KADIR HAS UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

MODELS FOR LONG-TERM ELECTRICITY PRICE FORECASTING FOR TURKISH ELECTRICITY MARKET

SİRUN ÖZÇELİK

SEPTEMBER, 2017

Sirun Özçelik M.S. Thesis 2017

MODELS FOR LONG-TERM ELECTRICITY PRICE FORECASTING FOR TURKISH ELECTRICITY MARKET

SİRUN ÖZÇELİK

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science
in
Industrial Engineering

KADIR HAS UNIVERSITY
2017

KADIR HAS UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

MODELS FOR LONG-TERM ELECTRICITY PRICE FORECASTING FOR TURKISH ELECTRICITY MARKET

SİRUN ÖZÇELİK

APPROVED BY:

Asst. Prof. Dr. Emre Çelebi (Kadir Has University)

Asst. Prof. Dr. Gökhan Kirkil (Kadir Has University)

Prof. Dr. Ayşe H. Bilge (Kadir Has University)

Asst. Prof. Dr. Burak Barutçu (Istanbul Technical University)

APPROVAL DATE: 13/09/2017

"I Sirun ÖZÇELİK, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in this thesis."

Sirun ÖZÇELİK

MODELS FOR LONG-TERM ELECTRICITY PRICE FORECASTING FOR TURKISH ELECTRICITY MARKET

Abstract

In this study, we have developed models for long-term electricity price forecasts for Turkish electricity market using multiple regression and time series forecasting methods. For the regression models, we have firstly obtained the monthly data for demand weighted average of market-clearing electricity price (dependent variable), electricity demand, hydro power production, wind power production, and population as well as yearly gross domestic product (GDP) and human development index (HDI) as independent variables for Turkey between December 2009 and September 2016 from the market operator's transparency database and other data sources. Secondly, we have examined the effect of each of these independent variables on market-clearing electricity price and then, by using time-series models, long-term forecasts are obtained for all independent variables. Finally, multiple-linear regression models are used to obtain forecasts for the monthly demand weighted average of electricity prices. In addition to multiple regression models, several time series models such as exponential moving average (Holt-Winters model), seasonal autoregressive integrated moving average (SARIMA) and Artificial Neural Network (ANN) models are also developed. In setting up forecasting models, R statistical packages and forecast tools as well as MATLAB (for ANN) are used. Long-term forecasts are made for the next 24 months starting from October 2016. Model results are evaluated according to mean absolute percentage error (MAPE), mean square error (MSE) and mean error (ME), which are commonly used error measures for evaluating forecasting results. We have found that on average around 8% of MAPE can be achieved through ANN method. This study would be useful for producers' investment decisions as well as market operator's long-term policy decisions.

Keywords: Long-term electricity price forecasting, multiple regression, time series models, Turkish electricity market

TÜRKIYE ELEKTRIK PIYASASI ICIN UZUN DONEM ELEKTRIK FIYAT TAHMIN MODELLERI

Özet

Bu çalışmada, Türkiye elektrik piyasasındaki uzun dönem fiyat tahminlerini belirlemek için çoklu regresyon ve zaman serileri modelleri geliştirildi. Regresyon modeli için Aralık 2009- Eylül 2016 tarihleri arasında elektrik piyasası takas fiyatının aylık ortalamalarını (bağımlı değişken), elektrik talebini, hidroelektrik üretimini, rüzgar enerjisi üretimini, Gayri Safi Yurtici Hasila(GDP) ve Insani Gelişim Endeksini(HDI) gösteren bağımsız değişken verileri EPIAŞ(Enerji Piyasaları İşletme Anonim Şirketi) Şeffaflık Platformu ve diğer veri kaynaklarından sağlandı. Regresyon modellerini kullanarak elektrik takas fiyatındaki bu bağımsız değişkenlerin etkileri incelendi. Ek olarak zaman serileri modelini kullanarak, bütün bağımsız değişkenler için uzun vadeli tahminler elde edildi. Çoklu Regresyon modeline ek olarak, üssel hareketli ortalama(Holt Winters), SARIMA (mevsimsel birleştirilmiş otoregresif hareketli ortalama modeli) ve yapay sinir ağları modelinden de yararlanıldı.
Tahmin modellerini kurarken, R istatistik paketleri ve tahmin araçları kullanılırken yapay sinir ağları modelinde ise MATLAB’den yararlanıldı. Ekim 2016'dan 24 ay sonrasına orta vadeli tahminleri yapıldı. Tahminlerdeki hata ölçümlerini hesaplamak için ortalama mutlak yüzde hata, ortalama hata kare ve mutlak hata formülüne göre modellerin sonuçları değerlendirildi. Bu çalışmanın sonucunda ANN yöntemi ile ortalama %8 mutlak yüzde hata gözlemlendi. Bu çalışmayla piyasadaki operatörlerin uzun dönem politika kararlarının yanı sıra üreticilerin yatırım kararları için de faydalı olacaktır.

Anahtar Kelimeler : Uzun dönem elektrik fiyat tahmini, çoklu regresyon, zaman serisi modelleri, Türkiye elektrik piyasası.

Acknowledgements

There are many people who helped to make my years at the graduate school most valuable. First, I thank Asst. Prof. Dr. Emre Çelebi and Asst. Prof. Dr. Gökhan Kirkil, my major professors and dissertation supervisors. Having the opportunity to work with them over the years was intellectually rewarding and fulfilling. They have always encouraged me and recommended to trust myself. They have always shown me the right ways to reach success in my research.

Many thanks to Asst. Prof. Dr. Atilla ÖZMEN, who patiently answered many of my questions about MATLAB. I would also like to thank to my graduate student colleagues who helped me all through the years full of class work and exams. My special thanks go to Elif and Şeyma whose friendship I deeply value.

The last words of thanks go to my family. I thank my parents for their patience and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

Electricity is an asset, which has special characteristics and can be sold, bought or traded. From the 1990s to deregulated period, most of the developed and developing countries deal with electricity generation and trading through organized electricity market. These markets have some special characteristics that are explained below:

1. Unlike other goods, electricity cannot be stored economically. Although electricity can be stored in batteries or by pumping water to storage, these methods are usually expensive, difficult and have limited capacity.
2. Electricity flows through transmission and distribution lines follows Kirchhoff's Laws (i.e., current and voltage laws). Kirchhoff's current law predicates that the algebraic sum of currents entering and leaving any point in a circuit is equal to zero. Second law supports that the algebraic sum of all voltages around a closed loop equals zero [1]. With respect to these two main laws, any inflow (outflow) to (from) the electricity system has an impact on the overall system flows.
3. Instant supply (generation) and demand (load) balance is required in electricity systems.

The necessity for price forecasting model is driven out of the typical characteristic of the power sector which includes long term capital investment and evident characteristics of electricity compared to other commodities. Governments would prefer predictability as electricity is directly linked to the economic development; customers prefer hedging from price fluctuations; and investors prefer confidence in the markets. Hence, the objective of price forecasting model is to project future electricity market dynamics to assist in present decision making process [2].

In the last two decades, with deregulation and introduction of competition in Turkish electricity market, a new challenge has emerged for power market participants. Utmost price volatility, that can be even twice as much higher than other assets, has forced generation firms and wholesale consumers to resist not only against volume risk but also against price changes.

Hence, price forecasts have become a key input for market participants’ strategy development and decision-making. In order to realize the intended investments, companies (fund seekers) require project loans from banks and/or financial institutions (funders). Every energy project has its own dynamics; however, project funders have certain expectations from fund seekers, e.g., having reliable and sustainable cash flows that can pay for interest and capital [3].

In order to achieve project funder’s objective, electricity load and price forecasting in medium and long term is very important. Therefore, the main goal of long-term price forecasting is investment profitability analysis and planning, such as determining the future generation expansions [2].

1.1 Capacity

OECD member countries’ capacity increase rate is given between year 1974-2000 and 2000-2014 in Figure 1. Highest increase is recorded for Solar PV followed by wind after 2000. There is no increase in nuclear power share in the 2000-2014 period. Hydropower and combustible fuel have increased by 1.5% on average [4].

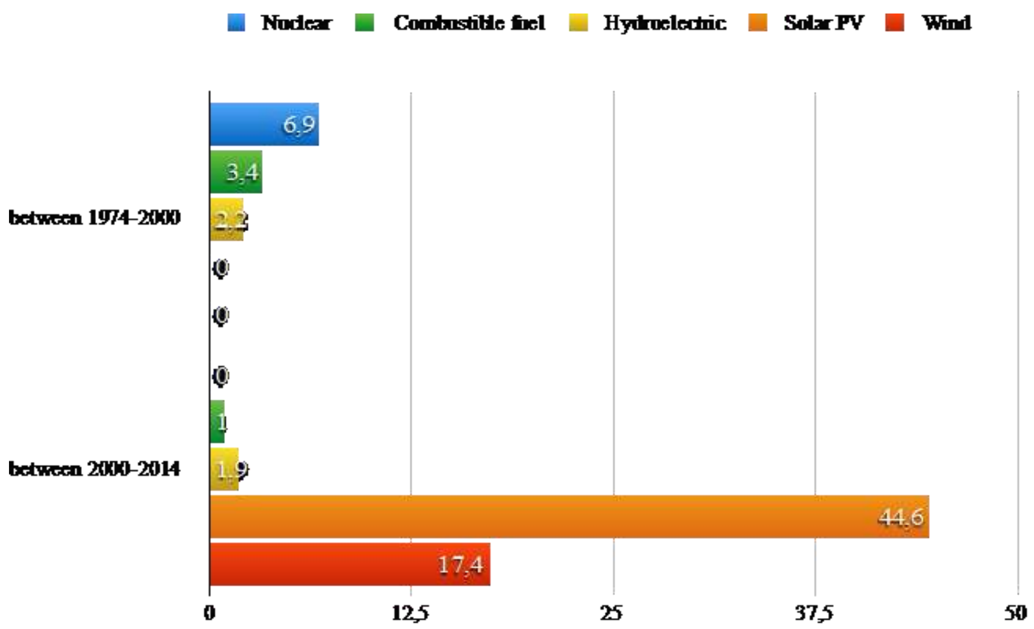

Figure 1. Average annual capacity increase rate between 1974-2000 and 2000-2014 [4]

1.2 Consumption

Since 1974 combined share of total electricity consumption in OECD member countries increased 15% on average. On the other hand, electricity consumption in industry decreased 17% on average during this 40- year period [4].

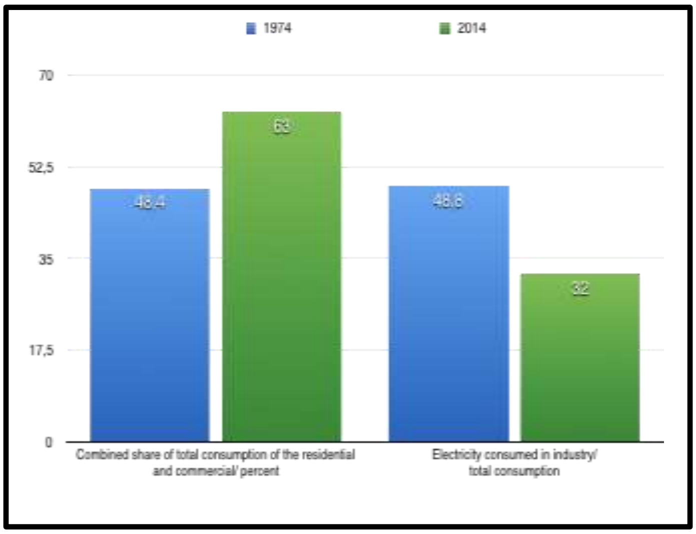

Figure 2. Electricity consumption shares in industry, residential and commercial in OECD countries [4]

Non-OECD countries’ share of world electricity final consumption increased from 26.0% in 1973 to 53.0% in 2014.

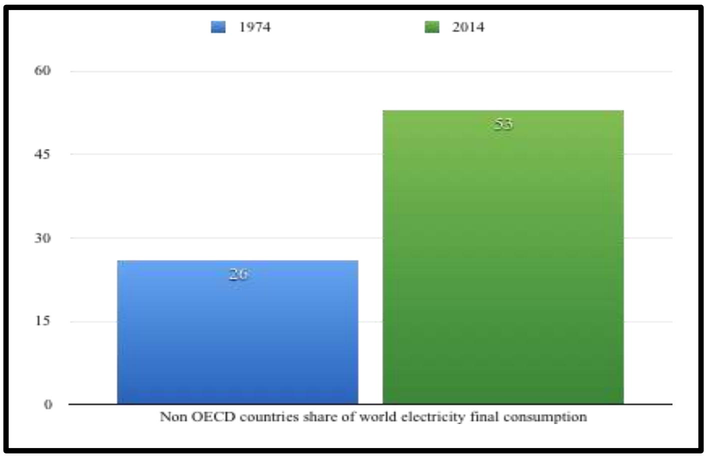

Figure 3. Non-OECD countries share of world electricity final consumption [4]

1.3 Investment

In 2016, worldwide energy investments are about 1.7 trillion dollars, which is equal to 2.2% of GDP and 10% of global gross capital formation. Due to a decrease in oil and gas investments in 2016, the total investment was 12% lower compared to 2015. From 2014 to 2016, the total energy investment worldwide dropped by 36%, but the oil and gas sector had a share of two fifths of total investment in this sector. On the other hand, electricity has the largest share of total energy investments by 40%. Examining Figure 4 and Figure 5 in details reveals that the investments in network assets reached 40% of electricity investments by an increase of 6% from 2015 [5].

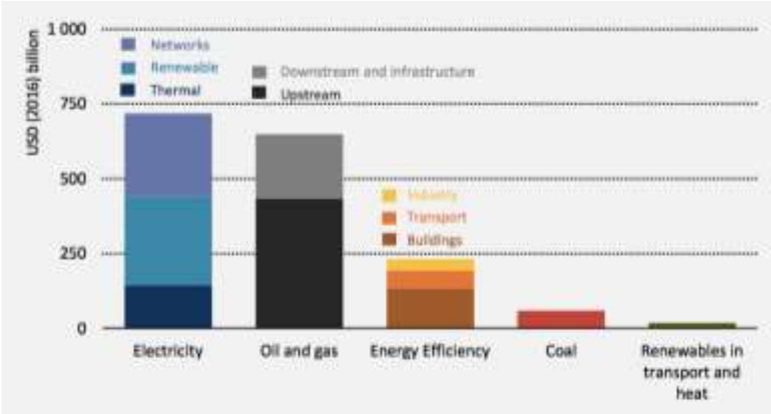

Figure 4. Global energy investment in 2016 [5]

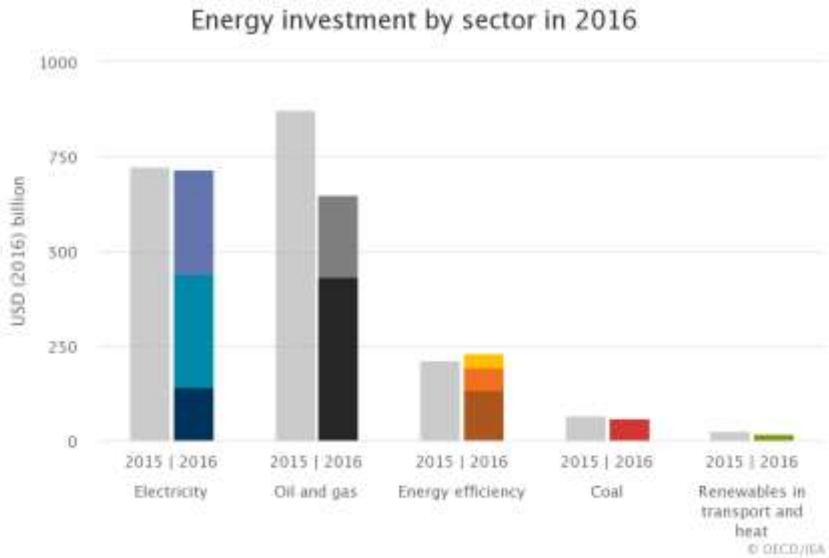

Figure 5. Comparison of energy investments in 2015- 2016

1.3.1 Electricity and Renewables Investment

Investment in generation of power worldwide decreased by 5% to 441 billion dollars in the electricity sector. Because of several factors, dispatchable plants have suffered. The reasons are uncertain demand, robust solar photovoltaic (PV) and wind additions (especially in China), policies to counter local air pollution (especially in the case of coal plants). Furthermore, the trend in investments will be downwards due to less large-scale dispatchable generation capacity investment plans for hydropower and others in the world electricity sector [5].

Global electricity investments have decreased by just under 1% to $718 billion, with an increase in spending on networks. New renewables-based power capacity investment, at $297 billion, stayed the largest area of electricity spending, despite a fall by 3%. Global investment in energy supply between 200-2016 has shown in Figure 6 [5]. Renewables investment was 3% lower than five years ago, but capacity additions were 50% higher and expected output from this capacity is about 35% higher, thanks to declines in unit costs and technology improvements in solar PV and wind [5].

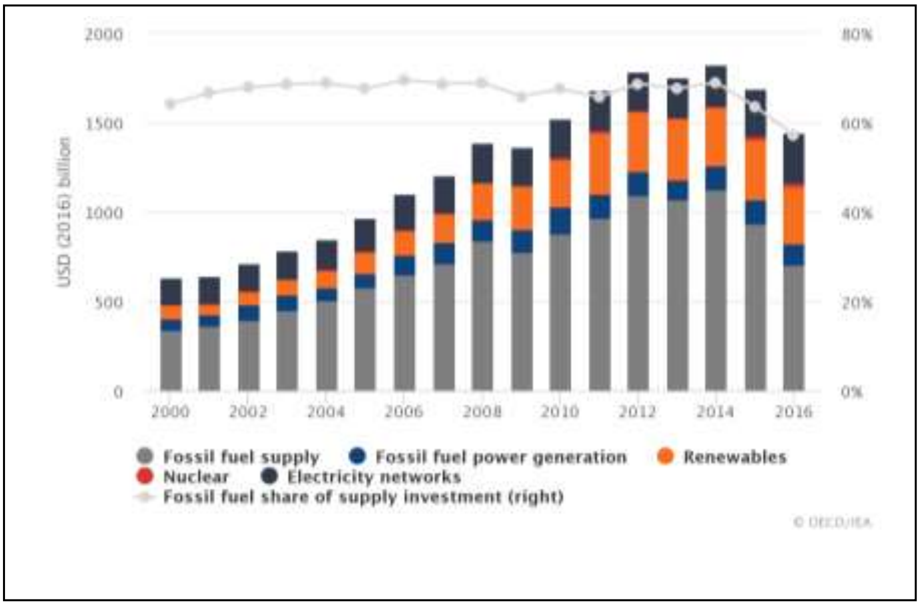

Figure 6. Global power sector investment [5]

1.3.2 Regional trends in investment

Although the investments structure of China has been changing, it remained the largest area of energy investment, which is 21% of the global total (Figure 7). In 2016, there was 25% decline in commissioning of new coal-fired power plants. China increases its investments in low-carbon electricity supply and networks as well as in energy efficiency [5].

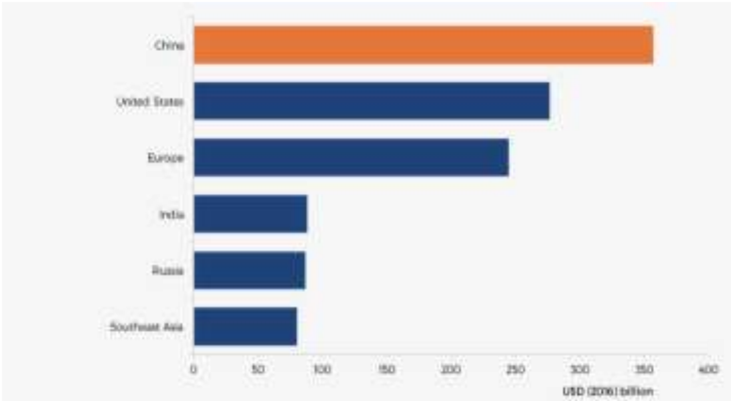

Figure 7. World energy investment by region [5]

India increased its energy investments by 7% and becomes the third-largest country after the United States (U.S.). The country got this position by a strong government push to improve its power system and increase electricity supply access.

The proportion of the U.S. in energy investment worldwide rose to 16%, while there has been a sharp decline in gas and oil investments. The U.S. still has higher share in investments than Europe where a decline of 10% is observed due to renewable investments [5].

1.3.3 Investments in Turkey

Turkey requires great investments in energy infrastructure, particularly in electricity and natural gas to match its citizens' ever growing energy demand and sustain economic growth. Since the majority of the consumption is gathered around its 3 big cities (İstanbul, Ankara, İzmir) that are quite far away from the energy resources – located in either far east or far west in Turkey, Turkish electricity sector encounters certain challenges. It is estimated that at least USD 260 billion is necessary in the energy industry by 2030, around two-thirds before 2023 (OME, 2014).

Competitive market and legal framework to be set up by the government are of great significance to make this industry appealing to investors [6].

There has been a sharp increase in electricity generation of Turkey for several decades (Figure 8) [6]. In 2015, 259,7 terawatt-hours (TWh) amount of electricity is generated, which stands for a record for Turkey. In Figure 8, it can be observed that there is an increase by 33.3% from 2009 to 2015. 67.7% of total generation in 2015 is from fossil fuels, which have the highest proportion compared to other sources [6]. However, it is 11.3% lower than 2014 levels due to higher contribution from renewable sources. For electricity, natural gas is the major fuel (38.6%) followed by coal (28.3%) [6].

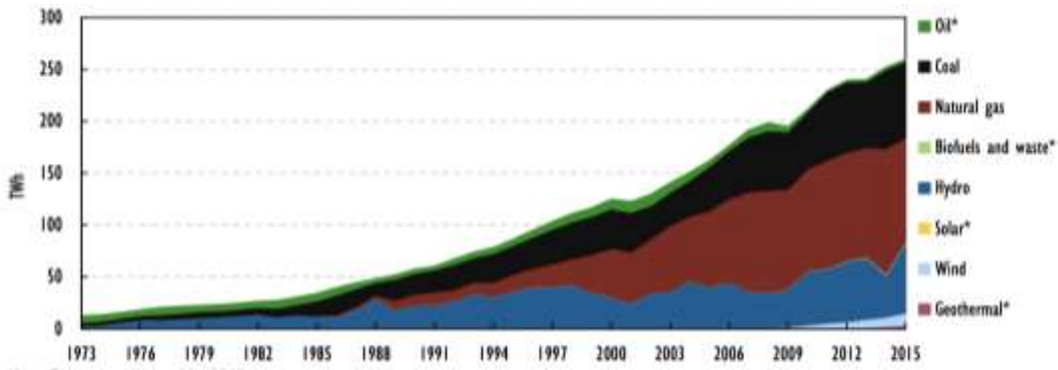

Figure 8. Electricity generation by source, 1973-2015 [6]

The electricity generation mix in Turkey can change year by year, owing to the seasonality of hydro supply and unavailability of old lignite plants [6]. In 2015, hydropower production is at a record high of 66.9 TWh, 64.6% higher than in 2014. Hydropower production averaged 47.1 TWh for the period 2005-15, or 22% of total generation [6]. The share of hydro and natural gas in total generation has been volatile. Since 2005, the use of coal in power generation has increased, while electricity production from oil declined by 60.1% [6]. Natural gas has increased its share from 45.3% in 2005 to 47.9% in 2014 [6]. However, in 2015 the share of natural gas fell to 38.6% of total, while the share of coal in the electricity mix increased from 26.7% to 28.3% [6].

