819, sep 2014.

[80] K. Maciejowska and R. Weron. Short- and Mid-Term Forecasting of Baseload Electricity Prices in the U.K.: The Impact of Intra-Day Price Relationships and Market Fundamentals. *IEEE Transactions on Power Systems*, pages 1–12, 2015.

[81] D. Masri, H. Zeineldin, and W. L. Woon. Electricity price and demand forecasting under smart grid environment. In *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, pages 1956–1960. IEEE, jun 2015.

[82] J. Mei, D. He, R. Harley, T. Habetler, and G. Qu. A random forest method for real-time price forecasting in New York electricity market. In *2014 IEEE PES General Meeting | Conference & Exposition*, volume 2014-Octob, pages 1–5. IEEE, jul 2014.

[83] A. Misiorek, S. Trueck, and R. Weron. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Studies in Nonlinear Dynamics and Econometrics*, 10(3), 2006. cited By 47.

[84] F. Moscetti, S. Paoletti, and A. Vicino. Analysis and models of electricity prices in the Italian ancillary services market. In *IEEE PES Innovative Smart Grid Technologies, Europe*, volume 2015-Janua, pages 1–6. IEEE, oct 2014.

[85] F. Nogales, J. Contreras, A. Conejo, and R. Espinola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17(2):342–348, may 2002.

[86] J. Nowotarski, E. Raviv, S. Trück, and R. Weron. An empirical comparison of alternative schemes for combining electricity spot price forecasts. *Energy Economics*, 46:395–412, nov 2014.

[87] J. Nowotarski and R. Weron. Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices. In *11th International Conference on the European Energy Market (EEM14)*, pages 1–5. IEEE, may 2014.

[88] G. Osório, J. Matias, and J. Catalão. Electricity prices forecasting by a hybrid evolutionary-adaptive methodology. *Energy Conversion and Management*, 80:363–373, apr 2014.

[89] G. J. O. Osorio, J. C. O. Matias, and J. P. S. Catalao. Hybrid evolutionary-adaptive approach to predict electricity prices and wind power in the short-term. In *2014 Power Systems Computation Conference*, pages 1–7. IEEE, aug 2014.

[90] T. Papadimitriou, P. Gogas, and E. Stathakis. Forecasting energy markets using support vector machines. *Energy Economics*, 44:135–142, jul 2014.

[91] F. Paraschiv, S.-E. Fleten, and M. Schürle. A spot-forward model for electricity prices with regime shifts. *Energy Economics*, 47:142–153, jan 2015.

[92] H. Qian and Z. Qiu. Feature selection using C4.5 algorithm for electricity price prediction. In *2014 International Conference on Machine Learning and Cybernetics*, volume 1, pages 175–180. IEEE, jul 2014.

[93] E. Raviv, K. E. Bouwman, and D. van Dijk. Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, 50:227–239, jul 2015.

[94] B. Ricky Rambharat, A. E. Brockwell, and D. J. Seppi. A threshold autoregressive model for wholesale electricity prices. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(2):287–299, apr 2005.

[95] T. A. Robinson. Electricity pool prices: A case study in nonlinear time-series modelling. *Applied Economics*, 32(5):527–532, 2000.

[96] D. Ruppert. *The Work of Raymond J. Carroll.* Springer International Publishing, Cham, 2014.

[97] K. B. Sahay and M. M. Tripathi. An analysis of short-term price forecasting of power market by using ANN. In *2014 6th IEEE Power India International Conference (PIICON)*, pages 1–6. IEEE, dec 2014.

[98] A. Samitas and A. Armenatzoglou. Regression tree model versus Markov regime switching: a comparison for electricity spot price modelling and forecasting. *Operational Research*, 14(3):319–340, jun 2014.

[99] M. Shahidehpour and L. Wu. A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid. *IET Generation, Transmission & Distribution*, 8(12):1937–1950, dec 2014.

[100] H. Shayeghi, A. Ghasemi, M. Moradzadeh, and M. Nooshyar. Day-ahead electricity price forecasting using WPT, GMI and modified LSSVM-based S-OLABC algorithm. *Soft Computing*, aug 2015.

[101] H. Shayeghi, A. Ghasemi, M. Moradzadeh, and M. Nooshyar. Simultaneous day-ahead forecasting of electricity price and load in smart grids. *Energy Conversion and Management*, 95:371–384, may 2015.

