SHORT TERM LOAD FORECASTING IN ELECTRICTY MARKETS IN TURKEY

ZEYNEP DUYGU TEKLER
B.S., Industrial Engineering, IŞIK UNIVERSITY, 2015

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

IŞIK UNIVERSITY
2017

IŞIK UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

Short Term Load Forecasting in Electricity Markets in Turkey

ZEYNEP DUYGU TEKLER

APPROVED BY:

Assist. Prof. Dr. Kemal Sarıca (Thesis Supervisor) Işık University ________________

Assist. Prof. Dr. Demet Özgür Ünlüakın Işık University ________________

Assist. Prof. Dr. Burak Çavdaroğlu Kadir Has University ________________

APPROVAL DATE: 16../8/..17

SHORT TERM LOAD FORECASTING IN ELECTRICITY MARKETS IN TURKEY

Abstract

As energy consumption rises, forecasting electricity parameters becomes a significant advantage on efficient power system applications, planning and decision making in deregulated power markets. An accurate short term load and price forecasting model is crucial for efficient decision making, management and utilization to gain economic optimization and avoiding unprofitable operations as well as inefficiencies in generation,transmission and distribution from both consumers and producers perspective in competitive electricity markets like Turkish power industry.

In this study, time series analysis including lagged variables that have been presented in autoregressive models in combination of specific categorical variables (hours of day, days of the week, months of the year and special events of Turkey) and environmental indicators as hourly temperature data in terms of heating-cooling degree hours have been applied in short term load and price forecasting, the day ahead hourly forecast of electricity load and market price.With using AR parameters for load and ARIMA parameters for price, 4 different load models in years 2011 and 2012 and a price model for 2015 constructed and with the effect of categorical variables and environmental indicators, new composite models proposed by applying multiple linear regression to forecast future loads and prices with high accuracy. As a result, the comparison of actual and observed data is studied and the power of model is tested with illustrating on various regression tests. Consequently, the results have shown that proposed models gave low percent of errors with extremely accurate day ahead forecasts considering Turkeys electricity load and price profile.

Keywords: Short term load and price forecasting, time series analysis, multiple linear regression analysis

TÜRKİYE ELEKTRİK PİYASALARINDA KISA DÖNEM YÜK VE FİYAT TAHMİNİ

Özet

Enerji tüketimi artışıyla beraber serbest piyasa içerisinde elektrik parametreleri tahmini, etkin güç sistemi uygulamalarında, planlamada ve karar mekanizmaları üzerinde önemli bir avantaj sağlamaktadır. Hata oranı düşük bir kısa dönem yük ve fiyat tahmin modeli, yönetimde, etkili karar ve fayda mekanizmalarında ekonomik güç optimizasyonunu sağlamak ve kârsız operasyonlardan kaçınmak, bunun yanında enerji üretimi, aktarımı ve dağıtımı üzerinde Türk enerji piyasası gibi rekabetçi piyasalarda hem üretici hem de tüketici perspektifinden verimli kararlar almak adına büyük önem arz etmektedir.

Bu makalede, belirli kategorik değişkenlerle (günün saatleri, haftanın günleri, yılın ayları, Türkiye'nin özel günleri) birlikte ısıtma ve soğutma derecelerine bağlı saatlik sıcaklık verisi gibi çevresel değişkenlerin olduğu otoregresif terimleri içeren zaman serileri analiziyle saatlik kısa dönem yük ve fiyat tahmini yapılmıştır. Yük için AR modeli fiyat içinse geleneksel ARIMA modeli kurulmuş, parametreler yardımıyla kategorik değişkenler ve çevresel faktörlerin kombinasyonu sağlanılarak 2011-2012 yılları için yük tahmini adına 4 farklı ve 2015 yılı için fiyat üzerinde çoklu lineer regresyon yöntemi kullanılarak kompozit modeller oluşturulmuştur. Bu modellerin kesinliğinin arttırılması amaçlanmıştır. Çeşitli regresyon testleriyle birlikte gerçek ve gözlemlenen değerlerle karşılaştırılarak incelenmiş, modelin gücü test edilmiştir. Sonuçlar, önerilen modellerin Türkiye'nin elektrik yük ve fiyat profiline göre hata oranının önemli ölçüde düşük olduğunu göstermektedir.

Anahtar kelimeler: Kısa dönem yük ve fiyat tahmini, zaman serileri analizi, regresyon analizi, otoregresif model

Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I want to thank the Department of Industrial Engineering and the Işık University community and I furthermore thank to my thesis supervisor Assist. Prof. Dr. Kemal Sarıca who encouraged me to go ahead with his guidance,contributions, motivation and valuable hints for my thesis.

I wish to thank Assist. Prof. Dr. Burak Çavdaroğlu for sharing his valuable energy, knowledge and for serving on my thesis committee. I am thankful to Assist. Prof. Dr. Demet Özgür Ünlüakın for her support as a member of my thesis committee.

Especially, I would like to give my special thanks to my family, my mother, my grandmother and dearest friends whose patient love enabled me to complete this work and invaluable support and encouragement not only during this study but also throughout my life.

To my mother and grandmother...

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AR	**A**uto**R**egressive
ARIMA	**A**uto**R**egressive **I**ntegrated **M**oving **A**verage
ACF	**A**uto **C**orrelation **F**unction
CDH	**C**ooling **D**egree **H**ours
EPIAS	**T**urkish **E**lectricity **M**arkets **M**anagement **C**ompany
EXIST	**E**nergy **E**xchange **I**stanbul
GDP	**G**ross **D**omestic **P**roduct
GOP	**D**ay **A**head **M**arket
HDH	**H**eating **D**egree **H**ours
IBM	**I**nternational **B**usiness **M**achines
MAPE	**M**ean**A**bsolute **P**ercentage**E**rror
OECD	**O**rganization for **E**conomic **C**ooperation and **D**evelopment
PACF	**P**artial **A**uto **C**orrelation **F**unction
PMUM	**F**inancial **M**arket**S**ettlement **C**enter
PPP	**P**ower **P**urchasing **P**arity
RMSE	**R**oot **M**ean **S**quare**E**rror
SPSS	**S**tatistical **P**ackage for the **S**ocial **S**ciences
TEIAS	**T**urkish **E**lectricity **T**ransmission **C**ompany
WU	**W**eather **U**nderground

Chapter 1

INTRODUCTION AND BACKGROUND

1.1 Introduction

Industry and technology continue to evolve rapidly in today's world. Energy as one of the most important sources of this development, is among the indispensable parts of daily life. A very significant amount of this energy is electricity energy that can be obtained from many different sources and used in various areas.

Planning of energy production, transmission and distribution has a critical role in order to meet increased energy need. On the other hand, environmental factors such as global warming and political external dependence on energy problems emphasize the necessity of savings in electricity consumption. Therefore it is required to provide the balance between energy production and consumption for operationally efficient market operations.

In Turkish electricity markets, considering supply side, alternative energy sources are enabled when its necessary as well as in order to keep down the contribution of low efficient traditional energy. Hence energy must be kept in the network against the sudden increases in demand side, electricity consumptions must be followed up. According to demand forecasts given by Energy Exchange Istanbul (EXIST), market is balanced under the errors of these estimates, which incurs extra costs for balancing operations. Accurate, short-term load and price forecasting is a necessary instrument that provides crucial information for power producers and consumers to develop accurate bidding strategies in order to maximize their profit.

Energy demand forecast analysis is done by Turkish Electricity Transmission Company (TEIAS) and published by Financial Market Settlement Center (PMUM) in Turkey. However, these general estimates are not enough according to regional planning since the coverage of these distribution areas and the population dynamics demonstrate that general forecasts are insufficient. Hence, according to changes in settlement areas, usage habits and user profiles, the first step of planning energy systems is efficient short term load and price forecasts.

Price and load forecasts can be categorized into short term, medium term and long term according to the time inherited in the models. Especially, short term price and load forecasting aimed to predict system over an interval ranging from one hour to one week, provides considerable saving potential for economic, accurate and secure power operations. In load and price predictions, error forecasting has a favorable effect on operational decisions like economic scheduling of generating capacity and providing adequate electricity supply, system security assessment and blackout risk minimization. Hence, it is exceedingly important to forecast the electric price and load for short term with a small margin.

This thesis is organized as follows. In chapter 2, studies in the literature to short term load and price forecasting are summarized and the short term load and price forecasting techniques are elaborated. In chapter 3, load and price data types with categorical variables according to Turkish calendar and environmental indicators are presented. Time series analysis and traditional AR and ARIMA models for parameter estimation are examined and the methodology for load and price models employed in this study is presented in chapter 4. Load and price forecasting results and performances are shared in chapter 5. Finally, in chapter 6, results of the study are discussed and evaluated.

1.2 Basic Facts of Turkish Energy Profile

1.2.1 Developments in Turkish Electricity Markets

Turkey's electricity supply industry is ruled by large, publicly owned companies. The first larger scale power plant was built in Istanbul in 1913. In 1935, several governmental institutions with electricity planning authority were established such as Electric Power Resources Survey and Development Administration (EIEI). In 1950's -mostly publicly owned entities - power plants on a larger scale construction were begun. On the other hand, private operators were Cukurova Electric Company (CEAS) and KEPEZ electric company operating under state concession. By 1970, with growing consumption and capacity government established the Turkish Electricity Authority(TEK) as a whole state-owned and run entity. As a result, all electricity production, transmission and distribution activities were handled within TEK and by 1982, all plants and unions were transferred to TEK.

The first wave of the liberalization process began in 1980's, mechanisms have been developed that allowed private and foreign participation in the power industry without outright privatization. Within the framework of privatization and reform, TEK has been split into two separate state-owned companies as Turkish Electricity Generation-Transmission Company (TEAS) and Turkish Electricity Distribution Company (TEDAS)

Between 2001 and 2004 changes in power industry brought new reforms and as a result, TEAS has been split in to three companies: Turkish Electricity Transmission Company(TEIAS), Electricity Generation Company(EUAS) and Turkish Electricity Trading and Contracting Company

With the electricity market balancing and settlement regulation in 2004, system of Day-ahead Planning system has gone into effect for the first time to simplify the management of real time balance system and improve system security and reliability with targeting generation optimization in Turkish electricity markets.

However electricity prices have remained stable for a long time and input costs have remained to be constant, producers have begun to seek solutions for more efficient trade structure, along with other changes in the direction of intensive demands of thermal power plant manufacturers, which have a significant share of the production scale.

With a new balancing and settlement regulation published in 2009, a reconciliation with a balanced market structure was carried out every hour. This market, which is run under the name of Day-ahead Planning , can be regarded as a transition process. It aimed to ensure that the Market Participants were trained and facilitated in the transition to the Day Ahead Market(GOP), which was defined as the final structure.

The necessary agreements have been signed with the participants before the transition from Day Ahead Planning to Day Ahead Market, and various trainings have been given at different times and places regarding Day Ahead Market. Moreover, it has been supported through a demo that the participants of the market have to make transactions in a fictitious system which does not carry any financial value in a real sense and prepare them for the Day Ahead Market. The aim here is to increase the number of participants in the Day-Ahead Market, which will be a voluntary market, when the market becomes operational. Unfortunately, targeted participation has not been achieved for a long time.

Market participants have been away from this new system because of prejudice. In order for the market participants to break this resistance in some way, some great players have been trying to enter the demo system and test the system as much as possible. On the other hand, the trainings given were made more frequent and the differences between the aim, structure and situation of Day Ahead Market were continued to be explained.

Turkish energy sector is proceeding with determined steps towards liberalization under the lead of electricity markets. Significant progress has been made, especially with the steps taken since 2009. Including Day-ahead Planning and

Electricity Market Balancing and Settlement Regulation, first phase of "organized electricity markets" has been begun and as a result, all of the electricity generation assets apart from the large-scale and strategically valued hydropower plants have been privatized. In 2011, second phase of the "organized electricity markets" has started. Day-ahead Markets and mechanism of warrant have been implemented. In order to increase market confidence and transparency, new electricity market law has entered into force. Important step of electricity wholesale have been taken towards the market.

Most significantly, it is moved to a stage in which the investment risk is overwhelmed by the producers over the consumer. Along with the completion of the privatizations, the authorities and responsibilities of all relevant parties, in particular the market operator, have been redefined in order to establish an effective wholesale market. In addition, the privatization process of distribution companies has been successfully completed; Distribution and retail activities are separated. As of July 1, 2015, the Day-ahead Market has become operational. EPİAS (Energy Electricity Markets Management Company Operations) which was established in charge of planning, establishing, developing and operating the energy market in an effective, transparent, reliable way and satisfying market needs started its operations with a Market Operation license as of September 1, 2015. After 2015, TETAS has begun to purchase the electricity generated in the Domestic Coal Fired Power Plants with the quantity and price determined by the Council of Ministers.

1.2.2 Electricity Demand in Turkey

Turkish economy and population have been growing rapidly correspond its demand for energy, especially for electricity, has been increasing fast. After1990s, energy consumption increased about 4.4% per year, with electricity consumption growing at an average annual rate of about 8.5%. [1] According to Central Bank of Turkey researches, in years between 2004 and 2012, Turkey has the highest average annual growth rate of GDP and the fifth highest average annual growth rate of, PPP-adjusted, per capita income among all the OECD countries. [2] Across the globe, Turkey has the second highest energy consumption growth after China, making it one of the fastest growing energy markets. Figure 1 shows the relationship between GDP and electricity demand for the years between 2004 and 2010.

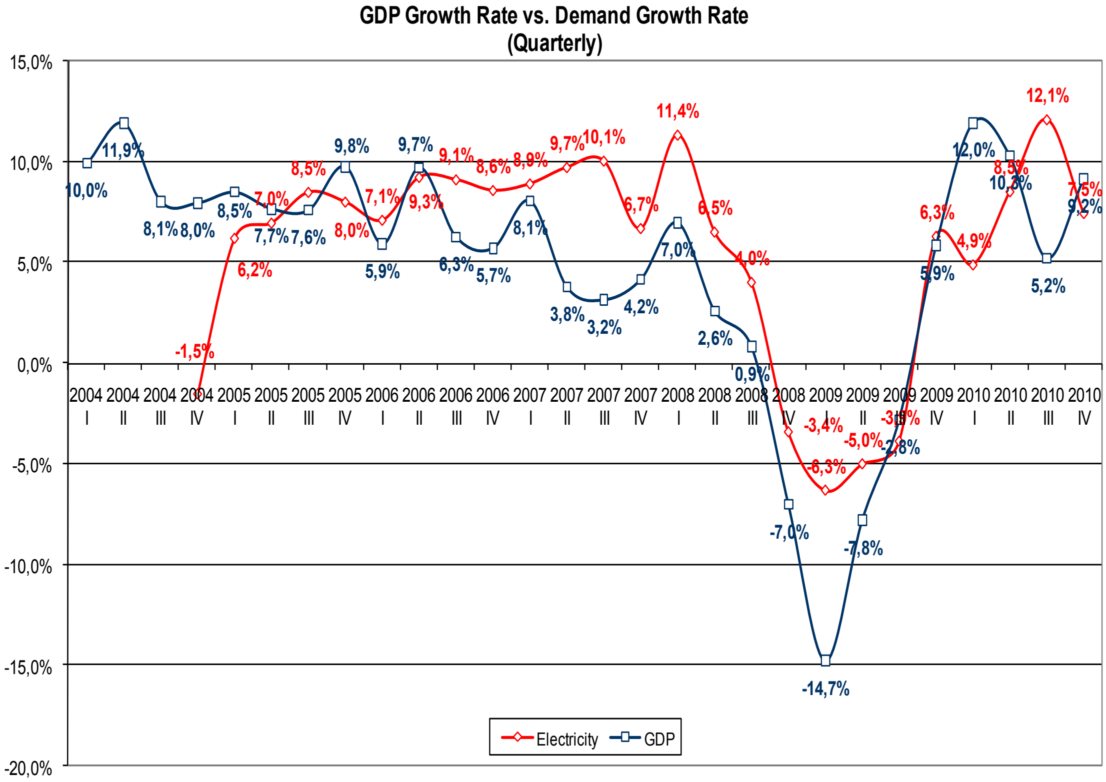

Figure 1.1: GDP and Electricity Demand growth rates between 2004 and 2010

Currently, in Turkey, electricity generation increased by 4.4% and consumption increased by 4.7% in 2016 compared to 2015. Comparing years 2010-2016, it is seen that the generation increased by 29.4% and the consumption increased by 32.3%. In 2016, peak demand increased by 3.3% , installed power increased by 7.3% compared to the previous year.

As a result of economic growth, annual electric energy consumption growth has been around 4.6% on average for many years. Due to the formation of the interconnection requirement with Europe in the import of electricity, the increase has started in 2011. However, in the year 2016, electricity imports decreased by 10.3% and exports decreased by 54.9% compared to the previous year.[3]

Table 1.1: Table of amount of electrical energy change in years 2010-2016

	Unit	2010	2015	2016	2010-2016(%)	2015-2016(%)
Installed Power	MW	49.524	73.147	78.497	58.5	7.3
Peak Demand	MW	33.392	43.289	44.734	34.0	3.3
Generation	GWh	211.208	261.783	273.387	29.4	4.4
Import	GWh	1.144	7.135	6.400	459.4	-10.3
Export	GWh	1.918	3.194	1.442	-24.8	-54.9
Consumption	GWh	210.434	265.724	278.345	32.3	4.7

Energy demand forecast projections for Turkey indicate that there will be a chronic increase in demand for energy, particularly for electricity, in the next two decades.With this massive growth and consumption, electricity energy turns into powerful input for social, environmental,technical and economic improvements. In that case, many elements become vital in electric power generation, distribution and consumption with the deregulation of the Turkish power industry. Due to the deregulation and competition in the electricity markets, sudden and unexpected changes in the industry create the necessity for accurate load and price forecasts even more. Consequently, to contribute economically efficient operations and power requirements, price and load forecasting has a substantial key.

Chapter 2

LITERATURE SURVEY

Short term load forecasting methods are generally classified in two categories, traditional statistical methods and modern artificial intelligence methods.The difference between two major methods is, in traditional statistical techniques equations are achieved by using the relationship between load and factors which affect the historical data after training on the other hand modern artificial intelligence techniques assimilate the way humans think and estimate using the knowledge of past data to forecast future load. While statistical methods mainly consist of regression methods, least square estimations and time series analysis, artificial intelligence methods are artificial neural networks, fuzzy logic, genetic algorithms so on. Especially in recent years, many load forecasting studies have been applied on the improvement of forecasting procedures. Some of these studies and methods listed below

First published short term load forecasting work was done by Heinemann et al. In early 1960's, the relationship between temperature and load has been analyzed for the first time [4]. In 1971, first load forecasting system was developed by Lijesen and Rosing using statistical approach. [5]. In 1987, Hagan and Behr forecasted load using a time series model which takes the load at any time depends mainly on previous load patterns [6]. After 90's instead of statistical approaches, artifical intelligence techniques were implemented. Artificial neural network is used in short term load forecasting in by 1991 Park et al.[7] Furthermore, in 1995

Srinivasan et al applied hybrid methods such as fuzzy logic and neural network were used for short term load forecasting systems.[8]

Considering studies in Turkey, short term load forecasting system was developed by Erkmen and Ozdogan in 1997 using artifical neural network method for the first time and temperature was not included as an input variable with the improvement of the first research [9].Topalli and Erkmen used artificial neural network method by using hourly load and calendar as input variables in 2003 [10].With the combination of these two studies Topalli et al. developed an artificial neural network model that included both hourly load, calendar data and temperature data as input variables in 2006 [11].Hamzaçebi et al. made a signficant study to forecast electricity load until 2010 in Turkey by using regression analysis and artificial neural network in 2004 [12].Similarly, Ceylan and Demirören used actual temperature and load data of 2002-2003 to forecast hourly loads of Gölbaşı, Turkey [13]. Furthermore, Yalçınöz et. al. forecasted month-ahead loads between 2001 to 2004 in Niğde, Turkey with 5 different forecasting methods. [14]

2.1 Short Term Load Forecasting Methods

2.1.1 Traditional Statistical Techiques

2.1.1.1 Multiple Regression Method

Multiple regression analysis is one of the most commonly used statistical methods. It uses the technique of weighted least-squares estimation for load forecasting. It is useful for analyzing the statistical relationship between total load and categorical variables like weather conditions, day types and special events influences. Using historical data the regression coefficients are calculated by an equally or exponentially weighted least-squares estimation.

The multiple regression model can be written as:

$$Y(t) = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i(t) + r(t) \tag{2.1}$$

where

$Y(t)$ is the dependent total load value at time t, $X_i(t)....X_n(t)$ are indepedent explanatory variables like day types, weather or special events, $r(t)$ is the residual load at time t means the unexplained white noise component, α_0 represents the constant standard load $\alpha_i...\alpha_n$ terms are regression parameters that are estimated varying coefficients

After obtaining estimated varying coefficients and fitting the linear regression model, goodness of fit tests can be applied for the model through observed data in order to analyze the power of the regression model. Most commonly used measurement factor is R-square as the square of multiple correlation coefficient that measures the proportion of variation in the dependent variable explained by independent variables. It assumes that every single variable explains the variation in the dependent variable whereas adjusted R-square assumes the percentage of variation explained by only indepedent variables that actually affect the dependent variable. Even high value of R-square shows a good sign, it is not enough to say the data fits the model well.

Also root mean square error (RMSE) is the square root of the variance of the residuals. It shows the absolute fit of the model to the data how close the observed data points are to the models predicted values. It is calculated as the square root of the average of squared differences between forecasted and actual observations. The lower value of RMSE points out better fit. Similarly, mean absolute percentage error measures (MAPE) the size of error in percentage terms in other words the quantity used to measure how close forecasts are to eventual outcomes. It is calculated as the average oversample of absolute the differences between actual and forecasted observations.

Next step is to check the significance of overall model and each explanatory variable. If results found are valid and have meaning then residual terms are needed to be analyzed. Residual analysis can be done through histograms and normal probability plots in order to assess the quality of the regression. Histogram can be used to check whether the variance is normally distributed with a mean of zero and the residuals should exhibit a symmetric bell-shaped distribution which is evenly distributed around zero. Also normal probability plot is another way to learn whether the error terms are normally distributed.

The relationship between explanatory variables can be judged by examining a quality called the variance inflation factor (VIF). A value of variance inflation factors greater than 10 is often taken as a signal that the data have collinearity problems.

The Durbin-Watson statistic d is widely known test in autocorrelation of regression analysis. Evidence of autocorrelation is indicated by the deviation of d from the numerical value 2. It reports a test statistic, with a value from 0 to 4 where;

- if d=2 no autocorrelation

-if $0 < d < 2$ is positive autocorrelation

-if $d > 4$ is negative autocorrelation

However according to a rule of thumb, test statistic values in the range of 1.5 to 2.5 are relatively normal. Values outside of this range could be a cause for concern.

2.1.1.2 Stochastic Time Series Method

Time series analysis is a dynamic type of forecasting method that is also a popular technique used in load forecasting. Although it is complex to use, requires more time and historical data for prediction, efficient results are obtained in the context of energy generation and demand by using time series analysis. This method based on the time series is converted into a stationary time series by differentiation while the white noise is filtered from the other series. In this technique load patterns are time series signals with seasonal, weekly and daily forecasts.

In time series, autocorrelation and partial autocorrelation coefficients are extremely beneficial in identifying and modeling patterns.

Autocorrelation Function (ACF) Autocorrelation Function is the correlation of a time series with its own past and future values, in other words it shows the correlation of series with itself at different lags. It is one of the common used tools in time series analysis that used to determine stationary and seasonality. Therefore, identification of an moving average(MA) model is more useful with autocorrelation function.