Although Turkey's big potential led by its interconnections with neighboring countries, trade volumes are quite low. In 2015, for instance, net electricity imports coming from neighboring countries to Turkey was about 4.4 TWh, or put it differently, 2.1% of all electricity supply in the country [6]. Yet, imports have been in an increasing trend since 2010 and reached to 7.4 TWh, with exports of 3 TWh in 2015. A great deal of the increase in imports has been after 2011 when the trial connection to the EU electricity grid has been made [6].

Around 66.7% of imports is from Bulgaria in 2014 and the rest is from Iran (28.3%), Georgia (3.7%), Azerbaijan (1.3%) and Greece (0.05%), whilst electricity is exported mainly to Greece (70.8%) and Iraq (29.1%) [6].

Turkey's electricity consumption reached 207.4 TWh in 2014, a record high. Since 2004, consumption only contracted once by 3.1% in 2009, after seven years of steady growth. Electricity consumption by sector has illustrated in Figure 9. In 2014, consumption was 71.2% higher than in 2004. At the same time, annual instantaneous peak demand in 2014 reached a historic high of 41 gigawatts (GW), 198% up from in 2004, higher than in any previous year since 2009 [6].

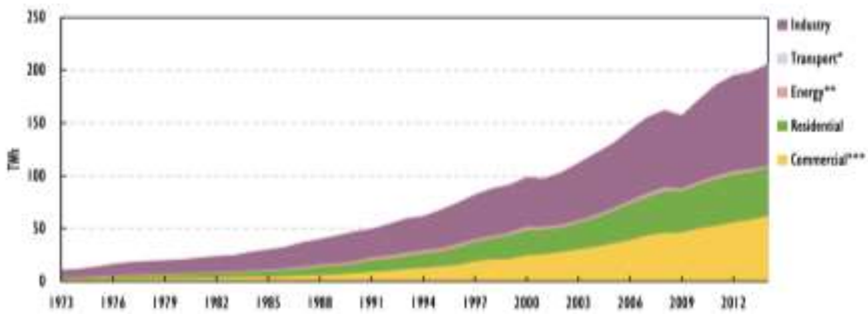

Figure 9. Electricity consumption by sector 1973-2014 [6]

Industry is the largest final consumer of electricity, accounting for 46.2% of total consumption. Industry demand has experienced a 65% growth over the past decade. During the economic crisis, industry demand contracted by 5.6% during 2009 but rebounded in 2010 with a 12.8% jump [6]. The industry share in total demand contracted slightly from 47.9% in 2004 as total demand grew slightly faster [6].

The sector of commercial and public services (including agriculture) and the residential sector accounted for 30.1% and 22.3% of demand in 2014, respectively [6]. Demand in both sectors increased faster than the total (effectively being the main drivers of demand growth), up by 88.1% and 67.2% compared to 2004. Consequently, they have gained a larger share of total demand, up from 27.4% for commercial and 22.8% for residential. The energy sector (including coal mining, oil and gas extraction and refining) consumed 0.03% of total electricity demand; transport consumed a negligible 0.4% [6].

During 2008-14, investment in generation has risen, and the global macroeconomics parameters and financial crisis in 2008-09 resulted in a decrease in energy demand, particularly in the manufacturing industries that are electric intensive [6].

With the peak demand of 41 GW and an installed capacity of 70 GW, the Turkish supply/demand capacity margin in 2014 was around 69%, forming the peak of a period of high investment and oversupply in the Turkish electricity market [6]. This is a positive change from periods when the margins had averaged around 15% before 2008 [6].

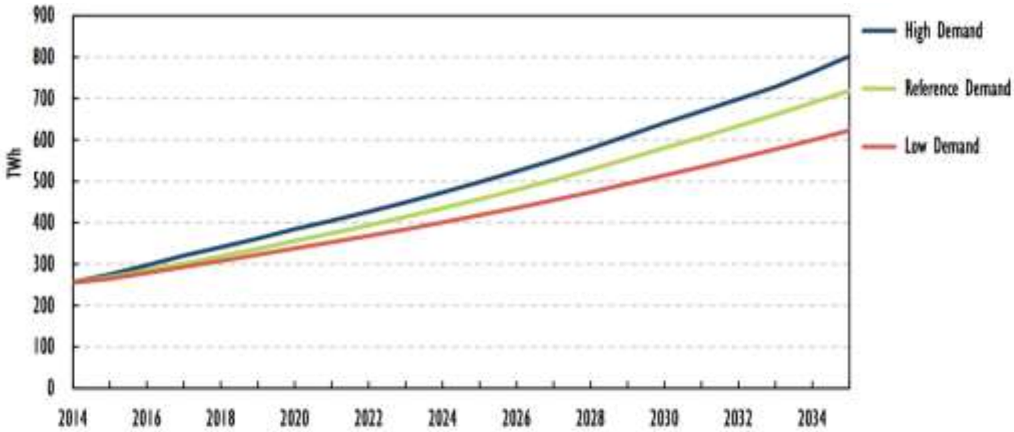

Figure 10. Electricity demand projection [6]

Demand projection scenarios are represented by IEA in Figure 10. IEA supports that in the medium term, overcapacity will occur in the Turkish market. The short-term system operation needs to arouse more interest to the management of peaks in demand and the less availability of plants in the network [6].

Analyzing the future demand/supply adequacy of the Turkish power system, Ministry of Energy and Natural Resources (MENR) estimates electricity demand to rise more than three times over the next 20 years, according to the “Electricity Demand Projection 2014-2035” report [6]. The government’s reference demand scenario foresees power demand growth to gear up and to come up with 581 TWh in 2030, a 127% increase over year 2014. The government ́s projections for demand foresee an annual increase by around 5.5% until 2023 when demand could reach 450 TWh [6]. In 2035, total demand will reach 719 TWh (with the highest forecast at 802

TWh and the lowest at 622 TWh, as shown in Figure 10), mainly driven by industrialisation and urbanisation along with population growth. Electricity demand is set to increase annually by 6.7% (low-case scenario) or 7.5% (high-case scenario) until 2020, according to the government projections [6].

1.4 A brief history of Turkish electricity market

In 1923, the Economic Congress that has been held in Izmir led to studies about electricity in Turkey. During those years, in the state of Turkey Republic, there were only three cities using electricity, which are Adapazarı, Istanbul and Tarsus.

Established by the state in 1935, Mineral Research and Exploration (MTA), Etibank, Electrical Power Resources Survey and Development Administration (EIE) with the opening of the State Hydraulic Works (DSİ) and the Iller Bank in the following years, work on electricity infrastructure has become more regular and planned [7].

On 15.07.1970, when the Turkish Electricity Authority Law No. 1312 entered into force, The Turkish Electricity Authority (TEK) was established in order to produce, transmit, distribute and trade the electricity required by the country in accordance with the government's general energy and economic policies. Thanks to this law all municipalities and Iller Bank are connected to TEK [8].

With the law number 3096, which entered into force in 1984, the electricity private sector road has been introduced. The model called Build - Operate - Transfer (BOT), that we hear frequently today, has begun to be used with this law. The purpose of the enactment of the law is to eliminate the monopoly of the TEK by providing the opportunity to devolve the private sector of the operating rights of the existing public electricity facilities.

As a result of Law No. 3096, ten companies were identified to distribute electricity generation and transmission in their locations. General overview of Turkish electricity market is shown in Figure 11.

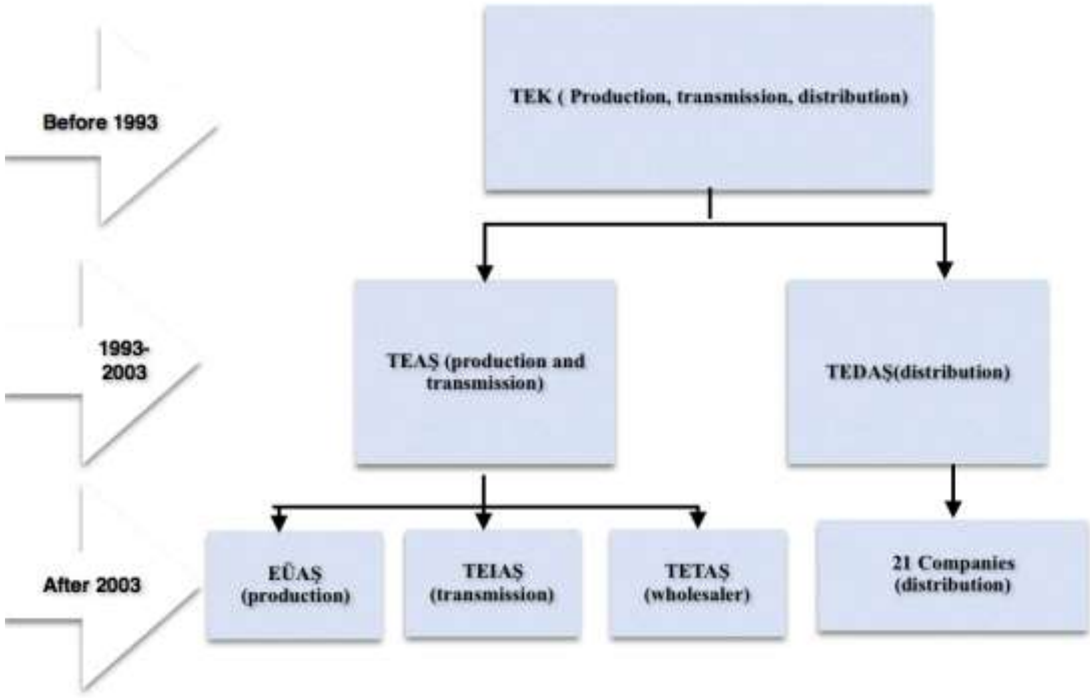

Figure 11. Turkish Electricity Market Overview [9]

The increasing use of electricity in those years has increased the work load of the Turkish Electricity Authority (TEK), which has undertaken the production, transmission and distribution of electricity, and has caused great difficulties in management [6]. According to scope of TEK by a decree, which was issued before, Turkish Electricity Corporation (TEAS) and Turkish Electricity Distribution Corporation (TEDAŞ) are founded. TEAŞ's duty responsibilities were "Electricity Generation and Transmission", TEDAŞ's duty responsibilities were "Power Distribution" as the name implies [6].

2001 is one of the most important years in which radical decisions about electricity are taken. First of all, with the decision of the Council of Ministers, TEAS was broken into four companies; Turkish Electricity Transmission Company (TEIAS), Turkish Electricity Trading and Contracting Incorporation Company (TETAŞ) and Electricity Generation Corporation (EÜAŞ) [6]. With this new structure, as its name implies, TETAS will deal with "Electricity Wholesale", TEİAŞ will deal with "Electricity Transmission" and EÜAŞ will deal with "Electricity Generation [6].

With the Electricity Market Law is published in 2001, TEAŞ has left its place to EÜAŞ, TEİAŞ and TETAS, which are newly structured [6]. With the help of TEDAS, electricity trading has prepared the ground for long-term investment incentives and has been carried out with bilateral agreements in a system that targets supply security [6].

On 3 March 2003, the electricity market was opened. Large customers are directly tied to the transmission system. Customers who consume more than 3.6 MWh per year (as of January 2016) are counted as eligible customers. [6] These type of customers have the right to choose their own suppliers. EMRA (Electricity Market Regulatory Authority) aimed to reduce these limits to increase the competitive market structure. In 2016, eligible limit value was 3,600 while it was 9,000,000 in 2002 [6].

As a result, distribution regions have been reestablished and Turkey has been divided into 21 distribution regions. The distribution network of Turkey is divided into 21 distribution regions taking into account the geographical proximity, administrative structure, energy demand and other technical / financial factors [6]. The PPA (Prime ministry privatization administration) has established a distribution company in each of the 20 regions where TEDAŞ owns the TEDAŞ privatization program [6].

The first Balancing and Settlement Regulation was published in November 2004 and the day-ahead balancing system targeting production optimization in the Turkish Electricity Market was introduced [10]. This system aims to simplify the management of real-time balance and improve system security and reliability. The reconciliation carried out under this regulation is between 06.00 - 17.00 day-time, between 17.00 - 22.00 peak and between 22.00 - 06.00 night-time [10].

Between 2003 and 2007, the increase of electricity demand was bigger than investment in new generating capacity [6]. After reserve margins decreased to the level of 5%, High Planning Council affirmed the 2nd Electricity Market and Security of Supply Strategy in 2009, which was a crucial reform in electricity market. For privatization and full market opening by 2015, the Strategy set out demonstrative goals for the use of energy resources in generation of electricity by 2023; at least a

30% share of renewables, using full potential of lignite and hydro, and beginning of nuclear energy [6].

With a new balancing and settlement regulation published in 2009, a reconciliation of the market structure to an unrestricted reconciliation was carried out on every hour. This market, which is carried out under the name of day ahead planning, can be regarded as a transition process.

On 30 March 2013, thanks to the Parliament, the New Electricity Market Law (EML, No. 6446) comprised the reform of Turkey's electricity market which assigned the rights and necessities of all participants that are associated with electricity generation, transmission, distribution, wholesale and retail sale, import, export and market operation activities [6]. On the other hand, transmission operator TEİAŞ informs EMRA available capacity for the next 5-10 years [6].

After unsuccessful results in the electricity sector to increase private investment, the government decided to privatize and restructure the distribution sector (TEDAŞ) [6]. In 2015, all distribution firms are passed to private parties. In addition, EMRA regulated these companies to reduce losses and theft. By this privatization, Turkish government got around USD 12.75 billion. Between 2008 and 2012, private sector's share in total installed capacity increased to 68% in 2014 thanks to privatization [6].

After the EML No. 6446 provided the operation of an organized wholesale electricity market and the financial settlement activities with EPIAŞ as new market operator, the Turkish electricity wholesale market has seen a major reform [6].
EPIAŞ was established in March 2015. Next, day-ahead and intraday market are passed to EPIAŞ from Electricity Market Financial Settlement Centre (PMUM). TEİAŞ operates the balancing power market (BPM) [6]. In the BPM, market participants submit capacity for both up/down regulation that can be realized within maximum 15 minutes [6].

Chapter 2

2 Electricity Price Forecasting

2.1 Literature Review

Forecasting is an essential part of decision making under uncertainty. A need for forecasting arises only when there is uncertainty about the future and some aspects of the future cannot be controlled [11]. The literature on short-term price forecasting is vast, however, for long-term forecasts, there are limited number of studies. There are several review papers that have been published in the last two decades.

Reference [12] has studied the short-term (daily) price and load forecasting methods in competitive power markets. Conventional time series method, neural network and combination of different forecast models are utilized in order to forecast the demand side. At the price forecast side, simulated artificial agents model was used.

Fundamentals of electricity pricing and forecasting are reviewed in [13] and artificial neural network (ANN) based price forecasting methods are also introduced. In this book, market power analysis based on game theory are introduced. Several game theory problems are handled.

Similarly, short-term price and load forecasting models are surveyed in [14]. In this article, forecasts for hourly CalPX (California electricity market) market clearing prices for both normal and highly volatile days are considered. Especially statistical methods including ARMA, ARMAX, GARCH, p-ARX, TAR and regime switching models are considered along with quantitative models and derivatives valuation. After model applications, forecast errors are evaluated for each model and p-ARX model has provided the best results. However, for days with price spikes, TAR type models have provided the best results.

In [15], the authors have pointed out the requirements for short-term price forecasts and reviewed challenges related to electricity price forecasting (EPF). In this study, Spanish and Californian markets are examined. The models considered are ARIMA, GARCH and genetic algorithm based on neural network. Although, they have justified the use of artificial intelligence, hybrid approaches turned out to be more beneficial.

In [16], the authors have briefly reviewed short term EPF methods and focused on artificial intelligence-based methods, in particular feature selection techniques and hybrid forecast engines. They have also discussed forecast error measures, the fine-tuning of model parameters, and price spike predictions.

Point and interval forecasting using exponential smoothing methods are examined in [17, 18]. In [17], the author provides information about early history of exponential smoothing methods. In the following sections, formulation of exponential smoothing model and some equivalent models are treated. At the last part of the research, model selection, model fitting and empirical studies are explained. Reference [18] is an intensive book about forecasting with exponential smoothing.

In [19], a method called THETA is used for forecasting daily or monthly electricity prices. In this study, main impression was changing the time series' local curvature through a coefficient 'Theta' which is the second difference of the data. In this study, THETA model is ideal for microeconomic data and monthly series.

There are several studies on EPF methods in which regression models have been used. Wavelet decomposition method coupled with multiple regressions is used in [20]. In this paper, the regression coefficients are calculated using the wavelet decomposition and the forecasted next days' system marginal price.

Another application of short term EPF study is performed in [21] using hourly PJM data. In this study, time series analysis, neural network and wavelets are studied. After model implementations, time series techniques are found to be more preferable than neural network or wavelet transform models. Dynamic regression models and transfer function algorithms are the most powerful models among time series models.

Additionally, in [22], the authors have studied on general seasonal periodic regression models with ARIMA, ARFIMA and GARCH disturbances for the analysis of daily electricity spot prices. Several methods are applied for European electricity markets such as Nordpool (in Norway), EEX (in Germany), Powernext (in France) and APX (in Netherlands).

Three complementarity modeling procedures are studied in [23]. Firstly, economic basics, strategic and market design effects on daily prices has been mentioned. Then, residual volatility is assigned to some significant fundamentals. After all evaluations, it is found that if all of the volatility sources are clarified, GARCH effects decreases in short term EPF forecasting.

Reference [24] has suggested a general model, which has contained both AR and MA components and explicitly included differencing in the formulation.

While making forecasts in short-term EPF, [25] utilized variants of AR (1) and general ARMA processes (including ARMA with jumps) in the German Leipzig Power Exchange market. In this paper, univariate time series performances are compared. The study presented that i) an hourly modeling strategy for electricity spot-prices significantly improves the forecasting power of linear univariate time-series models, and ii) evaluating the process of arrival of price spikes, even if it is in a simple behavior, can also provide better forecasts.

In another study, [26], the authors have used different auto regression schemes for modeling and forecasting short term spot prices in the California and Nordic market. Since California market is freely accessible and includes extreme price spikes it is a good idea to choose this market. Secondly Nordic market has less volatile prices and provides most of the electricity generation from hydro power plants. In this research twelve time series models are applied for two different markets with including several different conditions. After evaluations SNAR/SNARX semi parametric models have performed better than the other ones.

In [27], AR/ARX, GARCH, TAR/TARX and Markov regime has been applied in order to forecast spot prices in California Power Exchange System. In this paper, some special applications of linear autoregressive time series models with additional fundamental variables are used. The study is utilized for hourly electricity system price for California power market.

On the other hand, [28] has proposed Markov regime switching model with long memory for the Nord Pool area price forecasts in hourly data. In this empirical study, a new regime switching characterization with a potentially deterministic state is mentioned, and observable regimes appear to be really important to reach correct electricity price forecasts.

In [29], the authors have described an attractive methodology that combines elements of time series and multi-agent modeling in the Iberian power market. In this research, Naïve, CV (conjectural variation) ARIMA and Price ARIMA models are implemented in order to find best short-term forecasting results. For this 24-hour working market, CV ARIMA model has presented the most accurate results.

In [30], there is a new Hybrid Intelligence System (HIS) proposed for short-term EPF. This new model consists of NN, RCGA, cross validation, repetitive training, and archiving. New HIS has been investigated on the Spanish electricity market and compared with ARIMA, Wavelet-ARIMA, GARCH and FNN models.

On the other hand, in [31], support vector regression model (SVR) and ARIMA model are combined in order to make the most accurate analysis for short term EPF. This new model is called SVRARIMA, which utilizes nonlinear pattern via SVR and residual analysis via ARIMA model. In this study, mainly highlighted point is that the two most known and individually good working models sometimes can not work well together. While making predictions, it is crucial to make good assumptions and combinations.

Several methods applied in [32] to short term EPF in Spanish market. In this research, ARIMA, Holt Winters, regime switching, dynamic regression models utilized in order to reach most accurate results. Wind generation that is included in dynamic regression model has performed well, since wind has been the most significant factor in Spanish electricity market.

Another hybrid model in short term EPF is studied in [33]. After wavelet transform, ARIMA and GARCH models are utilized in Spanish and PJM electricity market. After evaluations, ARIMA model with GARCH error components has performed better than the other models.

In another comprehensive study [34], three best-known electricity price-forecasting methods (casual models, stochastic time series models and artificial based models) are applied. They have concluded that because of the limited data more studies should be done.

In another reference book [35], the authors give general information about modeling and forecasting electricity loads and prices. For long-term (annual time scale) EPF, [36] suggested an algorithm which is based on some pre-specified rules that switches between the predictions of different models (neural networks, fuzzy regressions and a standard regression). After implementations to Iran power market, fuzzy regression models gave best results. Author has mentioned that, applied and selected models are flexible and appropriate for long term electricity price forecasting.

2.2 Electricity Price Forecasting Procedure

Various methods have been proposed for electricity price forecasting. In the literature, there are a lot of short-term price forecasting studies proposed, however at the medium and long term there are not so much illustrations introduced. There is a significant gap occurred between short and medium-long term electricity price forecasting applications. This study has been served for medium-long term electricity price forecasting implementations improvement.

All of the Forecasting methods should have some criteria;

- Forecasting method should be adaptive.
- Forecasting method should be flexible.
- Forecasting method must be constructive.

  Factors-causes and factors-results should be clearly detached in forecasting method, i.e. forecasting method can’t be contradictory.

2.2.1 Electricity Price Forecasting Flowchart

Electricity price forecast application progress is shown in Figure 12 [37]. The process of time series based forecasting is depicted in the flow chart. Input data is usually initial step of the process of forecasting [37]. Past market prices, wind generation, hydro generation, HDI, GDP and population are some major input data for the price forecasting. Basic statistical analysis on the input data set and later some required transformations give idea about coming steps, which are model selection and validation. Selection and design of forecasting models/techniques is directly related with the scope of forecast and the accuracy of results required [37]. Optimization of the parameters of models is applied to check the models’ performance.

The model validation is completed after optimization of the parameters step. The process of validation is repeated until reaching satisfactory result. The model is utilized to do the forecast when the validation is successful [37].

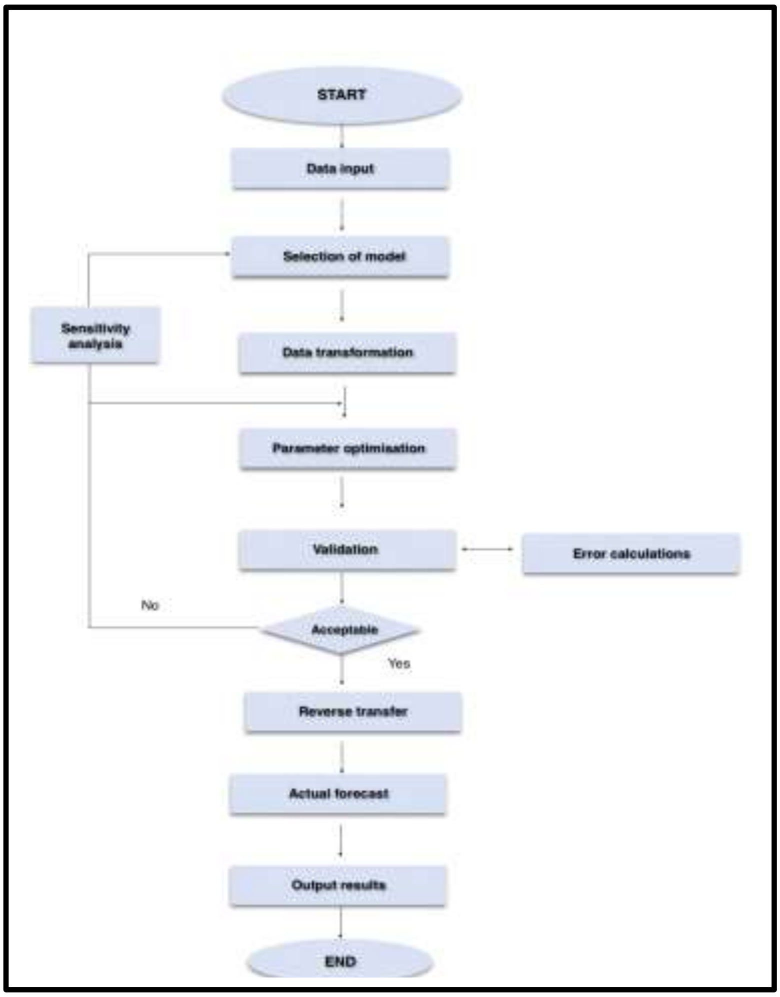

Figure 12. Electricity price forecasting flowchart

2.3 Planning Horizons

Electricity price forecasting (EPF) can be considered by all electricity market players. In the electricity market, aim of the electricity producers is to optimize the value of their portfolio during a certain period of time, while observing the risk within specific limits. Electricity market price should be determined depending on planned horizons, which can be divided into two main horizon; short-term forecasting, and medium-long term forecasting as shown in Figure 13.