[102] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi. Prediction interval estimation for electricity price and demand using support vector machines. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 3995–4002. IEEE, jul 2014.

[103] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi. Prediction Interval Estimation of Electricity Prices using PSO tuned Support Vector Machines. *IEEE Transactions on Industrial Informatics*, 11(2):1–1, 2015.

[104] N. A. Shrivastava and B. K. Panigrahi. A hybrid wavelet-ELM based short term price forecasting for electricity markets. *International Journal of Electrical Power & Energy Systems*, 55:41–50, feb 2014.

[105] N. A. Shrivastava and B. K. Panigrahi. Prediction interval estimations for electricity demands and prices: a multi-objective approach. *IET Generation, Transmission & Distribution*, 9(5):494–502, apr 2015.

[106] N. A. Shrivastava, B. K. Panigrahi, and M.-H. Lim. Electricity price classification using extreme learning machines. *Neural Computing and Applications*, pages 1–10, jan 2014.

[107] Z. Tan, J. Zhang, J. Wang, and J. Xu. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*, 87(11):3606–3610, 2010.

[108] J. W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152, jul 2010.

[109] Z. C. Taysi, G. Biricik, and O. O. Bozkurt. Evaluation of price forecast systems for Turkish Electric Market. In *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, pages 620–623. IEEE, may 2015.

[110] J. S. Vardakas and I. Zenginis. *Wireless Internet*, volume 146 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Springer International Publishing, Cham, 2015.

[111] V. Veera Anusuya, S. Sridevi, and G. Mahalakshmi. Prediction of time series data using SVM and ARMAX. *International Journal of Applied Engineering Research*, 10(20):19171–19175, 2015.

[112] S. Vijayalakshmi and G. P. Girish. Artificial neural networks for spot electricity price forecasting: A review. *International Journal of Energy Economics and Policy*, 5(4):1092–1097, 2015.

[113] J. M. Vilar, R. Cao, and G. Aneiros. Forecasting next-day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*, 39(1):48–55, jul 2012.

[114] S. Voronin and J. Partanen. Forecasting electricity price and demand using a hybrid approach based on wavelet transform, ARIMA and neural networks. *International Journal of Energy Research*, 38(5):626–637, apr 2014.

[115] S. Voronin, J. Partanen, and T. Kauranne. A hybrid electricity price forecasting model for the Nordic electricity spot market. *International Transactions on Electrical Energy Systems*, 24(5):736–760, may 2014.

[116] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong. A Hybrid Approach for Probabilistic Forecasting of Electricity Price. *IEEE Transactions on Smart Grid*, 5(1):463–470, jan 2014.

[117] J. Wang, L. Xiao, and J. Shi. The Combination Forecasting of Electricity Price Based on Price Spikes Processing: A Case Study in South Australia. *Abstract and Applied Analysis*, 2014:1–12, 2014.

[118] R. Weron. 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, oct 2014.

[119] R. Weron and A. Misiorek. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24(4):744–763, 2008. cited By 83.

[120] S. F. Xiong and X. Y. Zou. Value at risk and price volatility forecasting in electricity market: A literature review, 2014.

[121] X. Yan and N. a. Chowdhury. Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power and Energy Systems*, 53(1):20–26, 2013.

[122] X. Yan and N. A. Chowdhury. Mid-term electricity market clearing price forecasting: A multiple SVM approach. *International Journal of Electrical Power & Energy Systems*, 58:206–214, jun 2014.

[123] X. Yan and N. A. Chowdhury. Mid-term electricity market clearing price forecasting using multiple least squares support vector machines. *IET Generation, Transmission & Distribution*, 8(9):1572–1582, sep 2014.

[124] X. Yan and N. A. Chowdhury. Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input. *International Journal of Electrical Power & Energy Systems*, 63:64–70, dec 2014.

[125] Y. Yang, Y. Dong, Y. Chen, and C. Li. Intelligent Optimized Combined Model Based on GARCH and SVM for Forecasting Electricity Price of New South Wales, Australia. *Abstract and Applied Analysis*, 2014:1–9, 2014.

[126] D. Young, S. Poletti, and O. Browne. Can agent-based models forecast spot prices in electricity markets? Evidence from the New Zealand electricity market. *Energy Economics*, 45:419–434, sep 2014.