Given measurements, $Y_1, Y_2, ..., Y_N$ at time $X_1, X_2, ..., X_N$, the lag k autocorrelation function(r) is defined as,

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \tag{2.2}$$

Although the time variable, X, is not used in the formula for autocorrelation, the assumption is that the observations are equi-spaced. By definition, autocorrelation is a correlation coefficient. However, instead of correlation between two different variables, the correlation is between two values of the same variable at times X_i and X_{i+k} [15]

Partial Autocorrelation Function (PACF) Partial Autocorrelation Function is the correlation between time series and each of its intermediate lagged values. Identification of an autoregressive model(AR) is similarly more useful with partial autocorrelation function thereby it is functional to detect order of an autoregressive model by controlling for the values of time series at all shorter lags.

For a time series, the partial autocorrelation between X_t and X_{t-h} is defined as the conditional correlation between X_t and X_{t-h}, conditional on $X_{t-h+1}, ..., X_{t-1}$ the set of observations that come between the time points t and $t-h$. Second order(lag) partial autocorrelation is given,

$$\frac{\text{Covariance}(x_t, x_{t-2}|x_{t-1})}{\sqrt{\text{Variance}(x_t|x_{t-1})\text{Variance}(x_{t-2}|x_{t-1})}} \tag{2.3}$$

Basic form of this approach is known as ARMA modeling (autoregressive moving average) combine autocorrelation methods (AR) and moving averages (MA) into a composite model of the time series.Also when differencing is included in the context, ARIMA or Box-Jenkins modeling is developed. [16] [17] . Time series analysis has been employed in many fields like monitoring industrial processes, finance and in electrical load forecasting over the years.

Each model examined separately as Autoregressive Model (AR) , Moving Average Model (MA), Autoregressive Moving-Average(ARMA) and Autoregressive Integrated Moving-Average(ARIMA) Model as below;

Autoregressive(AR) Model If load is assumed to be a linear combination of previous loads then the autoregressive (AR) component of an ARMA model can be expressed in the form:

$$X_t = \alpha_1 X_{t-1} + ...\alpha_p X_{t-p} + Z_t \tag{2.4}$$

in closed form,

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t \tag{2.5}$$

where

X_t is the predicted load at time k and the α_i terms are autocorrelations coefficients, Z_t is a residual error term at lags 1,2..p In order to fit an autoregressive model to an observed dataset, the sum of squared errors are aimed to be minimized using the smallest number of terms that provide a satisfactory fit to the data. Models of this type are described as autoregressive.

Although in theory an autoregressive model might provide a good fit to an observed dataset, it would generally require previous elimination of any trend and periodic components, and even then might need a large number of terms in order to provide a good fit to the data. However, by combining the AR models with MA models, mixed models can be applied in a wide range of situations. These models are known as ARMA and ARIMA models, which are described in the following subsections.[18]

Moving Average (MA) Model Moving average (MA) models can be used to provide a good fit to some datasets, and variations on these models that involve double or triple exponential smoothing can handle trend and periodic components in the data. Furthermore, such models can be used to create forecasts that imitate the behavior of earlier periods.

Mean values computed over short periods, either preceding the current period or centered on the current period, are often more useful. Because such mean values will vary, or move, as the current period moves from time t=2, t=3, ... etc. they are known as moving averages (Mas). A simple moving average is (typically) the unweighted average of k prior values. An exponentially weighted moving average is essentially the same as a simple moving average, but with contributions to the mean weighted by their proximity to the current time.A range of models can be constructed using moving averages, and these are known as MA models.

A simple form of such models can be written as:

$$X_t = \beta_0 X_t + \beta_1 X_t - 1 + ...\beta_q X_{t-q} \tag{2.6}$$

in closed form,

$$X_t = \sum_{i=0}^{q} \beta_i X_{t-i} \tag{2.7}$$

where X_t is the predicted load at time t and the β_i terms are the weights applied to prior values in time series. The moving average value is estimated as a weighted average of the current and immediate past values.

Let Z_t be a set of independent and identically distributed random variables (a random process) with zero mean and known fixed variance, the process Xt can be written as a moving average of order q in terms of Z_t [18]

$$X_t = \beta_0 Z_t + \beta_1 Z_t - 1 + ...\beta_q Z_{t-q} \tag{2.8}$$

in closed form,

$$X_t = \sum_{i=0}^{q} \beta_i Z_{t-i} \tag{2.9}$$

Autoregressive Moving-Average(ARMA) Model ARMA models combine autocorrelation methods (AR) and moving averages (MA) into a composite model of the time series. Hence with the combination of equation 2.4 for p autoregressive term and equation 2.8 for q moving average term, ARMA model of order (p, q) can be constructed as:

$$X_t = \alpha_1 X_{t-1} + ...\alpha_p X_{t-p} + Z_t + \beta_1 Z_t - 1 + ...\beta_q Z_{t-q} \tag{2.10}$$

In the ARMA model the current value of the time series X_t is expressed linearly in terms of its values at previous periods $[X_{t-1}; X_{t-2}......]$ and in terms of previous values of a white noise $[Z_t, Z_{t-1}.......]$ [18]

On the whole, ARMA model emphasizes the estimated value at time t as the sum of q terms that represent the average variation of random variation over q previous periods (the MA component), plus the sum of p AR terms that compute the current value of x as the weighted sum of the p most recent values.

Nevertheless, this type of model assumes that the time series is stationary, which is infrequently the case. Practically, in many datasets trends and periodicity exists, hence it is necessary to eliminate these effects before applying sthat type of models. Elimination is typically accomplished by adding in the model an initial differencing stage, until the series is at least approximately stationary - exhibiting no obvious trends or periodicities. As with the moving average and autoregressive approach, the differencing process is described by the order of differencing, to illustrate; 1, 2, 3.... Together with these three elements make a triple: (p,d,q) which specifies the type of model applied. In this form, the model is described as an ARIMA model. The letter I in ARIMA responds to differentiation. ARIMA modeling is explained below. [18]

Autoregressive Integrated Moving-Average(ARIMA) Model As it is shown in the subsection 2.1.1.2 combining differencing of a non-stationary time series with the ARMA model provides a powerful models which is known as ARIMA also called Box-Jenkins Models.

Terms of non-seasonal ARIMA (p,d,q) refers to,

- Lags of the stationarized series are called autoregressive **AR** terms as p

-A series which needs to be differenced to be made stationary is an integrated **I** series as d

-Lags of the forecast errors are called moving average **MA** terms as q

Therefore,the first step in the Box-Jenkins procedure also ARIMA is to difference the time series until it is stationary, by that ensuring trend and seasonal components are removed. After pattern of autocorrelations and partial autocorrelations pattern studied in order to determine if lags of the stationarized series and/or lags of the forecast errors should be included in the forecasting equation, suggested model is fitted and examine its residual diagnostics, expecially the residual autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, to check if all coefficients are significant and all of the pattern has been explained. Also, patterns which remain in the ACF and PACF may suggest the need for additional AR or MA terms.

All in all, the decision on what these parameters should be can be guided by a number of basic principles:

(i) the constructed model should be as simple as possible, at least contain as few terms as possible, the values of p and q should be small;

(ii)Least squares principle that is the size of the squared differences between actual value and the estimated value at any past time period should be minimized as much as possible also the residuals from the selected model can then be examined to check if any remaining residuals are significantly different from zero

(iii)The order of autoregressive component (q) should be provided an indiciation by the measured partial autocorrelation at lags 1,2,3..

(iv) According to autocorrelation function plot shape, type of ARIMA can be suggested [18]

2.1.2 Modern Artificial Intelligence Techiques

2.1.2.1 Neural Networks

Neural networks which are also named as artificial neural networks have been an extensively researched load forecasting method for over a decade now and these powerful computational devices have been implemented in many applications because of their fast learning mechanisms. ANNs (artificial neural networks) are inspired by the biological nervous system which is made up of neurons that connected in parallel and feeding forward in several layers. There are different types of ANN that are ordered according to small number of connected layers of elements between network inputs and outputs. Multilayer Perceptron (MLP) is most widely used form for short term load forecasting that employs back-propagation learning algorithm.

2.1.2.2 Expert Systems

Expert Systems(ES) are computational model that based on rules from human experts, that includes 4 basic parts ; a knowledge base, a data base, an inference mechanism and a user interface. Knowledge base is the source which new information and rules are added in it. Rules are implemented into software by system afterwards without any expertise forecasts are made mechanically. This brings a huge advantage on making fast decisions without human assistance.

2.1.2.3 Fuzzy Inference Systems

Fuzzy model have been applied with many variations in the area of load forecasting. It is basically known as generalization of the boolean logic designed for digital circuit design. The advantage of fuzzy logic without precising inputs and most importantly without constructing mathematical model model mapping inputs to outputs and precising inputs appropriately designed fuzzy logic systems works for forecasting load. Under favor of centroid defuzzification process can be used to gain the precise output after the logical processing of fuzzy inputs.

Chapter 3

DATA ANALYSIS

3.1 Data Set

Short term load and price forecasting has been used to determine the forecasted load and price in Turkish Electricity Markets using IBM SPSS Statistics 23.0. The available hourly data for this research Turkey's total actual electric load for the years 2011-2012, total of 17522 load observations obtained through Turkish Electricity Transmission Company (TEIAS) [19] and Turkey's actual price hourly data for 2015, total of 8761 price observations obtained through Turkish Energy Markets Management Company (EPIAS) [20]. Moreover, categorical variables (such as hours of day, days of the week, months of the year and special days of Turkish calendar like festivals, eves of festivals, christmas, 1st of May etc.) and temperature measurements in terms of cooling degree hours (CDH) and heating degree hours (HDH) hourly taken from Istanbul, Turkey for both load and price models obtained through Weather Underground (WU) [21] are examined separately and used in composite models.

3.2 Load Characteristics

One of the vital requirements of short term load and price forecasting to obtain high forecasting accuracy and speed is to identify load characteristics and analyze the driving factors affecting load.

Hourly system load can be divided into four separate areas.

$L = L_n + L_w + L_s + L_r$

where

L represents the total system load

L_n stands for the normal part of the load that identifies type of the day and month of the year

L_w symbolizes the weather sensitive part of the load in terms of heating and cooling degree hours

L_s corresponds to the special event part of Turkish calendar

L_r represents the random part that is an unexplained factor

Lately, in competitive electricity markets system load may also be influenced by electricity prices. Prices vary depending on time and place. However, in this project price has not included as an independent variable, examined as a dependent variable for price model hence price is not be factored in the system load and its influencing factors.

$L = f(day, month, weather(cdh, hdh), special, random)$

Load model characteristics can be adapted to price model ,as a dependent variable, that is constructed as,

where P represents the market price

$P = f(day, month, weather(cdh, hdh), special, random)$

here $f(.)$ is a nonlinear function which is difficult to identify for this reason, traditional time series analysis is used with the combination of multiple linear regression to forecast accurate system load and price.

The input data for 24 hr in a day (ranging from 00:00 to 23:00) with dates used in the study include actual system load measured in megawatthour (MWh) for

2011 and 2012 and market price measured in Turkish lira per megawatthour (TL/MWh) for 2015 with categorical variables that are hours of day, days of month and months of the year, special events like Ramadan feast, 23th of April etc. and temperature measurements as cooling and heating hours are recorded.

3.2.1 Day Types

Each day and month types have their unique characteristics so forecasting models have to be established reflecting their inherit patterns. Hourly short term load forecasting models differs from each other according to load consumptions day by day. Hourly load consumptions in Turkey , January 2011 are given day by day for the first 2 weeks in table 3.1 and last 3 weeks in table 3.2

	WEEK 1							WEEK 2						
Hour/Day	1.1.2011	2.1.2011	3.1.2011	4.1.2011	5.1.2011	6.1.2011	7.1.2011	8.1.2011	9.1.2011	10.1.2011	11.1.2011	12.1.2011	13.1.2011	14.1.2011
00:00-01:00	21640	21901	22260	24361	24522	24650	24910	25282	24255	23105	24473	24914	24830	24955
01:00-02:00	20671	20415	20775	22898	22936	23192	23292	23598	22742	21264	23160	23313	23391	23355
02:00-03:00	19496	19219	19975	21995	22174	22221	22320	22622	21388	20321	22258	22416	22397	22308
03:00-04:00	18591	18692	19484	21596	21679	21681	21823	21916	20604	20055	21825	21771	21868	21985
04:00-05:00	18099	18421	19425	21398	21560	21506	21759	21814	20359	19771	21811	21769	21820	21955
05:00-06:00	17958	18491	19886	21794	21862	21817	22105	21985	20282	20258	22035	22075	22157	22280
06:00-07:00	18082	18685	21101	22756	22720	22889	23044	22379	20333	21363	23029	22950	23250	23102
07:00-08:00	17658	18420	22578	23868	23868	24120	24395	22657	19898	23140	24177	24148	24282	24448
08:00-09:00	18576	19451	26312	27266	27441	27324	27738	25455	20554	26800	27423	27353	27462	27637
09:00-10:00	20248	21259	29021	29629	29435	29757	30108	28309	22452	29397	29624	29708	29738	29888
10:00-11:00	21760	22877	30143	30822	30089	30605	31048	29561	24167	30265	30097	30193	30397	30632
11:00-12:00	22866	23844	30809	31269	30299	30851	31124	30114	24987	30444	30045	30353	30759	30817
12:00-13:00	23229	23908	29707	29959	29032	29718	28934	29060	24696	29143	28578	28815	29482	28407
13:00-14:00	23276	24081	30034	30417	29527	30083	30252	28731	24771	29295	28827	29268	29748	29468
14:00-15:00	23239	23949	30593	30655	29766	30459	30210	28564	24559	29503	28938	29213	29914	29533
15:00-16:00	23211	24232	30495	30401	29696	30179	30320	28013	24472	29334	29005	29125	29817	29562
16:00-17:00	24319	25234	31386	31286	30995	31376	31241	28515	25550	30391	30109	30475	30799	30299
17:00-18:00	25602	26746	31864	31880	31769	31943	31834	29631	27357	31771	31349	31506	31941	31368
18:00-19:00	25518	26729	30602	30713	30719	31067	30852	29266	27354	30786	30413	30574	30825	30565
19:00-20:00	25277	26588	29419	29785	29583	30093	29650	28515	27112	29717	29499	29747	29911	29390
20:00-21:00	24671	26277	28852	28961	28813	29321	28967	27937	26872	29031	28933	29143	29338	28656
21:00-22:00	24220	25964	28343	28357	28269	28713	28392	27391	26427	28421	28391	28410	28607	27954
22:00-23:00	24149	25867	28199	28388	28345	28495	28616	27501	26503	28639	28456	28370	28586	28323
23:00-24:00	23239,6	24365	26864	27036	27000	27487	27465	26444	25192	27055	27338	27114	27591	27056
Average	21899,82	22733,96	26588,625	27395,42	27170,79	27481,13	27516,63	26469,17	23870,25	26636,21	27074,71	27196,79	27454,58	27247,63
MAX	25602	26746	31864	31880	31769	31943	31834	30114	27357	31771	31349	31506	31941	31368
MIN	17658	18420	19425	21398	21560	21506	21759	21814	19898	19771	21811	21769	21820	21955
DAYS	Saturday	Sunday	Monday	Tuesday	Wednesd:	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesd:	Thursday	Friday

Table 3.1: First two weeks of hourly electricity load consumption in Turkey, January 2011

Table 3.3 represents total maximum and minimum electricity loads, average peak hour loads per day within 5 weeks. Also, the daily average electricity load consumtion pattern for each week in january, 2011 seen in table 3.2

WEEK 3							WEEK 4							WEEK 5		
15.1.2011	16.1.2011	17.1.2011	18.1.2011	19.1.2011	20.1.2011	21.1.2011	22.1.2011	23.1.2011	24.1.2011	25.1.2011	26.1.2011	27.1.2011	28.1.2011	29.1.2011	30.1.2011	31.1.2011
25063	23859	22763	24721	24934	25026	24858	25010	23898	22822	24443	24855	25275	25229	25440	24643	23363
23461	22428	21483	23386	23334	23533	23396	23443	22261	21429	23188	23352	23547	23562	23717	22837	21741
22660	21311	20529	22492	22495	22553	22575	22425	20997	20476	22372	22467	22543	22442	22572	21542	20731
21880	20493	20235	22031	21877	22099	21903	21810	20275	20119	21729	21981	21990	21861	22008	20719	20059
21615	20284	20070	21862	21833	22029	21915	21753	19966	19949	21593	21872	21874	21795	21724	20453	20007
21859	20227	20401	22178	22148	22452	22122	21900	20034	20367	22041	22189	22292	22135	21928	20347	20212
22247	20355	21665	23126	23120	23313	23305	22241	20095	21267	22806	22834	22888	22795	22239	20613	20876
22837	19813	23127	24403	24313	24500	24281	22627	19620	22815	23928	24162	23928	24103	22511	19758	22210
25542	20673	27026	27541	27528	27774	27669,7	25448	20586	26841	27417	27819	27571	27741	25456	20790	26560
28092	22550	29661	29849	29842	30013	29707,6	28028	22359	29309	29978	30314	30092	30353	28155	22551	29638
29233	23980	30733	30242	30492	30567	30538,4	28917	23773	30177	30871	31201	30836	31351	29536	24251	31158
29609	24813	31049	30430	30580	30382	30607,4	29510	24503	30662	31390	31676	31167	31627	30453	25320	31634
28564	24962	29772	28919	29022	29036	28234,9	28466	24287	29599	30150	30699	30105	29762	30041	25398	30763
28305	25073	30286	29299	29402	29236	29110	28333	24472	30119	30797	31022	30182	30724	30069	25536	30971
27923	24962	30269	29452	29474	29212	29205	27873	24417	30367	30938	31299	30149	31112	29844	25492	30929
27378	24944	30098	29339	29096	29093	29075	27614	24441	30231	30693	30907	29906	30892	29369	25296	30696
27966	25796	30797	30113	29908	30005	29930	27799	25075	30618	31239	31402	30203	31413	29294	25569	31012
29202	27163	31705	31408	31515	31501	31295	29115	26734	31366	32008	32072	31571	32135	30006	26686	32147
29103	27472	30975	30904	30914	30740	30575	28874	26963	30503	30948	31226	31231	31199	29687	27139	31404
28297	27052	29764	29875	29947	29644	29636	28038	26657	29325	29873	29974	30269	29954	28696	26807	30137
27566	26712	29168	29131	29356	29124	28789	27504	26290	28600	29252	29453	29502	29068	27991	26212	29195
27027	26156	28681	28466	28752	28579	28133	27109	25943	27870	28353	28587	28755	28529	27615	25910	28739
27066	26386	28597	28646	28792	28572	28288	27055	26084	27874	28418	28782	28751	28366	27686	26283	28867
26012	24993	27161	27323	27651	27788	27386	25809	24864	26811	27429	27758	27789	27402	26609	25244	27851
26187,79	23852,38	26917,29	27297,33	27346,88	27365,46	27188,958	26112,54	23524,75	26646,5	27577,25	27829,29	27600,667	27731,25	26776,92	23974,83	27120,83
29609	27472	31705	31408	31515	31501	31295	29510	26963	31366	32008	32072	31571	32135	30453	27139	32147
21615	19813	20070	21862	21833	22029	21903	21753	19620	19949	21593	21872	21874	21795	21724	19758	20007
Saturday	Sunday	Monday	Tuesday	Wednesd	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesd	Thursday	Friday	Saturday	Sunday	Monday

Table 3.2: Last three weeks of hourly electricity load consumption in Turkey, January 2011

Week 1-5				WEEK1	WEEK2	WEEK3	WEEK4	WEEK5
Day	Maximum	Minimum	Peak Hour	Average Loads (MW)				
Monday	32147	19425	17:00-18:00	26588,63	26636,21	26917,29	26646,5	27120,83
Tuesday	32008	21398	17:00-18:00	27395,42	27074,71	27297,33	27577,25	
Wednesday	32072	21560	17:00-18:00	27170,79	27196,79	27346,88	27829,29	
Thursday	31943	21506	17:00-18:00	27481,13	27454,58	27365,46	27600,67	
Friday	32135	21759	17:00-18:00	27516,63	27247,63	27188,96	27731,25	
Saturday	30453	17658	11:00-12:00	21899,82	26469,17	26187,79	26112,54	26776,92
Sunday	27472	18420	18:00-19:00	22733,96	23870,25	23852,38	23524,75	23974,83

Table 3.3: Maximum and minimum hourly electricity load consumption and peak hour loads in Turkey, January 2011

From the analysis of first month of year 2011 points out the load characteristics among the days in a week are different from each other. It can be concluded that average loads on weekends are lower than the weekdays. It is reasonable since electric consumption are reduced on off days comparing to working days. Mostly, on Saturdays electricity consumption is more than Sunday, this can be explained by extra operation hours in some private companies that people work.

Apparently, lots of people rest at night, this decreases the electrical consumption at late night hours. Moreover in the day, activities like lunch time, relaxation etc. occurs. The daily load consumption pattern tends to change.

Peak hour for week days are the same, from 17:00 to 18:00. This can be the last hour of their work. For Saturday it is 11:00-12:00 for common activities like cooking, washing the dishes or watching TV. Similarly, for Sunday it is 18:00-19:00, it can be the time for families to spend their time together and for some people it might be the time for preparing their works for Monday.

Also, the daily average electricity load consumption pattern for each week in january, 2011 as seen in Figure 3.1 points out the significant differences between the days. It can be concluded that during short-term load forecasting it is important to divide the days into several day types with each of them been known by its load pattern. It is obvious that the electricity load consumption pattern for Saturdays and Sundays is different compared to the weekdays.

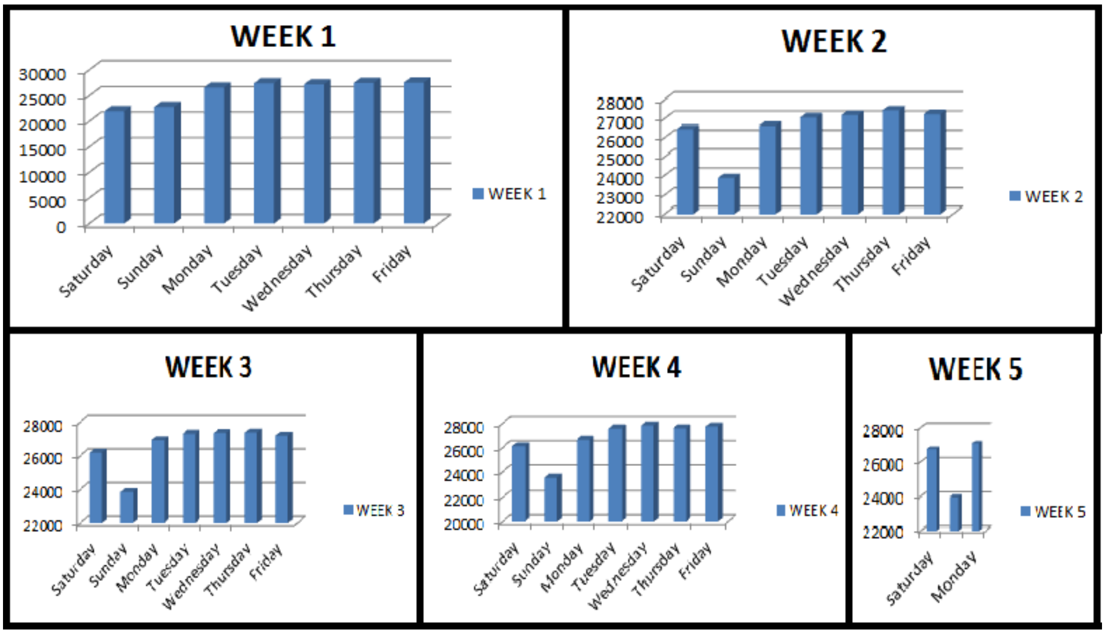

Figure 3.1: Weekly bar chart of electricity load consumption day by day in Turkey, January 2011

3.2.2 Special Events

Similarly, special events on special days have different load characteristics. Even sometimes special events in special days fall into the same class with weekends.