In the short term, market players must set up bids for the spot market. In the medium term, they have to define contract policies, and in the long term, expansion plans are defined.

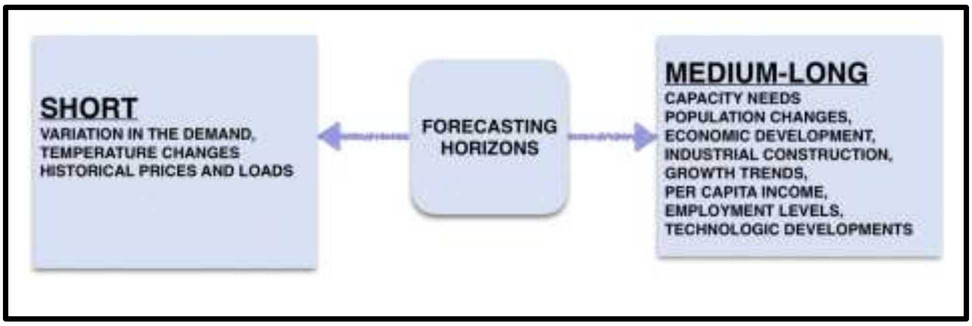

Figure 13. Planning Horizons of EPF

2.3.1 Short-Term Horizon

In the short-term, a producer needs to forecast electricity prices to derive its bidding strategy in the pool and to optimally schedule its electric energy resources [21]. Short-term forecasts include the period from a few minutes to about one week ahead. Short term trading means that short-term variations in load and the actual prices are only known after matching of bids and offers by the market operator [38]. The short-term planning horizon includes daily planning and weekly planning (e.g., 24 hours up to one week). The main uncertainties can be variation in the demand, weather conditions and water inflow for hydro power dams.

2.3.2 Medium-term and Long-term Horizon

The medium- term planning includes seasonally and yearly planning. The medium-term forecasting is appropriate for 3 months up to 12 months and the long-term planning is applicable from 2 years up to 30 years. The medium- and long-term forecasts considers the historical load and weather data, the number of customers, the appliances in the area and their characteristics including age, the economic and demographic data and their forecasts, the appliance sales data, and other factors [39]. Uncertainty on the inflows and demand exceptionally influence long term planning for electricity price forecasting, there are a lot of exogenous factors, which are affecting prices. In this circumstance, hydropower production, wind power production, GDP, population, weather conditions and many other agents should be considered. Long-term strategic decisions can involve investments such as buying or building a power plant, entering a new market, grid expansion, etc. but can also involve de-investments.

Chapter 3

3 Methodology

EPF models have been shown in Figure 14 [40]. In the literature, electricity price forecasting models are divided into three; namely game theory models, time series models and simulation models.

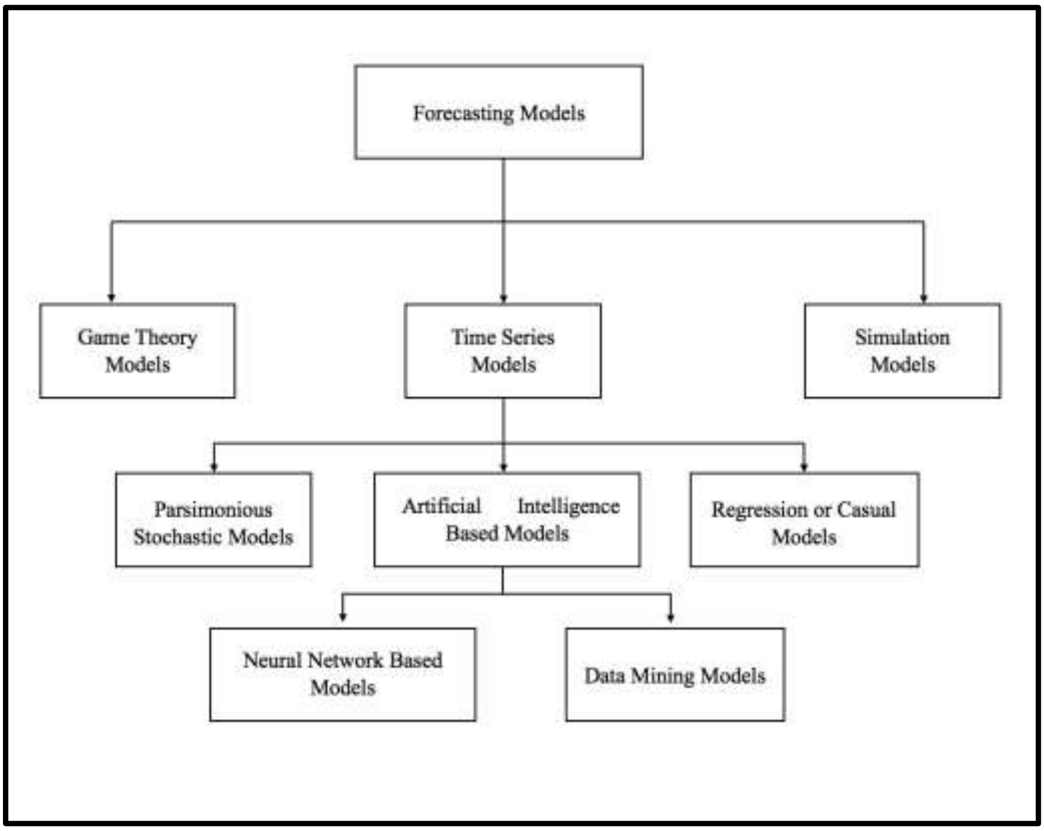

Figure 14. Electricity Price Forecasting Model

3.1 Game Theory Models

Game theory is a natural platform for market competition, which is between the producers for selling the electricity, and maximizing their own profits can be modeled as a game, in which electricity price will be estimated from the result of the game [41].

The models are utilized by the market operators for deciding the market strategies. In this group of models, equilibrium models, take the analysis of strategic market equilibrium.

There are several equilibrium models available like Nash equilibrium [42], Cournot model [43], Bertrand model [44] and supply function equilibrium model [45] are main models used for the electricity market based on the level of competition in the market [46].

The Game Theory models provide;

- Estimating prices in different equilibrium
- Estimated market shares of the agents
- Calculation of the best strategy for a particular agent
- Calculation of the ideal response to an agent
- Changes in market environment
- Information of the forecast of elements that characterize and maps pricing

3.2 Simulation Models

In the simulation approach, one creates a model of the system to calculate the price of the electricity based on the production cost using an optimal distribution of the load [47]. The simulation methods which are currently being used by the electric power industry range from the bubble-diagram type contract path models to production simulation models with full electrical representation, such as GE-MAPS software. There are popular simulation models such as MAPS (multi area production simulation) which has been developed and stands for market assessment and portfolio strategies, UPLAN software which is used to forecast electricity prices and to simulate the participants' behavior in the energy and other electricity markets like ancillary service market and emission allowance market [34].

Simulation models provides;

- Imitating the actual power flows in the system
- Simulating generator dispatch patterns over an extended period of times.
- Imitating the actual dispatch with system operating requirements and constraints
- High accuracy in calculations

3.3 Time Series Models

Time series models are utilized in many areas such as process control, economic forecasting, financing, marketing, population studies and biomedical science. Time series analysis is a method of forecasting which focuses on the past behavior of the dependent variable. Time series analysis prefers to understand the characteristics of a physical system that creates the time series and systematic approaches to extract information. There are various approaches to deal with time series analysis including dynamic model building and performing correlations. Most time series in practice are generally non-stationary [48].

There are three types of models, which can be listed as;

- Parsimonious stochastic models
- Regression or causal models
- Artificial intelligence (AI) models
    - ANN based models
    - Data-mining models

3.3.1 Parsimonious Stochastic Models

Stochastic time series can be divided into stationary process and non-stationary process. A stationary model preceded by data preprocessing can be used for tackling price modeling. In the short term, assuming that market agents do not change their strategies during these shorter periods [34].The term stationary time series is used to denote a time series whose statistical properties are independent of time [49]. In particular, this means that the process generating the data has a constant mean. The variability of the time series is constant over time. This argument cannot be applied when it comes to modeling large time series, often characterized by discrete changes in the agents' strategies. The major drawback of using stationary models is that non-stationary should be removed before adjusting the models. This is a nontrivial process when working with electricity price time series.

Non-stationary models can be classified into two groups: single global models, where a unique model is proposed to cope with all series data, and switching models, which are first concerned with identifying different arrangement in the time series and then with adjusting a different local model to each one.

Switching models introduce a new perspective to deal with non-stationary processes. Instead of considering a single global model, switching models are concerned with adjusting several local models for the different time series system. Switching models consists of econometric models and ANN-based models [34].
There are discrete time counterparts corresponding to the continuous-time stochastic models such as;

- Autoregressive (AR)
- Moving average (MA)
- Autoregressive integrated moving average (ARIMA)
- Generalized autoregressive conditional heteroskedastic (GARCH)

3.3.1.1 The Autoregressive (AR) Model

Autoregressive (AR) model specifies that the output variable depend linearly on its own previous values and on a stochastic term (an imperfectly predictable term). Autoregressive (AR) models of a time series can be used to forecast the value z_t of a time series based on a series of previous values z_{t-1}, z_{t-2}... z_{t-p}. An AR model can simple be defined as:

$$Z_t = C + \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varepsilon_t \quad (1)$$

$\emptyset_1, \emptyset_2, \emptyset_p$: Coefficient

ε_t: Forecast error

C: constant

Thus the model is in the form of a stochastic difference equation. An order of p and the current value “Z” depends on or related to previous values. The above equation can also be written equivalently as:

$$Z_t = C + \sum_{i=1}^{p} \emptyset_i Z_{t-i} + \varepsilon_t \quad (2)$$

3.3.1.2 The Moving-Average (MA) Model

Moving Average (MA) is one of the techniques used in the analysis of univariate time series. It is found by taking the average of sub sequences.

$$Z_t = \varepsilon_t + \sum_{j=1}^{q} Q_j \varepsilon_{t-j} \quad (3)$$

Q_j: Forecast of the time series for period t+1

ε_t : Actual value of the time series in period t

3.3.1.3 ARIMA Models

Autoregressive integrated moving-average (ARIMA) models are formed by combining AR and MA models [50, 51, 52]. ARIMA process is studied where the system load has been taken as the only exogenous variable.

In the ARIMA analysis, an identified underlying process is generated based on observations to a time series for generating a good model that shows the process-generating mechanism precisely [24]. The ARIMA technique includes identification [53], estimation [53], and diagnostic checking [53, 54]. Statisticians George Box and Gwilym Jenkins developed systematic methods for applying them to business & economic data in the 1970’s [24]. When there is no missing data in the within the time series which is stationary the ARIMA forecasting technique can be preferred.

If data has a stable or consistent pattern with a minimum number of outliers, ARIMA model is the best alternative.

Many statistical software packages can be used to construct the ARIMA model. Before applying ARIMA model, first autocorrelation (acf) and partial autocorrelation (pacf) functions should be determined.

Acf and pacf provide a statistical summary at a particular lag. Acf is a plot of the autocorrelation of a time series data values by varying time lags [35]. Pacf at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags [55].

The maximum number of lags is determined simply by dividing the number of observations by 4, for a series with less than 240 based on Box and Jenkins method. According to Box and Jenkins methods, the lag number is calculated as 20 where the number of observations in this study is 82. Autocorrelation and partial autocorrelation graphs, which provides information about the AR and MA orders, are then drawn, based on the specified lag number. Autoregressive (AR) process order is determined from the partial autocorrelation graph and similarly MA process order is determined from the autocorrelation graph.

3.3.1.3.1 Nonseasonal ARIMA Models

Table 1. Owerview of ARIMA models

ARIMA(p,d,q) forecasting equation	Explanation
ARIMA(1,0,0)	first-order autoregressive model
ARIMA(0,1,0)	random walk
ARIMA(1,1,0)	differenced first-order autoregressive model
ARIMA(0,1,1)	without constant simple exponential smoothing
ARIMA(0,1,1)	with constant simple exponential smoothing with growth
ARIMA(0,2,1) or (0,2,2)	without constant linear exponential smoothing
ARIMA(1,1,2)	with constant damped-trend linear exponential smoothing

A nonseasonal ARIMA model is classified as an "ARIMA (p, d, q) " model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationary, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The model may also include a constant term.

A non-seasonal ARIMA model can be written as:

$$(1-\emptyset_1 B-\cdots-\emptyset_p B^p)\nabla^d L_t = (1+\theta_1 B+\cdots+\theta_q B^q)\varepsilon_t \quad (4)$$

Where $\nabla x_t = (1-B)x_t$ is lag 1 differencing operator. $\emptyset_1 \ldots \emptyset_p$ is autoregressive parameters. B is the backward shift operator and $\theta_1 \ldots \theta_p$ is moving average parameters. L_t is the original data set.

3.3.1.3.2. Seasonal ARIMA Models (SARIMA)

If there is seasonal component in the ARIMA model, the model is called as the SARIMA. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models, and it is written as follows:

ARIMA *(p, d, q)*$(P, D, Q)_m$	
(Non-seasonal part of the model)	(Seasonal part of the model)
where m= number of periods per season	

ARIMA (p, d, q) × (P, D, Q)m notation, where:

p = non-seasonal AR order,

d = non-seasonal differencing,

q = non-seasonal MA order,

P = seasonal AR order,

D = seasonal differencing,

Q = seasonal MA order, and

m = time span of repeating seasonal pattern.

3.3.1.3.3. Auto.arima function in R statistical program

The auto.arima() function in R combines unit root tests, minimization of the AICc (bias-corrected Akaike's Information Criterion) and MLE (Maximum likelihood estimation) to obtain an ARIMA model. The algorithm follows these steps [56]:

1. With using repeated KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test, number of differencing d is determined.
2. After differencing the data d times, p and q values are chosen by minimizing the AICc. Instead of taking into account every possible p and q combination, the algorithm utilizes a stepwise search to traverse the model space.
   a) According to smallest AICc, the best model is selected from the following four:

      - ARIMA (2,d,2)
      - ARIMA (0,d,0)
      - ARIMA (1,d,0)
      - ARIMA (0,d,1)

   If d=0 then the constant c is included if d>=1 then the constant c is set to zero. This is called the current model.

   b) Variations on the current model are applied:
      - Change p and/or q from the current model by +-1;
      - Exclude and/or include c from the current model.

      New current model is the best model considered until this step.
   c) Until no lower AICc can be found Step 2(b) is repeated.

In ARIMA $c = \mu = 0$ when $d > 0$ which provides an estimate of μ when $d = 0$. μ is called the "intercept" in the R output. In auto.arima function included drift which allows $\mu \neq 0$ when $d = 1$.

3.3.1.3.4. Model selection criteria

The most famous information criteria are represented in the literature where FPE (Akaike's Final Prediction Error), AIC (Akaike's Information Criterion), AICc (bias-corrected Akaike's Information Criterion), BIC (Bayesian Information Criterion) and HQ(Hannan-QuinnCriterion) [61].

There is a small-sample (second-order bias correction) version of AIC called AICc, In a AR model the minimum AICc suggests to fit AR(p) process to the model residuals successively. The model fits well if the information criterion reaches its minimum value for p = 0 [61].

Identification of the model can be done by looking at Acf and Pacf plots or an automated iterative process. This process is formed by fitting several different possible model structures and using a goodness-of-fit statistic or information criterion to select the best model. Generally, in order to get an artificial improvement in fit, increasing the complexity of model structure could be the best idea, which can be done by increasing the number of parameters in the applied model [35].

Akaike's Final Prediction Error (FPE), bias-corrected Akaike's Information Criterion (AICC) and Bayesian Information Criterion (BIC; also known as Schwarz Information Criterion, SIC) are three of the most popular goodness-of-fit statistics:

$$FPE = V\frac{n+d}{n-d} \qquad (5)$$

$$AIC_i = -2\log L_i + 2V_i \qquad (6)$$

$$AICc = -2\log\mathcal{L} + \frac{2dn}{n-d-1} \qquad (7)$$

$$BIC = -2\log\mathcal{L} + d\log n \qquad (8)$$

where $V = \frac{1}{n}\sum\widehat{\varepsilon}_t^2$ is the variance of model residuals $\hat{\varepsilon}_t = L_t - \hat{L}_t$ where n is the sample size, d is the model size and $\log\mathcal{L}$ is the log-likelihood function and the data L = $(L_1, \dots, L_n)'$ are observations of a stationary Gaussian time series (L' denotes a transpose of the vector L).The best fit model is the one with the minimum value of information criterion. In the literature there is a prevalent mistake in the model selection in the use of AIC when AICc really should be used. Because in practice if n gets larger AICc converges to AIC.

3.3.1.4 Holt-Winters Exponential smoothing

The Holt-Winters exponential smoothing is used when the data exhibits both trend and seasonality. Holt-Winters models can be classified as additive model and multiplicative, which depends on the characteristics of the particular time series. Additive model for time series are including additive seasonality and multiplicative model for time series are exhibiting Multiplicative seasonality [58]. Holt-Winters method can be extended to deal with time series, which contain both trend and seasonal variations.

The component form for the additive Holt-Winters model is:

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+} \quad (9)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (10)$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \quad (11)$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \quad (12)$$

Where observed time series denoted by $y_1, y_2 \cdots y_n$ Forecast of y_{t+h}, h period ahead forecast is respresented by $\hat{y}_{t+h|t}$, m is the length of seasonality l_t represents the level of the series, b_t denotes the growth, s_t is the seasonal component. $h_m^+ = [(h-1) \bmod m] + 1$ is represents, estimation of the seasonal indices used for forecasting come from the final year of the sample.

3.3.1.4.1 Holt-Winters function in R statistical program

HoltWinters () function tries to find optimal smoothing parameters (α, β, γ) values by minimizing the average squared prediction errors, which is equivalent to minimizing in the case of additive errors.

3.3.2 Regression or Causal Models

Various problems in engineering and science consider finding the relationships between two or more variables. Regression analysis is a statistical technique that is very effective and beneficial for these types of problems. Regression type forecasting model consists of the theorized relationship between a dependent variable (electricity price in our study) and a number of independent variables that are known or can be estimated. Regression has wide range of applications including prediction and process control. In regression analysis, the aim is to model the dependent variable in the regression equation as a function of the independent variables, constants and an error term. The performance of the model depends on the estimate of the constants and coefficients. There are several types of regression models, which are Simple Linear Regression Models, Multiple Linear Regression Models, and Dynamic Regression Models.

3.3.2.1. Simple Linear Regression Models

Simple linear regression is a statistical method that can be summarize and study relationships between two continuous quantitative variables. The model concerns two-dimensional sample points with one independent variable and one dependent variable (conventionally, the *x* and *y* coordinates) and finds a linear function (a non-vertical straight line) that, predicts the dependent variable values as a function of the independent variables [59].A simple linear regression considers a single regressor or predictor *x* and a dependent or response variable *y*. Assuming the relationship between *y* and *x* is a straight line and that the observation y at each level of x is a random variable, the expected value of *y* for each value of *x* is:

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i \qquad (13)$$

$$i = 1, 2, 3\ldots,n$$

β_0 and β_1: unknown regression coefficients

ε_i : Random error.

when $\boldsymbol{x} = 0$ value of $\boldsymbol{y}$ is determined by the slope of the line is β_1 , and β_0 .

3.3.2.2. Multiple Linear Regression Models

If there are more than one independent regressor variables in a time series, the regression model can be called as multiple linear regression models. In general, the dependent variable or response y may be related to i independent or regressor variables. The general form of multiple regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i + \varepsilon \qquad (14)$$

β_0 .. β_i: unknown regression coefficients

ε_i : Random error.

The model attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. There is one variable to be forecasted and several predictor variables. Thanks to multiple linear regression analysis, how much will the dependent variable change when we change the independent variables can be analyzed. To estimate the parameters or regression coefficients the method of least squares can be used. In least squares method estimated the regression coefficients providing the best fit via minimization of the residual sum of squares [57]. Once the coefficients are estimated, the new value of the dependent variable can then be easily found [60].

3.3.2.1.1 Finding best regression subset in R statistical program

In R program there is a package which called ‘leaps’, that provides comparisons according to adjusted R square level. In this package regsubsets() function performs all subset regression, and chooses “nbest” model(s) for each number of predictors up to nvmax with respect to adjusted R square level .which.max(summary.out$adjr2) function gives the model with n variables (counting dummy variables separately) which has the highest adjusted R^2 . Variables marked with ‘TRUE’ are the ones should be chosen which shown as a result of summary.out$which[] function [61].

3.3.3 Artificial Intelligence (AI) Models

These may be considered as nonparametric models that map the input–output relationship without exploring the underlying process [34]. It is considered that AI models have the ability to learn complex and nonlinear relationships that are difficult to model with conventional models [48].

Artificial intelligence models can be divided into two categories which are artificial neural network (ANN) based models and data-mining models.

3.3.3.1 Artificial Neural Network

Artificial Neural Network (ANN) is inspired from working human brain cells (neurons). Therefore, artificial neural can learn by way of trial and can generalize information. Forecasting, is one of the most important area where artificial neural network used. Linear models are not suitable when problem is nonlinear and at this time ANN is the best alternative.

In neural network there are three types of units: input units, output units and hidden units. Input units are receive data from outside the neural network. Output units are send data out of the neural network and the hidden units whose input and output signals remain within the network [62]. Hidden layers numbers which can be determined with trial and error according to problem. Function of artificial neurons is similar with real neurons. When we have learnt a new information, a new way is formed between two neurons. In artificial neurons, a signal is sent from input neuron. Then it is multiplied with a weight and transferred to the output neuron.

In the neural network there is also an activation function $\mathcal{F}_k$ which takes the total input $s_k(t)$ and the current activation $y_k(t)$ and produces a new value of the activation of the unit k.

$$y_k(t+1) = \mathcal{F}_k(y_k(t), s_k(t)) \qquad (15)$$

There are some different threshold functions used according to models properties. These functions are sigmoid, hard limiting threshold function (sgn function), linear or semi-linear function and hyperbolic tangent functions [62]. In artificial neural network model there are two types of propagations. The first one is *feed forward* which makes data processing over multiple layers but there is no feedback connection. Second type of propagations are called as recurrent networks which implies feedback connections [62].

Training of artificial neural network is one of the most important parts. There are two learning situations, which called, supervised and unsupervised learning. In supervised learning input-output pairs provided by externally [62].

On the other hand, in unsupervised learning, output unit is trained to respond the cluster of pattern within the input [62].