[127] J. Zhang, Z. Tan, and C. Li. A Novel Hybrid Forecasting Method Using GRNN Combined With Wavelet Transform and a GARCH Model. *Energy Sources, Part B: Economics, Planning, and Policy*, 10(4):418–426, jan 2015.

[128] H. Zhou, F. Tang, X. Ye, and J. Jia. The dynamic model and prediction research on electricity forward price in competitive power market. *Journal of Computational Information Systems*, 11(11):4181–4189, 2015.

[129] F. Ziel, R. Steinert, and S. Husmann. Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, 47:98–111, jan 2015.

[130] F. Ziel, R. Steinert, and S. Husmann. Forecasting day ahead electricity spot prices: The impact of the EXAA to other European electricity markets. *Energy Economics*, 51:430–444, sep 2015.

[131] W. Zucchini. An Introduction to Model Selection. *Journal of mathematical psychology*, 44(1):41–61, mar 2000.

APPENDIX A

R CODES

```
# Installing required packages and reading the data -----------------------

setwd("/Users/apple/Google Drive/Thesis/Code")
install.packages("lmtest")
require("lmtest")
install.packages("TSA")
require(TSA)
install.packages("pastecs") # to use stat.desc() method for good descriptive statist
require("pastecs")
install.packages("forecast")
require("forecast")
install.packages("fGarch")
require(fGarch)
install.packages("stat")
require(stat)

data <- read.table("price-demand.csv", header=TRUE, sep=",")
price <- data[,3]
price[price==0] <- 0.01
plot(price, type="l",col="blue", xlab="Time (Hour)", ylab="Price (TL/MWh)")
lprice <- log(price)
plot(lprice, type="l",col="darkgreen", xlab="Time (Hour)", ylab="Log Price")
stat.desc(price)
```

```
new.par <- par(mfrow=c(2,1))
demand <- data[,4]
plot(demand, type="l",col="brown",
xlab="Time by Year", xaxt="n", ylab="Demand")
axis(side = 1, at = c(0, 8781,17545),
labels = c("1 Jan 2012","1 Jan 2013", "1 Jan 2014"))
plot(demand, type="l",col="brown",
xlab="Time by Week", ylab="Demand", xaxt = "n",
xlim=c(1,672))
axis(side = 1, at = c(0, 168, 336, 504, 672),
labels = c("1 Jan 2012","8 Jan 2012", "15 Jan 2012",
"22 Jan 2012", "29 Jan 2012"))
stat.desc(demand)
# Outlier detection -------------------------------------------------------

#removing the one 3*std greather than the mean
spikes <- rep(NA,length(price))
count = 0
pprice <- price
for(i in 1:length(pprice)){
M = mean(pprice)
S = sd(pprice) # calculate standard deviation
UL = M+3*S # upper limit
DL = M-3*S # lower limit
if(pprice[i]>= UL)
{
count = count + 1
spikes[i]=pprice[i]
pprice[i]= UL
}
if(pprice[i]<= DL)
{
```

```
count = count + 1
spikes[i]=pprice[i]
pprice[i]= DL
}
}
stat.desc(pprice)
# plot of the spikes
plot(spikes, ylab="Price (TL/MWh)", xlab="Time (hour)", main="Spikes in the Price")
# plot of the pre-processed price
plot(pprice, type="l",col="darkgreen", ylab="Price (TL/MWh)", xlab="Time (hour)",
main="Spikes Processed Price")
# price <- pprice
# Creating Dummy Variables  -----------------------------------------------

dates <- ISOdatetime(2012,1,1,0,0,0)+1:(length(price))*60*60
df = data.frame(dates)
df$day <- weekdays(as.Date(df$dates))
# maximum lag we need for modeling
maxlag <- 336
L <- length(price)-maxlag
d.sat <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Saturday")
d.sat[i]<-1
else
d.sat[i]<-0
}