According to Turkish calendar, special events of the year are marked as below. Also Figure ?? represents the hourly electricity load consumption on special events (christmas, national sovereignty and children' day, labour day and commemoration of Ataturk,youth and sports day) in 2011

- Christmas (New Year's): 1st of January

- National Sovereignty and Childrens Day :23th of April

- Labour Day :1st of May

- Commemoration of Ataturk, Youth and Sports Day :19th of May

- Victory Day:30th of August

- Ramadan Feast , Ramadan Feast Eve

- Republic Day , Republic Day Eve :29th of October

- Sacrifice Feast , Sacrifice Feast Eve

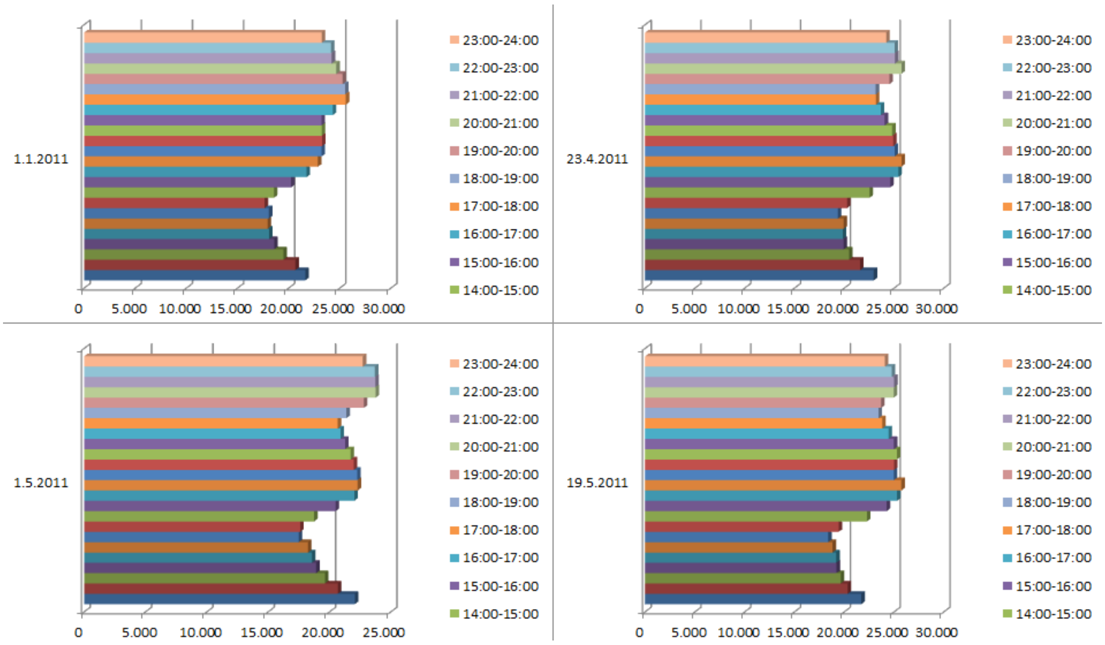

Figure 3.2: Hourly bar chart of electricity load consumption for special events (christmas, national sovereignty and children' day, labour day and commemoration of Ataturk,youth and sports day) in Turkey in 2011

3.2.3 Temperature

Temperature, as an environmental indicator is divided into two sub-measurement areas as heating and cooling degree hours.One of important meteorological variables which relate to energy consumption is cooling and heating hours. Hourly temperature records obtained from weather stations through weatherunderground (WU) were used to calculate heating degree hours with base temperature of 15 ^oC and cooling degree hours with base temperature of 22 ^oC It is defined as the sum of differences between hourly average temperatures and base temperature. The number of cooling degree hours (CDH) and heating degree hours (HDH) in a day are given below.

$$CDH_b = \sum_{i=1}^{N} (T_i - T_b)^+ \tag{3.1}$$

$$HDH_b = \sum_{i=1}^{N} (T_b - T_i)^+ \tag{3.2}$$

where

T_b is the base temperature , T_i average hourly temperature and N represents the number of hours in the day. The "+" sign on top of the equation points out that only positive values are considered in the summation. [22]. To give an example of calculations in detail , for HDH if the hourly temperature is less than or equal to 15 ^oC, 18 ^oC- T_i , otherwise 0 ^oC is applied. Similarly, for CDH if the hourly temperature is higher than 22 ^oC, T_i -22 ^oC, otherwise 0 ^oC is applied.

3.2.4 Time

Throughout the entire data analysis, continuous time variables ranging from the 1st hour to 8760th hours of the year as independent variables are integrated into the data sets to ensure that the residuals of the data are stationary. Time variables that are integrated in all models, are kept as long as their coefficients are higher than standard errors means they are significant. This approach, as seen in the residual analysis of the models, is very useful at the point where the error terms are stationary.

All in all, the sample of entire load and price data with all categorical variables are detailed in Table 3.4 and Table 3.5

Date	Time	Load (MwH)	Hour	Day	Month	Special Day	HDH	CDH
01.01.2011	0	21.640	00:00	Sat	Christ	January	14.25	0
01.01.2011	1	20.671	01:00	Sat	Christ	January	14.5	0
01.01.2011	2	19.496	02:00	Sat	Christ	January	18.0	0
01.01.2011	3	18.591	03:00	Sat	Christ	January	17.5	0
01.01.2011	4	18.099	04:00	Sat	Christ	January	17.0	0
01.01.2011	5	17.958	05:00	Sat	Christ	January	17.5	0
01.01.2011	6	18.082	06:00	Sat	Christ	January	18.5	0
01.01.2011	7	17.658	07:00	Sat	Christ	January	18.5	0
01.01.2011	8	18.576	08:00	Sat	Christ	January	17.0	0
01.01.2011	9	20.248	09:00	Sat	Christ	January	12.5	0
01.01.2011	10	21.760	10:00	Sat	Christ	January	11.0	0
01.01.2011	11	22.866	11:00	Sat	Christ	January	11.0	0
01.01.2011	12	23.229	12:00	Sat	Christ	January	11.0	0
01.01.2011	13	23.276	13:00	Sat	Christ	January	10.5	0
01.01.2011	14	23.239	14:00	Sat	Christ	January	11.0	0
01.01.2011	15	23.211	15:00	Sat	Christ	January	11.0	0
01.01.2011	16	24.319	16:00	Sat	Christ	January	11.5	0
01.01.2011	17	25.602	17:00	Sat	Christ	January	12.5	0
01.01.2011	18	25.518	18:00	Sat	Christ	January	13.0	0
01.01.2011	19	25.277	19:00	Sat	Christ	January	13.0	0
01.01.2011	20	24.671	20:00	Sat	Christ	January	12.67	0
01.01.2011	21	24.220	21:00	Sat	Christ	January	12.0	0
01.01.2011	22	24.149	22:00	Sat	Christ	January	11.5	0
01.01.2011	23	23.240	23:00	Sat	Christ	January	11.0	0

Table 3.4: The sample of hourly input load data with categorical variables and environmental indicators for 24 hr in a day in 01.01.2011

Date	Price(TL/MwH)	Hour	Day	Month	Special Day	HDH	CDH
23.04.2015	98.65	00:00	Thu	23Apr	April	11.67	0
23.04.2015	89.99	01:00	Thu	23Apr	April	12.00	0
23.04.2015	72.77	02:00	Thu	23Apr	April	13.00	0
23.04.2015	50.78	03:00	Thu	23Apr	April	13.33	0
23.04.2015	43.06	04:00	Thu	23Apr	April	13.00	0
23.04.2015	43.06	05:00	Thu	23Apr	April	12.50	0
23.04.2015	43.06	06:00	Thu	23Apr	April	12.33	0
23.04.2015	40.06	07:00	Thu	23Apr	April	11.50	0
23.04.2015	95.78	08:00	Thu	23Apr	April	10.00	0
23.04.2015	103.25	09:00	Thu	23Apr	April	10.25	0
23.04.2015	114.87	10:00	Thu	23Apr	April	10.50	0
23.04.2015	117.05	11:00	Thu	23Apr	April	8.50	0
23.04.2015	102.29	12:00	Thu	23Apr	April	8.00	0
23.04.2015	100.00	13:00	Thu	23Apr	April	8.50	0
23.04.2015	100.57	14:00	Thu	23Apr	April	9.00	0
23.04.2015	100.00	15:00	Thu	23Apr	April	9.00	0
23.04.2015	100.00	16:00	Thu	23Apr	April	8.50	0
23.04.2015	94.85	17:00	Thu	23Apr	April	8.00	0
23.04.2015	95.78	18:00	Thu	23Apr	April	7.67	0
23.04.2015	99.99	19:00	Thu	23Apr	April	8.50	0
23.04.2015	103.09	20:00	Thu	23Apr	April	9.00	0
23.04.2015	102.03	21:00	Thu	23Apr	April	9.00	0
23.04.2015	114.97	22:00	Thu	23Apr	April	9.00	0
23.04.2015	100.00	23:00	Thu	23Apr	April	9.50	0

Table 3.5: The sample of hourly input price data with categorical variables and environmental indicators for 24 hr in a day in 23.04.2015

Chapter 4

APPROACH and METHODOLOGY

4.1 Construction of Short Term Load Forecasting Model Achitecture

This section presents the details of methodology used in this study, discussing the analysis of historical data obtained from TEIAS in 2011-2012 and EPIAS in 2015, what is done with it using the 24hr lagged variables that has been presented in autoregressive electricity load models and merged them with categorical variables as day types, special events and environmental indicators as cdh and hdh for temperature to come up with a better hourly load and price forecasting models.

The research work basically concludes and analyzes the implementation of the following statistical tests;

-Regression standardized residual analysis through histogram and normal probability plot

-Variance inflation factor(vif) analysis

- Autocorrelation and partial autocorrelations function plot,

-Durbin-Watson,

-Forecasting power indicators as Mean Absolute Percentage Error(MAPE), R-Square etc.

-Cross validation tests.

4.1.1 Initial Data Classification

In chapter 3 initial load and price data is discussed and categorical variables and environmental indicators are analyzed in section 3.2. Based on our initial data for 24 hour period lagged variables with past load and price data and categorical variables; hours, days, special events, months and environmental indicators; cdh and hdh (as temperature) are assigned for year 2011 and 2012 . According to different data sets, measurements are done through 4 different load models and a price model that are given below,

- Data Set 1: The hourly data of electricity load consumption is lagged through lag 1,2....24 with categorical variables and environmental indicators in 2011

- Data Set 2: The hourly data of electricity load consumption is lagged through lag 1 and lag 2 with categorical variables and environmental indicators in 2011

- Data Set 3: The hourly data of electricity load consumption is lagged through lag 1,2....24 with categorical variables and environmental indicators in 2012

- Data Set 4: The hourly data of electricity load consumption is lagged through lag 1 and lag 2 with categorical variables and environmental indicators in 2012

- Data Set 5: The hourly data of electricity price is lagged through lag 1,2....24 with categorical variables and environmental indicators in 2015

4.1.2 Proposed Autoregressive Electricity Load Model for 2011

In the beginning of the study, initial data obtained from TEIAS in 2011 with 24 hr lagged variables are examined through autocorrelation and partial autocorrelation functions.

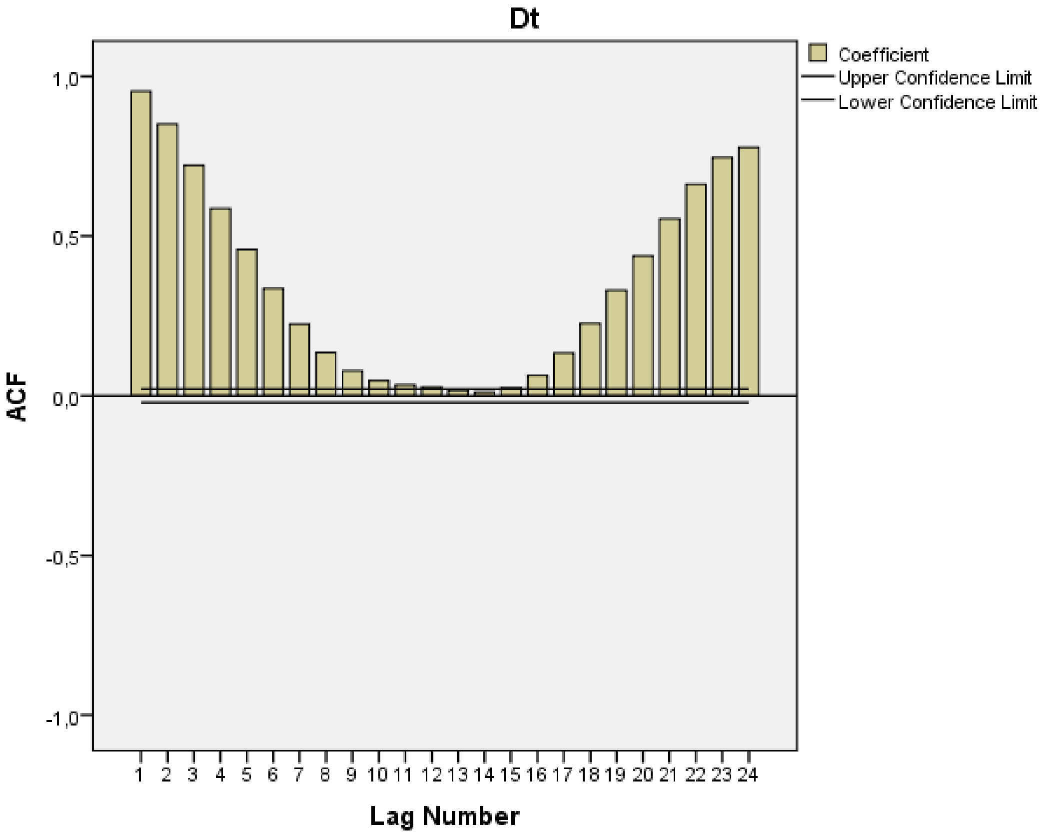

Figure 4.1: Autocorrelation function plot of electricity load in Turkey, 2011

Table 4.1: Autocorrelations of electricity load through 24 lagged variables in 2011

Lag	Autocorr.	Std. Error	Box-Ljung	df	Sig.b
1	0,953	0,011	7979,599	1	0
2	0,85	0,011	14329,01	2	0
3	0,722	0,011	18905,08	3	0
4	0,586	0,011	21924,76	4	0
5	0,458	0,011	23769,4	5	0
6	0,336	0,011	24762,94	6	0
7	0,224	0,011	25205,13	7	0
8	0,136	0,011	25367,02	8	0
9	0,078	0,011	25420,86	9	0
10	0,047	0,011	25440,42	10	0
11	0,034	0,011	25450,82	11	0
12	0,027	0,011	25457,47	12	0
13	0,017	0,011	25460	13	0
14	0,012	0,011	25461,26	14	0
15	0,025	0,011	25466,68	15	0
16	0,064	0,011	25502,32	16	0
17	0,133	0,011	25659,08	17	0
18	0,227	0,011	26110,85	18	0
19	0,33	0,011	27068,55	19	0
20	0,439	0,011	28762,98	20	0
21	0,554	0,011	31465,48	21	0
22	0,662	0,011	35327,73	22	0
23	0,746	0,011	40231,93	23	0
24	0,777	0,011	45554,8	24	0

According to autocorrelation table 4.1 and the plot 4.1 shown above, all autocorrelation values are higher than standard error which means AR(24) model is significant as a whole. On the authority of AR(24) model results are shown that,

Table 4.2: Residual statistics and summary of AR(24) Model, 2011

Residuals Statistics					
	Min	Max	Mean	Std. Dev.	N
Predicted Value	-11538,5	36370,617	26327,752	4109,281	8784
Residual	-24141,7	11538,474	0	692,03	8784
Std. Predicted Value	-9,215	2,444	0	1	8784
Std. Residual	-34,838	16,651	0	0,999	8784
Model Summary					
Model	R	R Square	Adj. R Squ.	Std. Err.	Durb.Wat
	0,986	0,972	0,972	692,978	2,392

Table 4.3: Electricity load forecasting model fit of AR(24) Model

Fit Statistic	Mean
Stationary R-squared	0,972
R-squared	0,972
RMSE	692,978
MAPE	1,69
MaxAPE	15,546
MAE	456,842
MaxAE	24151,29
Normalized BIC	13,096

The R-value indicates high degree correlation which is 0,986 and the R-square indicates how much of the total variation in the dependent variable which is electricity load can be explained by the independent variable as its lagged variables. In this case it is %97.2 in 2011. For the autocorrelation, durbin-watson statistics is 2.392 which is quite normal since its between 1.5 and 2.5 range. MAPE is very low that is % 1,69 indicates good fit for the model.

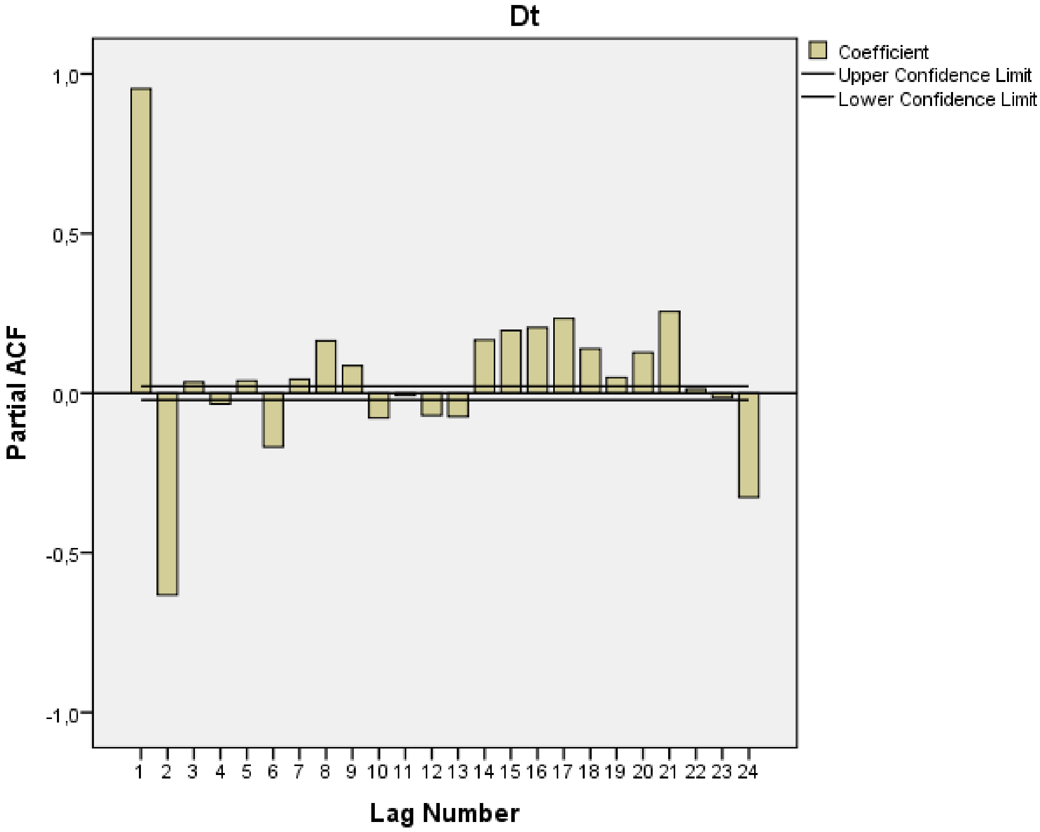

Figure 4.2: Partial autocorrelation function plot of electricity load in Turkey, 2011

Furthermore, in order to detect the correlation between time series and its lagged values, partial autocorrelation function should be applied. Consistent with partial autocorrelation plot lag1 and lag2 are statistically significant, whereas partial autocorrelations for all other lags are not statistically significant.