In Figure 15, ANN models used artificial neural neurons has shown.

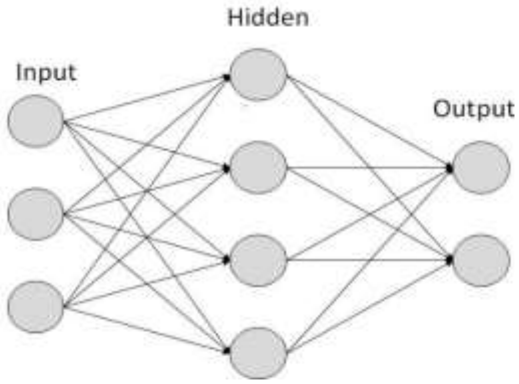

Figure 15. Artificial Neural Network model

.

Several researchers used ANN based models especially in short term price and load forecasting [63, 64, 65]

Chapter 4

4 Model Implementation & Results

In this study, four forecasting methods are used for long-term EPF for the upcoming 24 months (October 2016-September 2018) period. These methods are multiple regression (MR), Holt-Winters exponential smoothing (HW-ES), seasonal autoregressive integrated moving average (SARIMA) and artificial neural networks (ANN). In the subsequent sections, data sources, analyses and transformations are discussed and then each method is applied and analyzed in details.

4.1 Data Sources, Analyses and Transformations

Our main data source for the day-ahead average (weighted by demand) monthly electricity prices are from EPİAŞ between December 2009 and September 2016 (EPİAŞ Transparency Platform, 2016). Two alternative electricity prices are used in order to interpret the correlation between dependent (electricity price) and independent variables (electricity demand, hydro power production, wind production, population, human development index –HDI and gross domestic product –GDP): electricity prices in i) Turkish Lira (TL), and ii) U.S. Dollars (USD). Figure 16 and Figure 17 illustrate the time-series for monthly electricity prices in TL and USD, respectively.

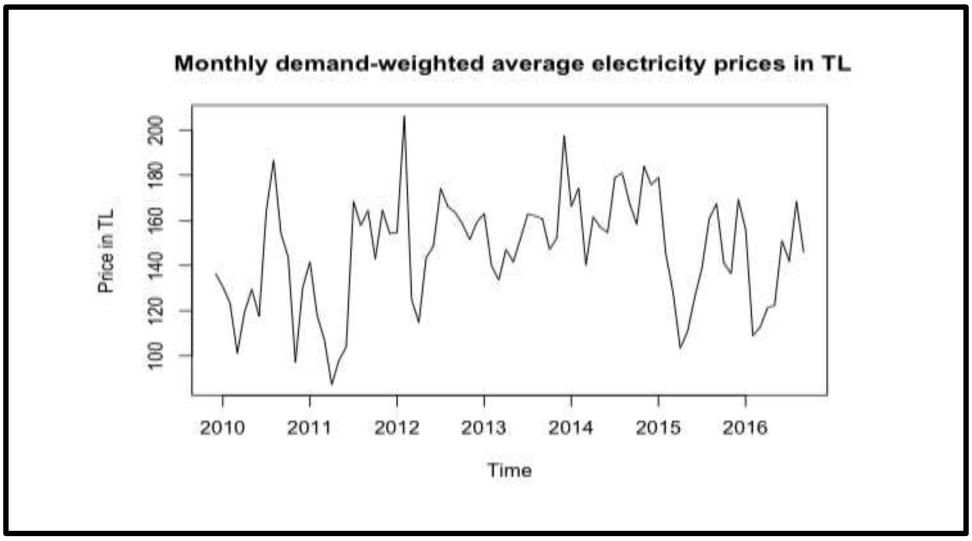

Figure 16. Monthly demand-weighted average electricity prices between December 2009 and September 2016 in TL

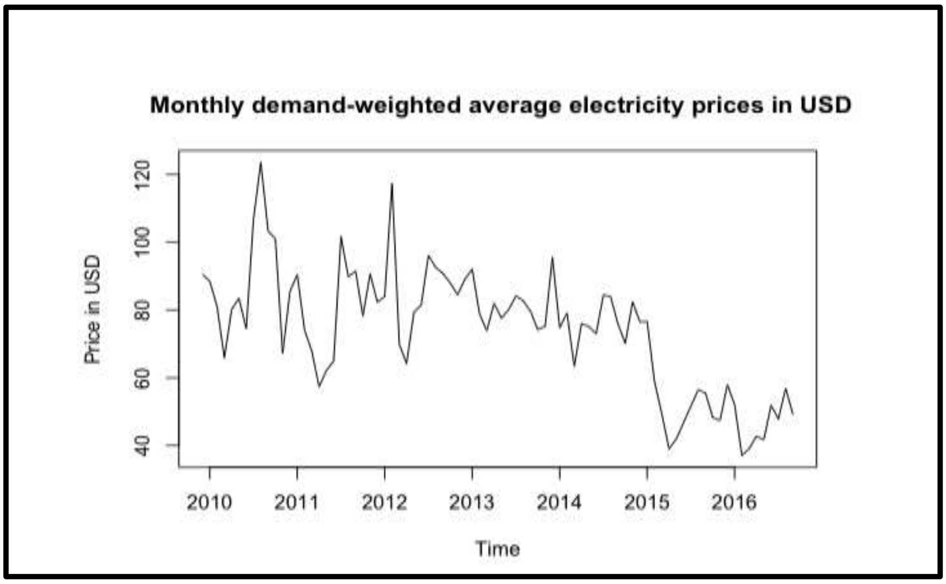

Figure 17. Monthly demand-weighted average electricity prices between December 2009 and September 2016 in USD

As observed in Figure 16, the electricity prices in TL fluctuates around a mean of 145 TL/MWh. However, due to fluctuations and an increase in USD/TL exchange rate in Turkey in recent years, the electricity price in USD in Figure 17 has a downward trend. In this study, monthly average electricity prices in USD is preferred, since it can reflect the real prices more accurately than nominal prices in TL.

.

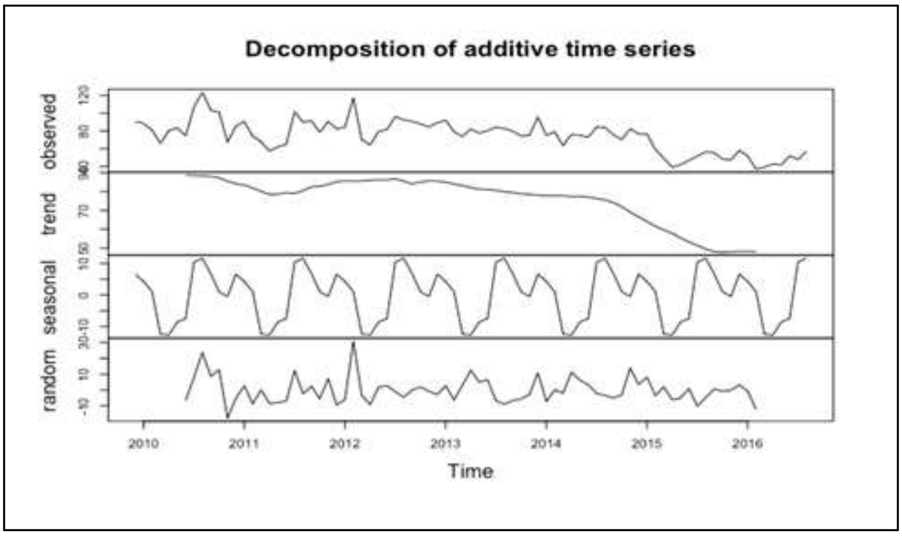

Figure 18. Decomposition of additive time series for electricity price in USD

With a more detailed analysis using additive time series decomposition of electricity prices in USD by trend, seasonality and random components are presented in Figure 18, and it is observed that prices have a clear downward trend. Also a high seasonality is observed for different months of the year, especially for spring and summer months. There are also high deviations observed in 2010 and 2012 in the random component of the decomposed data.

In this study, we have assumed that electricity prices in Turkish electricity market can be influenced by several factors including electricity demand, hydro and wind production, GDP, population and HDI. We have employed a backward elimination method for selecting independent variables for our multiple regression models.

Data sources for these independent variables are as follows. Electricity demand, hydro production and wind production data are monthly and from EPİAŞ Transparency Platform [66]; GDP and population data are yearly and from TURKSTAT [67]. HDI data are yearly and from human development reports [68]. HDI, GDP and population data are annual and in order to reach monthly values the data is interpolated (see Appendix A for the dataset).

Figure 19 shows the time-series data for the independent variables. Demand and wind production have similar time series, both having upward trend throughout the years. Hydro production has high seasonality and fluctuates during all years. HDI, GDP and population increases by each year.

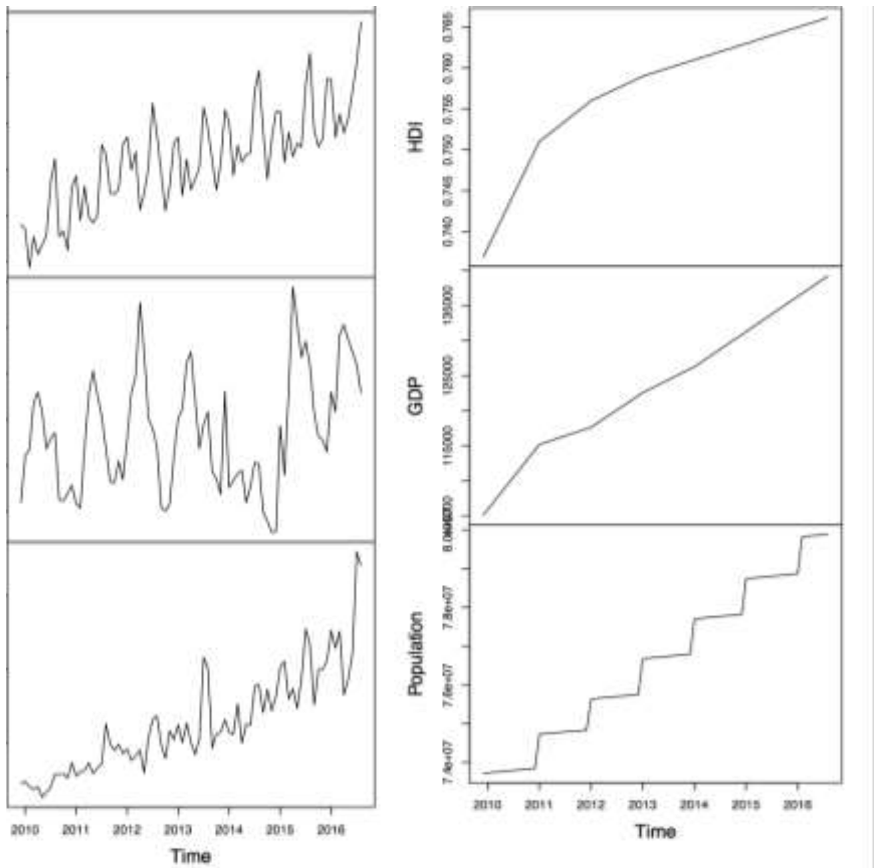

Figure 19. Independent variables between 2009 and 2016

Figure 20 depicts the scatterplot matrix between all variables and also a best-fit trend line is shown with a red line.

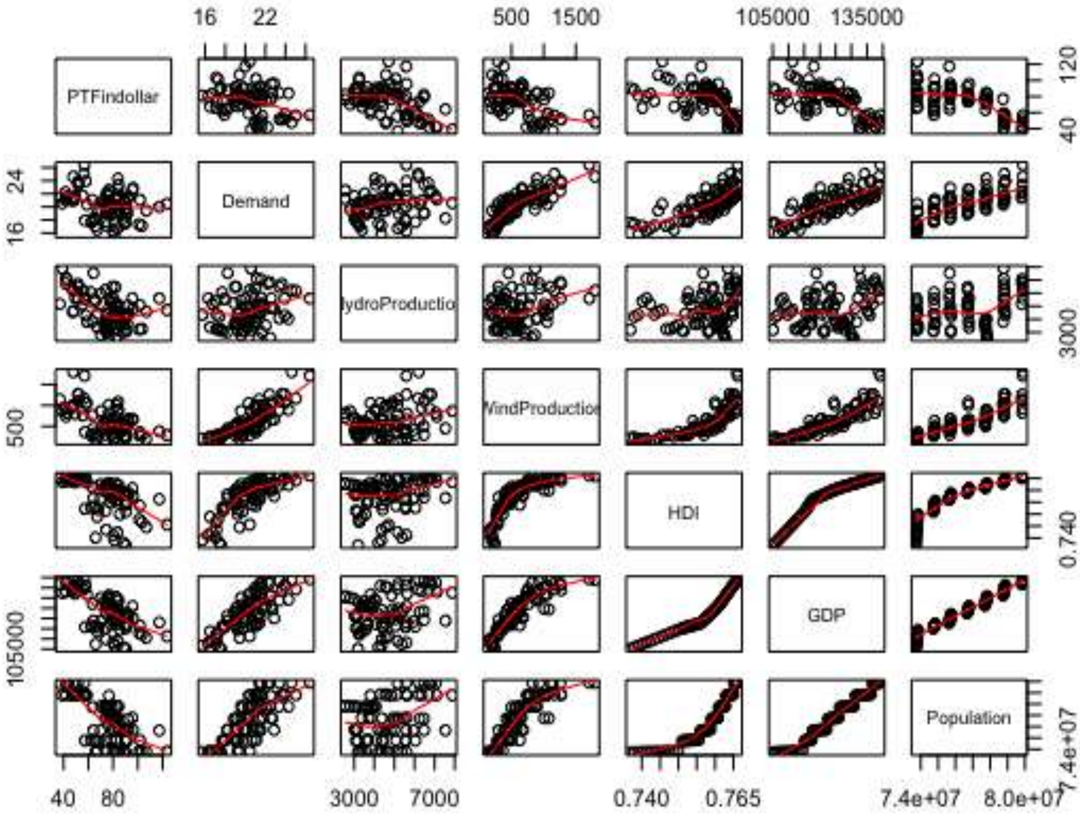

Figure 20. Relationships between dependent and independent variables before transformation of the data

In order to change the measurement scale, some mathematical operations are performed for a variable and this is called as "transformation". In the literature, transformations are classified into five main methods and can be listed as exponential, quadratic, logarithmic, reciprocal and power transformations. Logarithmic transformation is commonly used to estimate percentage change effects [69] and the slope shows the ratio of the percentage changes in Y (dependent) and x (independent) variables, i.e., for every 1% increase in x, the model predicts β_1 % increase in Y (for really small β_1).

$$\%\Delta Y \cong \beta_1 \times \%\Delta x \quad (14)$$

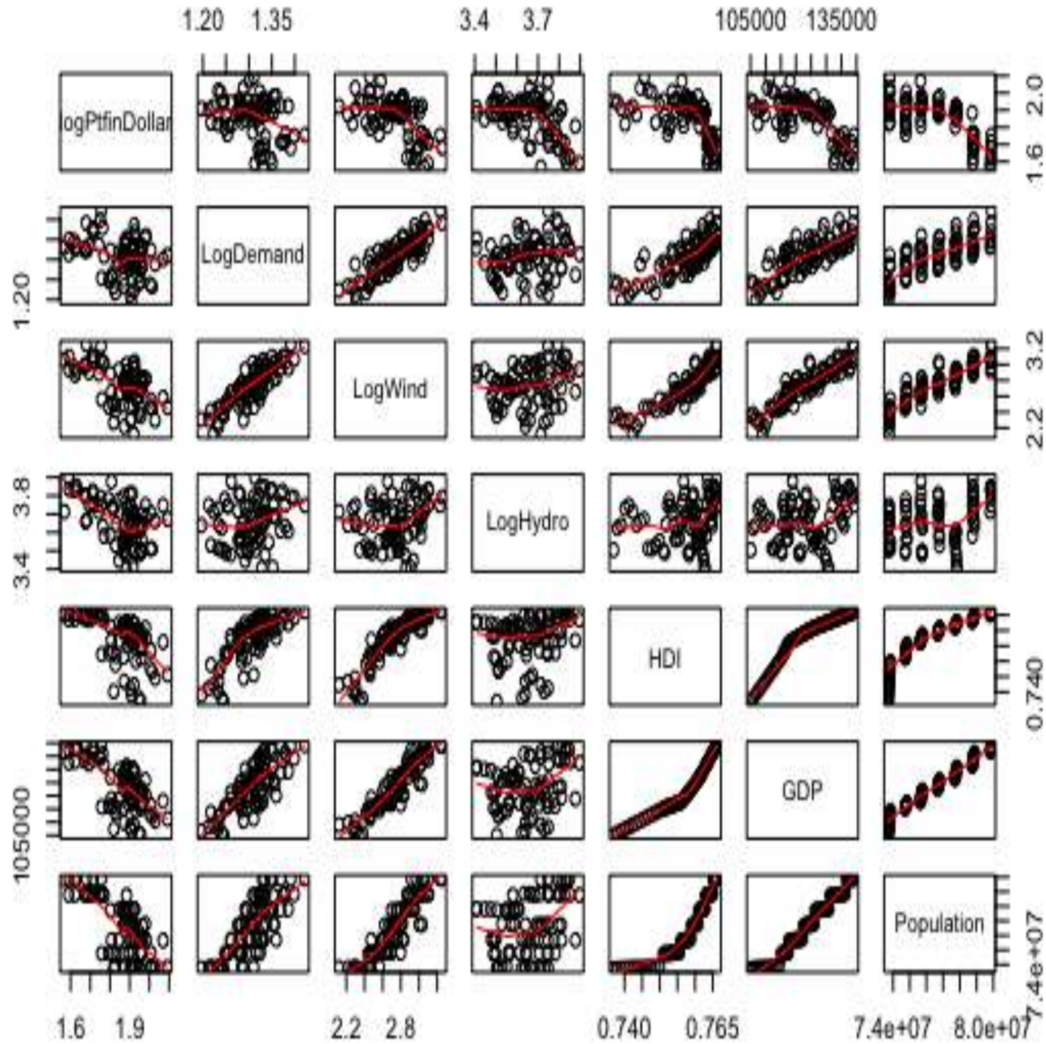

Figure 21. Relationships between dependent and independent variables after logarithmic transformation

It is assumed that a model form that is non-linear (as observed in Figure 20 trend lines) can be transformed to a linear model by taking logarithm of the variables (see Figure 21 and some of the linear trend lines). After logarithmic transformation on electricity demand, hydro production and wind production, many non-linear trend lines become linear and the multiple regression models are expected to provide better goodness of fit measures.

4.2 Regression Model

Regression models have a wide range of applications including forecasting and process control. In regression analysis, the aim is to model the dependent variable (electricity price) in the regression equation as a function of the independent variables, a constant and an error term. The performance of the model depends on the estimate of the constant, coefficients and error terms (i.e., residuals).

In Figure 22, transformed dependent variable (logarithm of electricity price) versus each of the transformed independent variables (logarithm of demand, logarithm of hydro production, logarithm of wind production, HDI, GDP and population) are plotted. According to this figure, all transformed independent variables have downward trend against transformed dependent variable between 2009 and 2016.

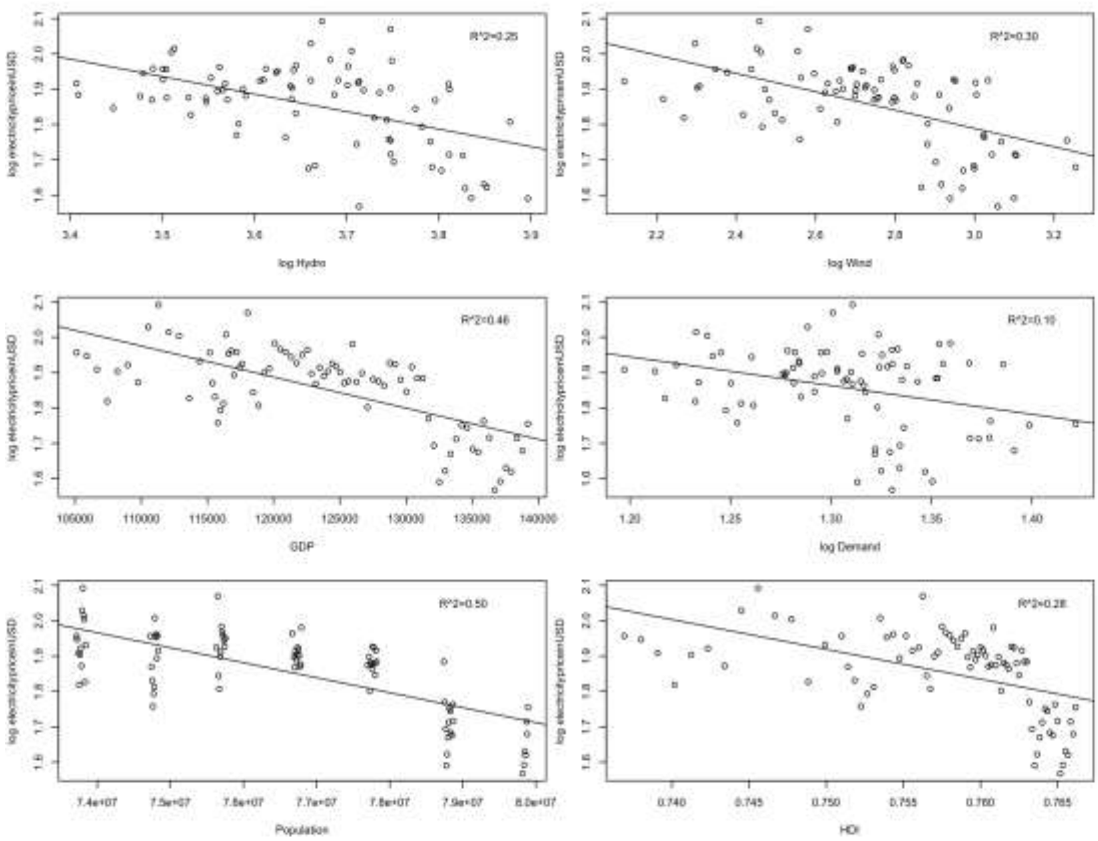

Figure 22. Scatter diagrams between dependent and independent variables

Several multiple linear regression models are analyzed for seeking the best form and summary results are shown below in Table 2.

Table 2. Multiple linear regression results

Regression Number	Dependent Variable	Independent Variables	Insignificant Independet Variable	Adjusted R^2	p-value
1	Price in TL	D, HP, WP, HDI, GDP, P	WP & HDI	0.5291752	3.75e-12
2	Log Price in TL	D, HP, WP, HDI, GDP, P	WP & HDI & GDP & P	0.5467353	2.80e-12
3	Log Price in TL	Log D, HP, WP, HDI, GDP	WP & HDI & GDP	0.5459142	1.00e-12
4	Log Price in TL	Log D, HP, HDI, GDP	WP & HDI & GDP	0.5372481	2.00e-12
5	Log Price in TL	Log D, Log HP, WP, HDI, GDP,P	P	0.5366154	2.10e-12
6	Price in USD	D, HP, WP, HDI, GDP	WP	0.7208305	< 2.22e-16
7	Log Price in USD	Log D, HP, WP, HDI, GDP	WP	0.777	< 2.22e-16
8	Log Price in USD	Log D, Log HP, WP, HDI, GDP	WP	0.7751298	< 2.22e-16
9	Log Price in USD	Log D, Log HP, Log WP, HDI, GDP	WP	0.7753303	2.22e-16
10	Log Price in USD	Log D, Log HP, HDI, GDP	-	0.7751298	< 2.22e-16
		(D=Demand HP=Hydro Production WP=Wind Production P=Population)			

We have firstly analyzed the regression model for electricity prices in TL as dependent variable and electricity demand, hydro production, wind production, GDP, HDI and population as independent variables. In this regression model, the adjusted R^2 value is 0.529 and p-value is 3.75e-12. Except for wind production and HDI, all the independent variables are significant. Despite having only two insignificant variables, adjusted R^2 value is low. After some transformations for a better regression model, we have used backward elimination by using best.model() function in R. This function provides the best multiple regression model with the highest adjusted R^2 level (see Appendix B1 for R code).