d.sun <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Sunday")
```

```
d.sun[i]<-1
else
d.sun[i]<-0
}
d.mon <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Monday")
d.mon[i]<-1
else
d.mon[i]<-0
}
d.tue <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Tuesday")
d.tue[i]<-1
else
d.tue[i]<-0
}
d.wed <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Wednesday")
d.wed[i]<-1
else
d.wed[i]<-0
}
d.thu <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Thursday")
d.thu[i]<-1
```

```
else
d.thu[i]<-0
}
d.fri <- rep(NA,L)
for(i in 1:L)
{
if(df$day[i]=="Friday")
d.fri[i]<-1
else
d.fri[i]<-0
}
# find and create minimum price of yesterday
mins <- rep(NA, L)
i <- 1
while(i<=L){
mins[i:(i+23)] <- min(price[i:(i+23)])
i <- i+24
}
# finding the mean price of yesterday
means <- rep(NA, L)
i <- 1
while(i<=L){
means[i:(i+23)] <- mean(price[i:(i+23)])
i <- i+24
}

# Creating a data frame for regressors ------------------------------------

x <- data.frame(cbind(as.ts(price),lag(price,k=-1),lag(price,k=-2),
lag(price,k=-3), lag(price,k=-4),lag(price,k=-5),
lag(price,k=-18),lag(price,k=-24), lag(price,k=-36),
lag(price,k=-48),lag(price,k=-72),lag(price,k=-168),
lag(price,k=-maxlag),lag(mins,k=-24), lag(means,k=-24),
```

```
lag(demand,k=-maxlag),lag(d.mon,k=-maxlag),
lag(d.tue,k=-maxlag),lag(d.wed,k=-maxlag),lag(d.thu,k=-maxlag),
lag(d.fri,k=-maxlag),lag(d.sat,k=-maxlag),lag(d.sun,k=-maxlag)))
colnames(x) <- c("t","t1","t2","t3","t4","t5","t18","t24","t36","t48",
"t72","t168","t336","pmins","pmeans","dmnd"
,"d_mon","d_tue","d_wed","d_thu","d_fri","d_sat","d_sun")
xs <- x[maxlag+1:(length(price)-2*maxlag),]
attach(xs)

# length(xs$t)

# modeling the price ------------------------------------------------------

model1 <- lm(t ~ -1 + t1 + t24 + t48 + t168 + t336 + dmnd + pmins +
d_mon + d_tue + d_wed + d_fri + d_sat + d_sun)
summary(model1)

model2 <- lm(t ~ -1 + t1 + t24 + t48 + t168 + dmnd + pmins +
d_mon + d_tue + d_wed + d_fri + d_sat + d_sun)
summary(model2)
model3 <- lm(t ~ -1 + t24 + t48 + t168 + t336 + dmnd + pmins +
d_mon + d_tue + d_wed + d_fri + d_sat + d_sun)
summary(model3)

model4 <- lm(t ~ -1 + t24 + t48 + t168 + dmnd + pmins +
d_mon + d_tue + d_wed + d_fri + d_sat + d_sun)
summary(model4)

model5 <- lm(t ~ -1 + t24 + t48 + t168 + dmnd + pmins +
d_mon + d_tue + d_wed + d_thu + d_fri + d_sat + d_sun)
summary(model5)

model6 <- lm(t ~ -1 + t1 + t24 + t48 + t168 + dmnd + pmins +
```

```
d_mon + d_sat + d_sun)
summary(model6)

model7 <- lm(t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 + t36 +
t48 + t72 + t168 +
dmnd + pmins + d_mon + d_tue + d_wed + d_fri + d_sat + d_sun)
summary(model7)

model8 <- lm(t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 +
t48 + t72 + t168 + dmnd +
d_sat + d_sun)
summary(model8)
model <- model8
accuracy(model)
# normality test and q-q plot for residuals

# tests for serial correlation
bgtest(model)
dwtest(model)
# tests for heteroscedasticity
bptest(model)

# Modeling Residuals -------------------------------------------------------

res <- resid(model)
plot(res,type="l")
ndiffs(res)
par(mfrow=c(2,1))
acf(res, lag.max=500)
pacf(res, lag.max=500)
# modeling residuls of the residuals
rr <- data.frame(cbind(as.ts(res),lag(res,k=-1),lag(res,k=-2),lag(res,k=-3),