All in all, according to autocorrelation and partial autocorrelation results two models were constructed using the dataset 1 and dataset 2 as mentioned in section 4.1.1

4.1.3 Multiple Linear Regression of Load Model using data set1

As the general form of the multiple linear regression model discussed in section 2.1; to obtain regression coefficients, regression analysis is applied and variables that are insignificant means according to t-distribution whose alpha value are higher than 0.1 are dropped by using backward elimination method. Sample of unstandardized coefficients (B) of entered independent variables of linear regression load model lagged through lag 1,2....24 with categorical variables and environmental indicators in 2011 are given below

Table 4.4: Multiple linear regression coefficients for entered variables using data set1

	B	Std. Error	Beta	t	Sign.	Tolerance	VIF
(Constant)	1817,9	134,535		13,512	0		
Time	-0,424	0,015	-0,258	-27,764	0	0,02	50,086
Dt-lag(1)	1,22	0,01	1,217	116,542	0	0,016	63,18
Dt-lag(2)	-0,402	0,016	-0,4	-24,527	0	0,006	154,293
Dt-lag(3)	0,038	0,017	0,037	2,263	0,024	0,006	158,116
Dt-lag(4)	-0,04	0,016	-0,04	-2,552	0,011	0,007	140,626
Dt-lag(5)	0,052	0,014	0,052	3,674	0	0,009	115,755
Dt-lag(6)	-0,021	0,011	-0,021	-1,997	0,046	0,016	62,131
Dt-lag(9)	-0,027	0,007	-0,026	-3,823	0	0,037	27,054
Dt-lag(11)	0,082	0,012	0,08	6,598	0	0,012	85,362
Dt-lag(12)	-0,051	0,017	-0,049	-3,064	0,002	0,007	149,262
Dt-lag(13)	0,04	0,017	0,039	2,385	0,017	0,007	152,38
Dt-lag(14)	-0,071	0,012	-0,069	-5,724	0	0,012	84,116
Dt-lag(16)	0,061	0,008	0,059	7,257	0	0,027	37,646
Dt-lag(18)	-0,038	0,013	-0,037	-3,026	0,002	0,012	86,209
Dt-lag(19)	0,061	0,016	0,058	3,689	0	0,007	143,783
Dt-lag(20)	-0,049	0,013	-0,047	-3,908	0	0,012	82,687
Dt-lag(22)	-0,047	0,013	-0,045	-3,699	0	0,012	83,796
Dt-lag(23)	0,277	0,016	0,263	17,089	0	0,007	136,633
Dt-lag(24)	-0,13	0,01	-0,123	-12,59	0	0,018	55,364
Hour_1	-252,5	48,816	-0,012	-5,173	0	0,315	3,173
Hour_2	-276,5	48,666	-0,013	-5,683	0	0,317	3,153
Hour_3	-340	48,274	-0,016	-7,043	0	0,322	3,103
Hour_4	-98,69	39,732	-0,005	-2,484	0,013	0,476	2,102
Hour_6	-123,3	46,315	-0,006	-2,661	0,008	0,35	2,856
Hour_7	402,8	52,359	0,019	7,693	0	0,274	3,65
Hour_8	1983	52,271	0,095	37,937	0	0,275	3,638
Hour_9	783,92	55,417	0,038	14,146	0	0,245	4,089
Hour_10	306,55	55,031	0,015	5,571	0	0,248	4,032
Hour_11	822,97	55,422	0,039	14,849	0	0,245	4,09
Hour_12	-588,2	53,748	-0,028	-10,944	0	0,26	3,846
Hour_13	1047,7	51,968	0,05	20,16	0	0,278	3,596
Hour_14	525,03	44,048	0,025	11,919	0	0,387	2,583
Hour_16	611	42,937	0,029	14,23	0	0,407	2,455
Hour_17	590,03	45,841	0,028	12,871	0	0,357	2,798
Hour_18	326,87	49,533	0,016	6,599	0	0,306	3,267
Hour_19	346,63	53,047	0,017	6,534	0	0,267	3,747
Hour_20	499	50,818	0,024	9,819	0	0,291	3,438
Hour_21	-95,13	49,132	-0,005	-1,936	0,053	0,311	3,214
Hour_22	529,75	42,666	0,025	12,416	0	0,413	2,424
Hour_24	-1018	44,757	-0,049	-22,738	0	0,375	2,667

	B	Std. Error	Beta	t	Sign.	Tolerance	VIF
Mon	749,33	29,409	0,063	25,479	0	0,28	3,571
Tue	472,32	24,905	0,04	18,965	0	0,397	2,521
Wed	451,02	24,797	0,038	18,188	0	0,4	2,499
Thu	441,65	24,66	0,037	17,909	0	0,405	2,472
Fri	409,45	24,185	0,034	16,93	0	0,421	2,377
Sat	232,36	22,658	0,019	10,255	0	0,479	2,086
Christ.	3535	163,92	0,044	21,565	0	0,41	2,441
May.01	-197,5	107,823	-0,002	-1,832	0,067	0,947	1,056
May.19	297,22	108,377	0,004	2,742	0,006	0,937	1,067
30 Agu	-575,2	131,912	-0,007	-4,36	0	0,633	1,581
Ram. B	-263,1	89,241	-0,006	-2,948	0,003	0,463	2,159
Ram. B. Ar.	-726,5	151,962	-0,006	-4,781	0	0,952	1,05
Kurb. B	-579,3	70,089	-0,014	-8,265	0	0,565	1,771
Kurb. B. Ar	-877,8	153,396	-0,008	-5,723	0	0,934	1,07
January	-1092	48,336	-0,074	-22,601	0	0,161	6,215
February	-781,7	42,88	-0,05	-18,229	0	0,231	4,331
March	-486	35,946	-0,032	-13,52	0	0,299	3,34
April	-248,1	29,391	-0,016	-8,44	0	0,461	2,167
June	345,39	30,346	0,023	11,382	0	0,433	2,31
July	821,35	45,257	0,055	18,148	0	0,189	5,294
August	1083,2	52,219	0,072	20,743	0	0,142	7,049
September	1368,2	56,263	0,09	24,317	0	0,126	7,942
October	1635,6	64,155	0,109	25,494	0	0,094	10,639
November	2088,6	79,963	0,137	26,12	0	0,062	16,042
December	2373,6	90,37	0,159	26,265	0	0,047	21,11
CDH	12,007	4,289	0,006	2,8	0,005	0,391	2,559

With the combination of 24 hr lagged and categorical variables and environmental indicators' coefficients multiple linear regression model is constructed by using equation 2.1. Model results are given below

Table 4.5: Electricity load forecasting model summary of data set1

R	R-Sq.	Adj-R Sq.	Std. Err.	Durb.Wat
0,992	0,985	0,985	513,2728	2,044

Table 4.6: Electricity load forecasting model fit of data set1

Fit Statistic	Mean
Stationary R-squared	0,985
R-squared	0,985
RMSE	513,273
MAPE	1,277
MaxAPE	12,251
MAE	335,88
MaxAE	20085,25
Normalized BIC	12,551

According to table 4.5 and 4.14 R-square is 0,985 means total variation in forecasted electricity load is explained %98,5 by singificant 24hr lagged variables,categorical variables and environmental indicators. For the autocorrelation, durbin-watson statistics is 2.044 which is quite normal since its between 1.5 and 2.5 range. MAPE is very low that is % 1,277 indicates good fit for the model.

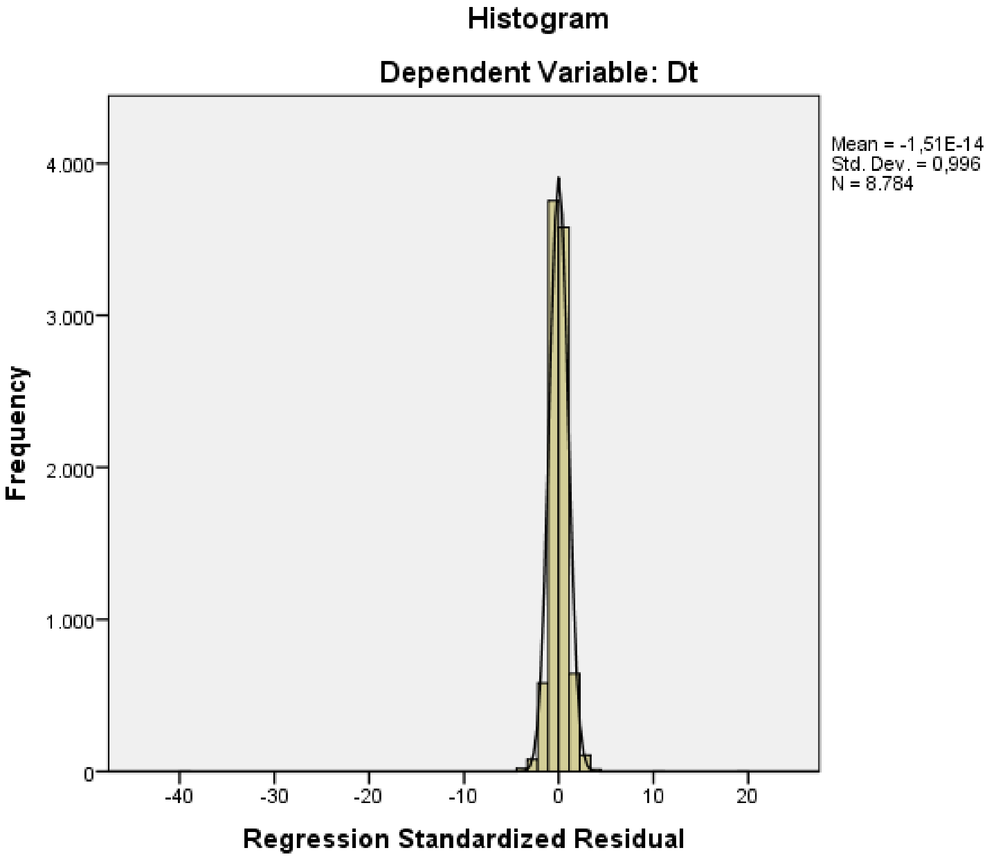

Figure 4.3: Histogram of forecasted load as a dependent variable of data set1

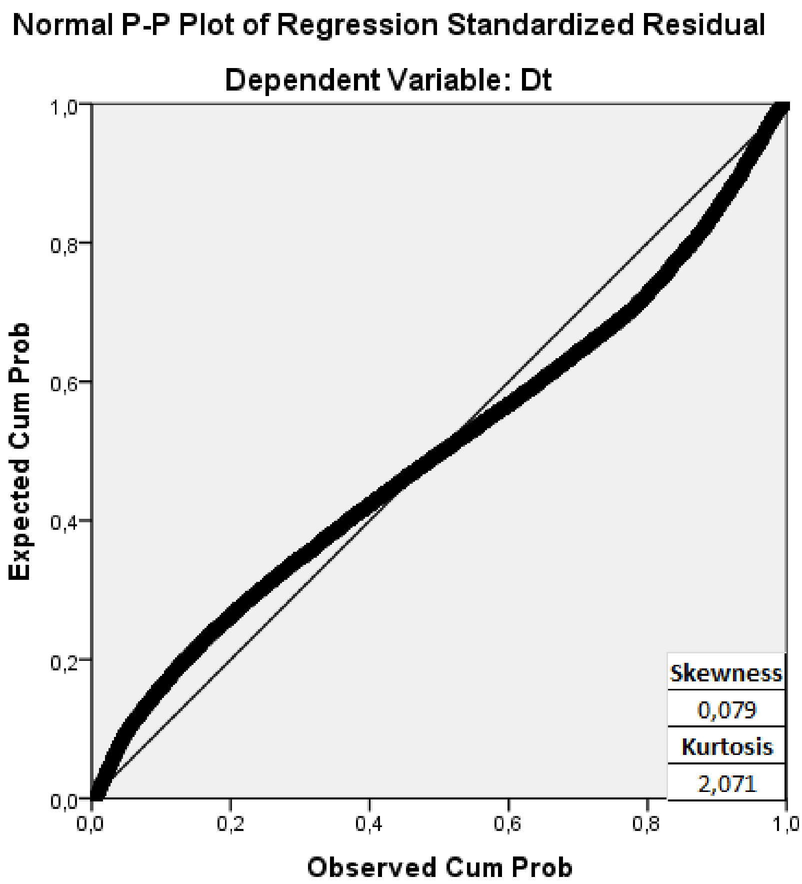

Figure 4.4: Normal probability plot of regression standardized residual for data set1

Figure 4.3 and 4.4 illustrate a normal distribution of residuals produced by a model for load forecasting. Residual normal probability plot and histogram shows that standardized residuals randomly distributed around zero. Skewness value is 0,079 that shows the amount and direction of skew departure from horizontal symmetry. Since its in the range between and +, the distribution is approximately symmetric.Also, kurtosis value is 2,071 that represents how tall and sharp the central peak is relative to a standard bell curve. Since it is less than 3, the sample very likely has negative excess kurtosis means platykurtic. For the collinearity check, VIF values are higher for lagged variables. It is expected since autocorrelation and partial autocorrelation functions indicate high correlations between actual loads and its lagged variables. Also, some independent variables like October, November and December represent the winter months have highly influental effects on electricity load consumption since economic activities are high in 2011 (e.g. quarterly GDP of Turkey)

Table 4.7: Crossvalidation test on %20 and %80 of the sample of data set1

Correlations				
			Dt	Model
20%sample	Dt	Pearson Correlation	1	,994**
		Sig. (2-tailed)		0
		N	1764	1764
	Model	Pearson Correlation	,994**	1
		Sig. (2-tailed)	0	
		N	1764	1764
80%sample	Dt	Pearson Correlation	1	,994**
		Sig. (2-tailed)		0
		N	7020	7020
	Model	Pearson Correlation	,994**	1
		Sig. (2-tailed)	0	
		N	7020	7020
** Correlation is significant at the 0.01 level (2-tailed).				

Last of all, according to cross validation test on %20 and %80 of the sample of data set1, correlation coefficients are highly correlate with each other that are for both %20 and %80 of the samples are 0,994.

4.1.4 Multiple Linear Regression of Load Model using dataset2

As the general form of the multiple linear regression model discussed in section 2.1; to obtain regression coefficients, regression analysis is applied and variables that are insignificant means according to t-distribution whose alpha value are higher than 0.1 are dropped by using backward elimination method. Sample of unstandardized coefficients (B) of entered independent variables of linear regression load model lagged through lag1 and lag2 with categorical variables and environmental indicators in 2011 are given below

Table 4.8: Multiple linear regression coefficients for entered variables using data set2

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	1213,94	69,004		17,592	0		
Time	0,053	0,009	0,031	6,118	0	0,072	13,932
Dt-lag(1)	1,284	0,01	1,282	129,531	0	0,018	54,308
Dt-lag(2)	-0,358	0,01	-0,357	-36,529	0	0,019	52,822
Hour_1	-434,542	39,634	-0,02	-10,964	0	0,552	1,812
Hour_2	-290,214	35,65	-0,013	-8,141	0	0,682	1,466
Hour_3	-156,044	33,975	-0,007	-4,593	0	0,751	1,331
Hour_7	1085,952	33,458	0,05	32,457	0	0,775	1,291
Hour_8	2841,03	36,087	0,13	78,728	0	0,666	1,502
Hour_9	1450,577	48,314	0,066	30,024	0	0,371	2,692
Hour_10	662,585	45,207	0,03	14,657	0	0,424	2,357
Hour_11	664,721	42,053	0,03	15,807	0	0,49	2,039
Hour_12	-624,903	42,102	-0,028	-14,843	0	0,489	2,044
Hour_13	1157,153	41,432	0,053	27,929	0	0,505	1,98
Hour_14	674,056	40,825	0,031	16,511	0	0,52	1,922
Hour_15	157,219	41,186	0,007	3,817	0	0,511	1,956
Hour_16	689,956	40,578	0,031	17,003	0	0,527	1,899
Hour_17	443,921	40,383	0,02	10,993	0	0,532	1,881
Hour_18	194,829	40,054	0,009	4,864	0	0,54	1,85
Hour_19	441,111	39,543	0,02	11,155	0	0,555	1,803
Hour_20	597,095	39,041	0,027	15,294	0	0,569	1,758
Hour_21	137,96	39,062	0,006	3,532	0	0,568	1,76
Hour_22	460,768	38,605	0,021	11,935	0	0,582	1,719
Hour_23	-605,514	38,32	-0,028	-15,801	0	0,591	1,693
Hour_24	-1379,86	38,177	-0,063	-36,143	0	0,595	1,681

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Mon	252,809	23,034	0,02	10,976	0	0,527	1,898
Tue	291,867	24,531	0,023	11,898	0	0,472	2,118
Wed	295,87	24,563	0,024	12,045	0	0,471	2,124
Thu	292,388	24,596	0,023	11,887	0	0,47	2,13
Fri	271,641	24,255	0,022	11,199	0	0,483	2,071
Sat	170,846	23,232	0,014	7,354	0	0,526	1,9
Christ.	-2249,51	133,005	-0,027	-16,913	0	0,718	1,392
30 Agu	-281,985	140,299	-0,003	-2,01	0,044	0,646	1,549
Ram. B	-609,975	85,81	-0,013	-7,108	0	0,578	1,729
Ram. B. Ar.	-441,75	162,773	-0,004	-2,714	0,007	0,958	1,044
Kurb. B	-608,313	65,977	-0,014	-9,22	0	0,736	1,359
Kurb. B. Ar	-338,704	162,813	-0,003	-2,08	0,038	0,957	1,044
April	-135,083	24,322	-0,008	-5,554	0	0,778	1,286
May	-229,858	27,072	-0,015	-8,491	0	0,609	1,641
June	-180,188	28,263	-0,011	-6,375	0	0,576	1,736
July	61,925	27,349	0,004	2,264	0,024	0,597	1,675
September	-169,249	33,12	-0,011	-5,11	0	0,419	2,384
October	-304,985	42,924	-0,019	-7,105	0	0,242	4,125
November	-229,566	54,169	-0,014	-4,238	0	0,157	6,377
December	-235,563	59,888	-0,015	-3,933	0	0,125	8,03
HDH	9,761	2,334	0,012	4,181	0	0,231	4,335
CDH	8,052	4,451	0,004	1,809	0,07	0,419	2,388

As a result of linear regression load model lagged through lag1 and lag2 with categorical variables and environmental indicators in 2011,

Table 4.9: Electricity load forecasting model summary of dataset2

R	R-Sq.	Adj. R-Sq.	Std. Err.	Durb.Wat
0,992	0,984	0,984	551,4963	1,883

Table 4.10: Electricity load forecasting model fit of dataset2

Fit Statistic	Mean
Stationary R-squared	0,984
R-squared	0,984
RMSE	551,518
MAPE	1,348
MaxAPE	13,015
MAE	361,484
MaxAE	19947,9
Normalized BIC	12,675

In table 4.9 and 4.10 R-square is 0,984 means total variation in forecasted electricity load is explained %98,4 by lag1 and lag2 ,categorical variables and environmental indicators. For the autocorrelation, durbin-watson statistics is 1.883 which is quite normal since its between 1.5 and 2.5 range. MAPE is also very low that is % 1,348 indicates good fit for the model.

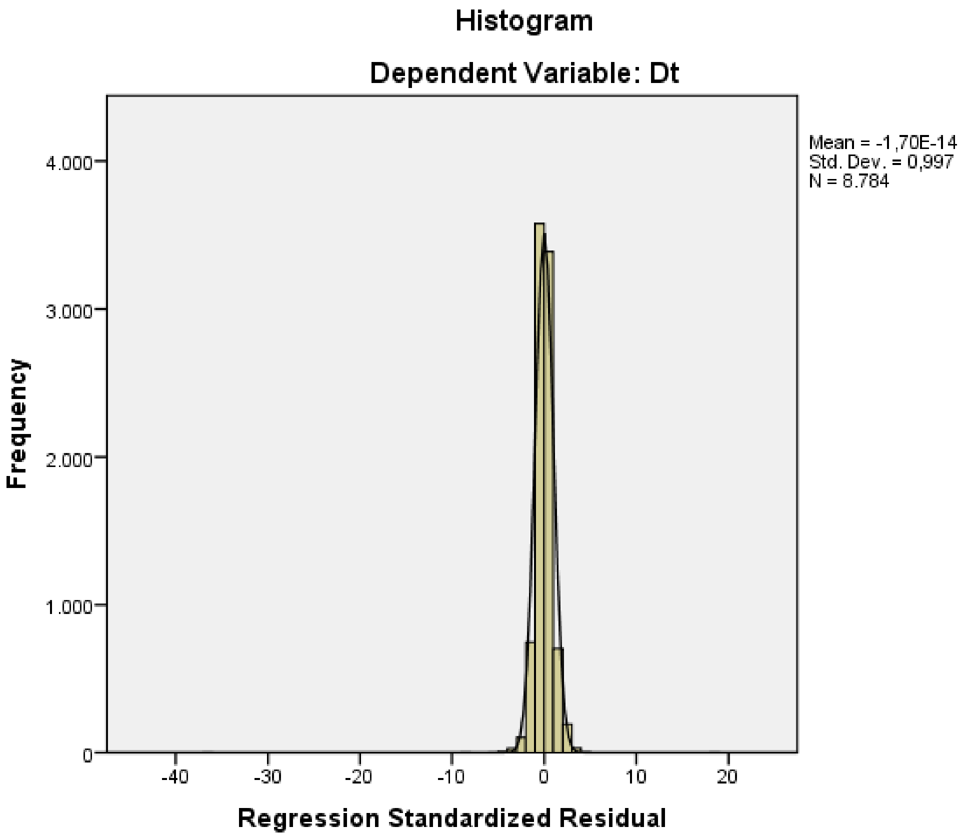

Figure 4.5: Histogram of forecasted load as a dependent variable for dataset2

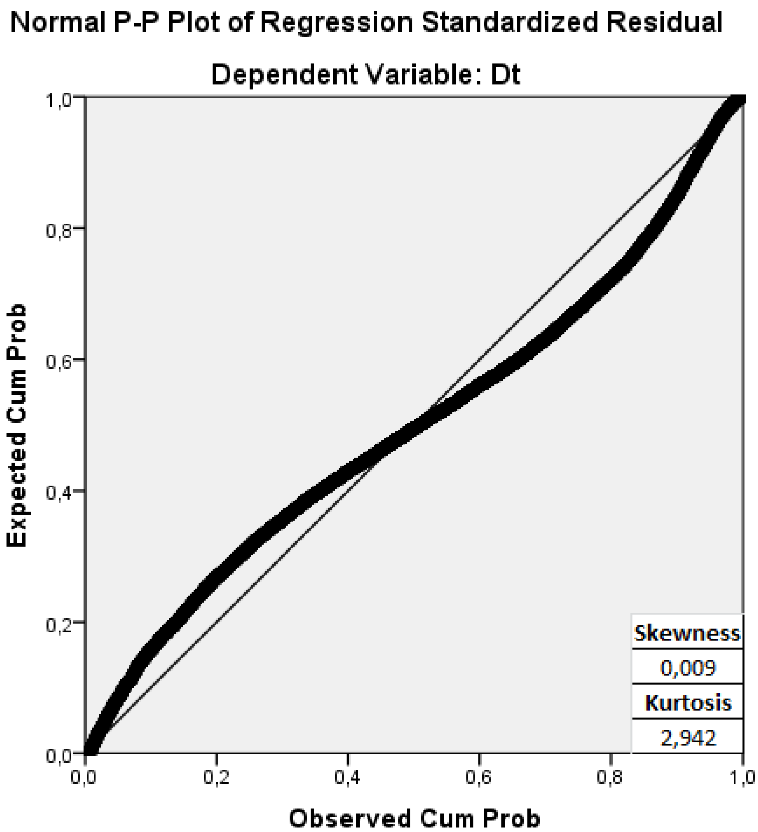

Figure 4.6: Normal probability plot of regression standardized residual for data set2

In figure 4.5 and 4.6 illustrate a normal distribution of residuals produced by a model for load forecasting.Residual normal probability plot and histogram shows that standardized residuals randomly distributed around zero.Skewness value is 0,009 that shows the amount and direction of skew departure from horizontal symmetry. Since its in the range between and +, the distribution is approximately symmetric.Also, kurtosis value is 2,942 that represents how tall and sharp the central peak is relative to a standard bell curve. Since it is less than 3, the sample very likely has negative excess kurtosis means platykurtic. For the collinearity check, VIF values are higher for lagged variables. It is expected since autocorrelation and partial autocorrelation functions indicate high correlations between actual loads and its lagged variables which means there is no collinearity problem detected.

Table 4.11: Crossvalidation test on %20 and %80 of the sample of data set2

Correlations				
			Dt	Model
20%sample	Dt	Pearson Correlation	1	,993**
		Sig. (2-tailed)		0
		N	1732	1732
	Model	Pearson Correlation	,993**	1
		Sig. (2-tailed)	0	
		N	1732	1732
80%sample	Dt	Pearson Correlation	1	,992**
		Sig. (2-tailed)		0
		N	7052	7052
	Model	Pearson Correlation	,992**	1
		Sig. (2-tailed)	0	
		N	7052	7052
** Correlation is significant at the 0.01 level (2-tailed).				

Finally, according to cross validation test on %20 and %80 of the sample of data set2, correlation coefficients are highly correlate with each other that are for %20 of the sample is 0,993 and %80 of the samples are 0,992.

4.1.5 Proposed Autoregressive Electricity Load Model for 2012

Similarly, initial data obtained from TEIAS in 2012 with 24 hr lagged variables are examined through autocorrelation and partial autocorrelation functions.