As shown in Table 3, the last (10^{th}) regression model results provides the best regression with an adjusted R^2 value around 77.5% and an overall significant model with very small p-values for each independent variable.

Table 3. 10th model regression results

	Estimate	*Std. Error*
*(Intercept)***	-5,149	1,768
*Log Demand ****	1,371	0,224
*Log Hydro ****	-0,323	0,054
*GDP ****	-2,268E-05	2,208E-06
*HDI****	12,143	2,655

*(Significance code : 0 '*** 0,001 '**')*

4.2.1 Residual Analyses for the Best Model

According to error terms, Mean Error (ME) of the regression model is -1.91e-17 where Root Mean Square Error (RMSE) is 0.053. Another widely used error term, Mean Absolute Error (MAE), is 0.042 and Mean Percentage Error (MPE) of the prediction is -0.085. Mean Absolute Percentage Error (MAPE) is 2.291 and Mean Absolute Scaled Error is 0.458.

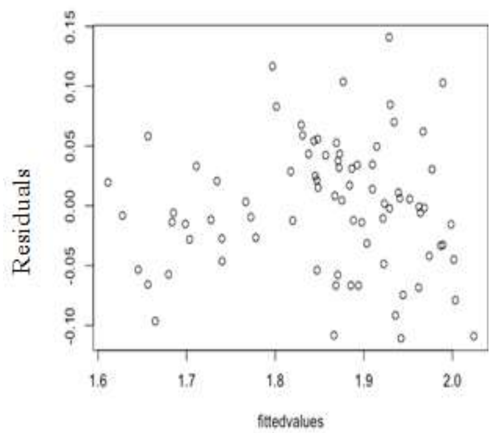

Figure 23.Residuals vs. Fitted values

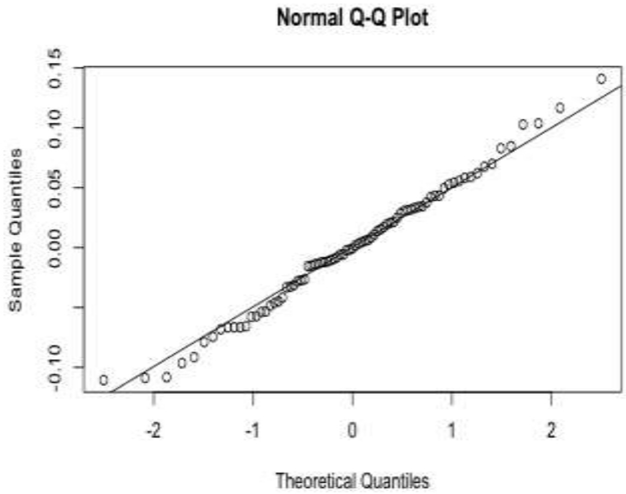

Figure 24. Normal residual plots

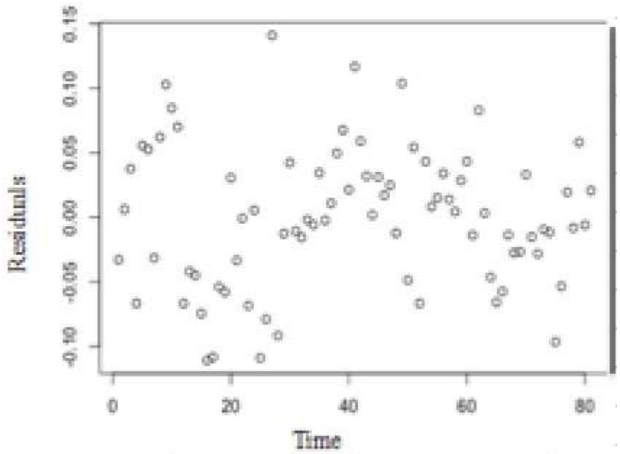

Figure 25. Residuals plot

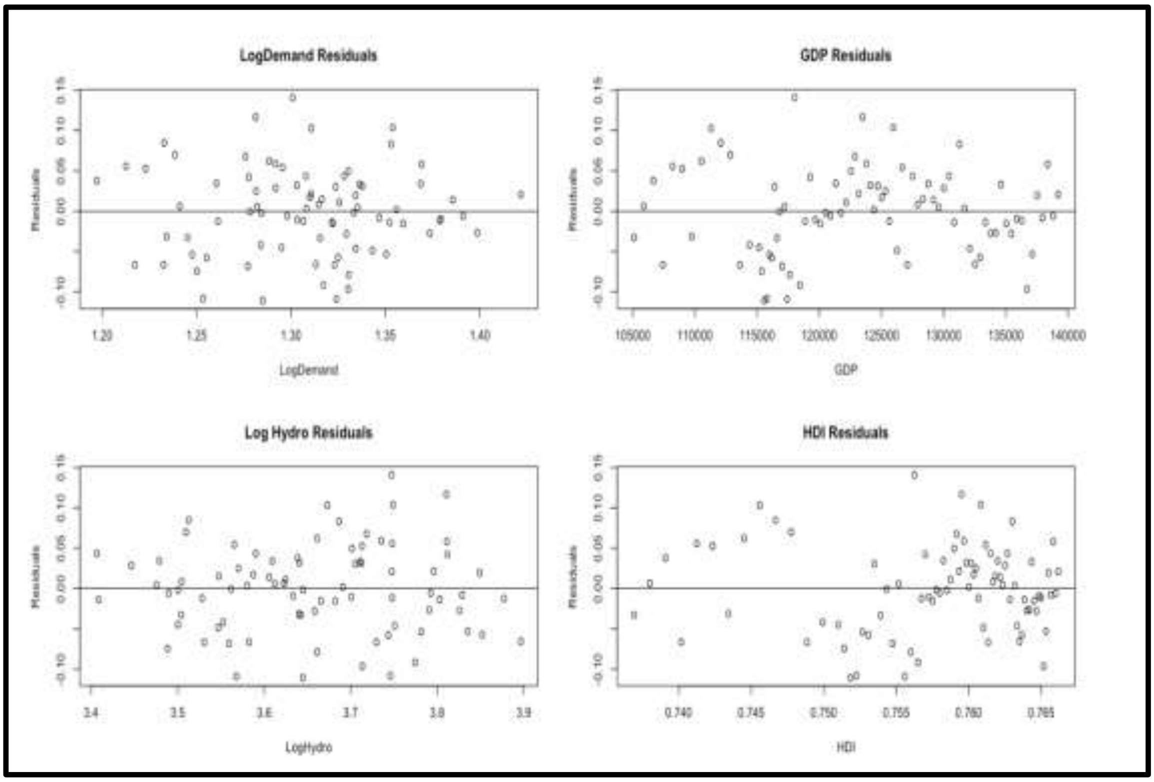

Figure 26. Residual graphs for dependent and each of independent variables

Residual analyses provide most accurate information about the error terms in a regression model and they are normally distributed around zero mean and a constant variance in Figure 23 and Figure 25. Figure 24 indicates the normal residual plot that is also supporting the previous finding. In Figure 26 residual graphs for independent variables are depicted and they are also normally distributed.

Before making any forecasts using the multiple regression model, it is required to forecast the independent variables. In order to forecast the electricity demand, hydro production, wind production, GDP and HDI, ARIMA model is used. Then, forecasts using these values with multiple regression model are depicted in Figure 27 for the next 24 months.

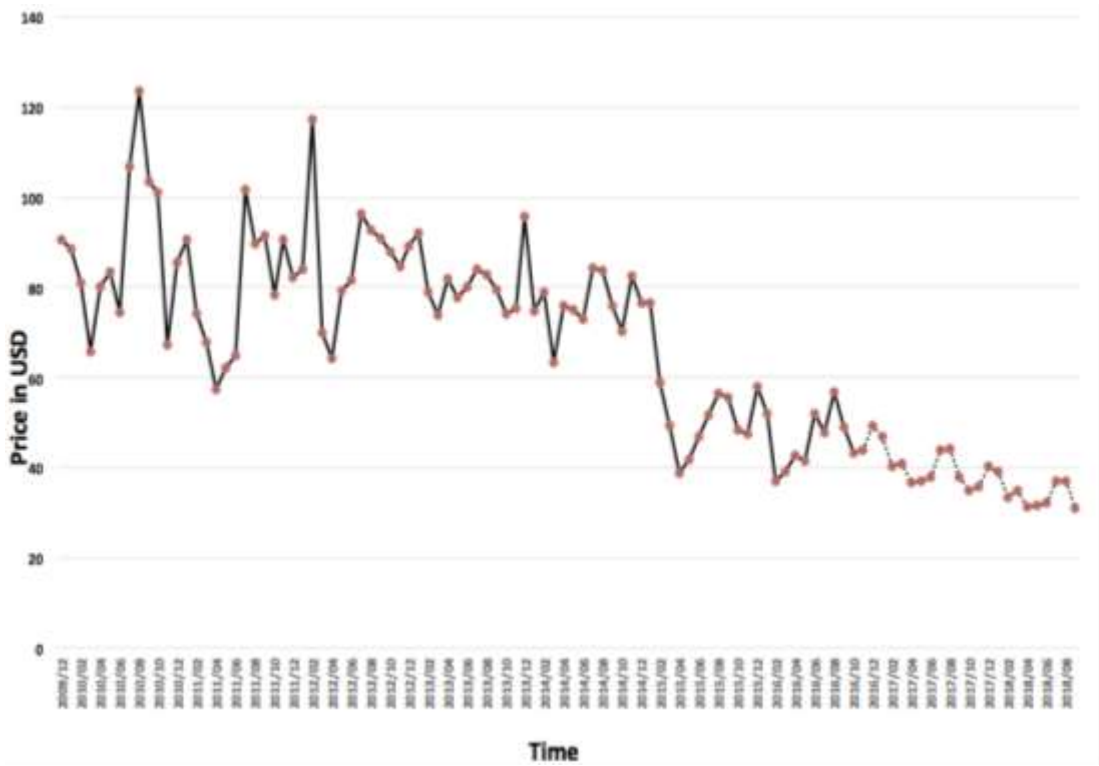

Figure 27. Multiple regression model forecast results

4.3 SARIMA Model

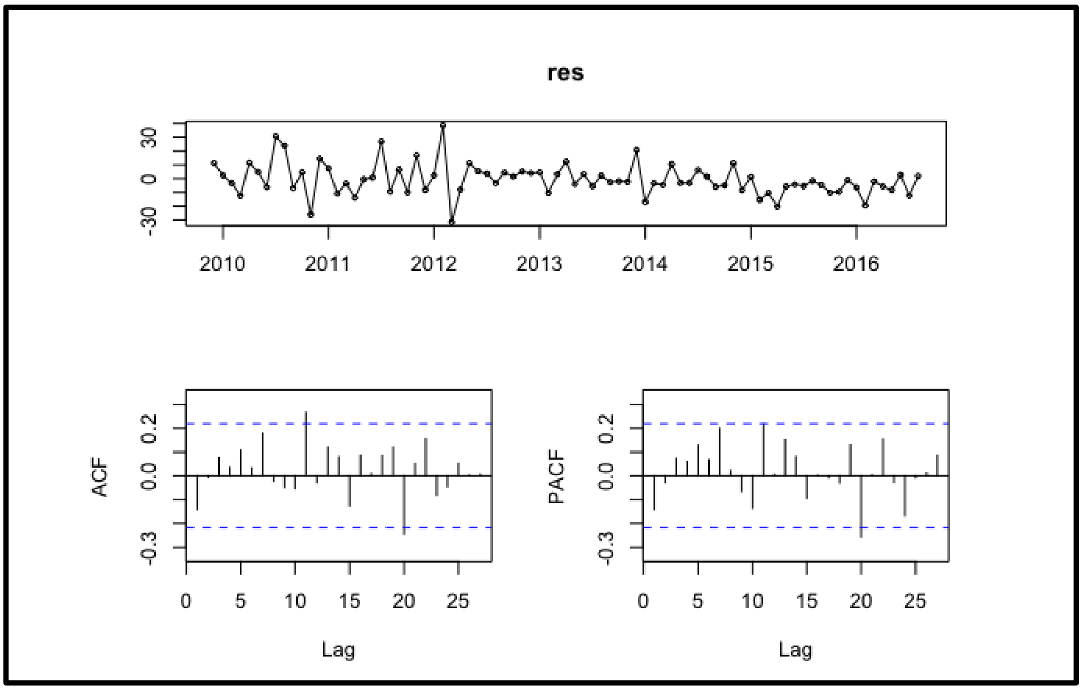

Figure 28. Residuals, auto-correlation (acf) and partial auto- correlation (pacf) graphs of the ARIMA model

In Figure 28, residuals, ACF (auto-correlation) and PACF (partial auto-correlation) are shown. According to decomposed data in Figure 18 there is a clear seasonality component. Therefore, a SARIMA model seems to be an appropriate approach.

In this method, we have utilized R's forecasting package "auto.arima" that provides the best SARIMA result for a given dataset (see Appendix B2 for R code). This forecasting tool in R operates using a trial and error approach for all possible ARIMA results and compare all the possibilities according to AICc level. Desired model has the least AICc value. All applied models with AICc values are shown in Table 4.

Table 4. All Applied ARIMA Models in R

ARIMA MODEL	AICc
ARIMA(2,1,2)(1,0,1)[12] with drift	: Inf
ARIMA(0,1,0) with drift	: 640.6209
ARIMA(1,1,0)(1,0,0)[12] with drift	: 635.9375
ARIMA(0,1,1)(0,0,1)[12] with drift	: 632.2449
ARIMA(0,1,0)	: 638.6009
ARIMA(0,1,1)(1,0,1)[12] with drift	: Inf
ARIMA(0,1,1) with drift	:635.1823
ARIMA(0,1,1)(0,0,2)[12] with drift	:630.7593
ARIMA(1,1,1)(0,0,2)[12] with drift	**: 627.123**
ARIMA(1,1,0)(0,0,2)[12] with drift	: 636.7752
ARIMA(1,1,2)(0,0,2)[12] with drift	: 629.4601
ARIMA(0,1,0)(0,0,2)[12] with drift	: 640.3907
ARIMA(2,1,2)(0,0,2)[12] with drift	: Inf
ARIMA(1,1,1)(0,0,2)[12]	: 626.904
ARIMA(1,1,1)(1,0,2)[12]	: Inf *
ARIMA(1,1,1)(0,0,1)[12]	: 628.7443
ARIMA(0,1,1)(0,0,2)[12]	: 628.8104
ARIMA(2,1,1)(0,0,2)[12]	: 629.2394
ARIMA(1,1,0)(0,0,2)[12]	: 634.6236
ARIMA(1,1,2)(0,0,2)[12]	: 629.2389
ARIMA(0,1,0)(0,0,2)[12]	: 638.26
ARIMA(2,1,2)(0,0,2)[12]	: Inf

The best fitted SARIMA model was $\boldsymbol{ARIMA(1,1,1)(0,0,2)_{12}}$

Table 5. $\boldsymbol{SARIMA}(1,1,1)(0,0,2)_{12}$ forecast coefficients

	Ar1	Ma1	Sma1	Sma2
coefficient	0,398	-0.847	0.148	0.268
Std. error	0.151	0.0789	0.129	0.135

(Ar1: First order Auto-regressive process, Ma1: First order Moving Average, Sma1: First order Seasonal Moving Average, Sma2: Second order Seasonal Moving Average)

In Table 5, ARIMA forecast coefficients are displayed. In this model AIC is 626.09 where AICc is 626.9 and BIC is 638. According to error terms, Mean Error (ME) is -1.5192 and Root Mean Square Error (RMSE) is 11.129. ARIMA model's Mean Absolute Error (MAE) is 7.9811 where Mean Percentage Error (MPE) is 4.2696. Another commonly used error term Mean Absolute Percentage Error (MAPE) is 11.289 and Mean Absolute Scaled Error is 0.636. First Order Autocorrelation Coefficient (ACF1) is - 0.0264. Forecasts for the ARIMA model are shown in Figure 29 with 80% and 95% confidence intervals.

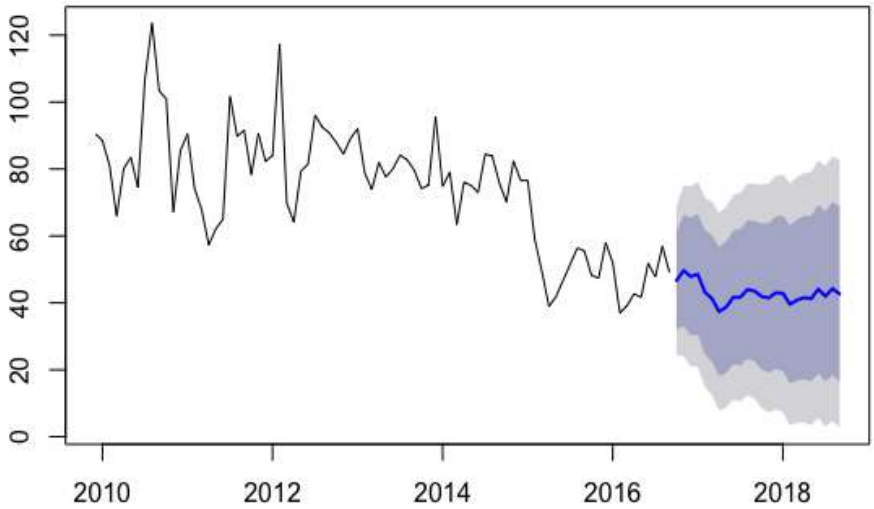

Figure 29. Forecast result for $\boldsymbol{SARIMA}(1,1,1)(0,0,2)_{12}$

4.4 Holt-Winters Exponential Smoothing

Holt-Winters exponential smoothing (HW-ES) method is also used to forecast the next 24 months electricity prices, where best smoothing parameters are determined. There are three important parameters in this model that should be optimized in order to reach the most accurate forecast results. Using R's forecast package, smoothing parameters are determined as follows (see Appendix B3 for R code);

α: 0.2576247

β: 0

γ: 0.6440515

where alpha is the smoothing factor of the level component, beta is the smoothing factor of the trend component and gamma is the smoothing factor of the seasonal component. The value of alpha (0.25) is relatively low, representing that the prediction of the level at the current time point is based upon some observations in the more distant past. The value of beta is 0, meaning that the estimate of the slope b of the trend component is not updated over the time series, and instead is set equal to its initial value. This makes good intuitive sense, as the level changes quite a bit over the time series, but the slope b of the trend component remains roughly the same. In contrast, the value of gamma (0.64) indicates that the estimate of the seasonal component at the current time point is based on both recent observations and some observations in the more distant past. But very recent observations plays more significant role.

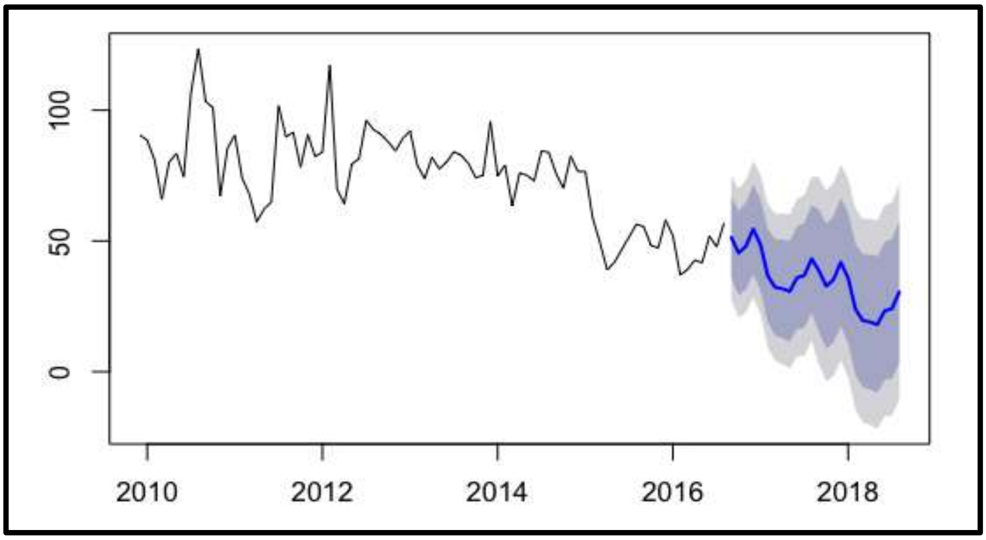

Figure 30. Holt -Winters Forecast Results

Forecasts for the HW-ES model are shown in Figure 30 with 80% and 95% confidence intervals for the next 24 months. In the model, AIC is 626.09, whereas AICc is 626.9 and BIC is 638. The error terms are as follows, Mean Error (ME) is 1.462 and Root Mean Square Error (RMSE) is 12.146. On the other hand, Mean Absolute Error (MAE) is 18.679 and Mean Percentage Error (MPE) is 1.282. Mean Absolute Percentage Error (MAPE) is 11.998, whereas Mean Absolute Scaled Error is 0.691. Finally, First Order Autocorrelation Coefficient (ACF1) is 0.086.

4.5 Artificial Neural Network Model

Artificial neural network (ANN) model is applied in MATLAB. Before applying ANN model, each of the input are normalized in order to obtain similar range of values. Therefore, the data are scaled within the range of 0 and 1. 92 data points for each independent variable from December 2009 to September 2016 is used. We have divided these data into training and test data in two different ways. The first one is the interpolated version where the whole set of data is divided in order to catch randomization, e.g., first three data points are used for training; the fourth data point is used for test and so on. The second one is the predicted version, where first 82 data point is used for training and the last 10 data points are used for testing (i.e., prediction).

In this study, in order to have more accurate forecasted results compared to actual values, a Multi Input Single Output (MISO), feed forward neural network is used. One of the most critical issues in constructing the ANN is the choice of the number of hidden layers and the number of neurons. The tangent sigmoid function is used in the hidden layer and linear activation function is used at the output layer. In this method, several tests are performed by different number of hidden layers, different number of neurons in hidden layers and different combination of input variables. After different implementations, removing hydro production achieved better forecast results. The architecture with one hidden layer and five neurons is found to be the most accurate. Best architecture of developed ANN model is shown in Figure 31.

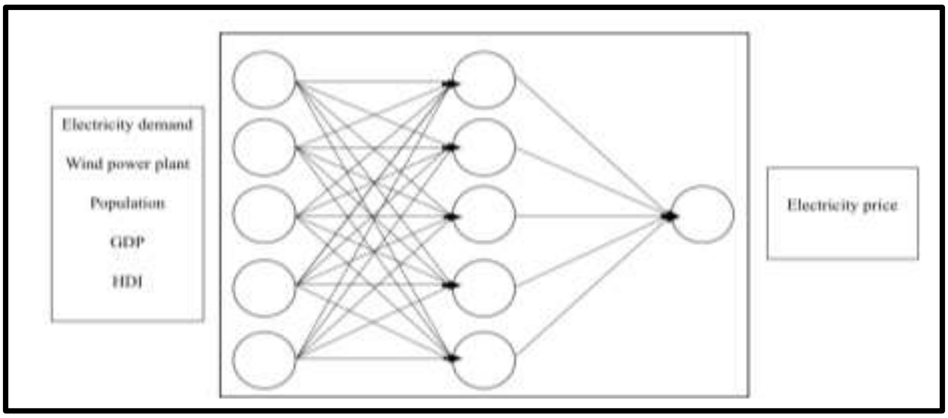

Figure 31. Most accurate ANN architecture

4.5.1. Interpolated Neural Network Model

Before starting our analysis, changes in electricity price are examined first. When electricity prices in USD are employed, there is a big decline between two parts of the data, where the reason is directly related to changes in USD/TL exchange rate. In order to reduce this inconsistency in our study, first three data points are used for training and the fourth data point is used for test and similarly for the subsequent periods.