lag(res,k=-4),lag(res,k=-5),lag(res,k=-6),lag(res,k=-7),
```

```
lag(res,k=-12),lag(res,k=-24),lag(res,k=-25),lag(res,k=-40),
lag(res,k=-45),lag(res,k=-46),lag(res,k=-48),lag(res,k=-72),
lag(res,k=-96),lag(res,k=-168)))
colnames(rr) <- c("r","r1","r2","r3","r4","r5","r6","r7","r12","r24","r25",
"r40","r45","r46","r48","r72","r96","r168")
resmaxlag <- 168
rs <- rr[resmaxlag+1:(length(rr$r)-2*resmaxlag),]
length(rs$r)
attach(rs)
model.res1 <- lm(r ~ r1 + r4 + r24)
summary(model.res1)
model.res2 <- lm(r ~ r1 + r24 + r48)
summary(model.res2)
model.res3 <- lm(r ~ r1 + r4 + r5 + r24 + r48)
summary(model.res3)
model.res4 <- lm(r ~ r1 + r2 + r3 + r4 + r5 + r6 + r7 + r24 + r48)
summary(model.res4)
model.res5 <- lm(r ~ -1 + r1 + r2 + r3 + r6 + r24 + r48 + r72 + r96 + r168)
summary(model.res5)
model.res6 <- lm(r ~ -1 + r1 + r2 + r3 + r5 + r6 + r12+ r24 + r25 + r45 +
r46+ r40 + r48 + r72)
summary(model.res6)

rmodel <- model.res5
rres <- resid(rmodel)
plot(rres, type="l")
acf(rres, main="Vt")

# q-q plot of residuals
par(mfrow=c(1,1))
qqnorm(rres)
qqline(rres)
# tests for serial correlation
```

```
bgtest(rmodel)
dwtest(rmodel)
# tests for heteroscedasticity
bptest(rmodel)


# ARMA(24,24)GARCH(24,1) --------------------------------------------------

garch.model1 = garchFit(formula= ~arma(24,24)  + garch(24,1),
res, cond.dist = "std", trace=FALSE)
summary(garch.model1)
garch.model1@fit$matcoef
res1 = residuals(garch.model1)
res1_std = res1 / garch.model1@sigma.t
par(mfrow=c(2,3))
plot(res1)
acf(res1,lag=100)
acf(res1^2,lag=100)
plot(res1_std)
acf(res1_std,lag=100)
acf(res1_std^2,lag=100)

fit1 <- lm(formula = t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 +
t48 + t72 + t168 + dmnd +
d_sat + d_sun, weights = 1/garch.model1@sigma.t^2)
summary(fit1)
accuracy(fit1)
preds1 <- predict(fit1)

# ARMA(24,4)GARCH(24,1) ---------------------------------------------------

garch.model2 = garchFit(formula= ~arma(24,4)  + garch(24,1), res,
cond.dist = "std", trace=FALSE)
```

```
summary(garch.model2)
garch.model2@fit$matcoef
res2 = residuals(garch.model2)
res2_std = res2 / garch.model2@sigma.t
par(mfrow=c(2,3))
plot(res2)
acf(res2,lag=100)
acf(res2^2,lag=100)
plot(res2_std)
acf(res2_std,lag=100)
acf(res2_std^2,lag=100)

fit2 <- lm(formula = t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 +
t48 + t72 + t168 + dmnd +
d_sat + d_sun, weights = 1/garch.model2@sigma.t^2)
summary(fit2)
accuracy(fit2)
preds2 <- predict(fit2)

# ARMA(5,4)GARCH(24,1) -----------------------------------------------------

garch.model3 = garchFit(formula= ~arma(5,4) + garch(24,1), res,
cond.dist = "std", trace=FALSE)
summary(garch.model3)
garch.model3@fit$matcoef
res3 = residuals(garch.model3)
res3_std = res3 / garch.model3@sigma.t
par(mfrow=c(2,3))
plot(res3)
acf(res3,lag=100)
acf(res3^2,lag=100)
plot(res3_std)
acf(res3_std,lag=100)
```

```
acf(res3_std^2,lag=100, main="ARMA(5,4)GARCH(24,1)")

fit3 <- lm(formula = t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 +
t48 + t72 + t168 + dmnd +
d_sat + d_sun, weights = 1/garch.model3@sigma.t^2)
summary(fit3)
accuracy(fit3)
preds3 <- predict(fit3)

# ARMA(24,24)GARCH(1,1) ---------------------------------------------------

garch.model4 = garchFit(formula= ~arma(24,24) + garch(1,1), res,
cond.dist = "std", trace=FALSE)