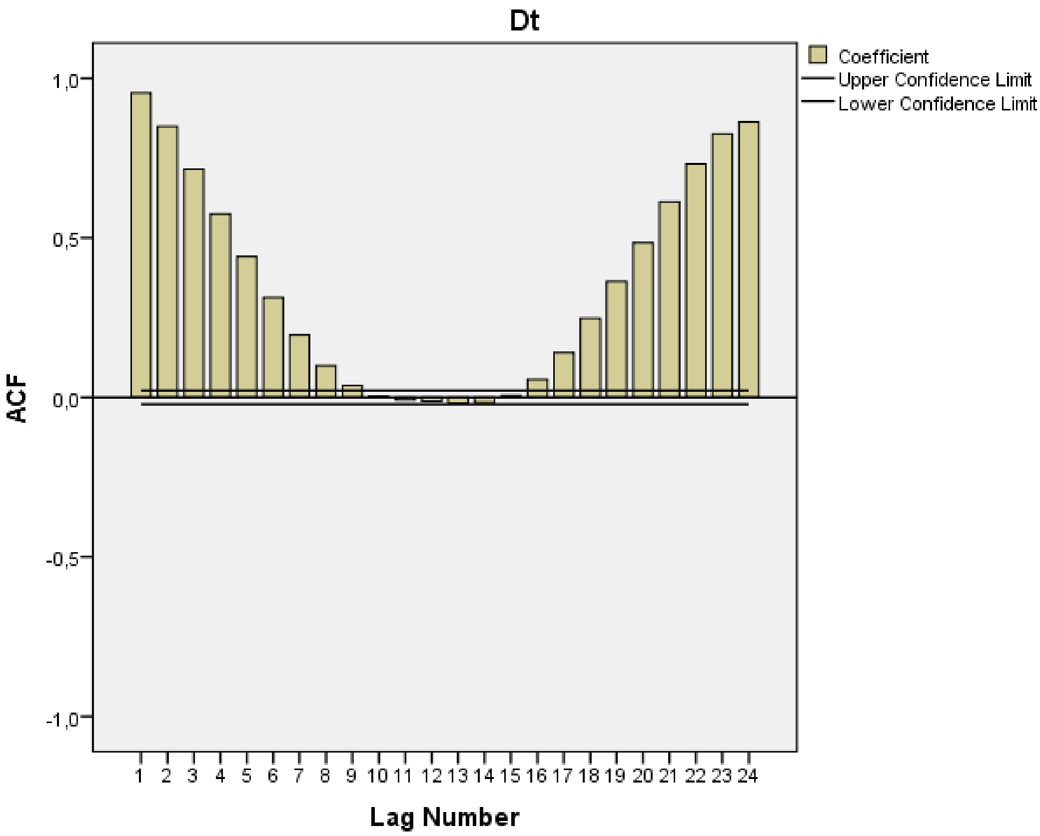

Figure 4.7: Autocorrelation function plot of electricity load in Turkey, 2012

Table 4.12: Autocorrelations of electricity load through 24 lagged variables in 2012

Lag	Autocorr.	Std. Error	Box-Ljung	df	Sig.b
1	0,954	0,011	8020,427	1	0
2	0,848	0,011	14364,27	2	0
3	0,716	0,011	18876,21	3	0
4	0,575	0,011	21788,12	4	0
5	0,441	0,011	23506,06	5	0
6	0,314	0,011	24372,93	6	0
7	0,195	0,011	24707,36	7	0
8	0,099	0,011	24794,57	8	0
9	0,037	0,011	24806,58	9	0
10	0,004	0,011	24806,71	10	0
11	-0,008	0,011	24807,27	11	0
12	-0,012	0,011	24808,62	12	0
13	-0,018	0,011	24811,35	13	0
14	-0,016	0,011	24813,65	14	0
15	0,006	0,011	24813,92	15	0
16	0,056	0,011	24841,85	16	0
17	0,14	0,011	25014,36	17	0
18	0,247	0,011	25553,46	18	0
19	0,363	0,011	26719,77	19	0
20	0,485	0,011	28792,8	20	0
21	0,612	0,011	32098,76	21	0
22	0,732	0,011	36824,96	22	0
23	0,825	0,011	42842,08	23	0
24	0,863	0,011	49422,07	24	0

According to autocorrelation table 4.12 and the plot 4.7 shown above, all autocorrelation values are higher than standard error which means AR(24) model is significant as a whole. AR(24) model results shows that,

Table 4.13: Residual statistics and summary of AR(24) Model, 2012

Residual Statistics					
Min	Max	Mean	Std. Dev.	N	
Predicted Value	13859,21	39030,86	27511,16	4176,365	8808
Residual	-4977,17	5053,004	2,66E-11	623,5618	8808
Std. Predicted Value	-3,269	2,758	0	1	8808
Std. Residual	-7,971	8,092	0	0,999	8808
Model Summary					
Model	R	R Squ.	Adj. R Squ.	Std. Err.	Durb.Wat
	0,989	0,978	0,978	624,4132	2,549

Table 4.14: Electricity load forecasting model fit of AR(24) Model

Fit Statistic	Mean
Stationary R-squared	0,978
R-squared	0,978
RMSE	624,413
MAPE	1,735
MaxAPE	21,42
MAE	473,249
MaxAE	5045,78
Normalized BIC	12,903

The R-value indicates high degree correlation which is 0,978 and the R-square indicates %97.8 means variation in electricity load can be explained by its lagged variables in 2012. For the autocorrelation, durbin-watson statistics is 2.549 which is normal since its between 1.5 and 2.5 range. MAPE is very low that is % 1,735 indicates good fit for the model.

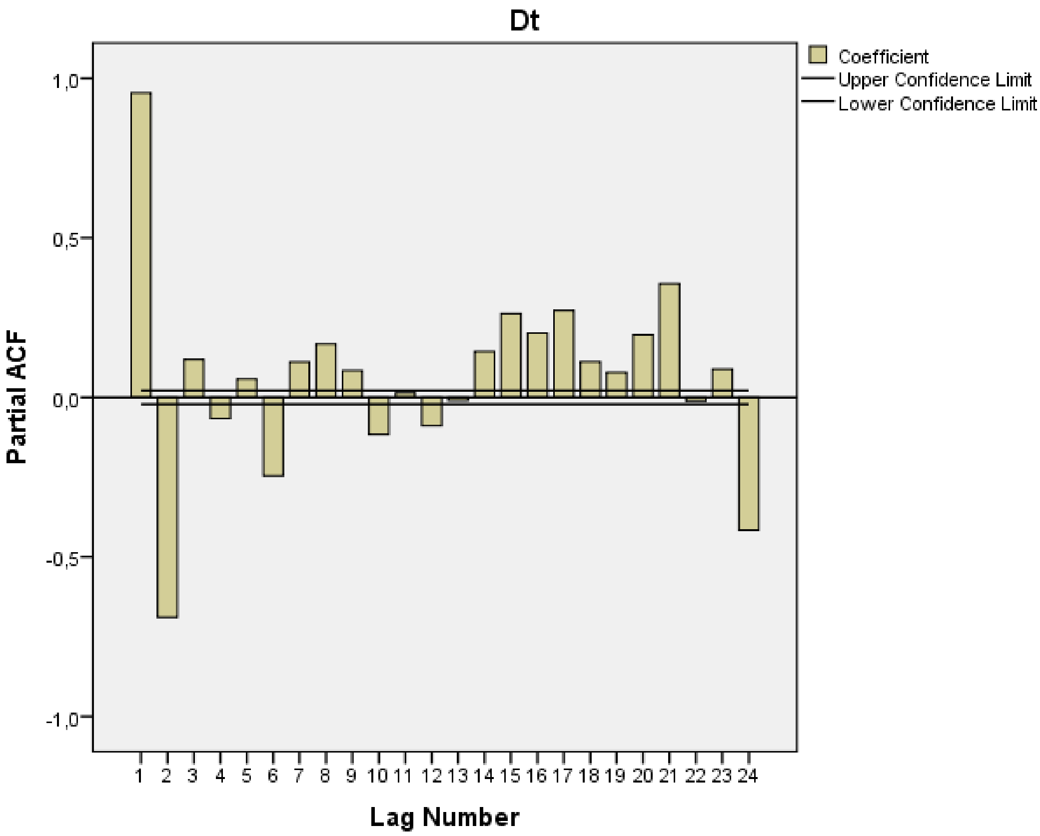

Figure 4.8: Partial autocorrelation function plot of electricity load in Turkey, 2012

Also similarly with 2011 data, in order to detect the correlation between time series and its lagged values, partial autocorrelation function should be applied as it is mentioned. Consistent with partial autocorrelation plot lag1 and lag2 are statistically significant, whereas partial autocorrelations for all other lags are not statistically significant.

All in all, two models were constructed according to the dataset 3 and dataset 4 as mentioned in section 4.1.1

4.1.6 Multiple Linear Regression of Load Model using data set3

As the general form of the multiple linear regression model discussed in section 2.1; to obtain regression coefficients, regression analysis is applied and variables that are insignificant means according to t-distribution whose alpha value are higher than 0.1 are dropped by using backward elimination method. Similar process with database1, after obtaining the unstandardized coefficients (B) entered independent variables of linear regression load model lagged through lag 1,2....24 with categorical variables and environmental indicators for 2012, multiple regression model is constructed and results of the model given below,

Table 4.15: Multiple linear regression coefficients for entered variables using data set3

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	1672,774	116,364		14,375	0		
Time	-0,046	0,008	-0,028	-5,797	0	0,057	17,467
Dt-lag(1)	1,375	0,01	1,375	135,126	0	0,013	78,422
Dt-lag(2)	-0,637	0,017	-0,637	-37,264	0	0,005	221,016
Dt-lag(3)	0,224	0,018	0,224	12,384	0	0,004	248,619
Dt-lag(4)	-0,205	0,018	-0,205	-11,274	0	0,004	250,297
Dt-lag(5)	0,194	0,019	0,194	10,467	0	0,004	260,536
Dt-lag(6)	-0,122	0,019	-0,122	-6,526	0	0,004	262,558
Dt-lag(7)	0,067	0,018	0,067	3,782	0	0,004	240,967
Dt-lag(8)	-0,076	0,018	-0,076	-4,288	0	0,004	240,176
Dt-lag(9)	0,037	0,019	0,037	1,999	0,046	0,004	262,219
Dt-lag(10)	-0,041	0,019	-0,041	-2,19	0,029	0,004	265,409
Dt-lag(11)	0,098	0,019	0,098	5,266	0	0,004	263,811
Dt-lag(12)	-0,056	0,018	-0,056	-3,017	0,003	0,004	256,486
Dt-lag(13)	0,051	0,018	0,051	2,801	0,005	0,004	248,904
Dt-lag(14)	-0,1	0,017	-0,1	-5,805	0	0,004	223,627
Dt-lag(15)	0,041	0,012	0,041	3,306	0,001	0,009	115,963
Dt-lag(17)	0,049	0,012	0,049	4,019	0	0,009	110,347
Dt-lag(18)	-0,062	0,017	-0,062	-3,653	0	0,005	215,717
Dt-lag(19)	0,092	0,018	0,092	5,089	0	0,004	249,437
Dt-lag(20)	-0,129	0,018	-0,129	-7,034	0	0,004	253,795
Dt-lag(21)	0,122	0,017	0,122	7,104	0	0,005	221,85
Dt-lag(22)	-0,194	0,018	-0,193	-10,897	0	0,004	238,339
Dt-lag(23)	0,448	0,017	0,448	26,741	0	0,005	212,117
Dt-lag(24)	-0,236	0,01	-0,236	-23,698	0	0,013	75,251

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Hour_1	-177,123	51,084	-0,008	-3,467	0,001	0,226	4,425
Hour_2	-157,117	48,008	-0,007	-3,273	0,001	0,256	3,908
Hour_3	-345,645	42,222	-0,016	-8,186	0	0,331	3,023
Hour_6	-76,354	44,966	-0,004	-1,698	0,09	0,292	3,429
Hour_7	292,061	45,037	0,014	6,485	0	0,291	3,44
Hour_8	1744,682	43,734	0,083	39,893	0	0,308	3,244
Hour_9	522,375	45,265	0,025	11,54	0	0,288	3,475
Hour_11	812,333	45,419	0,038	17,885	0	0,286	3,498
Hour_12	-626,047	46,137	-0,03	-13,569	0	0,277	3,61
Hour_13	1075,989	44,81	0,051	24,012	0	0,294	3,405
Hour_14	321,076	46,708	0,015	6,874	0	0,27	3,7
Hour_16	559,664	46,9	0,026	11,933	0	0,268	3,73
Hour_17	411,681	51,228	0,019	8,036	0	0,225	4,45
Hour_18	355,975	53,043	0,017	6,711	0	0,21	4,771
Hour_19	288,018	57,055	0,014	5,048	0	0,181	5,52
Hour_20	416,193	56,796	0,02	7,328	0	0,183	5,47
Hour_21	-260,413	57,862	-0,012	-4,501	0	0,176	5,678
Hour_22	731,649	57,914	0,035	12,633	0	0,176	5,688
Hour_23	-235,681	56,621	-0,011	-4,162	0	0,184	5,437
Hour_24	-531,561	53,824	-0,025	-9,876	0	0,204	4,913
Tue	-216,807	23,284	-0,018	-9,311	0	0,352	2,845
Wed	-242,231	24,358	-0,02	-9,944	0	0,326	3,065
Thu	-238,057	25,087	-0,02	-9,489	0	0,308	3,251
Fri	-278,64	25,304	-0,023	-11,012	0	0,302	3,307
Sat	-420,189	25,975	-0,035	-16,177	0	0,287	3,485
Sun	-685,426	28,187	-0,057	-24,317	0	0,24	4,169
23 April	-304,761	95,765	-0,004	-3,182	0,001	0,945	1,058
May.01	-191,61	96,256	-0,002	-1,991	0,047	0,935	1,069
30 Agu	-239,945	95,895	-0,003	-2,502	0,012	0,942	1,061
Ram. B	-604,32	65,999	-0,013	-9,156	0	0,667	1,5
Ram.B AR	-624,475	135,83	-0,005	-4,597	0	0,938	1,066
Cumh. B.	-289,255	97,396	-0,004	-2,97	0,003	0,913	1,095
Cumh. B. AR	548,235	142,183	0,005	3,856	0	0,856	1,168
Kurb. B	-435,973	64,613	-0,011	-6,747	0	0,523	1,911
Kurb. B. AR	-760,128	136,402	-0,007	-5,573	0	0,93	1,075
April	-35,62	21,244	-0,002	-1,677	0,094	0,695	1,439
June	204,446	27,901	0,013	7,328	0	0,403	2,482
July	401,561	38,458	0,026	10,442	0	0,206	4,858
August	390,196	40,297	0,026	9,683	0	0,188	5,333
September	290,388	39,335	0,019	7,382	0	0,203	4,932
October	189,547	40,808	0,012	4,645	0	0,183	5,469
November	271,808	47,008	0,018	5,782	0	0,142	7,044
December	363,595	57,318	0,024	6,343	0	0,093	10,79
HDH	6,799	1,604	0,01	4,239	0	0,261	3,828

Table 4.16: Electricity load forecasting model summary for data set3

R	R-Squ.	Adj.- R Squ.	Std. Err.	Dur.Wat
0,994	0,988	0,988	455,4063	2,232

Table 4.17: Electricity load forecasting model fit of data set3

Fit Statistic	Mean
Stationary R-squared	0,988
R-squared	0,988
RMSE	455,406
MAPE	1,229
MaxAPE	20,591
MAE	330,678
MaxAE	4786,282
Normalized BIC	12,314

In Table 4.16 and 4.17 R-square is 0,988 means total variation in forecasted electricity load is explained %98,8 by singificant 24hr lagged variables,categorical variables and environmental indicators. For the autocorrelation, durbin-watson statistics is 2.232 which is quite normal since its between 1.5 and 2.5 range. MAPE is also very low that is % 1,229 indicates good fit for the model.

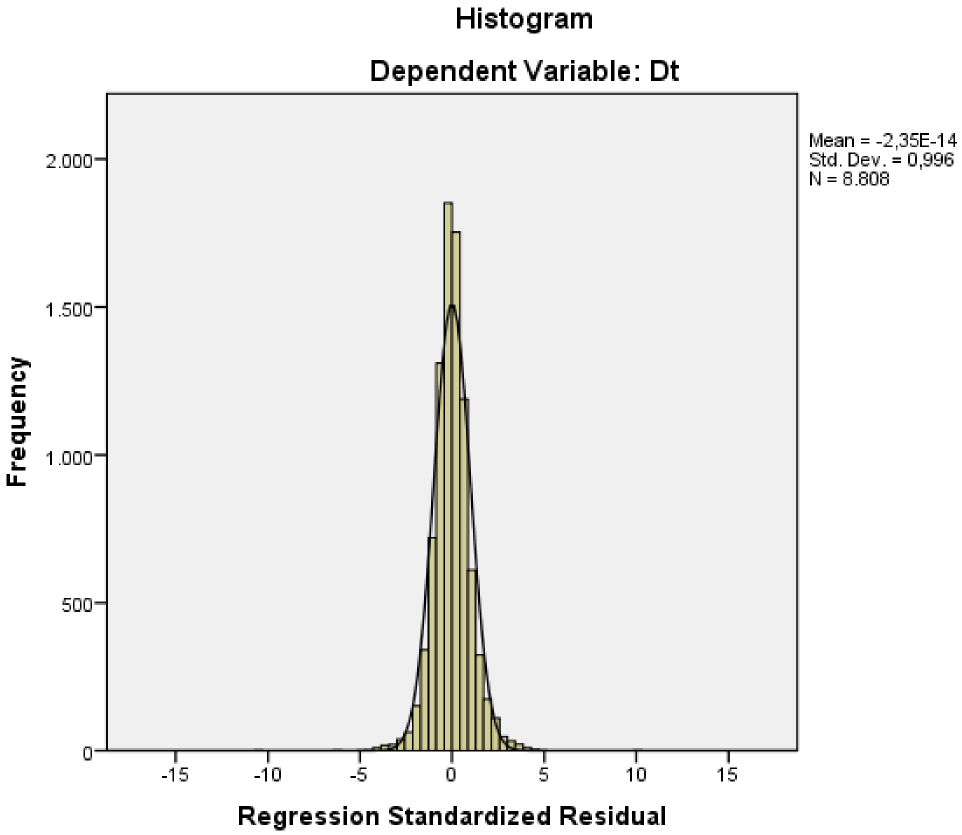

Figure 4.9: Histogram of forecasted load as a dependent variable of data set3

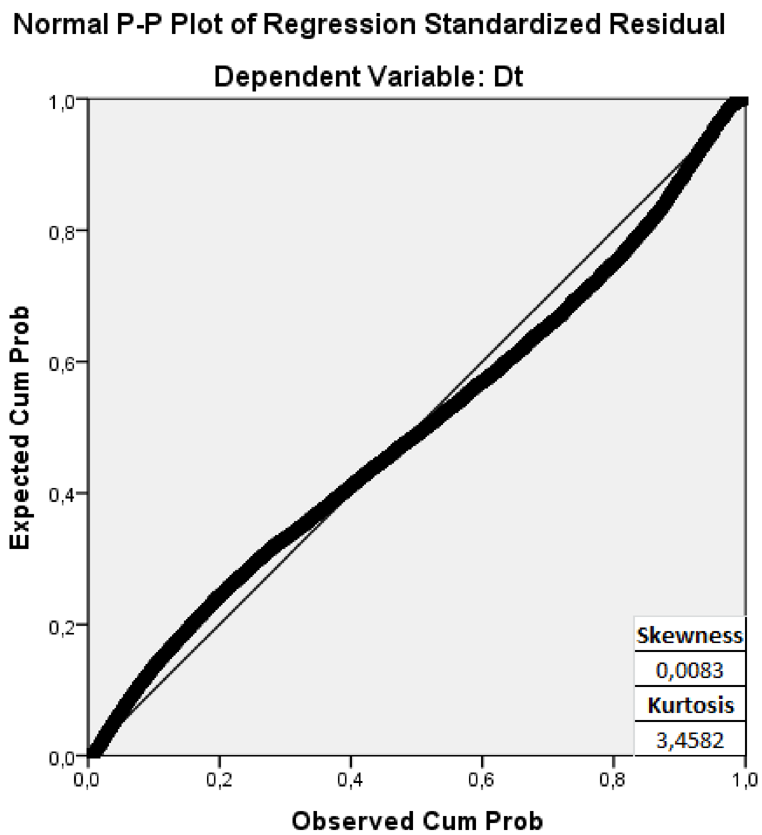

Figure 4.10: Normal probability plot of regression standardized residual for data set3

In Figure 4.9 and 4.10 illustrate a normal distribution of residuals produced by a model for load forecasting.Residual normal probability plot and histogram shows that standardized residuals randomly distributed around zero.Skewness value is 0,0083 that shows the amount and direction of skew departure from horizontal symmetry. Since its in the range between and +, the distribution is approximately symmetric. Also, kurtosis value is 3,4582 that represents how tall and sharp the central peak is relative to a standard bell curve. Since it is more than 3, the sample very likely has positive excess kurtosis means leptokurtic. For the collinearity check, VIF values are higher for lagged variables. It is expected since autocorrelation and partial autocorrelation functions indicate high correlations between actual loads and its lagged variables. Also, vif of independent variable december is 10,79 so little bit higher than 10 value. It is also winter month that has influental effect on electricity load consumption since economic activities are high in December,2012 (e.g. quarterly GDP of Turkey) hence overall there is no collinearity problem detected.

Table 4.18: Crossvalidation test on %20 and %80 of the sample of data set3

Correlations				
			Dt	Model
20%sample	Dt	Pearson Correlation	1	,993**
		Sig. (2-tailed)		0
		N	1769	1769
	Model	Pearson Correlation	,993**	1
		Sig. (2-tailed)	0	
		N	1769	1769
80%sample	Dt	Pearson Correlation	1	,994**
		Sig. (2-tailed)		0
		N	6967	6967
	Model	Pearson Correlation	,994**	1
		Sig. (2-tailed)	0	
		N	6967	6967
** Correlation is significant at the 0.01 level (2-tailed).				

All in all, according to cross validation test on %20 and %80 of the sample of data set3, correlation coefficients are highly correlate with each other that are for %20 of the sample is 0,993 and %80 of the samples are 0,994.