The data for independent variables are used as input variables (monthly electricity demand, monthly hydro power production, and monthly wind power production) from Enerji Atlası website [70] for the last 10 data points (October 2016 to July 2017). For the same period, the other three input variables (HDI, GDP and population) are estimated by considering constant linear trends in recent data.

In this section of the study, we have used 92 data points (including the last 10 months for forecasting) where first of the three data points (75%) is used for training and the fourth data point (25%) is used for test. In many studies, using 75% of the data for training is suggested and therefore, we have employed every fourth data point (e.g., 23 data points) for testing and the rest (69 data points) used for training. This way of skipping data for testing provides interpolation in data and the data points that are used for testing have been predicted.

We have also tested different number of hidden neurons and the optimum architecture is achieved by using 5 neurons in input layer, 5 neurons in hidden layer and 1 neuron in output layer (e.g., 5-5-1) by trial and error approach while achieving a minimum root mean square error (RMSE) for testing results (see Appendix C1 for MATLAB code).In order to find the most accurate results, ANN model is repeated 5 times for each hidden neuron number. In Table 6, different number of hidden neurons and the corresponding RMSE results are shown. According to this table, the degree of RMSE for test data is more important than trained data. Since the trained data only memorizes the past data, decision making criteria was RMSE1, which has the lowest level when there are 5 hidden neurons

Table 6. RMSE values of tried neural network models

Hidden neuron numbers	**RMSE 1 (test data)**	**RMSE 2 (trained data)**
3	11,345	7,351
4	10,903	8,582
5	10,860	7,599
6	11,93	6,432
7	11,895	5,361
8	10,961	8,55
9	13,512	3,9
10	16,611	4,094
11	14,841	10,714
12	11,963	7,132

In Figure 32, training performance of neural network is shown through 10000 epochs.

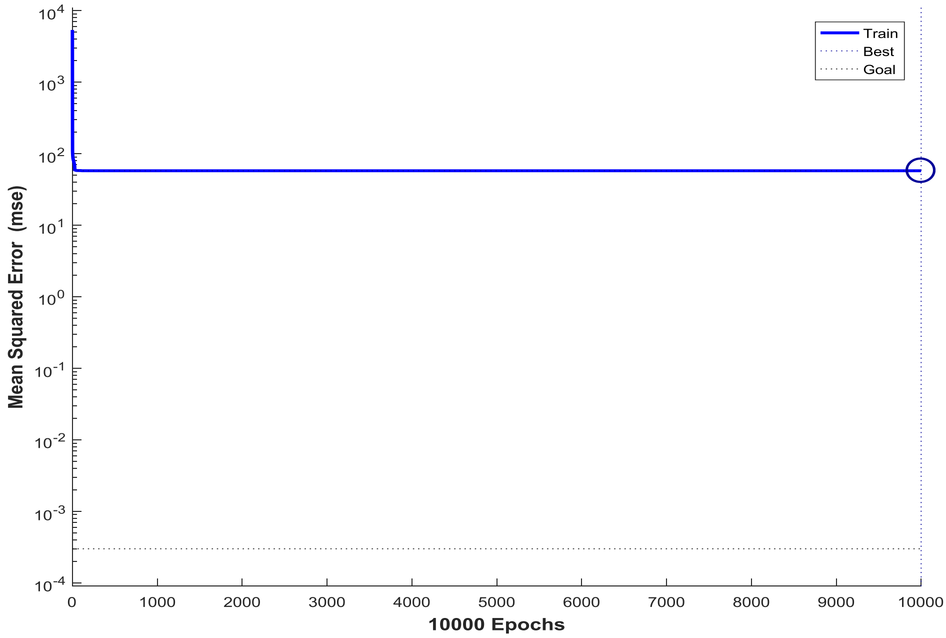

Figure 32. Training performance of the neural network

In Figure 33 and Figure 34, performance of trained and test data in our neural network model have been illustrated, respectively.

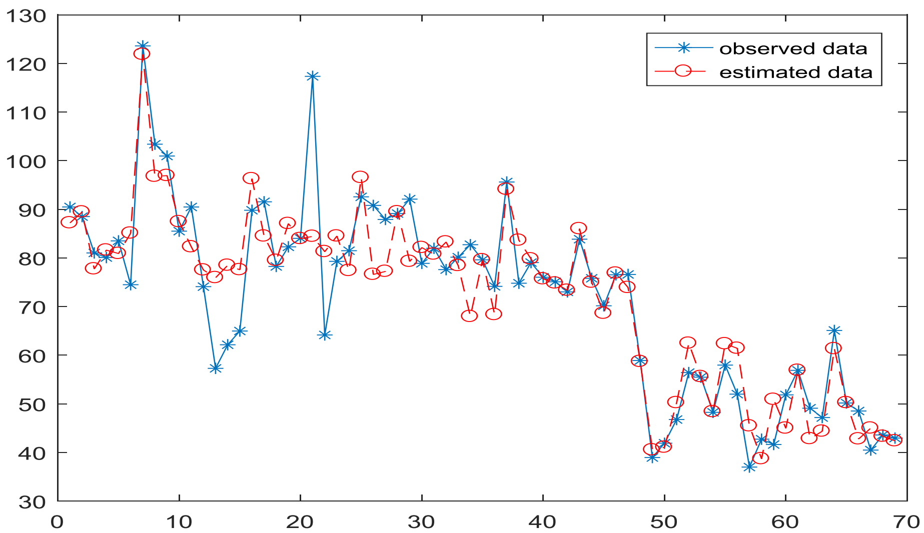

Figure 33. Trained data vs. observed electricity price

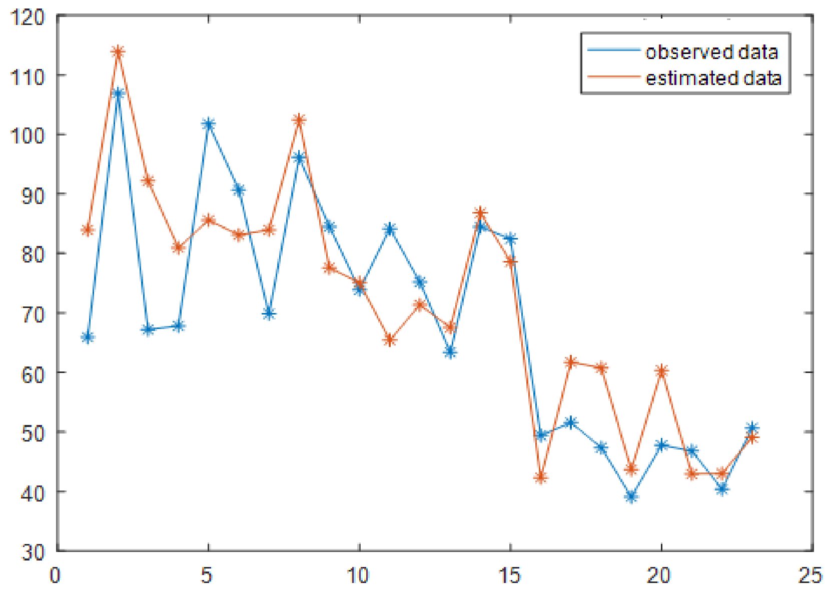

Figure 34. Test data vs. observed electricity price

The accuracy of the trained network is evaluated in two well-known error term, mean square error (MSE) and root mean square error (RMSE). Results are presented in Table 7.

Table 7. Error measures for the ANN model

Data Set	Root Mean Square Error(RMSE)	Mean Square Error(MSE)
Training	**7,599**	**57,744**
Test	**10,860**	**117,939**

4.5.2. Predicted Neural Network Model

In the second part of the neural network study, we have used all 92 data points, where first 82 data points are used for training and the remaining 10 data points are used for test.

Similar to interpolated neural network model in Section 4.5.1, we have tested different number of hidden neurons by using trial and error approach, and the optimum neural network architecture uses 5 neurons in input layer, 5 neurons in hidden layer and 1 neuron in output layer (5-5-1), where the root mean square error (RMSE) is minimized (see Appendix C2 for MATLAB code). Results are obtained after 5 repetitions of run for each hidden neuron number.

In Table 8, different number of hidden neurons and the corresponding RMSE results are shown. According to test data RMSE results, the best configuration is achieved with 5 neurons in hidden layer.

Table 8. RMSE values of tried neural network models

Hidden neuron numbers	**RMSE 1 (Test data)**	**RMSE 2 (Trained data)**
3	8,67	11,4
4	7,52	9,51
5	4,76	9,07
6	5,6	7,11
7	7,28	7,16
8	11,56	7,61
9	13,1	8,3
10	10,33	5,77
11	4,98	6,28
12	8,7	5,55

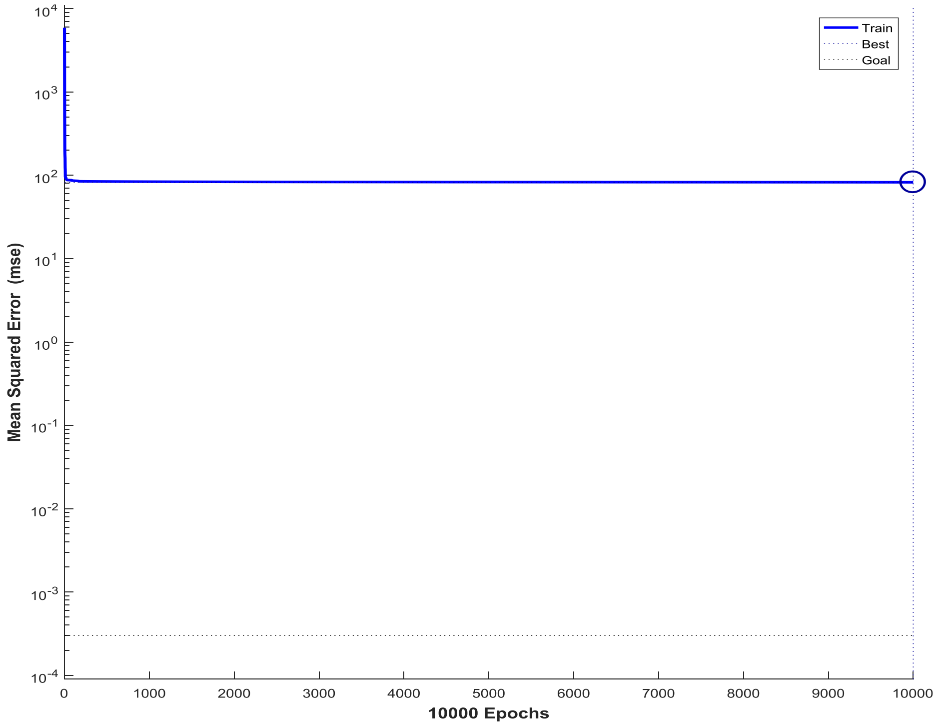

Figure 35 .Training performance of neural network

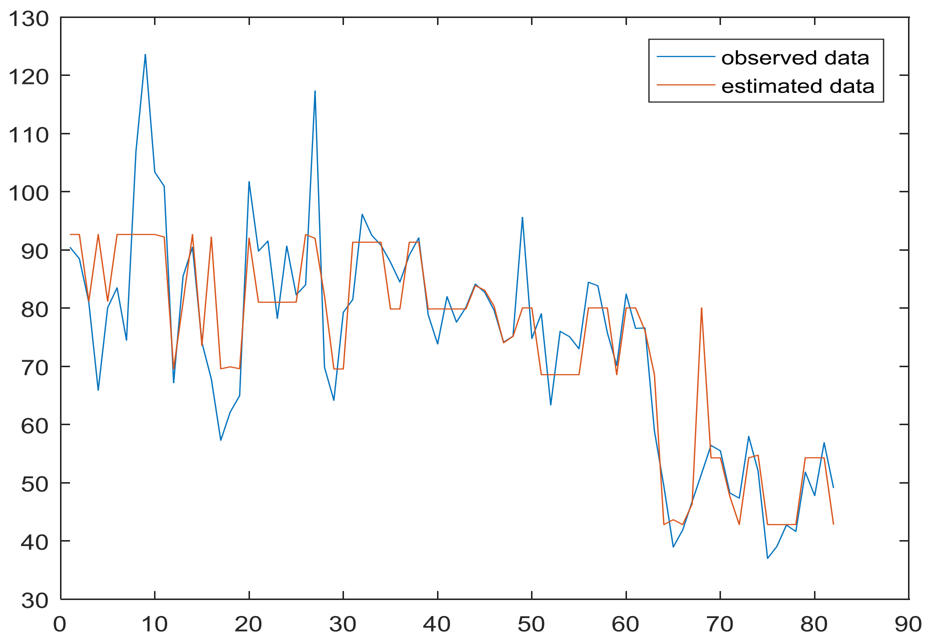

Figure 36. Trained data vs. observed PTF values

As can be observed in Figure 36, trained data generally catch up with the observed data, especially for the most recent data points.

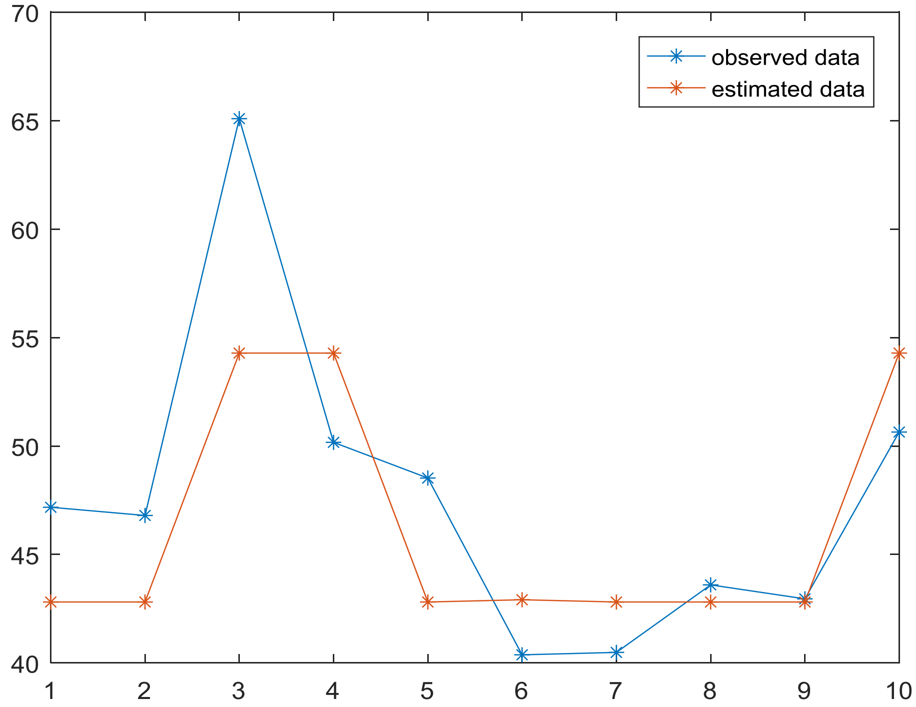

Figure 37. Estimated vs observed price

Figure 37 illustrates the predicted and the observed values. Even though it seems to be very irrelevant, these high differences occur because of the limited data. In fact, the patterns are following each other very closely.

4.6 Comparison of all applied models

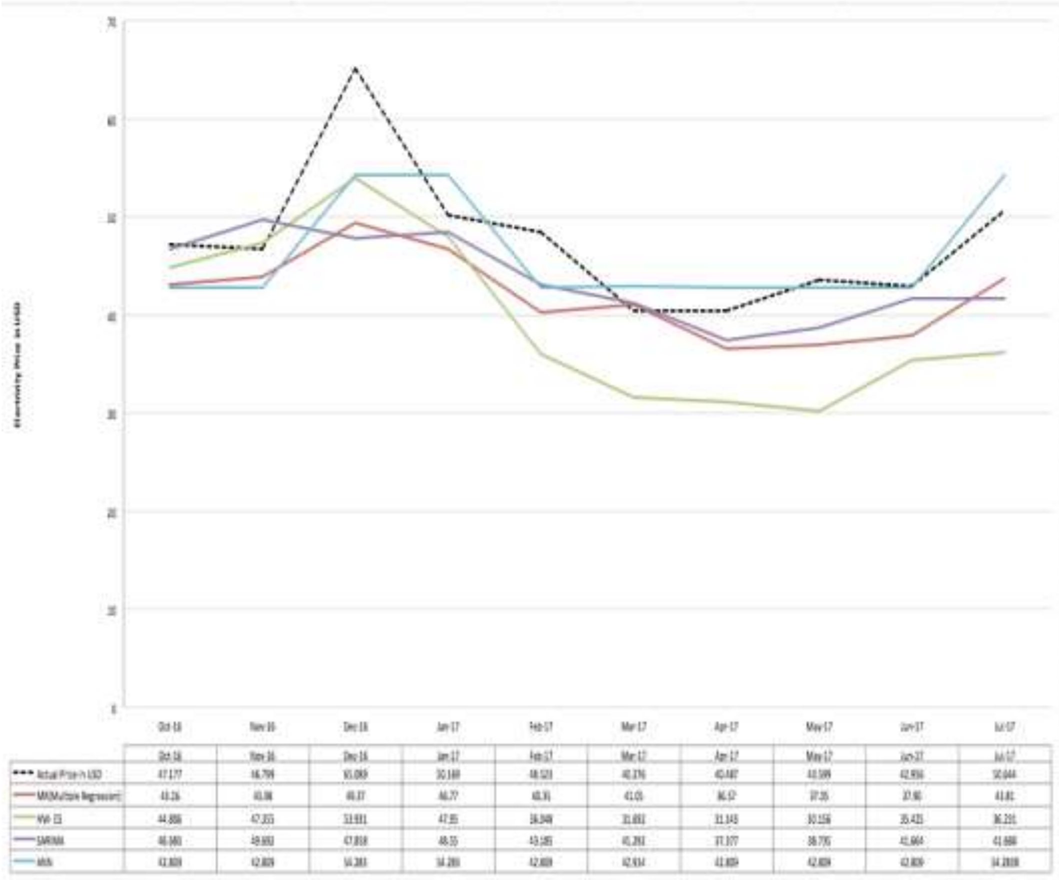

Figure 38. Comparison for all models

In Figure 38 and Figure 39 predicted values by all four models are compared with actual electricity prices in USD. According to this graph ANN obtains the best forecast results. In order to make analysis more accurate, error comparisons of all applied models are demonstrated in Table 9 (for more detailed information see Appendix D).

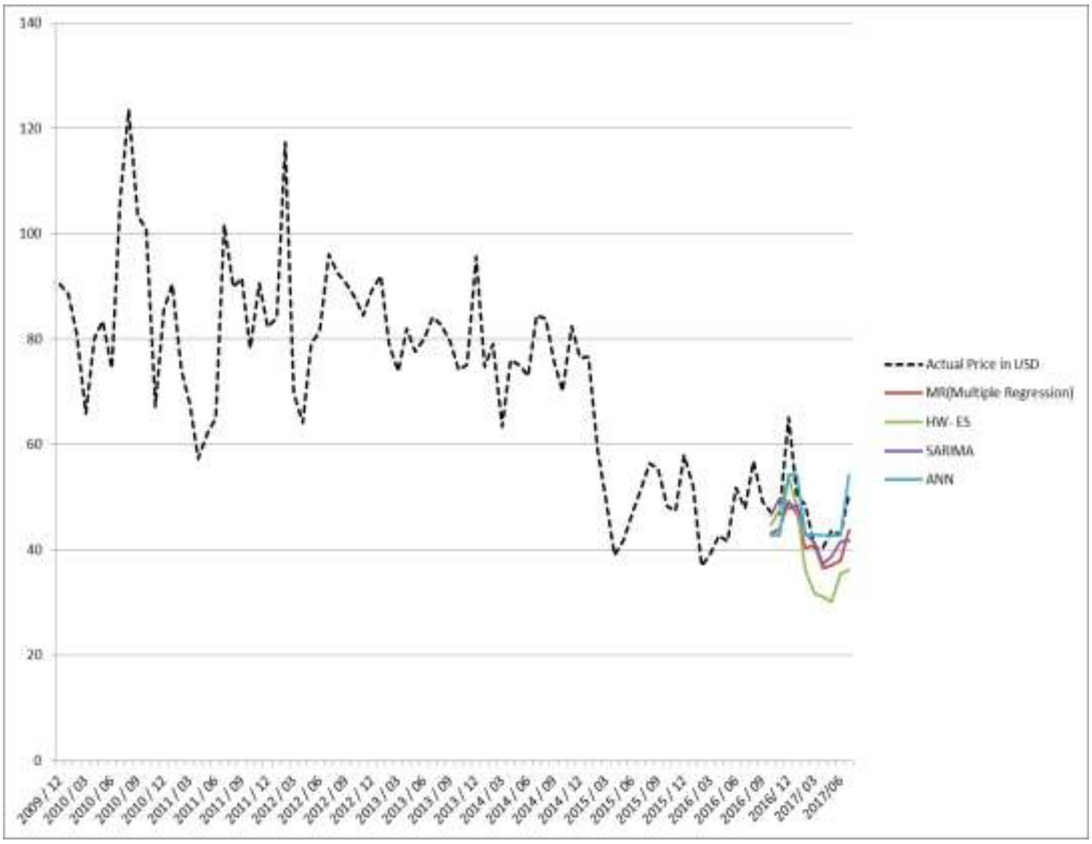

Figure 39. Price changes and forecasts (between 2009-2017)

Table 9. Error comparisons for all models

	MR	HW- ES	SARIMA	ANN
MAE	5,716	8,219	4,673	**3,843**
MSE	48,057	89,887	45,304	**22,707**
RMSE	6,932	9,481	6,731	**4,765**
MAPE	11%	17%	9%	**8%**
Mean BIAS	5,58	8,1	3,91	**1,32**

In Table 9, error measures for all models are summarized. ANN has the best performance over all other models, in terms of all calculated error measures, e.g., mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean bias. SARIMA and MR results are very close. On the other hand, HW-ES model has the worst performance in forecasting for this case study.

Chapter 5

5 Conclusion and Suggestions

In this thesis, multiple linear regression, SARIMA, Holt-Winters exponential smoothing and ANN methods are used in order to forecast electricity prices for the next 24 months. ANN model has been created using MATLAB. On the other hand, multiple linear regression, ARIMA and Holt-Winters exponential smoothing models, uses R for statistical computing. Despite the fact that there has been many studies that is applied to Turkish electricity load forecasting models, there is a few number of short and medium term price forecasting studies. Moreover, there is no long term price forecasting study encountered in the literature neither for Turkish nor for any other system. Hence, this research is a first attempt to apply long-term price forecasting concepts.

The best method in terms of error measures is the ANN model, with a MAPE of 8% followed by ARIMA and multiple linear regression model with MAPE of 9% and 11%, respectively. Holt-Winters Exponential Smoothing has the worst performance with a MAPE of 17%. ANN model has given the closest forecast results and the lowest error measures.

Nevertheless, in the literature many authors argued that the performance of neural network models is very dependent on the number of data points used. As the number of data points increases, better results are expected. In our case study, only 92 data points are available and this may not be enough for making good forecasts with ANN. However, it has outperformed the other methods in quality of forecasts due to inherent nonlinear relations of the input (independent) variables represented by ANN method. By increasing the number of data points (e.g., using daily or hourly resolution instead of monthly) and also the number of input variables (e.g., electricity production by natural gas and/or fuel-oil) may provide better results.