summary(garch.model4)
garch.model4@fit$matcoef
res4 = residuals(garch.model4)
res4_std = res4 / garch.model4@sigma.t
par(mfrow=c(2,3))
plot(res4)
acf(res4,lag=100)
acf(res4^2,lag=100)
plot(res4_std)
acf(res4_std,lag=100)
acf(res4_std^2,lag=100)

fit4 <- lm(formula = t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 +
t48 + t72 + t168 + dmnd +
d_sat + d_sun, weights = 1/garch.model4@sigma.t^2)
summary(fit4)
accuracy(fit4)
preds4 <- predict(fit4)
```

```
# ARMA(8,8)GARCH(24,1) ----------------------------------------------------

garch.model5 = garchFit(formula= ~arma(8,8)  + garch(24,1), res,
cond.dist = "std", trace=FALSE)
summary(garch.model5)
garch.model5@fit$matcoef
res5 = residuals(garch.model5)
res5_std = res5 / garch.model5@sigma.t
par(mfrow=c(2,3))
plot(res5, type="l")
acf(res5,lag=100)
acf(res5^2,lag=100)
plot(res5_std, type="l")
acf(res5_std,lag=100)
acf(res5_std^2,lag=100)

# ARMA(7,7)GARCH(24,1) modeling -------------------------------------------

garch.model6 = garchFit(formula= ~arma(7,7)  + garch(24,1),
res, cond.dist = "std", trace=FALSE)
summary(garch.model6)
garch.model6@fit$matcoef
res6 = residuals(garch.model6)
res6_std = res6 / garch.model6@sigma.t
par(mfrow=c(2,3))
plot(res6, type="l")
acf(res6,lag=100)
acf(res6^2,lag=100)
plot(res6_std, type="l")
acf(res6_std,lag=100)
acf(res6_std^2,lag=100)

# ARMA(5,2)GARCH(24,1) modeling -------------------------------------------
```

```
garch.model7 = garchFit(formula= ~arma(5,2)  + garch(24,1),
res, cond.dist = "std", trace=FALSE)
summary(garch.model7)
garch.model7@fit$matcoef
res7 = residuals(garch.model7)
res7_std = res7 / garch.model7@sigma.t
par(mfrow=c(2,3))
plot(res7, type="l")
acf(res7,lag=100)
acf(res7^2,lag=100)
plot(res7_std, type="l")
acf(res7_std,lag=100)
acf(res7_std^2,lag=100)

# ARMA(1,1)GARCH(24,1) modeling -------------------------------------------

garch.model8 = garchFit(formula= ~arma(1,1)  + garch(24,1),
res, cond.dist = "std", trace=FALSE)
summary(garch.model8)
garch.model8@fit$matcoef
res8 = residuals(garch.model8)
res8_std = res8 / garch.model8@sigma.t
par(mfrow=c(2,3))
plot(res8, type="l")
acf(res8,lag=100)
acf(res8^2,lag=100)
plot(res8_std, type="l")
acf(res8_std,lag=100)
acf(res8_std^2,lag=100)

# ARMA(1,1)GARCH(1,1) modeling --------------------------------------------
```

```
garch.model9 = garchFit(formula= ~arma(1,1)  + garch(1,1),
res, cond.dist = "std", trace=FALSE)
summary(garch.model9)
garch.model9@fit$matcoef
res9 = residuals(garch.model9)
res9_std = res9 / garch.model9@sigma.t
par(mfrow=c(2,3))
plot(res9, type="l")
acf(res9,lag=100)
acf(res9^2,lag=100)
plot(res9_std, type="l")
acf(res9_std,lag=100)
acf(res9_std^2,lag=100)

# ARMA(1,1)GARCH(1,0) modeling --------------------------------------------

garch.model10 = garchFit(formula= ~arma(1,1)  + garch(1,0),
res, cond.dist = "std", trace=FALSE)
summary(garch.model10)
garch.model10@fit$matcoef
res10 = residuals(garch.model10)
res10_std = res10 / garch.model10@sigma.t
par(mfrow=c(2,3))
plot(res10, type="l")
acf(res10,lag=100)
acf(res10^2,lag=100)
plot(res10_std, type="l")
acf(res10_std,lag=100)
acf(res10_std^2,lag=100)

# refitting the model -----------------------------------------------------

fit5 <- lm(formula = t ~ -1 + t1 + t2 + t3 + t4 + t5 + t18 + t24 +