4.1.7 Multiple Linear Regression of Load Model using data set4

As the general form of the multiple linear regression model discussed in section 2.1; to obtain regression coefficients, regression analysis is applied and variables that are insignificant means according to t-distribution whose alpha value are higher than 0.1 are dropped by using backward elimination method. Sample of unstandardized coefficients (B) of entered independent variables of linear regression load model lagged through lag1 and lag2 with categorical variables and environmental indicators in 2012 are given below

Table 4.19: Multiple linear regression coefficients for entered variables using data set4

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	2926,599	90,388		32,378	0		
Time	0,05	0,008	0,03	6,027	0	0,066	15,175
Dt-lag(1)	1,354	0,009	1,354	153,798	0	0,021	47,846
Dt-lag(2)	-0,484	0,008	-0,484	-60,542	0	0,025	39,497
Hour_4	106,817	29,114	0,005	3,669	0	0,852	1,173
Hour_7	1203,435	30,013	0,057	40,098	0	0,802	1,247
Hour_8	2984,766	33,537	0,141	88,999	0	0,642	1,557
Hour_9	1399,412	46,993	0,066	29,779	0	0,327	3,056
Hour_10	776,236	44,925	0,037	17,279	0	0,358	2,793
Hour_11	1134,913	41,149	0,054	27,58	0	0,427	2,343
Hour_12	-166,275	41,186	-0,008	-4,037	0	0,426	2,347
Hour_13	1818,023	36,996	0,086	49,141	0	0,528	1,894
Hour_14	1137,888	39,218	0,054	29,015	0	0,47	2,128
Hour_15	635,437	39,589	0,03	16,051	0	0,461	2,169
Hour_16	1195,813	37,915	0,057	31,539	0	0,503	1,989
Hour_17	722,755	38,267	0,034	18,887	0	0,493	2,026
Hour_18	655,423	37,035	0,031	17,697	0	0,527	1,898
Hour_19	937,713	35,805	0,044	26,19	0	0,564	1,774
Hour_20	986,75	35,6	0,047	27,718	0	0,57	1,754
Hour_21	515,597	35,831	0,024	14,39	0	0,563	1,777
Hour_22	1064,569	34,497	0,05	30,86	0	0,607	1,647
Hour_23	-392,702	35,132	-0,019	-11,178	0	0,585	1,708
Hour_24	-700,698	32,107	-0,033	-21,824	0	0,701	1,427

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Tue	51,005	20,42	0,004	2,498	0,013	0,568	1,762
Wed	58,283	20,545	0,005	2,837	0,005	0,561	1,783
Thu	85,805	20,696	0,007	4,146	0	0,553	1,809
Fri	42,819	20,605	0,004	2,078	0,038	0,558	1,794
Sat	-105,158	20,461	-0,009	-5,139	0	0,565	1,769
Sun	-442,307	22,621	-0,037	-19,553	0	0,455	2,196
23 April	-353,561	105,761	-0,004	-3,343	0,001	0,947	1,056
May.01	-351,485	105,812	-0,004	-3,322	0,001	0,946	1,057
30 Agu	-343,929	106,23	-0,004	-3,238	0,001	0,938	1,066
Ram. B	-1105,24	69,473	-0,024	-15,909	0	0,735	1,36
Ram.B AR	-695,986	149,168	-0,006	-4,666	0	0,951	1,052
Cumh. B.	-610,277	107,051	-0,008	-5,701	0	0,924	1,082
Cumh. B. AR	618,733	156,59	0,005	3,951	0	0,863	1,159
Kurb. B	-1089,69	65,585	-0,027	-16,615	0	0,621	1,611
Kurb. B. AR	-907,121	149,826	-0,008	-6,054	0	0,942	1,061
February	-58,961	25,588	-0,004	-2,304	0,021	0,604	1,656
March	-159,936	24,513	-0,011	-6,524	0	0,619	1,615
April	-405,644	27,847	-0,026	-14,567	0	0,494	2,022
May	-456,956	28,509	-0,03	-16,029	0	0,458	2,184
June	-121,282	28,734	-0,008	-4,221	0	0,464	2,153
July	178,703	26,43	0,012	6,761	0	0,533	1,877
September	-265,62	34,431	-0,017	-7,715	0	0,323	3,092
October	-575,043	42,958	-0,038	-13,386	0	0,202	4,958
November	-520,684	49,702	-0,034	-10,476	0	0,155	6,442
December	-387,324	59,055	-0,026	-6,559	0	0,107	9,37
HDH	13,085	1,866	0,018	7,012	0	0,236	4,234
CDH	15,454	4,352	0,007	3,551	0	0,391	2,558

As a result of linear regression load model lagged through lag1 and lag2 with categorical variables and environmental indicators in 2012,

Table 4.20: Electricity load forecasting model summary for data set4

R	R Squ.	Adj.R-Squ.	Std. Err.	Durb.Wat.
0,993	0,986	0,986	503,438	1,883

Table 4.21: Electricity load forecasting model fit of data set4

Fit Statistic Mean	
Stationary R-squared	0,986
R-squared	0,986
RMSE	503,438
MAPE	1,358
MaxAPE	21,286
MAE	367,962
MaxAE	4947,751
Normalized BIC	12,495

In table 4.20 and 4.21 R-square is 0,986 means total variation in forecasted electricity load is explained %98,6 by lag1 and lag2 variables,categorical variables and environmental indicators. For the autocorrelation, durbin-watson statistics is 1,883 which is quite normal since its between 1.5 and 2.5 range. MAPE is also very low that is % 1,358 indicates good fit for the model.

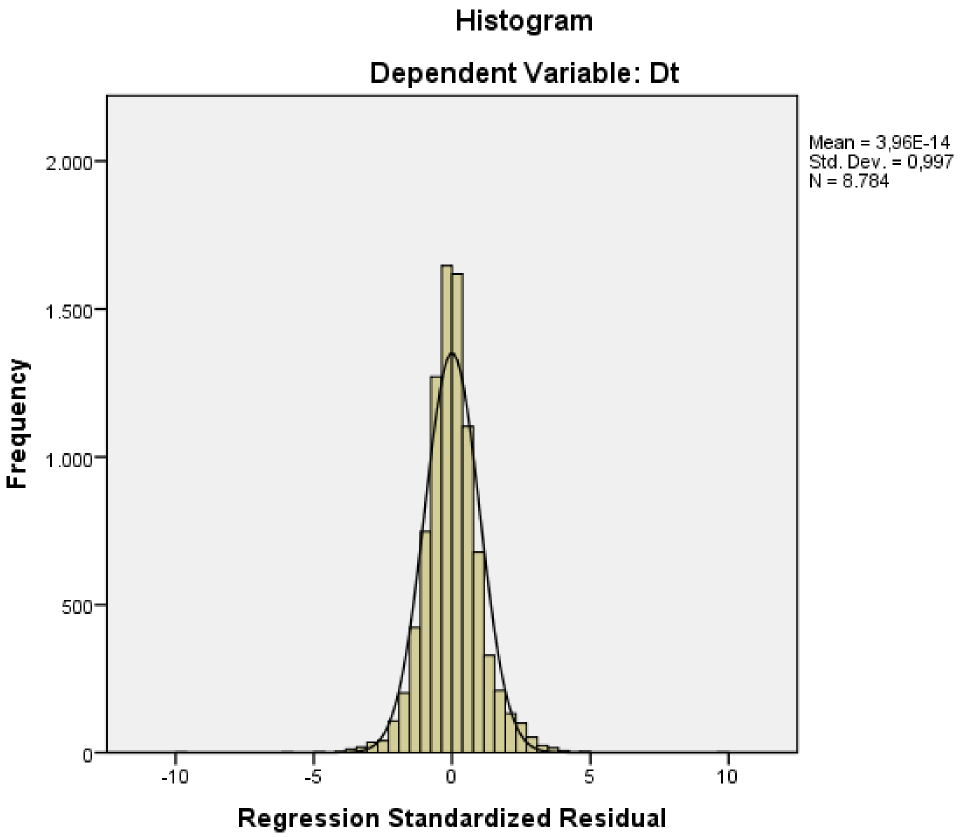

Figure 4.11: Histogram of forecasted load as a dependent variable of data set4

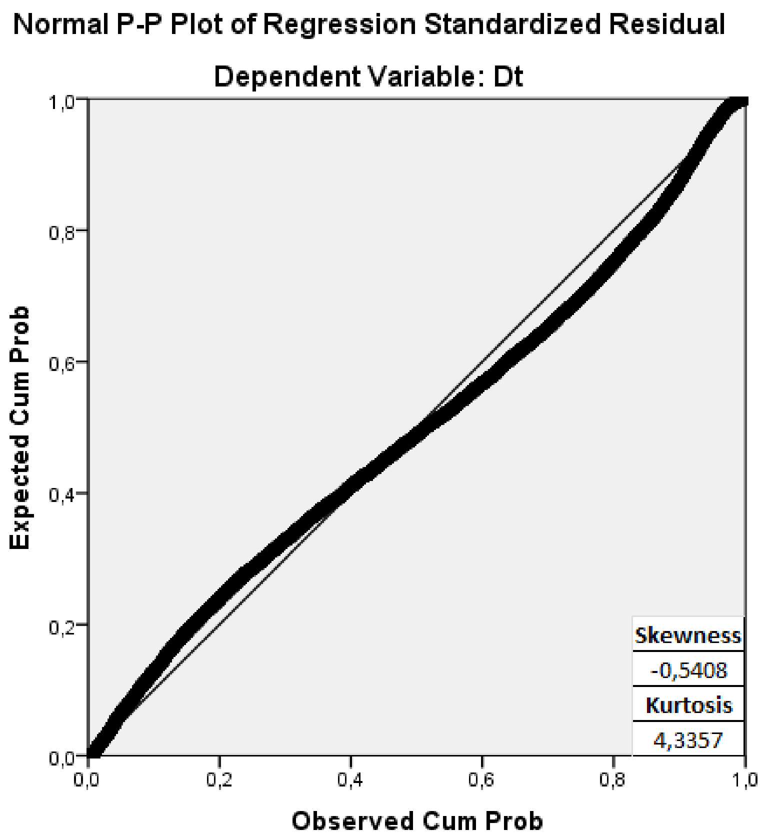

Figure 4.12: Normal probability plot of regression standardized residual for data set4

In figure 4.11 and 4.12 illustrate a normal distribution of residuals produced by a model for load forecasting.Residual normal probability plot and histogram shows that standardized residuals randomly distributed around zero.Skewness value is -0,5408 that shows the amount and direction of skew departure from horizontal symmetry. Since its in the range between 1 and , the distribution is moderately skewed. However since the value is pretty close to , its very close to be approximately symmetric Also, kurtosis value is 4,3357 that represents how tall and sharp the central peak is relative to a standard bell curve. Since it is more than 3, the sample very likely has positive excess kurtosis means leptokurtic. For the collinearity check, VIF values are only higher for lagged variables which means there is no collinearity problem detected.

Table 4.22: Crossvalidation test on %20 and %80 of the sample of data set4

Correlations				
			Dt	Model
20%sample	Dt	Pearson Correlation	1	,992**
		Sig. (2-tailed)		0
		N	1765	1765
	Model	Pearson Correlation	,992**	1
		Sig. (2-tailed)	0	
		N	1765	1765
80%sample	Dt	Pearson Correlation	1	,992**
		Sig. (2-tailed)		0
		N	6971	6971
	Model	Pearson Correlation	,992**	1
		Sig. (2-tailed)	0	
		N	6971	6971
** Correlation is significant at the 0.01 level (2-tailed).				

Consequently, according to cross validation test on %20 and %80 of the sample of data set4, correlation coefficients are highly correlate with each other that are for both %20 and %80 of the samples are 0,992.

4.2 Construction of Short Term Price Forecasting Model Architecture

4.2.1 Proposed Autoregressive Electricity Price Model for 2015

In the beginning of the study, initial data obtained from EPIAS in 2015, Turkey with 24 hr lagged variables are examined through autocorrelation and partial autocorrelation functions.

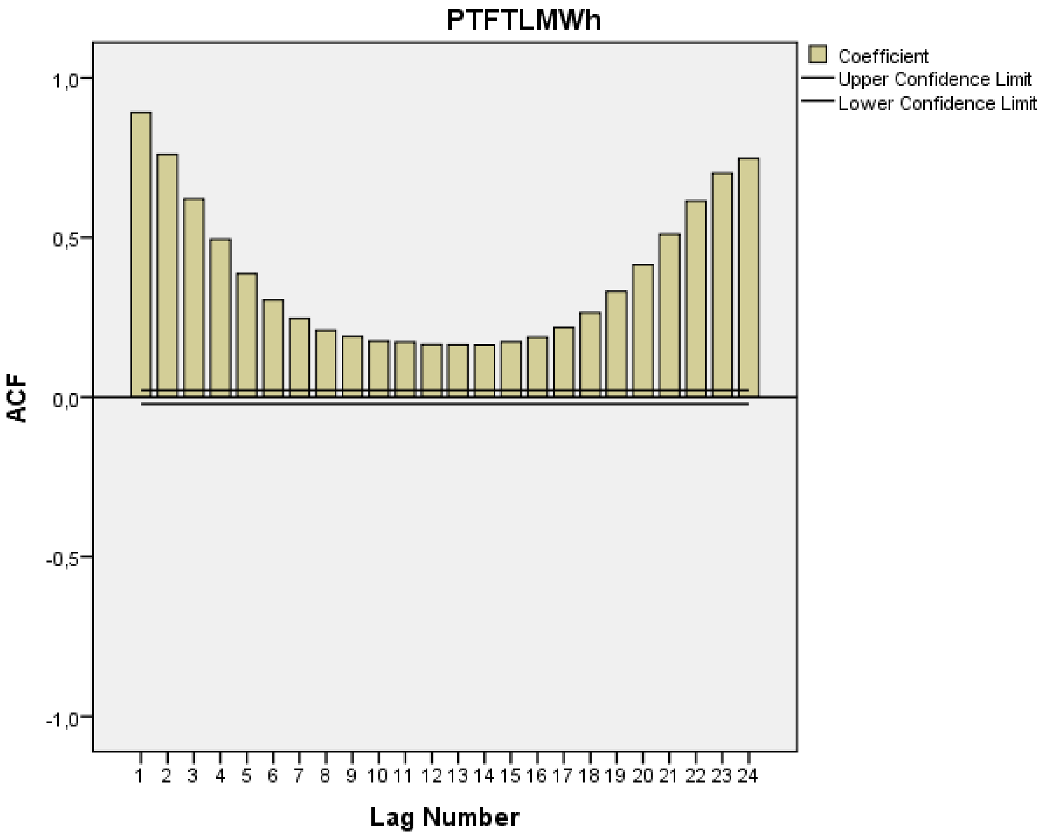

Figure 4.13: Autocorrelation function plot of electricity prices in Turkey

Table 4.23: Autocorrelations of electricity prices of 24hr lagged variables

Lag	Autocorr.	Std. Error	Box-Ljung	df	Sig.b
1	0,891	0,011	6952,99	1	0
2	0,76	0,011	12009,28	2	0
3	0,62	0,011	15379,46	3	0
4	0,493	0,011	17514,3	4	0
5	0,387	0,011	18824,99	5	0
6	0,304	0,011	19636,63	6	0
7	0,245	0,011	20164,27	7	0
8	0,209	0,011	20545,9	8	0
9	0,189	0,011	20860,05	9	0
10	0,175	0,011	21128,37	10	0
11	0,172	0,011	21387	11	0
12	0,164	0,011	21623,46	12	0
13	0,164	0,011	21858,37	13	0
14	0,163	0,011	22090,52	14	0
15	0,172	0,011	22351,28	15	0
16	0,188	0,011	22660,68	16	0
17	0,218	0,011	23076,7	17	0
18	0,264	0,011	23690,45	18	0
19	0,331	0,011	24652,54	19	0
20	0,415	0,011	26163,33	20	0
21	0,511	0,011	28452,48	21	0
22	0,614	0,011	31759,26	22	0
23	0,702	0,011	36082,63	23	0
24	0,747	0,011	40987,07	24	0

As stated in autocorrelation table 4.23 and the plot 4.13, all autoregressive term coefficients are higher than standard error which means AR(24) model is significant as a whole.

Also, in order to detect the correlation between time series and its lagged values, partial autocorrelation function should be applied.

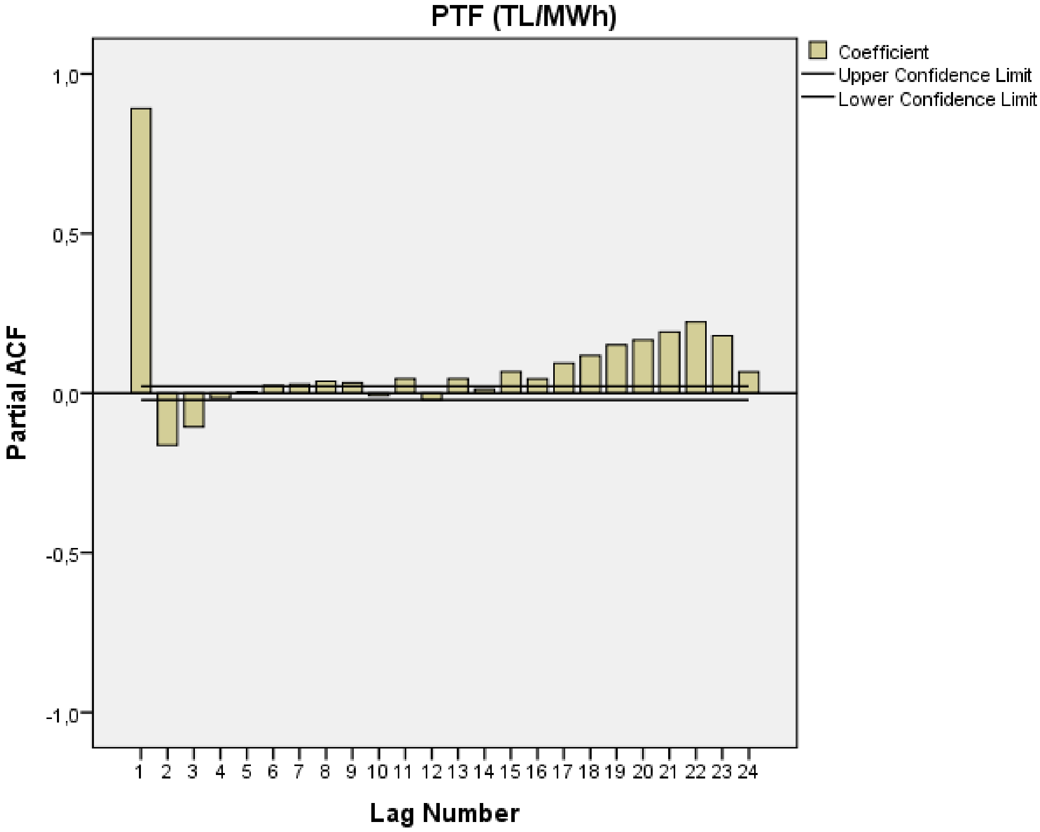

Figure 4.14: Partial autocorrelation function plot of electricity prices in Turkey

Consistent with partial autocorrelation plot, lag1 is definitely statistically significant, lag2 might also seem significant, and if model does not give the better result with two of these lags,then lag 3 can be added as an autoregressive term whereas partial autocorrelations for all other lags are not statistically significant. So other price model types can be tested with AR(2) or AR(3) as previous load models.

The AR(24) Model Summary is given

Table 4.24: Electricity price forecasting model for AR(24)Model, 2015

Fit Statistic	Mean
Stationary R-squared	0,839
R-squared	0,839
RMSE	19,668
MAPE	64,668
MaxAPE	87323,7
MAE	13,499
MaxAE	136,194
Normalized BIC	5,984

R-square indicates how much of the total variation in the dependent variable which is electricity load can be explained by the independent variable by its lagged variables. In this case it is %83,9 and MAPE:%64,668 are not sufficient enough to for an accurate model. Hence step by step AR(24) model is transformed into powerful hyrid model with existing autoregressive term(AR),by combining integration(I) and moving average(MA) parameters, ARIMA model is constructed.

Estimates of integration and moving average parameters of ARIMA are made by using residual autocorrelation and partial autocorrelation function charts of ARIMA(24,0,0)

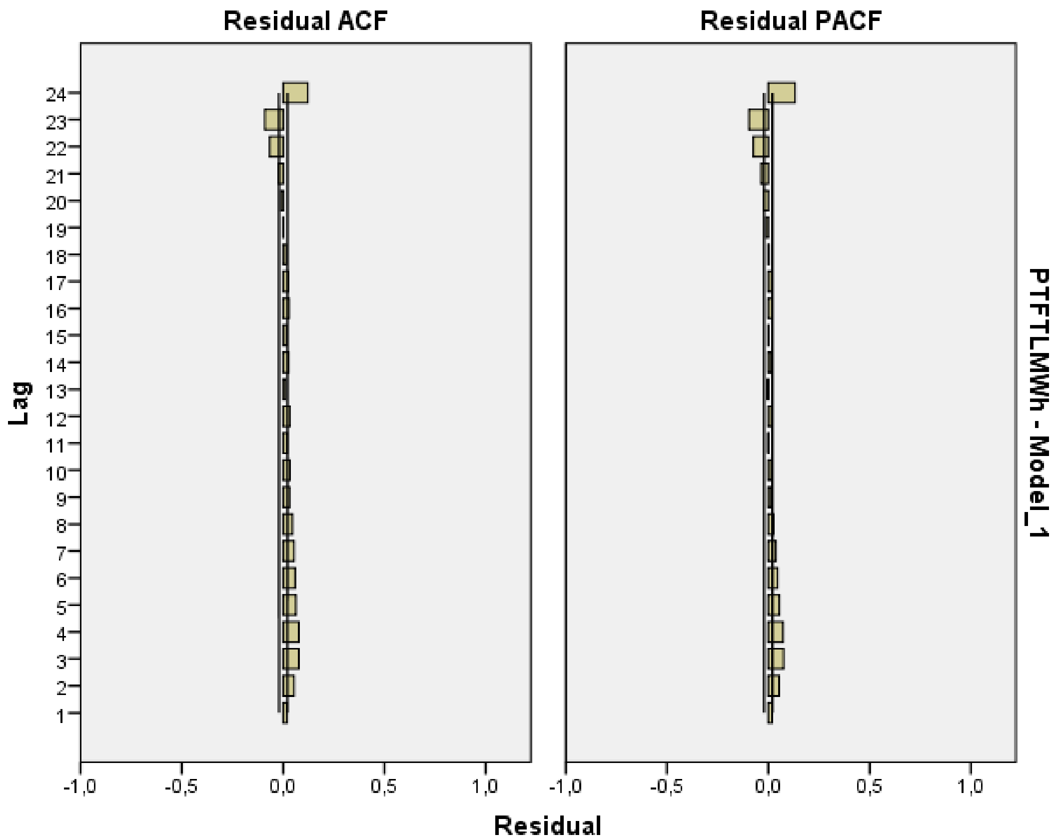

Figure 4.15: Autocorrelation and partial autocorrelation function of the residuals for ARIMA(24,0,0)

According to the charts, the differencing order refers to successive first differences, hence difference order seems 1, means the variable analyzed is $D_t - D_{t-1}$ by analyzing partial autocorrelation. Also,the ideal Autocorrelation function for residuals is all autocorrelations are 0 means.So here, moving average term might be 3 by checking cycles.

4.2.2 Proposed Autoregressive Integrated Moving Average(ARIMA) Price Model in 2015

Residual results of ARIMA(24,0,0) pointed out that p=24 d=1 and q=3 therefore ARIMA(24,1,3) is constructed.

Better results obtained as ,

Table 4.25: Summary of ARIMA(24,1,3) price forecasting model, 2015

Fit Statistic	Mean
Stationary R-squared	0,341
R-squared	0,856
RMSE	18,627
MAPE	49,742
MaxAPE	34516,1
MAE	12,46
MaxAE	137,809
Normalized BIC	5,878

Updated results of R-square: 0,856 and MAPE:%49,742 Hence with ARIMA parameters R-square increased and MAPE decreased which is a sign of better results.