Moreover, in order to make more accurate forecasts, further studies are required to make hybrid models. It is possible to reach more accurate results when several methods are combined instead of individual models. A new hybrid model can be developed by assuming some weights to different applied models (MR, ARIMA, ANN and HW-ES) and finding best combination with weights can have more accurate forecasts.

References

[1] Halliday, D., Walker, J., & Resnick, R. (2010). Fundamentals of Physics, Chapters 33-37. John Wiley & Sons.

[2] Soni, J. S. (2014) Electricity Price Forecasting Model-Defining the Need and Approach for India Market

[3] Türkiye Enerji Forumu, Enerji Şirketlerinin Gelecek Planları; 'Rekabet ve işbirliği' (2008) retrieved from http://www.enerjiforumu.com/gorseller/forumlar/3337365818876587.pdf

[4] IEA Key Electricity Trends 2016

[5]International Energy Agency World energy investment (2017)

[6] International Energy Agency Energy policies of IEA countries (2016) TURKEY 2016 Review

[7] Özdemir N., (2011), Türkiye' de Elektriğin Tarihsel Gelişimi, Ankara

[8] Türk Mühendis ve Mimar Odaları Birliği, Elektrik Mühendisleri Odası Ankara şubesi EMO Ankara Şubesi, 2009, 5.Bülten retrieved from http://www.emo.org.tr/ekler/c48ca73fc192d0c_ek.pdf?dergi=.

[9] T.C. Özelleştirme Idaresi Başkanlığı Biçen, Y. (2016). Türkiye Elektrik Enerjisi Piyasası Gelişim Süreci: Gün Öncesi ve Dengeleme Güç Piyasası Özellikleri. Karaelmas Fen ve Mühendislik Dergisi, 6(2), 432-438.

[10] Enerji Piyasası Denetleme Kurumu retrieved from http://www.epdk.org.tr.

[11] Armstrong, J. S., ed. (2002). Principles of forecasting: A handbook of researchers and practitioners, Kluwer Academic, Boston, MA.

[12] Bunn, D. W. (2000). Forecasting loads and prices in competitive power markets. Proceedings of the IEEE, 88(2), 163

[13] Shahidehpour, M., Yamin, H., & Li, Z. (2002). Market operations in electric power systems: forecasting, scheduling, and risk management. Wiley.

[14] Weron, R., & Misiorek, A. (2006). Short-term electricity price forecasting with time series models: A review and evaluation. In W. Mielczarski (Ed.), Complex electricity markets (pp. 231–254). Łódź: IEPŁ& SEP. –169.

[15] Amjady, N., & Hemmati, M. (2006). Energy price forecasting. IEEE Power and Energy Magazine, March/April, 20–29.

[16] Amjady, N. (2012). Short-term electricity price forecasting. In J. P. S. Catalão (Ed.), Electric power systems: advanced forecasting techniques and optimal generation scheduling. CRC Press.

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

[18] Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: the state space approach. Springer.

[19] Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. International Journal of Forecasting, 16, 521–530.

[20] Kim, C.-I., Yu, I.-K., & Song, Y. H. (2002). Prediction of system marginal price of electricity using wavelet transform analysis. Energy Conversion and Management, 43, 1839–1851.

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

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

[23] Karakatsani, N. V., & Bunn, D. W. (2010). Fundamental and behavioural drivers of electricity price volatility. Studies in Nonlinear Dynamics and Econometrics, 14(4), art. no. 4.

[24] Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: forecasting and control. San Francisco: Holden-Day.

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

[26] Weron, R., & Misiorek, A. (2005). Forecasting spot electricity prices with time series models. IEEE Conference Proceedings – EEM05 (pp. 133–141).

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

[28] Haldrup, N., & Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. Journal of Econometrics, 135, 349–376.

[29] Lagarto, J., De Sousa, J., Martins, A., & Ferrão, P. (2012). Price forecasting in the day-ahead Iberian electricity market using a conjectural variations ARIMA model. IEEE Conference Proceedings — EEM12, art. no. 6254734.

[30] Amjady, M., & Hemmati, M. (2009). Day-ahead price forecasting of electricity markets by a hybrid intelligent system. European Transactions on Electrical Power, 19(1), 89–102.

[31] Che, J., & Wang, J. (2010). Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. Energy Conversion and Management, 51(10), 1911–1917.

[32] Cruz, A., Muñoz, A., Zamora, J. L., & Espinola, R. (2011). The effect of wind generation and weekday on Spanish electricity spot price forecasting. Electric Power Systems Research, 81(10), 1924–1935

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

[34] Aggarwal, S. K., Saini, L. M., & Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. International Journal of Electrical Power & Energy Systems, 31(1), 13-22.

[35] Weron, R. (2007). Modeling and forecasting electricity loads and prices: A statistical approach (Vol. 403). John Wiley & Sons.

[36] Azadeh, A., Moghaddam, M., Mahdi, M., & Seyedmahmoudi, S. H. (2013). Optimum long-term electricity price forecasting in noisy and complex environments. Energy Sources, Part B: Economics, Planning and Policy, 8(3), 235–244

[37] Niimura, T. (2006) Forecasting Techniques for Deregulated Electricity Market Prices—Extended Survey. 2006 IEEE PES Power Systems Conference and Exposition, Atlanta, 29 October-1 November 2006, 51-56.

[38] Singh, N., & Mohanty, S. R. (2015). A Review of Price Forecasting Problem and Techniques in Deregulated Electricity Markets. Journal of Power and Energy Engineering, 3(09), 1.

[39] Chow, J., Wu, F., & Momoh, J. (2005). Applied mathematics for restructured electric power systems. *Applied Mathematics for Restructured Electric Power Systems*, 1-9.

[40] Hu, L., Taylor, G., Wan, H.-B. and Irving, M. (2009) A Review of Short-Term Electricity Price Forecasting Tech- niques in Deregulated Electricity Markets. Universities Power Engineering Conference (UPEC), 2009 Proceedings of the 44th International, Glasgow, 1-4 September 2009, 1-5.

[41] Ventosa, M., Baıllo, A., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends. Energy policy, 33(7), 897-913.).

[42] Nash, J. F. (1950). Equilibrium points in n-person games. Proceedings of the national academy of sciences, 36(1), 48-49.

[43] Barquín, J., & Vazquez, M. (2008). Cournot equilibrium calculation in power networks: An optimization approach with price response computation. IEEE Transactions on Power Systems, 23(2), 317-326.

[44] Younes, Z., & Ilic, M. (1998). Transmission networks and market power. In Power Systems Restructuring (pp. 335-384). Springer US.

[45] Green, R. J., & Newbery, D. M. (1992). Competition in the British electricity spot market. Journal of political economy, 100(5), 929-953.

[46] Day, C. J., Hobbs, B. F., & Pang, J. S. (2002). Oligopolistic competition in power networks: a conjectured supply function approach. IEEE Transactions on power systems, 17(3), 597-607.

[47] Safarinejadian, B., Gharibzadeh, M., & Rakhshan, M. (2014). An optimized model of electricity price forecasting in the electricity market based on fuzzy timeseries. Systems Science & Control Engineering: An Open Access Journal, 2(1), 677-683.

[48] Vera, V. D. G. (2015). Forecasting Electricity Demand for Small Colombian Populations. CUADERNO ACTIVA, 7(7), 111-120.

[49] Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2015). An introduction to management science: quantitative approaches to decision making. Cengage learning.

[50] Blanchard, M., & Desrochers, G. (1984). Generation of autocorrelated wind speeds for wind energy conversion system studies. Solar Energy, 33(6), 571-579.

[51] Kamal, L., & Jafri, Y. Z. (1997). Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Solar Energy, 61(1), 23-32.

[52] Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. Computers & industrial engineering, 35(1-2), 213-216.

[53] Abdel-Aal, R. E., & Al-Garni, A. Z. (1997). Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. Energy, 22(11), 1059-1069.

[54] Brockwell, P. J., & Davis, R. A. (2002). Introduction to time series and forecasting.

[55] Cowpertwait, P. S., & Metcalfe, A. V. (2009). Introductory time series with R. Springer Science & Business Media.

[56] Hyndman, R. J., & Khandakar, Y. (2007). Automatic time series for forecasting: the forecast package for R (No. 6/07). Monash University, Department of Econometrics and Business Statistics.

[57] Berk, K. (2015). Modeling and Forecasting Electricity Demand: A Risk Management Perspective. Springer.

[58] Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi School of Information Technology, 4329008, 1-13.

[59] "What is Simple Linear Regression?". Pennsylvania State University. Retrieved 2016-10-17.

[60] Mazengia, D. H. (2008, November). Forecasting spot electricity market prices using time series models. In Sustainable Energy Technologies, 2008. ICSET 2008. IEEE International Conference on (pp. 1256-1261). IEEE.

[61] Thomas Lumley based on Fortran Code by Allan Miller,(2017), Package 'leaps'

[62] Kröse, B., Krose, B., van der Smagt, P., & Smagt, P. (1993). An introduction to neural networks.

[63] Szkuta, B. R., Sanabria, L. A., & Dillon, T. S. (1999). Electricity price short-term forecasting using artificial neural networks. IEEE transactions on power systems, 14(3), 851-857.

[64] Chen C-C, Kuo Y-C, Huang C-H, Chen A-P. Applying market profile theory to forecast Taiwan index futures market. Expert Syst Appl 2014;41:4617–4624.

[65] Chen C-C, Chen A-P, Yeh P-Y. Modeling and simulation of the open-end equity mutual fund market in Taiwan by using self-organizing map. Simulat Model Pract Theory 2013;36:60–73.

[66] EPİAŞ Transparency Platform, 2016 retrieved from https://seffaflik.epias.com.tr/transparency/

[67] Turkey Statistical Intiutive retrieved from (http://www.turkstat.gov.tr/)

[68] Human development reports retrieved from (http://hdr.undp.org/en/data)

[69] Sheather, S. (2009). A modern Approach to Regression with R. Springer Science & Business Media

[70] Enerji Atlası retrieved from http://www.enerjiatlasi.com

APPENDICES

Appendix A: In applied models used data set

Period	Demand(TWh)	Hydro Production (GWh)	Wind Production (GWh)	Human Development Index (HDI)	Gross Domestic Product (GDP in TL)	Population	Price in USD	USD/TRY	Price in TL
2009 / 12	17,59	3.192,2	222,90	0,736916667	105.111,55	73.716.312,79	90,48124924	1,507992	136,445
2010 / 01	17,42	4.199,0	238,50	0,738	105.885,64	73.722.988,00	88,50022763	1,476064	130,632
2010 / 02	15,75	4.352,8	203,40	0,739083333	106.659,73	73.743.013,62	81,02367702	1,521222	123,255
2010 / 03	17,08	5.368,4	185,70	0,740166667	107.433,82	73.753.026,43	65,90413979	1,534547	101,133
2010 / 04	16,31	5.596,7	200,30	0,74125	108.207,91	73.763.039,24	80,07842338	1,492412	119,51
2010 / 05	16,71	5.167,0	131,40	0,742333333	108.982,00	73.773.052,05	83,50677234	1,550497	129,477
2010 / 06	17,14	4.366,5	164,70	0,743416667	109.756,09	73.783.064,86	74,49571982	1,576802	117,465
2010 / 07	19,43	4.582,1	304,30	0,7445	110.530,18	73.793.077,67	106,8872403	1,538715	164,469
2010 / 08	20,45	4.709,7	287,60	0,745583333	111.304,27	73.803.090,48	123,6043437	1,510311	186,681
2010 / 09	17,09	3.255,7	283,20	0,746666667	112.078,36	73.813.103,29	103,3847444	1,492609	154,313
2010 / 10	17,32	3.233,1	289,60	0,74775	112.852,45	73.823.116,10	100,95683	1,424391	143,802
2010 / 11	16,49	3.394,1	261,80	0,748833333	113.626,54	73.833.128,91	67,18792725	1,444575	97,058
2010 / 12	19,23	3.570,2	365,70	0,749916667	114.400,63	73.843.141,72	85,50375361	1,522401	130,171
2011 / 01	19,72	3.164,7	274,30	0,751	115.174,72	74.724.269,00	90,47516021	1,566861	141,762
2011 / 02	17,79	3.079,1	304,70	0,751416667	115.378,92	74.742.331,30	74,09321045	1,588985	117,733

Period	Demand(TWh)	Hydro Production (GWh)	Wind Production (GWh)	Human Development Index (HDI)	Gross Domestic Product (GDP in TL)	Population	Price in USD	USD/TRY	Price in TL
2011 / 03	19,28	4.412,0	314,50	0,751833333	115.583,11	74.751.362,45	67,80803648	1,578279	107,02
2011 / 04	17,92	5.568,7	363,40	0,75225	115.787,30	74.760.393,60	57,2917968	1,520846	87,132
2011 / 05	17,69	6.048,2	292,20	0,752666667	115.991,49	74.769.424,75	62,14420169	1,576736	97,985
2011 / 06	18,00	5.539,3	327,90	0,753083333	116.195,68	74.778.455,90	64,97211524	1,59944	103,919
2011 / 07	21,07	5.074,1	358,70	0,7535	116.399,87	74.787.487,05	101,7427044	1,655932	168,479
2011 / 08	20,67	4.387,3	628,50	0,753916667	116.604,06	74.796.518,20	89,79864084	1,757109	157,786
2011 / 09	18,99	3.645,9	494,00	0,754333333	116.808,26	74.805.549,35	91,53879008	1,795807	164,386
2011 / 10	18,93	3.627,1	448,30	0,75475	117.012,45	74.814.580,50	78,24848104	1,827077	142,966
2011 / 11	19,15	4.095,2	490,60	0,755166667	117.216,64	74.823.611,65	90,65889242	1,814196	164,473
2011 / 12	21,09	3.697,0	426,80	0,755583333	117.420,83	74.832.642,80	82,28212281	1,871731	154,01
2012 / 01	21,41	4.582,2	457,90	0,756	117.625,02	75.627.384,00	84,02148006	1,842267	154,79
2012 / 02	19,99	5.588,1	380,70	0,75625	118.035,97	75.648.193,60	117,3348215	1,757901	206,263
2012 / 03	20,76	5.950,0	409,10	0,7565	118.446,93	75.658.598,40	69,85124861	1,789832	125,022
2012 / 04	18,25	7.538,2	451,90	0,75675	118.857,88	75.669.003,20	64,16071426	1,787371	114,679
2012 / 05	18,95	6.481,6	296,90	0,757	119.268,83	75.679.408,00	79,29348934	1,809909	143,514
2012 / 06	20,10	5.020,3	533,80	0,75725	119.679,79	75.689.812,80	81,48296472	1,822835	148,53
2012 / 07	22,88	4.809,6	661,50	0,7575	120.090,74	75.700.217,60	96,14247681	1,811551	174,167
2012 / 08	21,54	4.414,2	682,10	0,75775	120.501,69	75.710.622,40	92,56080282	1,794777	166,126
2012 / 09	19,86	3.088,2	488,00	0,758	120.912,65	75.721.027,20	90,80193262	1,801074	163,541
2012 / 10	18,22	3.012,4	394,90	0,75825	121.323,60	75.731.432,00	87,91272211	1,80206	158,424

Period	Demand(TWh)	Hydro Production (GWh)	Wind Production (GWh)	Human Development Index (HDI)	Gross Domestic Product (GDP in TL)	Population	Price in USD	USD/TRY	Price in TL
2012 / 11	19,24	3.165,9	581,10	0,7585	121.734,55	75.741.836,80	84,48342491	1,793263	151,501
2012 / 12	21,16	4.214,2	522,80	0,75875	122.145,51	75.752.241,60	89,12466398	1,787485	159,309
2013 / 01	21,40	5.028,7	622,00	0,759	122.556,46	76.667.864,00	92,08995121	1,770204	163,018
2013 / 02	18,87	5.228,7	502,50	0,759166667	122.864,91	76.688.424,80	78,88386353	1,775661	140,071
2013 / 03	20,45	6.252,3	634,20	0,759333333	123.173,35	76.698.705,20	73,84840078	1,807595	133,488
2013 / 04	19,11	6.463,8	504,80	0,7595	123.481,80	76.708.985,60	81,98069896	1,795447	147,192
2013 / 05	19,58	5.438,7	420,00	0,759666667	123.790,24	76.719.266,00	77,59595597	1,825108	141,621
2013 / 06	20,10	4.373,3	529,90	0,759833333	124.098,69	76.729.546,40	80,14654586	1,897836	152,105
2013 / 07	22,69	4.906,8	1.080,80	0,76	124.407,14	76.739.826,80	84,14761648	1,934933	162,82
2013 / 08	21,77	5.156,7	1.006,10	0,760166667	124.715,58	76.750.107,20	82,72012568	1,959233	162,068
2013 / 09	20,42	3.867,2	467,00	0,760333333	125.024,03	76.760.387,60	79,60959281	2,01943	164,473
2013 / 10	19,12	3.719,1	558,80	0,7605	125.332,47	76.770.668,00	74,1533863	1,983254	154,01
2013 / 11	20,26	3.375,0	572,80	0,760666667	125.640,92	76.780.948,40	75,18734976	2,024689	154,79
2013 / 12	22,59	5.610,3	658,70	0,760833333	125.949,36	76.791.228,80	95,6096938	2,067259	206,263
2014 / 01	22,04	3.523,1	570,40	0,761	126.257,81	77.695.904,00	74,79722383	2,221045	125,022
2014 / 02	19,75	3.677,6	559,20	0,761166667	126.675,72	77.716.806,98	79,04535562	2,206518	114,679
2014 / 03	21,04	3.824,0	762,70	0,761333333	127.093,63	77.727.258,47	63,37471069	2,21216	143,514
2014 / 04	20,32	3.892,8	499,90	0,7615	127.511,53	77.737.709,96	76,02928523	2,124484	148,53
2014 / 05	20,64	3.196,7	627,70	0,761666667	127.929,44	77.748.161,45	75,09356918	2,090432	156,978
2014 / 06	20,72	3.529,5	619,90	0,761833333	128.347,35	77.758.612,94	73,01640796	2,118606	154,693
2014 / 07	23,38	4.068,6	884,80	0,762	128.765,26	77.769.064,43	84,46120043	2,119044	178,977

Period	Demand(TWh)	Hydro Production (GWh)	Wind Production (GWh)	Human Development Index (HDI)	Gross Domestic Product (GDP in TL)	Population	Price in USD	USD/TRY	Price in TL
2014 / 08	24,31	4.033,9	894,90	0,762166667	129.183,16	77.779.515,92	83,85187518	2,15926	181,058
2014 / 09	21,65	2.989,7	705,60	0,762333333	129.601,07	77.789.967,41	75,80639551	2,213771	167,818
2014 / 10	19,58	2.795,4	863,00	0,7625	130.018,98	77.800.418,90	70,15992482	2,25606	158,285
2014 / 11	21,29	2.550,6	717,00	0,762666667	130.436,89	77.810.870,39	82,42082337	2,232761	184,026
2014 / 12	22,51	2.562,7	814,90	0,762833333	130.854,80	77.821.321,88	76,5318206	2,29694	175,789
2015 / 01	22,54	4.859,4	1.010,30	0,763	131.272,70	78.741.053,00	76,60240778	2,33742	179,052
2015 / 02	20,33	3.804,6	1.053,20	0,763166667	131.690,61	78.751.791,18	58,904378	2,465029	145,201
2015 / 03	21,59	5.635,7	798,00	0,763333333	132.108,52	78.762.529,36	49,39290781	2,595899	128,219
2015 / 04	20,57	7.883,8	867,60	0,7635	132.526,43	78.773.267,54	38,95327154	2,651767	103,295
2015 / 05	21,14	7.111,9	734,50	0,763666667	132.944,33	78.784.005,72	41,89933678	2,644839	110,817
2015 / 06	20,98	6.351,9	936,80	0,763833333	133.362,24	78.794.743,90	46,77027249	2,695537	126,071
2015 / 07	23,64	6.697,2	1.272,80	0,764	133.780,15	78.805.482,08	51,56665781	2,697111	139,081
2015 / 08	25,05	6.178,4	1.166,50	0,764166667	134.198,06	78.816.220,26	56,42455946	2,85218	160,933
2015 / 09	21,69	5.138,2	761,10	0,764333333	134.615,96	78.826.958,44	55,47278875	3,019192	167,483
2015 / 10	20,99	4.629,2	995,90	0,7645	135.033,87	78.837.696,62	48,24878882	2,924633	141,11
2015 / 11	21,35	4.554,6	998,10	0,764666667	135.451,78	78.848.434,80	47,3500184	2,88002	136,369
2015 / 12	23,95	4.300,9	1057,60	0,764833333	135.869,69	78.859.172,98	57,96189244	2,92173	169,349
2016 / 01	23,93	5.595,1	1263,00	0,765	136.287,60	78.869.911,16	51,98962886	3,001791	156,062
2016 / 02	21,40	5.169,7	1144,00	0,765166667	136.705,50	79.814.871,00	36,99457173	2,945324	108,961
2016 / 03	22,41	6.842,4	1257,00	0,765333333	137.123,41	79.836.347,36	39,08593	2,882802	112,677
2016 / 04	21,59	7.062,9	825,00	0,7655	137.541,32	79.847.085,54	42,72964605	2,834917	121,135

Period	Demand(TWh)	Hydro Production (GWh)	Wind Production (GWh)	Human Development Index (HDI)	Gross Domestic Product (GDP in TL)	Population	Price in USD	USD/TRY	Price in TL
2016 / 05	22,23	6.739,1	931,00	0,765666667	137.959,23	79.857.823,72	41,63464507	2,940724	122,436
2016 / 06	23,40	6.474,4	1106,00	0,765833333	138.377,13	79.868.561,90	51,84029567	2,914586	151,093
2016 / 07	24,62	6.201,8	1797,00	0,766	138.795,04	79.879.300,08	47,78886872	2,969792	141,923
2016 / 08	26,41	5592,6	1707,00	0,766166667	139.212,95	79.890.038,26	56,88074882	2,960826	168,414
2016 / 09	21,36	4158,3	1314	0,766333333	139.630,86	79.900.776,44	49,09546361	2,968482	145,74
2016 / 10	21,8	3986	1115	0,7665	140.048,77	80.278.369,90	47,17732757	3,0751	145,08
2016 / 11	22,7	3721	1339	0,766666667	140.466,67	80.409.473,31	46,79882748	3,3091	154,86
2016/12	25,2	5998	1614	0,766833333	140.884,58	80.540.576,73	65,08865603	3,5023	227,96
2017/01	25,1	5618	1632	0,767	141.302,49	80.671.680,14	50,16914786	3,7482	188,044
2017/02	22,5	4254	1550	0,767166667	141.720,40	80.802.783,56	48,52335633	3,6671	177,94
2017/03	23,6	5680	1359	0,767333333	142.138,30	80.933.886,98	40,37567376	3,6734	148,32
2017/04	22	7130	983	0,7675	142.556,21	81.064.990,39	40,48722438	3,6554	148,00
2017/05	22,9	7368	1302	0,767666667	142.974,12	81.196.093,81	43,5993166	3,5704	155,67
2017/06	22,3	5265	971	0,767833333	143.392,03	81.327.197,22	43,10863763	3,5218	151,82
2017/07	27,8	4720	1977	0,768	143.809,93	81.458.300,64	50,64369366	3,5607	180,33