```

```
t48 + t72 + t168 + dmnd + d_sat + d_sun,
weights = 1/garch.model5@sigma.t^2)
summary(fit5)
accuracy(fit5)
preds5 <- predict(fit5)

# Plotting the Predicted values ----------------------------------------

price1 <- price[(maxlag):length(price)]

par(mfrow=c(1,1))
plot(price1, xlim=c(1001,1024),ylim=c(0,300),
xaxp=c(0,length(preds5),length(preds5)/1),
type="l", xlab="Time (hour)", ylab="Price (TL/MWh)", lty=2)
lines(preds5, col="blue")
abline(h=mean(price1), col = "gray10", lty = 3)
legend("topright",c("Actual Price","Predicted Price"),
lty=c(2,1), col=c("black","blue"),lwd=c(1.5,2.5),
cex=0.85, inset=0.02,text.width=3, bty="o", xjust=0, seg.len=2)


plot(price1, xlim=c(1001,1168),ylim=c(0,300),
xaxp=c(0,length(preds5),length(preds5)/24),
type="l", xlab="Time (hour)", ylab="Price (TL/MWh)", lty=2)
lines(preds5, col="blue")
abline(h=mean(price1), col = "gray10", lty = 3)
legend("topright",c("Actual Price","Predicted Price"),
lty=c(2,1), col=c("black","blue"),lwd=c(1.5,2.5),
cex=0.85, inset=0.02,text.width=20, bty="o", xjust=0, seg.len=2)

xaxp1 <- c(0,length(preds5)-length(preds5)%%168,
(length(preds5)-length(preds5)%%168)/168)
plot(price1, xlim=c(1001,1672),ylim=c(0,300),
```

```
xaxp=xaxp1, type="l", xlab="Time (hour)", ylab="Price (TL/MWh)", lty=2)
lines(preds5, col="blue")
abline(h=mean(price1), col = "gray10", lty = 3)
legend("topright",c("Actual Price","Predicted Price"),
lty=c(2,1), col=c("black","blue"),lwd=c(1.5,2.5),
cex=0.85, inset=0.02, text.width=70,bty="o", xjust=0, seg.len=2)

# Forecasting Demand ------------------------------------------------------

# subsetting the demand into training and test sets
demand <- data[,4]
demand.t <- demand[1:17544]
# using ets() method
demand.fit.ets <- ets(demand.t, model="ZZZ", damped=NULL, alpha=NULL,
beta=NULL, gamma=NULL,
phi=NULL, additive.only=FALSE, lambda=NULL,
lower=c(rep(0.0001,3), 0.8), upper=c(rep(0.9999,3),0.98),
opt.crit=c("lik","amse","mse","sigma","mae"), nmse=3,
bounds=c("both","usual","admissible"),
ic=c("aic","aicc","bic"), restrict=TRUE)
f.demand.ets <- forecast.ets(demand.fit.ets,h=24*1)
plot(f.demand.ets, xlim=c(17520,17544+24*1), xaxp=c(17520,17520+24*2,2),
ylab="Demand (MWh)", xlab="Time (hour)")
abline(v=17544, lty=2)
lines(demand, lty=2, type="l")

# using auto.arima() method
demand.fit.autoarima <- auto.arima(demand.t, d=NA, D=NA, max.p=24, max.q=24,
max.P=2, max.Q=2, max.order=25, start.p=2, start.q=2,
start.P=1, start.Q=1, stationary=FALSE, seasonal=TRUE,
ic=c("aicc","aic", "bic"), stepwise=FALSE, trace=FALSE,
approximation=(length(demand.t)>100 | frequency(demand.t)>12), xreg=NULL,
test=c("kpss","adf","pp"), seasonal.test=c("ocsb","ch"),
```

```
allowdrift=TRUE, lambda=NULL, parallel=TRUE, num.cores=NULL)
f.demand.fit.autoarima <- forecast(demand.fit.autoarima,h=24)
length(f.demand.fit.autoarima)
length(f.demand.fit.autoarima$mean)
accuracy(f.demand.fit.autoarima$mean, demand[17545:17568])
cor(f.demand.fit.autoarima$mean, demand[17545:17568])
plot(f.demand.fit.autoarima,
xlim=c(17520,17544+24*1), xaxp=c(17520,17520+24*2,2),
ylab="Demand (MWh)", xlab="Time (hour)")
abline(v=17544, lty=2)
lines(demand, lty=2, type="l")

summary(demand.fit.autoarima)
accuracy(demand.fit.autoarima)


old.par <- par()