According to residual autocorrelation and partial autocorrelation function charts of ARIMA(24,1,3)

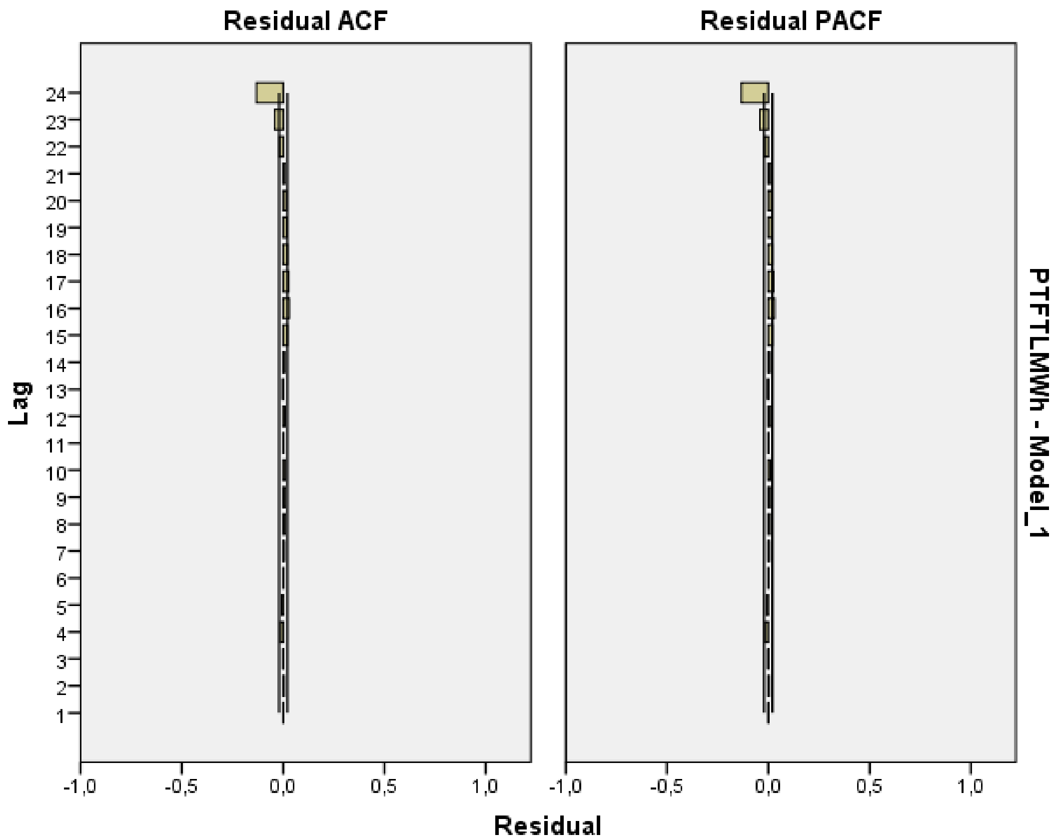

Figure 4.16: Autocorrelation and partial autocorrelation function of the residuals for ARIMA(24,1,3)

Autocorrelation function for residuals is almost all autocorrelations are zero means that is quite likely there is no need to change the parameters of ARIMA. So, the ideal ARIMA parameters are defined. Final process is combining the model with categorical variables and environmental indicators by multiple linear regression and update model one last time

4.2.3 Multiple Linear Regression of Price Model using data set5

As the general form of the multiple linear regression model discussed in section 2.1; to obtain regression coefficients, regression analysis is applied and variables that are insignificant means according to t-distribution whose alpha value are higher than 0.1 are dropped by using backward elimination method. Sample of unstandardized coefficients (B) of entered independent variables of linear regression price model lagged through lag 1,2....24 with categorical variables and environmental indicators in 2015 are given below

Table 4.26: Multiple linear regression coefficients for entered variables using data set5

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	6,815	1,452		4,694	0		
Dt(1)	0,749	0,008	0,749	89,032	0	0,212	4,719
Dt(3)	-0,032	0,008	-0,032	-3,857	0	0,218	4,58
Dt(8)	-0,021	0,011	-0,021	-1,992	0,046	0,136	7,371
Dt(9)	0,019	0,011	0,019	1,672	0,095	0,12	8,337
Dt(11)	0,038	0,008	0,038	4,564	0	0,22	4,552
Dt(14)	-0,023	0,008	-0,023	-2,932	0,003	0,253	3,955
Dt(17)	0,026	0,009	0,026	3,004	0,003	0,207	4,842
Dt(19)	0,024	0,009	0,024	2,564	0,01	0,167	5,975
Dt(21)	-0,023	0,01	-0,023	-2,393	0,017	0,165	6,062
Dt(23)	0,096	0,012	0,096	8,264	0	0,111	9,012
Dt(24)	0,063	0,011	0,063	5,942	0	0,133	7,539
Hour_1	-11,689	1,315	-0,048	-8,886	0	0,518	1,929
Hour_2	-20,27	1,306	-0,083	-15,526	0	0,526	1,9
Hour_3	-18,109	1,345	-0,074	-13,464	0	0,496	2,016
Hour_4	-8,463	1,357	-0,035	-6,235	0	0,487	2,054
Hour_5	-6,885	1,365	-0,028	-5,042	0	0,481	2,078
Hour_6	-6,687	1,385	-0,027	-4,828	0	0,468	2,138
Hour_7	13,005	1,335	0,053	9,742	0	0,503	1,986
Hour_8	20,289	1,272	0,083	15,949	0	0,554	1,804
Hour_9	12,984	1,276	0,053	10,175	0	0,551	1,815
Hour_10	4,063	1,239	0,017	3,28	0,001	0,585	1,71
Hour_11	4,839	1,223	0,02	3,955	0	0,599	1,668
Hour_12	-7,188	1,213	-0,029	-5,925	0	0,61	1,641
Hour_13	6,547	1,186	0,027	5,521	0	0,638	1,567
Hour_14	6,3	1,189	0,026	5,297	0	0,634	1,577
Hour_16	4,027	1,097	0,016	3,671	0	0,746	1,341
Hour_17	-5,064	1,101	-0,021	-4,598	0	0,74	1,352
Hour_21	-5,906	1,132	-0,024	-5,217	0	0,7	1,428
Hour_22	6,56	1,208	0,027	5,431	0	0,615	1,626
Hour_23	-9,735	1,265	-0,04	-7,695	0	0,561	1,784
Hour_24	7,317	1,273	0,03	5,747	0	0,553	1,807

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Mon	9,748	0,777	0,07	12,549	0	0,485	2,062
Tue	6,547	0,747	0,047	8,765	0	0,524	1,907
Wed	5,49	0,739	0,039	7,431	0	0,536	1,866
Thu	6,165	0,745	0,044	8,27	0	0,527	1,899
Fri	4,635	0,735	0,033	6,309	0	0,542	1,845
Sat	4,565	0,728	0,033	6,27	0	0,552	1,811
23 April	-6,536	3,708	-0,007	-1,762	0,078	0,951	1,052
May.01	-8,804	3,707	-0,009	-2,375	0,018	0,951	1,051
Ram. B	-8,666	2,344	-0,016	-3,696	0	0,797	1,254
Ram.B AR	-10,933	5,2	-0,008	-2,103	0,036	0,966	1,035
Kurb. B. AR	-13,48	5,155	-0,01	-2,615	0,009	0,983	1,018
January	3,575	0,811	0,02	4,408	0	0,72	1,389
March	-1,297	0,747	-0,007	-1,737	0,082	0,825	1,213
April	-2,7	0,837	-0,015	-3,226	0,001	0,676	1,479
May	-2,378	0,801	-0,014	-2,968	0,003	0,716	1,397
June	-2,018	0,763	-0,011	-2,644	0,008	0,813	1,23
July	-2,597	0,895	-0,015	-2,903	0,004	0,574	1,741
August	-3,545	0,941	-0,02	-3,769	0	0,52	1,924
December	2,183	0,764	0,012	2,859	0,004	0,789	1,268
CDH	0,933	0,141	0,042	6,627	0	0,377	2,65

With the combination of 24 hr lagged and categorical variables and environmental indicators' coefficients multiple linear regression model is constructed by using equation 2.1. Model results are given below

Table 4.27: Electricity price forecasting model summary of data set5

R	R Squ.	Adj R-Squ.	Std. Err.	Durb.Wat
0,933	0,87	0,869	17,70828	1,984

Table 4.28: Electricity load forecasting model fit of data set5

Fit Statistic	Mean
Stationary R-squared	0,87
R-squared	0,87
RMSE	17,708
MAPE	55,585
MaxAPE	36150,48
MAE	12,408
MaxAE	114,867
Normalized BIC	5,835

In table 4.31 and 4.28 R-square is 0,870 means total variation in forecasted electricity price is explained %87,0 by significant lagged variables,categorical variables and environmental indicators.Comparing with load models R-square is lower than previous models. For the autocorrelation, durbin-watson statistics is 1,984 which is still normal since its between 1.5 and 2.5 range. MAPE is also very high that is % 55,585 might not indicate a good fit for the model.

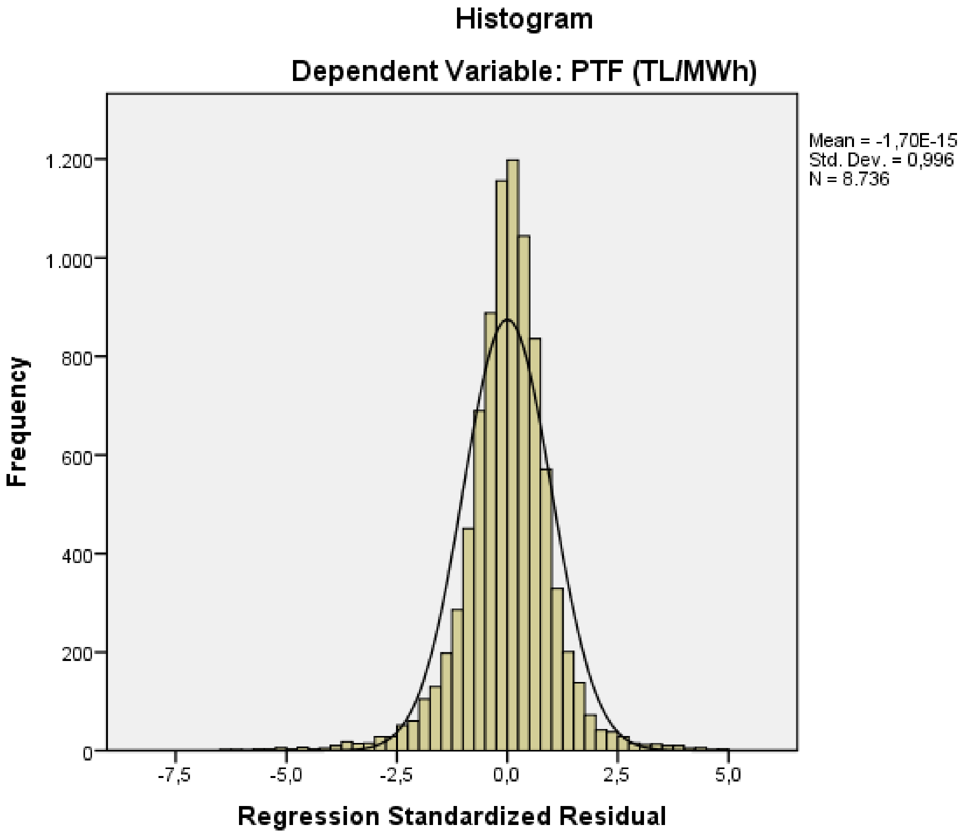

Figure 4.17: Histogram of forecasted load as a dependent variable of data set5

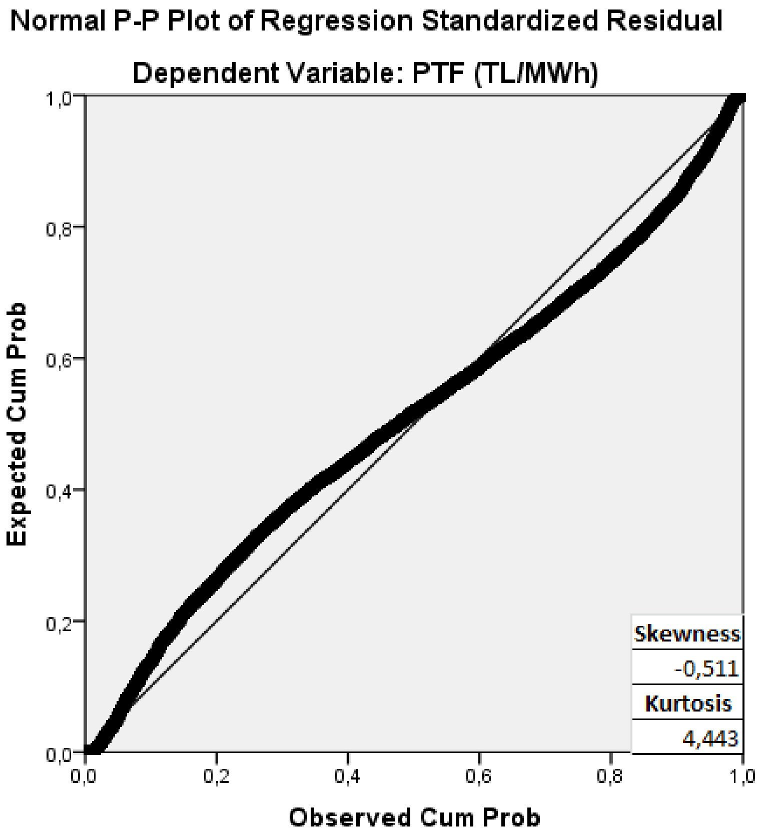

Figure 4.18: Normal probability plot of regression standardized residual for data set5

In figure 4.17 and 4.18 illustrate a normal distribution of residuals produced by a model for price forecasting.Residual normal probability plot and histogram shows that standardized residuals randomly distributed around zero.Skewness value is -0,511 that shows the amount and direction of skew departure from horizontal symmetry. Since its in the range between 1 and , the distribution is moderately skewed. However since the value is pretty close to , its very close to ve approximately symmetric Also, kurtosis value is 4,443 that represents how tall and sharp the central peak is relative to a standard bell curve. Since it is more than 3, the sample very likely has positive excess kurtosis means leptokurtic. For the collinearity check, VIF values are only higher for lagged variables which means there is no collinearity problem detected.

Table 4.29: Crossvalidation test on %20 and %80 of the sample of data set5

Correlations				
			Dt	Model
20%sample	Dt	Pearson Correlation	1	,934**
		Sig. (2-tailed)		0
		N	1749	1749
	Model	Pearson Correlation	,934**	1
		Sig. (2-tailed)	0	
		N	1749	1750
80%sample	Dt	Pearson Correlation	1	,932**
		Sig. (2-tailed)		0
		N	6987	6987
	Model	Pearson Correlation	,932**	1
		Sig. (2-tailed)	0	
		N	6987	7010
** Correlation is significant at the 0.01 level (2-tailed).				

Finally, according to cross validation test on %20 and %80 of the sample of data set5, correlation coefficients are highly correlate with each other that are for %20 of the sample is 0,934 and %80 of the samples are 0,932.

4.2.4 Multiple Linear Regression of Hybrid Price Model

In multiple linear regression, each term of ARIMA p=24,d=1,q=3 as independent variables by combining categorical variables and environmental indicators therefore the new proposed model counted as composite hybrid regression model.

Table 4.30: Multiple linear regression coefficients for entered variables for hybrid price model

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	12,954	1,523		8,508	0		
Dt(3)	-0,023	0,011	-0,023	-2,183	0,029	0,135	7,43
Dt(8)	-0,021	0,011	-0,021	-1,982	0,048	0,136	7,371
Dt(9)	0,019	0,011	0,019	1,669	0,095	0,12	8,337
Dt(11)	0,038	0,008	0,038	4,556	0	0,22	4,553
Dt(14)	-0,022	0,008	-0,022	-2,88	0,004	0,253	3,958
Dt(17)	0,026	0,009	0,026	3,029	0,002	0,207	4,84
Dt(19)	0,024	0,009	0,024	2,549	0,011	0,167	5,976
Dt(21)	-0,023	0,01	-0,023	-2,366	0,018	0,165	6,066
Dt(23)	0,095	0,012	0,096	8,214	0	0,111	9,018
Dt(24)	0,063	0,011	0,063	5,903	0	0,133	7,525
I1	0,388	0,011	0,182	35,93	0	0,586	1,707
MA2	0,74	0,011	0,72	68,404	0	0,136	7,378
Hour_1	-11,944	1,324	-0,049	-9,018	0	0,511	1,955
Hour_2	-20,146	1,311	-0,082	-15,364	0	0,522	1,917
Hour_3	-18,14	1,344	-0,074	-13,495	0	0,497	2,014
Hour_4	-8,615	1,359	-0,035	-6,339	0	0,486	2,059
Hour_5	-7,096	1,371	-0,029	-5,176	0	0,477	2,095
Hour_6	-6,789	1,386	-0,028	-4,9	0	0,467	2,14
Hour_7	12,929	1,336	0,053	9,679	0	0,503	1,989
Hour_8	20,043	1,286	0,082	15,583	0	0,542	1,844
Hour_9	12,933	1,277	0,053	10,129	0	0,55	1,817
Hour_10	4,162	1,24	0,017	3,356	0,001	0,583	1,715
Hour_11	4,956	1,226	0,02	4,042	0	0,597	1,676
Hour_12	-7,19	1,213	-0,029	-5,926	0	0,61	1,641
Hour_13	6,668	1,189	0,027	5,61	0	0,635	1,575
Hour_14	6,136	1,196	0,025	5,129	0	0,627	1,596
Hour_16	4,075	1,097	0,017	3,713	0	0,745	1,342
Hour_17	-5,107	1,102	-0,021	-4,636	0	0,739	1,353
Hour_21	-5,922	1,132	-0,024	-5,232	0	0,7	1,428
Hour_22	6,6	1,208	0,027	5,462	0	0,614	1,627
Hour_23	-9,886	1,269	-0,04	-7,793	0	0,557	1,794
Hour_24	7,435	1,277	0,03	5,821	0	0,55	1,819

	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Mon	3,673	0,632	0,026	5,808	0	0,731	1,367
Fri	-1,429	0,586	-0,01	-2,436	0,015	0,851	1,176
Sat	-1,504	0,584	-0,011	-2,577	0,01	0,859	1,164
Sun	-6,065	0,619	-0,043	-9,803	0	0,764	1,308
23 April	-6,411	3,684	-0,007	-1,74	0,082	0,963	1,038
May.01	-8,781	3,707	-0,009	-2,369	0,018	0,951	1,051
Ram. B	-8,657	2,344	-0,016	-3,693	0	0,797	1,254
Ram.B AR	-10,807	5,183	-0,008	-2,085	0,037	0,972	1,029
Kurb. B. AR	-14,024	5,139	-0,011	-2,729	0,006	0,989	1,011
January	3,582	0,811	0,02	4,416	0	0,72	1,389
March	-1,286	0,747	-0,007	-1,723	0,085	0,825	1,212
April	-2,732	0,837	-0,015	-3,265	0,001	0,676	1,478
May	-2,389	0,801	-0,014	-2,981	0,003	0,716	1,397
June	-2,01	0,763	-0,011	-2,634	0,008	0,813	1,23
July	-2,63	0,894	-0,015	-2,941	0,003	0,575	1,74
August	-3,555	0,941	-0,02	-3,779	0	0,52	1,924
December	2,188	0,763	0,012	2,865	0,004	0,789	1,268
CDH	0,936	0,141	0,042	6,647	0	0,377	2,649

As a results of regression,

Table 4.31: Electricity price forecasting hybrid model summary of data set5

R	R Squ.	Adj R-Squ.	Std. Err.	Durb.Wat
0,933	0,874	0,873	17,77488	1,990

Table 4.32: Electricity price forecasting hybrid model fit, 2015

Fit Statistic	Mean
Stationary R-squared	0,423
R-squared	0,874
RMSE	17,475
MAPE	52,66
MaxAPE	34738,52
MAE	12,121
MaxAE	134,724
Normalized BIC	5,809

Hybrid model gives MAPE:%52,66 and R-Square:%87,4 , According to residual autocorrelation and partial autocorrelation function charts

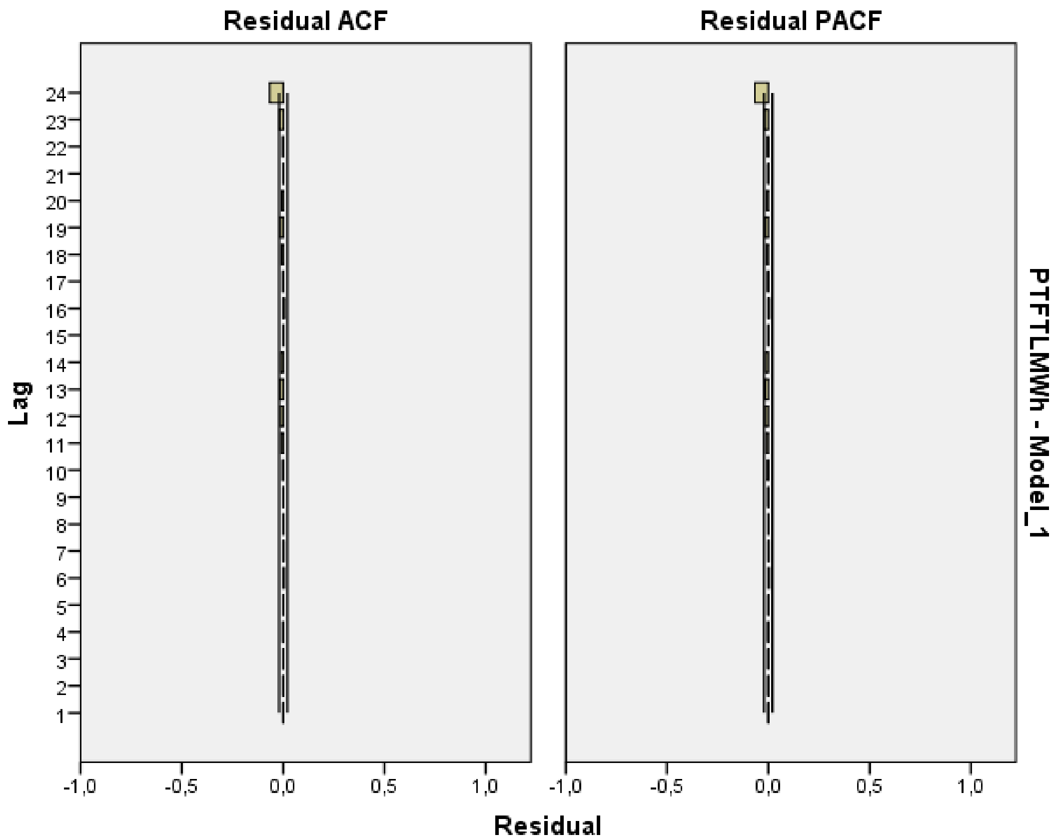

Figure 4.19: Autocorrelation and partial autocorrelation function of the residuals for hybrid model

Since residuals are lower than standard error, it means model has reached the end of improvement steps.

4.3 Comparison of Short Term Load and Price Forecasting Models

4.3.1 Model Classification

For every data set (DS), different models are constructed that are named, descripted and given model results in Table 4.33 in order to compare with each other as below. Time Series Models including AR and ARIMA models consist of only model parameters without other categorical variables and environmental indicators in their specific datasets. On the other hand, multiple linear regression models combines all independent variables with significant lagged variables, categorical variables and environmental indicators together.

Table 4.33: Summary of all load and price forecasting models descriptions

	Name	Description	R-square	MAPE
Load	DS1-M1	Time Series- AR(24) Load Model without other variables for 2011	0,972	1,69
	DS1-M2	Multiple Linear Regression Load Model lagged through 1..24 for 2011	0,985	1,277
	DS2-M1	Multiple Linear Regression Load Model lagged through 1 and 2 for 2011	0,984	1,348
	DS3-M1	Time Series- AR(24) Load Model without other variables for 2012	0,978	1,735
	DS3-M2	Multiple Linear Regression Load Model lagged through 1..24 for 2012	0,988	1,229
	DS4-M1	Multiple Linear Regression Load Model lagged through 1 and 2 for 2012	0,986	1,358
Price	Data Set5-M1	Time Series- AR(24) Price Model without other variables for 2015	0,839	64,668
	DS5-M2	Time Series- ARIMA(24,1,3) Price Model without other variables for 2015	0,856	49,742
	DS5-M3	Multiple Linear Regression Price Model lagged through 1..24 for 2015	0,87	55,585
	DS5-M4	Multiple Linear Regression Hybrid Price Model lagged through 1..24 for 2015	0,874	52,66

4.3.2 Comparison of the best results given Load Models in 2011 and 2012

Among the all load forecasting models Data set1-M2 in 2011 and Data set3-M2 in 2012 give better results according to highest R-square and lowest MAPE values. Here two models are compared with each other.