Appendix B: R Codes

Appendix B1: Multiple Regression R Codes

```
install.packages("leaps")
library(leaps)
regsubsets.out <-
  regsubsets( PTF~ Demand + Hydro + Wind + HDI+ GDP ,
             data = LinearRegression,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
             force.in = NULL, force.out = NULL,
             method = "exhaustive")
regsubsets.out
summary.out <- summary(regsubsets.out)

as.data.frame(summary.out$outmat)
plot(regsubsets.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:2, ncol = 2))
## Adjusted R2
res.legend <-
  subsets(regsubsets.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
## Mallow Cp
res.legend <-
  subsets(regsubsets.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)

res.legend #abbrevation
which.max(summary.out$adjr2)
summary.out$which[5,]

best.model <- lm(PTF ~ Demand + Hydro+ Wind + GDP + HDI ,
data=LinearRegression)
summary(best.model)


#PriceLOG
pairs(ptflog, panel = panel.smooth)

regsubsets2.out <-
  regsubsets(  logPTF ~ Demand + HydroProduction + WindProduction +
HDI+ GDP+ Population,
              data = ptflog,
              nbest = 1,       # 1 best model for each number of predictors
              nvmax = NULL,    # NULL for no limit on number of variables
              force.in = NULL, force.out = NULL,
```

```
            method = "exhaustive")
regsubsets.out
summary.out2 <- summary(regsubsets2.out)
as.data.frame(summary.out$outmat)
plot(regsubsets2.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend <-
  subsets(regsubsets2.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
## Mallow Cp
res.legend <-
  subsets(regsubsets2.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
which.max(summary.out2$adjr2)
summary.out2$which[6,]

best.modellogptf <- lm(logPTF ~ Demand + HydroProduction  +
WindProduction + GDP + HDI + Population , data=ptflog)
summary(best.modellogptf)

#PriceLOG and DemandLOG
pairs(logptfandlogdemand, panel = panel.smooth)
plot(logptfindollarlogdemandloghydro$logPtfinDollar,
logptfindollarlogdemandloghydro$logDemand)
abline(0,0)
library(lattice)
layout(matrix(2:2, ncol = 2))
par(mfrow=c(2,2))
xyplot(logptfindollarlogdemandloghydro$logDemand~
logptfindollarlogdemandloghydro$logPtfinDollar  ,
      col='black',
      xlab="LogDemand", ylab="LogPTFinDollar")
xyplot(logptfindollarlogdemandloghydro$logHydro~
logptfindollarlogdemandloghydro$logPtfinDollar , type=c("smooth", "p") ,
      col='black',
      xlab="LogHydro", ylab="LogPTFinDollar")

%Scatterplots
plot(logptfindollarlogdemandloghydro$logPtfinDollar~
logptfindollarlogdemandloghydro$logHydro, xlab="log Hydro", ylab="log
electricitypriceinUSD")
second <-lm(logPtfinDollar~logHydro,
data=logptfindollarlogdemandloghydro)
abline(second)
text(x=3.86, y=2.05 , label= "R^2=0.25")
```

```
xyplot(logptfindollarlogdemandloghydro$HDI~
logptfindollarlogdemandloghydro$logPtfinDollar , type=c("smooth", "p") ,
      col='black',
      xlab="HDI", ylab="LogPTFinDollar")

plot(logptfindollarlogdemandloghydro$logPtfinDollar~
logptfindollarlogdemandloghydro$HDI, xlab="HDI", ylab="log
electricitypriceinUSD")
third <- lm(logPtfinDollar~ HDI , data=logptfindollarlogdemandloghydro)
abline(third)
text(x=0.764, y=2.05 , label= "R^2=0.28")

xyplot(logptfindollarlogdemandloghydro$GDP~
logptfindollarlogdemandloghydro$logPtfinDollar , type=c("smooth", "p") ,
      col='black',
      xlab="GDP", ylab="LogPTFinDollar")

fourth<- lm(logPtfinDollar~GDP , data=logptfindollarlogdemandloghydro)
plot(logptfindollarlogdemandloghydro$GDP ,
logptfindollarlogdemandloghydro$logPtfinDollar , xlab="GDP", ylab="log
electricitypriceinUSD")
abline(fourth)
text(x=136000, y=2.05 , label= "R^2=0.46")

nine<- lm(logPtfinDollar~LogWind ,
data=logptfindollarlogdemandloghydrologwind)
plot(logptfindollarlogdemandloghydrologwind$LogWind ,
logptfindollarlogdemandloghydrologwind$logPtfinDollar , xlab="log
Wind", ylab="log electricitypriceinUSD")
abline(nine)
text(x=3.1, y=2.05 , label= "R^2=0.30")

par(mfrow=c(3,2))
 plot(logptfindollarlogdemandloghydrologwind$LogDemand,
logptfindollarlogdemandloghydrologwind$logPtfinDollar, xlab="log
Demand", ylab="log electricitypriceinUSD")
first<-lm( logPtfinDollar~LogDemand,
data=logptfindollarlogdemandloghydrologwind)
abline(first)
text(x=1.4, y=2.05 , label= "R^2=0.10")
plot(logptfindollarlogdemandloghydrologwind$Population,
logptfindollarlogdemandloghydrologwind$logPtfinDollar,
xlab="Population", ylab="log electricitypriceinUSD")
sixth<-lm(logPtfinDollar ~Population,
data=logptfindollarlogdemandloghydrologwind)
abline(sixth)
text(x=79008880, y=2.05 , label= "R^2=0.50")
```

```
rsq <- function (Population,logPtfinDollar ) cor(Population, logPtfinDollar)
^ 2
summary(sixth)
legend(2,3)

abline(logptfindollarlogdemandloghydrologwind$logPtfinDollar,
logptfindollarlogdemandloghydrologwind$LogDemand)
par(mar=c(4,4,2,0.5),mfrow=c(1,1))
plot(logptfandlogdemand$logPTF, logptfandlogdemand$logDemand,
panel=panel.smooth)
abline(0,0)
abline(lm(logptfindollarlogdemandloghydrologwind$logPtfinDollar ~
logptfindollarlogdemandloghydrologwind$))
abline(logptfandlogdemand$logPTF, logptfandlogdemand$logDemand ,
col='red')
regsubsets3.out <-
  regsubsets(  logPTF ~ logDemand+ HydroProduction + WindProduction +
HDI+ GDP+ Population,
             data = logptfandlogdemand,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
             force.in = NULL, force.out = NULL,
             method = "exhaustive")
regsubsets.out
summary.out3 <- summary(regsubsets3.out)
as.data.frame(summary.out$outmat)
plot(regsubsets.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:2, ncol = 2))
## Adjusted R2
res.legend <-
  subsets(regsubsets3.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend <-
  subsets(regsubsets3.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
which.max(summary.out3$adjr2)
summary.out3$which[5,]

best.modellogptflogdemand <- lm(logPTF ~ logDemand +
HydroProduction  + WindProduction + GDP + HDI ,
                          data=logptfandlogdemand)
summary(best.modellogptflogdemand)
```

```
#logpricelogdemandloghydro
pairs(logptflogdemandloghydro_, panel = panel.smooth)

regsubsets4.out <-
  regsubsets(  logPTF ~ logDemand+ loghydro + WindProduction + HDI+
GDP+ Population,
             data = logptflogdemandloghydro_,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
             force.in = NULL, force.out = NULL,
             method = "exhaustive")
regsubsets4.out
summary.out4 <- summary(regsubsets4.out)
as.data.frame(summary.out$outmat)
plot(regsubsets4.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend4<-
  subsets(regsubsets4.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend5 <-
  subsets(regsubsets4.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend4
which.max(summary.out4$adjr2)
summary.out4$which[5,]
best.modellogptflogdemandloghydro <- lm(logPTF ~ logDemand +
loghydro  + WindProduction + GDP + HDI ,
                          data=logptfandlogdemand)
summary(best.modellogptflogdemandloghydro)



#LogPrice LogDemand LogHydro LogWind

pairs(ptfindollar, panel = panel.smooth)

regsubsets5.out <-
  regsubsets(  logPTF ~ logDemand+ loghydro + logwind + HDI+ GDP+
Population,
             data = logptflogdemandloghydroandlogwind,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
```

```
            force.in = NULL, force.out = NULL,
            method = "exhaustive")
regsubsets5.out
summary.out5 <- summary(regsubsets5.out)
as.data.frame(summary.out$outmat)
plot(regsubsets5.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend6<-
  subsets(regsubsets5.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend7 <-
  subsets(regsubsets4.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend4

which.max(summary.out5$adjr2)
summary.out5$which[5,]

best.modellogptflogdemandloghydronowind<- lm(logPTF ~ logDemand +
loghydro + GDP + HDI+ Population,
                                data=logptflogdemandloghydroandlogwind)
summary(best.modellogptflogdemandloghydronowind)

#Price IN  DOLLARS
pairs(ptfindollars, panel = panel.smooth)

regsubsets6.out <-
  regsubsets(  PTFindollar ~ Demand + HydroProduction + WindProduction
+ HDI+ GDP+ Population,
            data = ptfindollars,
            nbest = 1,       # 1 best model for each number of predictors
            nvmax = NULL,    # NULL for no limit on number of variables
            force.in = NULL, force.out = NULL,
            method = "exhaustive")
regsubsets6.out
summary.out6 <- summary(regsubsets6.out)
as.data.frame(summary.out6$outmat)
plot(regsubsets6.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
```

```
## Adjusted R2
res.legend7<-
  subsets(regsubsets6.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend8 <-
  subsets(regsubsets6.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend4

which.max(summary.out6$adjr2)
summary.out6$which[5,]

best.modelptfindollars<- lm(PTFindollar ~ Demand + HydroProduction +
WindProduction+ GDP + HDI,
                                    data=ptfindollars)
summary(best.modelptfindollars)


#LogPrice IN  DOLLARS
pairs(logptfindollar, panel = panel.smooth)

regsubsets7.out <-
  regsubsets(  lofptfindollar ~ Demand + HydroProduction +
WindProduction + HDI+ GDP+ Population,
             data = logptfindollar,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
             force.in = NULL, force.out = NULL,
             method = "exhaustive")
regsubsets7.out
summary.out7 <- summary(regsubsets7.out)
as.data.frame(summary.out7$outmat)
plot(regsubsets7.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend9<-
  subsets(regsubsets7.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend10<-
```

```
  , statistic="cp", legend = FALSE,  main = "Mallow Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend10

which.max(summary.out7$adjr2)
summary.out6$which[5,]

best.modellogptfindollars<- lm(lofptfindollar ~ Demand + HydroProduction
+ WindProduction+ GDP + HDI,
                    data=logptfindollar)
summary(best.modellogptfindollars)


#LogPriceDOLLARS and LogDemand
pairs(logptfindollarlogdemandloghydro, panel = panel.smooth)

regsubsets8.out <-
  regsubsets(  logptfindollar ~ logdemand + HydroProduction +
WindProduction + HDI+ GDP+ Population,
             data = logptfdollarlogdemand,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
             force.in = NULL, force.out = NULL,
             method = "exhaustive")

summary.out8 <- summary(regsubsets8.out)
as.data.frame(summary.out7$outmat)
regsubsets8.out

plot(regsubsets8.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend11<-
  subsets(regsubsets8.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend12<-
  subsets(regsubsets8.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend11

which.max(summary.out8$adjr2)
summary.out8$which[5,]
```

```
best.modellogptfindollarslogdemand<- lm(logptfindollar ~ logdemand +
HydroProduction + WindProduction+ GDP + HDI + population,
                          data=logptfdollarlogdemand)
summary(
  best.modellogptfindollarslogdemand)

k<-lm(logptfindollar ~ logdemand + HydroProduction + WindProduction+
GDP + HDI + Population,
   data=logptfdollarlogdemand)



#LogPrice in Dollars LogDemand LogHydro
pairs(logptfindollarlogdemandloghydro, panel = panel.smooth)

regsubsets9.out <-
  regsubsets(  logPtfinDollar ~ logDemand + logHydro + WindProduction +
HDI+ GDP+ Population,
             data = logptfindollarlogdemandloghydro,
             nbest = 1,       # 1 best model for each number of predictors
             nvmax = NULL,    # NULL for no limit on number of variables
             force.in = NULL, force.out = NULL,
             method = "exhaustive")

summary.out9 <- summary(regsubsets9.out)
as.data.frame(summary.out9$outmat)
regsubsets9.out

plot(regsubsets9.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend13<-
  subsets(regsubsets9.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend14<-
  subsets(regsubsets9.out, statistic="cp", legend = FALSE,  main = "Mallow
Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend13

which.max(summary.out9$adjr2)
summary.out8$which[4,]
```

```
best.modellogptfindollarslogdemandloghydro<- lm(logPtfinDollar ~
logDemand + logHydro + GDP + HDI,
                                    data=logptfindollarlogdemandloghydro)
summary(
  best.modellogptfindollarslogdemandloghydro)



#logptfindollarlogdemandloghydrologwind

pairs(logptfindollarlogdemandloghydrologwind, panel = panel.smooth)

regsubsets10.out <-
  regsubsets(  logPtfinDollar ~ LogDemand + LogHydro + LogWind +
HDI+ GDP+ Population,
              data = logptfindollarlogdemandloghydrologwind,
              nbest = 1,       # 1 best model for each number of predictors
              nvmax = NULL,    # NULL for no limit on number of variables
              force.in = NULL, force.out = NULL,
              method = "exhaustive")

summary.out10 <- summary(regsubsets10.out)
as.data.frame(summary.out10$outmat)
regsubsets10.out

plot(regsubsets10.out, scale = "adjr2", main = "Adjusted R^2")
#variables which have black box at the top of the table should be considered
#in our model
library(car)
layout(matrix(1:1, ncol = 1))
## Adjusted R2
res.legend15<-
  subsets(regsubsets10.out, statistic="adjr2", legend = FALSE, main =
"Adjusted R^2")
abline(a = 1 , b = 1, lty = 2)

## Mallow Cp
res.legend16<-
  subsets(regsubsets10.out, statistic="cp", legend = FALSE,  main =
"Mallow Cp")
abline(a = 1 , b = 1, lty = 2)
res.legend16

which.max(summary.out10$adjr2)
summary.out10$which[5,]

#Standard residuals Plot
best.modellogptfindollarslogdemandloghydronowind.stdres =
rstandard(best.modellogptfindollarslogdemandloghydronowind)
best.modellogptfindollarslogdemandloghydronowind.stdres
```

```
plot(best.modellogptfindollarslogdemandloghydronowind.stdres ,
     main = 'Normal Probability Plot of Standard Residuals',
     ylab='residuals'
     )
abline(0,0)  #horizontalline

library(car)
avPlots(best.modellogptfindollarslogdemandloghydronowind)


#kullanilan en iyi regresyon model
best.modellogptfindollarslogdemandloghydronowind<- lm(logPtfinDollar ~
LogDemand + LogHydro+ GDP + HDI,

data=logptfindollarlogdemandloghydronowind)
```

Appendix B2: ARIMA R Codes

```
tsarima<- ts(arimadata, frequency=12, start=c(2009,12) )
plot.ts(tsarima , main=" Monthly demand-weighted average electricity
prices in USD", ylab="Price in USD")
tsarima2<- ts(TL_, frequency=12, start=c(2009,12) )
plot.ts(tsarima2 , main=" Monthly demand-weighted average electricity
prices in TL", ylab="Price in TL")
arimamodel <- auto.arima( tsarima, trace=TRUE)
arimamodel <-Arima(timeseriesarima , order=c(1,0,0) , seasonal = c(2,0,0))
plot(forecast(arimamodel,h=24))
res <- residuals(arimamodel)
tsdisplay(res)
arimaindollar1 <- as.numeric(arimaindollar)
```

Appendix B3: Holt Winter Exponential Smoothing R Codes

```
HW2 <- HoltWinters(tsarima, seasonal="additive") #Estimated Holt
Winters
#plot the fitted values
F2<- forecast(HW2)
plot(F2, h=24)
plot(forecast(HW1,h=24))
```

Appendix C: MATLAB Codes

Appendix C1: 3 train 1 test ANN Model Matlab Code

```
close all; clear all; clc;
% data = [data(:,1) data(:,3:7)];
  data = xlsread('dataann') save datann
 data  load datann
data = [data(:,1) data(:,3:7)];
%outputtest2 = data(82:end,6);
%outputtrain=data(2:2:end,6)';
%outputtest=data(1:2:end,6)';
inx = 1:92;
outputtrain = data(find(mod(inx,4)~=0),6)';
 outputtest = data(find(mod(inx,4)==0),6)';
  data=data(:,1:5);
data=mapminmax(data');
data = data';  input = data(1:end,1:5);
%inputtrain=input(2:2:end,1:5)';
%inputtest = input(1:2:end,1:5)';
inx = 1:92;
inputtrain = input(find(mod(inx,4)~=0),1:5)';
 inputtest = input(find(mod(inx,4)==0),1:5)';
%inputtest2 = data(82:end,1:5)' %output = data(1:81,6);
%outputtrain=output(2:2:end)';
%outputtest = output(1:2:end)';
 %output_test=mapminmax(output_test);
%input_test=mapminmax(input_test);
net = newff(minmax(inputtrain), [5 1],{'tansig' 'purelin'},'trainlm');%'trainrp'
%net = feedforwardnet(4*length(giris_egit(:,1)),'trainrp');
net.trainParam.epochs = 10000; net.trainParam.goal = 0.0003;
net.trainParam.min_grad = 1e-30; net = train(net,inputtrain,outputtrain);
outputtrainest = net(inputtrain);
outputtestest = net(inputtest);
rmse1 = sqrt(immse(outputtrain, outputtrainest))
 rmse2 = sqrt(immse(outputtest, outputtestest))
 figure(1) plot(outputtrain,'*-');
hold on;
plot(outputtrainest,'o--r');
legend('observed data','estimated data')
hold off;  figure(2) plot(outputtest,'-*');
hold on; plot(outputtestest,'-*');
hold off;
[outputtest; outputtestest]'
```

Appendix C2: Forecast with ANN Model Matlab Code

```
close all
 %load
data data = xlsread('dataann')
```

```
save datann data  load datann
  data = [data(:,1) data(:,3:7)];
 outputtest = data(83:end,6)';
 outputtrain=data(1:1:82,6)';
data=data(:,1:5);
 %training data is from 2009/12-2016/09
 %test data is from 2016/10- 2017/07  data=mapminmax(data');
 data = data';
 input = data(1:82,1:5);
 % In input data : 1st column=Demand(TWh)
 %               2nd column=Hydro Production(GWh)
 %               3rd column=HDI
 %               4th column=GDP
 %               5th column= Population
inputtrain=input(1:1:end,1:5)';
 inputtest = data(83:end,1:5)'
%output = data(1:81,6);
%outputtrain=output(2:2:end)';
 %outputtest = output(:2:end)';   %output_test=mapminmax(output_test);
%input_test=mapminmax(input_test);
net = newff(minmax(inputtrain), [5 1],{'tansig' 'purelin'},'trainlm');%'trainrp'
%net =feedforwardnet(4*length(giris_egit(:,1)),'trainrp');
net.trainParam.epochs = 10000; net.trainParam.goal = 0.0003;
net.trainParam.min_grad = 1e-30;
 net = train(net,inputtrain,outputtrain);
outputtrainest = net(inputtrain);
outputtestest = net(inputtest);
rmse1 = sqrt(immse(outputtrain, outputtrainest))
rmse2 = sqrt(immse(outputtest, outputtestest)) plot(outputtrain);
hold plot(outputtrainest);
 figure plot(outputtest,'-*');
 hold plot(outputtestest,'-*');   [outputtestest;outputtest]
```

Appendix D: Forecast Results and Error Measures

	Price in USD	ANN	(At-Ft)ANN	Absolute(At-Ft)ANN	Square(At-Ft)ANN	Percentage ANN
Oct-16	47,17732757	42,809	4,368327567	4,368327567	19,08228573	9%
Nov-16	46,79882748	42,809	3,989827476	3,989827476	15,91872329	9%
Dec-16	65,08865603	54,283	10,80565603	10,80565603	116,7622022	17%
Jan-17	50,16914786	54,283	-4,113852142	4,113852142	16,92377945	8%
Feb-17	48,52335633	42,809	5,714356331	5,714356331	32,65386827	12%
Mar-17	40,37567376	42,914	-2,538326237	2,538326237	6,443100087	6%
Apr-17	40,48722438	42,809	-2,32177562	2,32177562	5,390642028	6%
May-17	43,5993166	42,809	0,790316603	0,790316603	0,624600333	2%
Jun-17	42,95587484	42,809	0,146874837	0,146874837	0,021572218	0%
Jul-17	50,64369366	54,2838	-3,640106344	3,640106344	13,2503742	7%

	Price in USD	SARIMA	(At-Ft)SARIMA	Absolute(At-Ft)SARIMA	Square(At-Ft)SARIMA	percentage SARIMA
Oct-16	47,17732757	46,683	0,494327567	0,494327567	0,244359743	1%
Nov-16	46,79882748	49,692	-2,893172524	2,893172524	8,370447255	6%
Dec-16	65,08865603	47,858	17,23065603	17,23065603	296,8955071	26%
Jan-17	50,16914786	48,55	1,619147858	1,619147858	2,621639785	3%
Feb-17	48,52335633	43,185	5,338356331	5,338356331	28,49804831	11%
Mar-17	40,37567376	41,292	-0,916326237	0,916326237	0,839653773	2%
Apr-17	40,48722438	37,377	3,11022438	3,11022438	9,673495696	8%
May-17	43,5993166	38,735	4,864316603	4,864316603	23,66157602	11%
Jun-17	42,95587484	41,664	1,291874837	1,291874837	1,668940594	3%
Jul-17	50,64369366	41,668	8,975693656	8,975693656	80,5630766	18%

	Price in USD	HW- ES	(At-Ft)HW	Absolute(At-Ft)HW	Square(At-Ft)HW	percentage HW
Oct-16	47,17732757	44,806	2,371327567	2,371327567	5,623194428	5%
Nov-16	46,79882748	47,355	-0,556172524	0,556172524	0,309327877	1%
Dec-16	65,08865603	53,931	11,15765603	11,15765603	124,493288	17%
Jan-17	50,16914786	47,95	2,219147858	2,219147858	4,924617214	4%
Feb-17	48,52335633	36,048	12,47535633	12,47535633	155,6345156	26%
Mar-17	40,37567376	31,692	8,683673763	8,683673763	75,40619002	22%
Apr-17	40,48722438	31,143	9,34422438	9,34422438	87,31452927	23%
May-17	43,5993166	30,156	13,4433166	13,4433166	180,7227613	31%
Jun-17	42,95587484	35,425	7,530874837	7,530874837	56,71407581	18%
Jul-17	50,64369366	36,231	14,41269366	14,41269366	207,7257384	28%

	Price in USD	MR(Multiple Regression)	(At-Ft)MR	Absolute(At-Ft)MR	Square(At-Ft)MR	percentage MR
Oct-16	47,17732757	43,156	4,021376773	4,021376773	16,17147115	9%
Nov-16	46,79882748	43,979	2,81983161	2,81983161	7,951450307	6%
Dec-16	65,08865603	49,367	15,72131441	15,72131441	247,1597267	24%
Jan-17	50,16914786	46,772	3,397017092	3,397017092	11,53972512	7%
Feb-17	48,52335633	40,350	8,172889397	8,172889397	66,7961211	17%
Mar-17	40,37567376	41,051	-0,675256863	0,675256863	0,455971831	2%
Apr-17	40,48722438	36,572	3,914746826	3,914746826	15,32524271	10%
May-17	43,5993166	37,053	6,546075361	6,546075361	42,85110263	15%
Jun-17	42,95587484	37,895	5,060476272	5,060476272	25,6084201	12%
Jul-17	50,64369366	43,809	6,834680388	6,834680388	46,71285601	13%