plot(f.demand.fit.autoarima2, xlim=c(17544,17544+24*7),
xaxp=c(17544,17544+24*7,7),
ylab="Demand (MWh)", xlab="Time (hour)")
abline(v=17544, lty=6)
lines(demand, lty=2, type="l")
legend("bottomright",legend=c("Actual Demand", "Forecasted Demand"),
col=c("black","blue"),lty=c(2,1),lwd=c(1,2),ncol=2,xpd=NA,bty="n",inset=0)


f.demand.fit.autoarima2 <- forecast(demand.fit.autoarima,h=24*7)

# using double seasonal holt-winter method
plot(demand.t,type="l")
demand.t[demand.t==0] <- mean(demand.t) # replace zero in demand with mean of demand
```

```
f.demand.holt <- dshw(demand.t,period1=24,period2=168)
names(f.demand.holt)
# f.demand.holt$model
# c(17544,17544+336,14)

plot(ts(f.demand.holt$mean), ylim=c(15000,40000), xaxp=c(0,336,14),lwd=2,
col="blue", ylab="Demand (MWh)", xlab="Time (hour)")
lines(ts(demand[17545:(17544+336)]),lty=2)
legend("bottomright",legend=c("Actual Demand", "Forecasted Demand"),
col=c("black","blue"),lty=c(2,1),lwd=c(1,2),ncol=2,xpd=NA,bty="n",inset=0)



accuracy(f.demand.holt)
accuracy(f.demand.holt$mean[1:24], demand[(17544+1):(17544+24)])
# Forecasting price -------------------------------------------------------

# fist we need to replace the demand values with its forecast

new.x <- data.frame(cbind(as.ts(price),lag(price,k=-1),lag(price,k=-2),
lag(price,k=-3), lag(price,k=-4),lag(price,k=-5),
lag(price,k=-18),lag(price,k=-24), lag(price,k=-36),
lag(price,k=-48),lag(price,k=-72),lag(price,k=-168),
lag(price,k=-maxlag),lag(mins,k=-24), lag(means,k=-24),
lag(demand,k=-maxlag),lag(d.mon,k=-maxlag),
lag(d.tue,k=-maxlag),lag(d.wed,k=-maxlag),lag(d.thu,k=-maxlag),
lag(d.fri,k=-maxlag),lag(d.sat,k=-maxlag),lag(d.sun,k=-maxlag)))
colnames(new.x) <- c("t","t1","t2","t3","t4","t5","t18","t24","t36","t48",
"t72","t168","t336","pmins","pmeans","dmnd"
,"d_mon","d_tue","d_wed","d_thu","d_fri","d_sat","d_sun")
new.xs <- new.x[maxlag+1:(length(price)-2*maxlag),]
length(fitted(f.demand.fit.autoarima))
length(dmnd)
```

```
# use auto.arima forecast for demand and predict price
new.xs$dmnd[17545:17568] <- f.demand.fit.autoarima$mean
length(fitted(f.demand.fit.autoarima))
new.xxs <-new.xs[(17544+1):(17544+24),]
# predicting 24 step ahead forecast after replacing demand with its forecast
f.price <- predict(fit5,24,newdata=new.xxs)
f.price$fit
summary(f.price)
accuracy(f.price$fit, new.xxs$t)

# use double seasonal holt winter forecast for demand and predict price
new.xs$dmnd[17545:17568] <- f.demand.holt$mean[1:24]
new.xxs <-new.xs[(17544+1):(17544+24),]
# predicting 24 step ahead forecast after replacing demand with its forecast
f.price2 <- predict(fit5,24,newdata=new.xxs)
f.price2$fit
summary(f.price2)
accuracy(f.price2$fit, new.xxs$t)
# comparing accuracy of the two forecasts for price
accuracy(f.price$fit, f.price2$fit)

# Plotting price forecast for one day ahead ----------------------------------------
# use demand forecast by auto.arima
plot(f.price2$fit, ylim=c(0,300), xaxp = c(0,24,24),
main="Price forecast for one day ahead",
type="l", col="blue", xlab="Time (hour)", ylab="Price (TL/MWh)", lty=1)
lines(new.xxs$t, lty=2)
#lines(f.price$fit, lty=5, col="red")
legend("topright",c("Forecast Price","Actual Price"),
lty=c(1,2), col=c("blue","black"),lwd=c(2.5,2.5),
cex=0.85, inset=0.02,bty="o", xjust=0, seg.len=2)
```