Table 4.34: Comparison of entering independent variables of DS1-M2 and DS3-M2

DS1-M2 (2011)				DS3-M2 (2012)			
Time	Hr-1	Mon	Jan	Time	Hr-1	Tue	Apr
Dt-lag(1)	Hr-2	Tue	Feb	Dt-lag(1)	Hr-2	Tue	June
Dt-lag(2)	Hr-3	Wed	Mar	Dt-lag(2)	Hr-3	Wed	July
Dt-lag(3)	Hr-4	Thu	Apr	Dt-lag(3)	Hr-6	Thu	Aug
Dt-lag(4)	Hr-6	Fri	June	Dt-lag(4)	Hr-7	Fri	Sep
Dt-lag(5)	Hr-7	Sat	July	Dt-lag(5)	Hr-8	Sat	Oct
Dt-lag(6)	Hr-8	Christ	Aug	Dt-lag(6)	Hr-9	Sun	Nov
Dt-lag(9)	Hr-9	May.01	Sep	Dt-lag(7)	Hr-11	23April	Dec
Dt-lag(11)	Hr-10	May.19	Oct	Dt-lag(8)	Hr-12	May.01	HDH
Dt-lag(12)	Hr-11	30Aug	Nov	Dt-lag(9)	Hr-13	30Aug	
Dt-lag(13)	Hr-12	RamB	Dec	Dt-lag(10)	Hr-14	RamB	
Dt-lag(14)	Hr-13	RamB.Ar	CDH	Dt-lag(11)	Hr-16	RamB.Ar	
Dt-lag(16)	Hr-14	KurbB		Dt-lag(12)	Hr-17	CumhB	
Dt-lag(18)	Hr-16	KurbB.Ar		Dt-lag(13)	Hr-18	CumhB.Ar	
Dt-lag(19)	Hr-17			Dt-lag(14)	Hr-19	KurbB	
Dt-lag(20)	Hr-18			Dt-lag(16)	Hr-20	KurbB.Ar	
Dt-lag(22)	Hr-19			Dt-lag(18)	Hr-21		
Dt-lag(23)	Hr-20			Dt-lag(19)	Hr-22		
Dt-lag(24)	Hr-21			Dt-lag(20)	Hr-23		
	Hr-22			Dt-lag(21)	Hr-24		
	Hr-24			Dt-lag(22)			
				Dt-lag(23)			
				Dt-lag(24)			

Generally, lagged variables from lag1 to lag24, time, categorical variables and environmental indicators have influences on electricity load consumption in both of these models in 2011 and 2012. The difference is, in 2012, Dt-lag(7),Dt-lag(8), Dt-lag(10),Dt-lag(21) variables have significant effect while in 2011 they are insignificant, means they have no effect on forecasting load consumption.Also, Dt-lag(15) is excluded from both of these models.It might be reasonable, obtaining coefficients from Table 4.4 and Table 4.15 shows the magnitude of coefficients are different from Dt-lag(14) to Dt-lag(16) and seems the reset point for positive and negative load consumption

Comparing with the hour variable, in 2012 hours from 03:00 to 04:00 and 09:00-10:00 are not significant while in 2011 it is, on the other hand hour from 22:00

to 23:00 is significant for 2012 model while it has no effect in 2011 .Also hours from 04:00 to 05:00 and 14:00-15:00 are excluded from both of these models. The change in load consumption shows the life style of the occupants: working time, leisure time, lunch time, and sleeping time and even overtime at work might change the consumption. Also, for Turkey, the daily saving time application might have highly effect on hour between 04:00 to 05:00.

Also, in day time it is observed that, base line is Sunday in 2011, means have no influental effect on load consumption likewise in 2012 it is Monday. With a simple calculation effects of day times can be observed. If it the calculation is made in 2011 in order to identify the the effect of similarity of these two models, to make Monday's coefficient is zero means decreasing its coefficient by 749,33, and it makes the Tuesday variable in 2011 pretty close to next years Tuesday which is in 2012. So, there is no huge difference of different days effect among 2011 and 2012. In other words, day types as independent variables have statistically pretty close results on both 2011 and 2012.

From the months perspective the change in economic activities, as well as the change in quarterly GDP activities can highly trigger the consumption magnitude and direction at the specific months. In this case, in 2011 the influental effect of cold weather is generally homogeneous in the winter months. For this reason, that effect can be seen very high during October, November and December, and can be explained by categorical variables. However, the effect of hot weather is remote, randomly distributed throughout the entire months. In this point, while the effect of cold weather can be seen with monthly categorical variables, but the effect of hot weather can only be modeled by CDH. Similarly the opposite effect is presented by HDH in 2012 model. Therefore while CDH can be explained as significant independent variable in the 2011, HDH can be explained as significant independent variable in the 2012 model.

Chapter 5

RESULTS and DISCUSSION

This section represents the results of the comparisons for actual and forecasted load and price values. The power of forecasted models are tested both within the year that the data is provided and the next year for the best result given models among each data set mentioned in section 4.3.2 All load units are MwH and price units are TL/MwH. Two types of days are randomly selected to compare load and price forecasts during spring (15th of April) and summer (28th of August) seasons. Even though forecasts are totally not always accurate as it always deviates from the actual values, it is observed the forecasted load curves fluctuates corresponded with the actual load curves. This shows that the tendency of the model to forecast load and price consumption is accurate.

5.1 Model Results of DS1-M2

According to Data Set1, multiple linear regression model lagged through lag 1 to lag 24 for 2011 has better results in according to highest R-square (0,985) and lowest MAPE(1,277) values. It is observed that forecast results for the year of data provided (2011), model fluctuates almost the same direction and magnitude with the actual load consumption. On the other hand, according to next year's (2012) electricity load consumption forecast ; for 15.04.2012 forecasted load and actual load are almost the same until the hour 06:00, after hour 06:00 with the initiation of shoulder formation a biased estimate error until hour 21:00. For 28.08.2012 it is observed that errors through the behavioral tendency of the estimates are similar with actual data. There is a varying error all across the day.

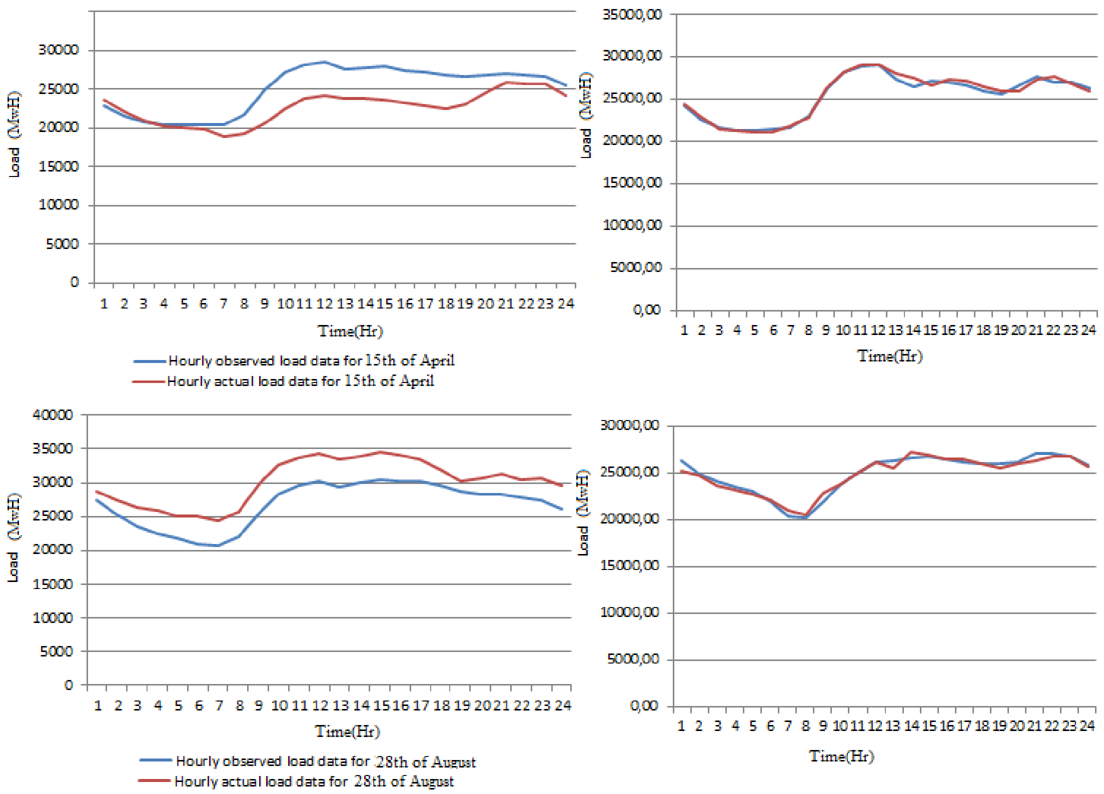

Figure 5.1: Actual vs observed forecasted load data for the selected dates within the year 2011 (right) and next year 2012 (left)

5.2 Model Results of DS2-M1

According to Data Set2, multiple linear regression model lagged through lag 1 and lag 2 for 2011 has better results in according to highest R-square (0,984) and lowest MAPE(1,348) values. It is observed that forecast results for the year of data provided, model fluctuates almost the same direction and magnitude with the actual load consumption. Similar with Data set1, according to next year's (2012) electricity load consumption forecast ; for 15.04.2012 forecasted load and actual load are almost the same until the hour 06:00, after hour 06:00 with the initiation of shoulder formation a biased estimate error until hour 21:00.

For 28.08.2012 it is observed that errors through the behavioral tendency of the estimates are similar with actual data. There is a varying error all across the day.

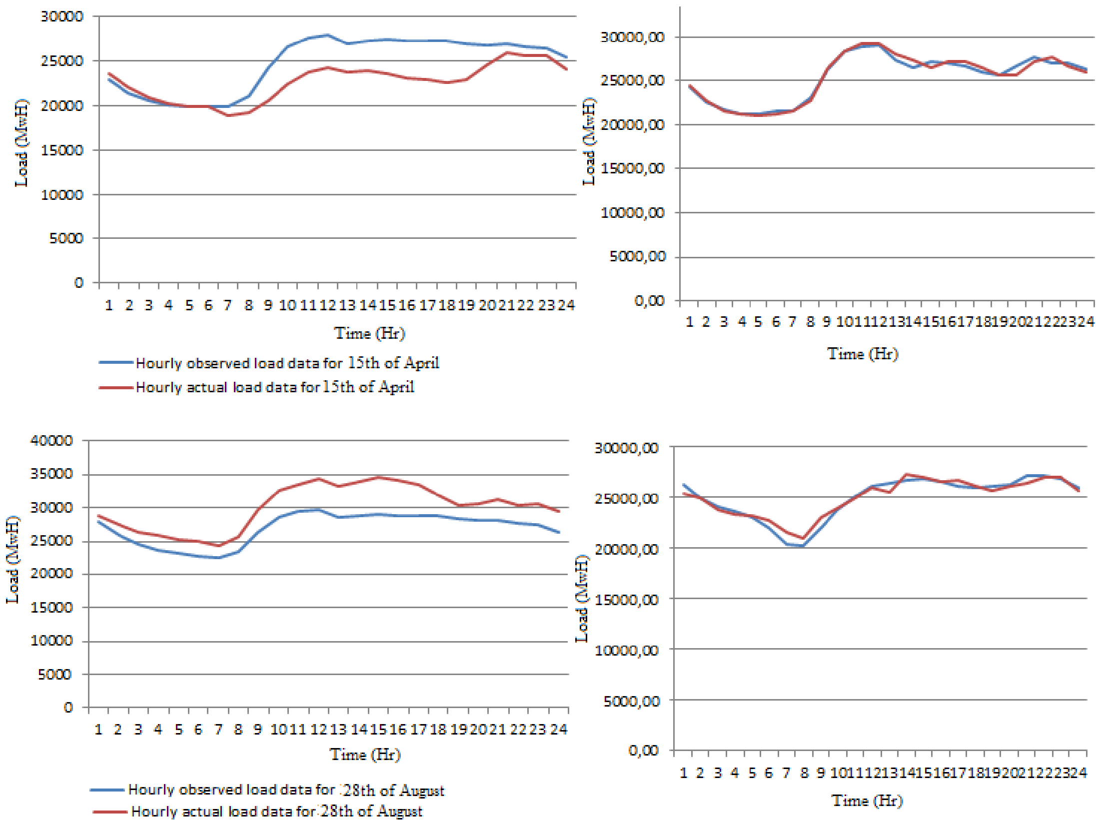

Figure 5.2: Actual vs observed forecasted load data for the selected dates within the year 2011 (right) and next year 2012 (left)

5.3 Model Results of DS3-M2

According to Data Set3, multiple linear regression model lagged through lag 1 to lag 24 for 2012 has better results in according to high R-square (0,988) and low MAPE(1,229) values. It is observed that forecast results for the year of data provided, model fluctuates almost the same direction and magnitude with the actual load consumption. . According to next year's (2013) electricity load consumption forecast gives better results; for 15.04.2013 forecasted load and actual load are almost the same until the hour 09:00, after 09:00 with the initiation of shoulder formation a biased estimate error. For 28.08.2012it is observed that errors through the behavioral tendency of the estimates are the most similar with actual data. Least varying error all across the day and the closest forecast results are observed among all other load models.

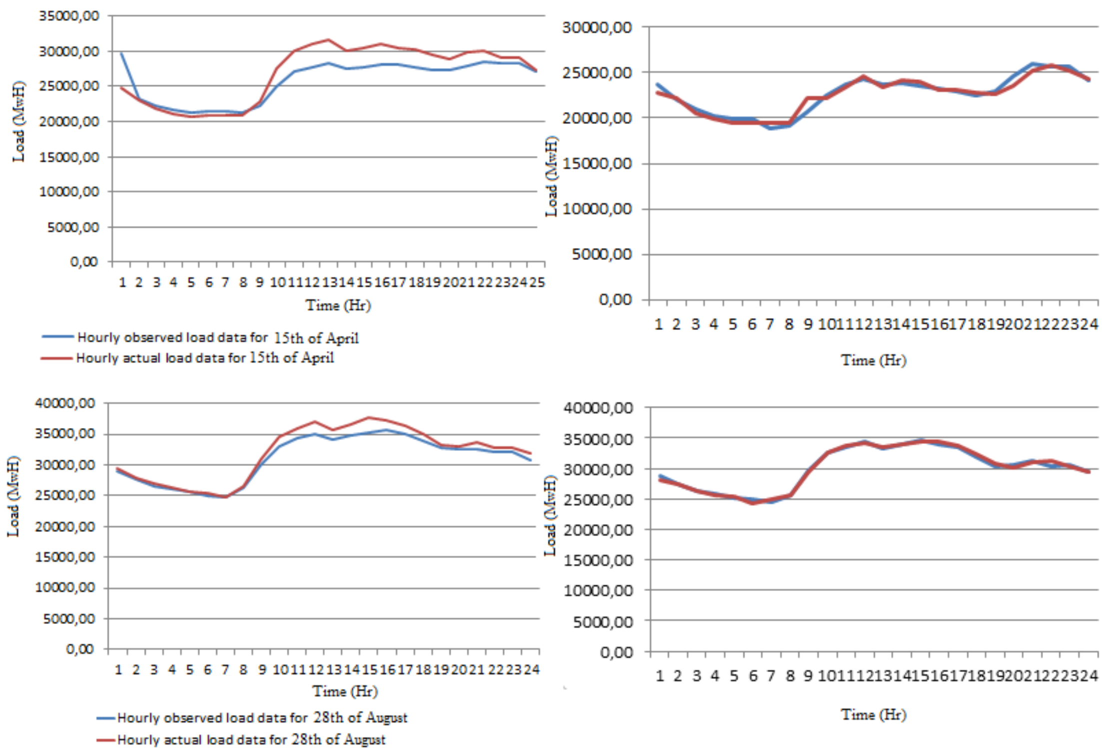

Figure 5.3: Actual vs observed forecasted load data for the selected dates within the year 2012 (right) and next year 2013 (left)

5.4 Model Results of DS4-M1

According to Data Set1, multiple linear regression model lagged through lag 1and lag 2 for 2012 has better results in according to high R-square (0,986) and low MAPE(1,358) values. It is observed that forecast results for the year of data provided, model fluctuates almost the same direction and magnitude with the actual load consumption. According to next year's (2013) electricity load consumption forecast gets much closer comparing with data set1 and 2 ; for 15.04.2013 forecasted load and actual load are almost the same until the hour 07:00, after hour 07:00 with the initiation of shoulder formation a biased estimate error until hour 20:00. For 28.08.2012 it is observed that errors through the behavioral tendency of the estimates are similar with actual data. There is a varying error all across the day.

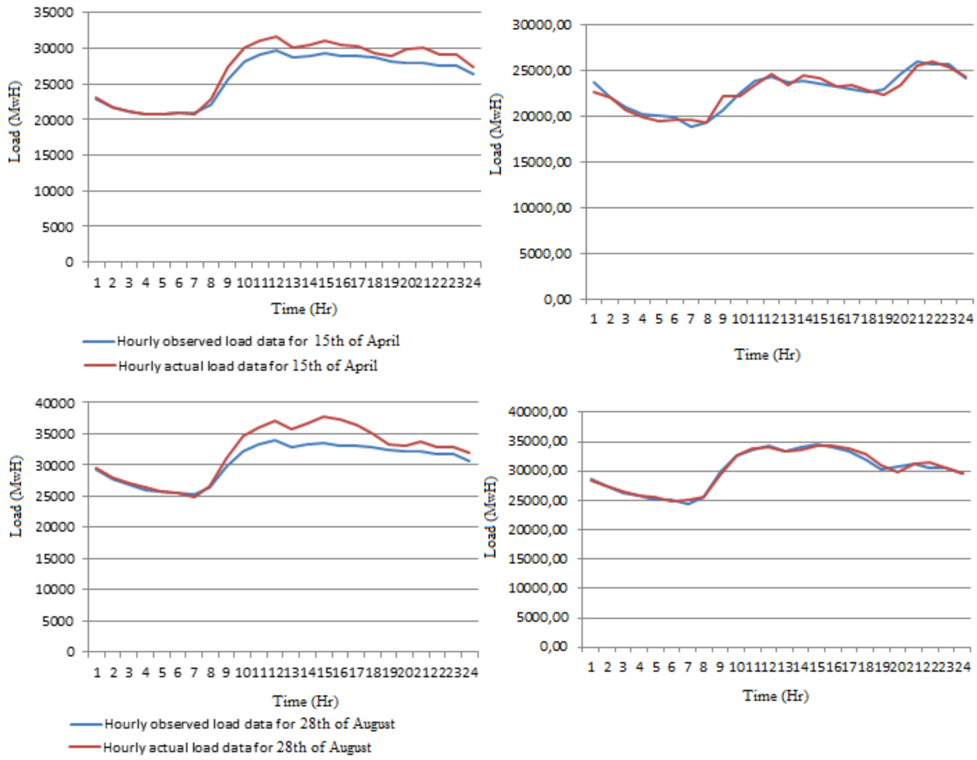

Figure 5.4: Actual vs observed forecasted load data for the selected dates within the year 2012 (right) and next year 2013 (left)

5.5 Model Results of DS5-M4

According to Data Set5, multiple linear regression model lagged through lag 1 to lag 24 for 2015 has better results in according to highest R-square (0,874) and lowest MAPE(52,66) values. It is observed that in forecast results for the year of data provided and next year's, there are lots of unstable fluctuations since errors are pretty high. It is cause by the structure of price that is very changeable. It seems lots of other independent variables might drive the architecture of price model as well.

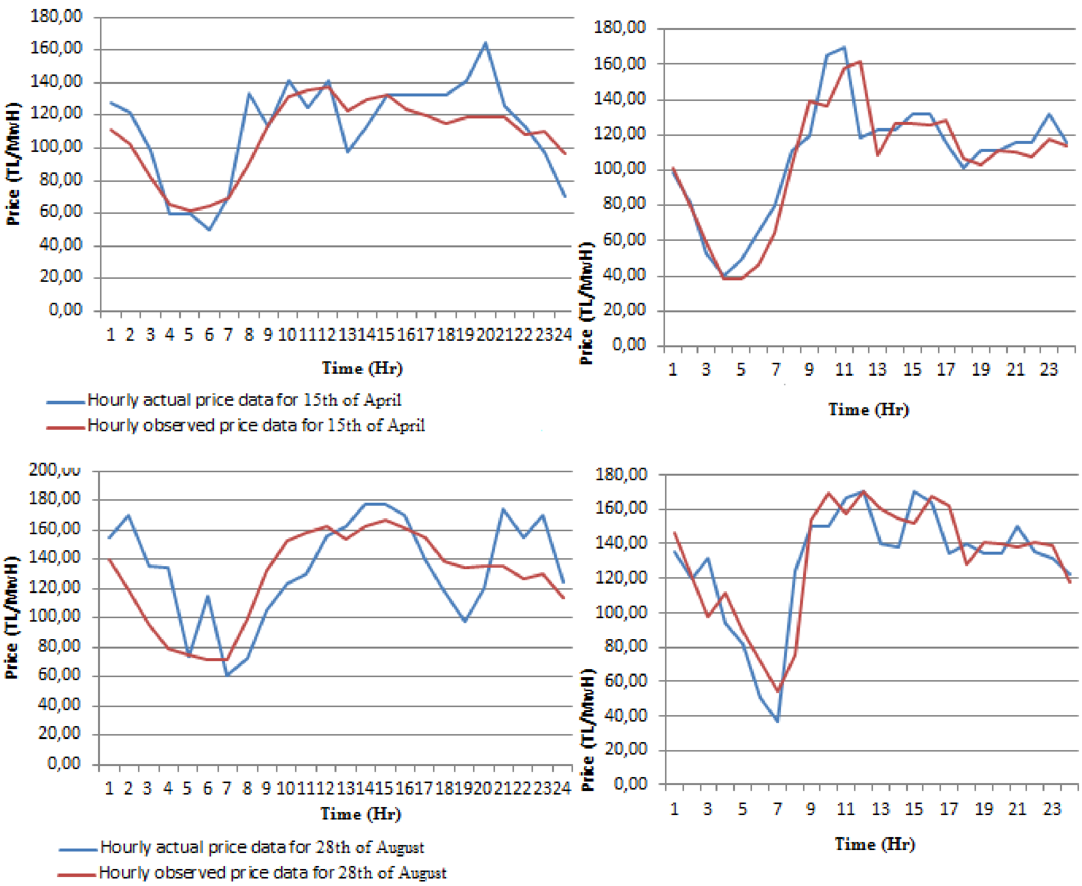

Figure 5.5: Actual vs observed forecasted load data for the selected dates within the year 2015 (right) and next year 2016 (left)

Chapter 6

CONCLUSION

For efficient power systems applications, electrical energy systems planning and economic applications, an accurate short term load and price forecasting model brings an important advantage to utilities with increasing energy consumption in competitive electricity markets like Turkish power industry.

The objective of this research has been the development of accurate short term load and price models, by doing that, time series analysis including lagged variables that have been presented in autoregressive models merged with specific categorical variables (hours of day, days of the week, months of the year and special events of Turkey) and environmental indicators as hourly temperature data in terms of heating-cooling degree hours have been applied. With 24hr lagged variables and significant lags as lag1 and lag2, AR and ARIMA models employed with the results of parameters the combination of those lags, 4 different load models in years 2011 and 2012, a price model for 2015 are constructed with using multiple linear regression in order to test each models accuracy. For load models determination coefficients, R-square results; %98,5 %98,4 %98,8 %98,6 and MAPE's are %1,277 %1.348 %1,229 %1,358 respectively. Load models are not improved to ARIMA models since load models residuals ACF and PACF plots are inconclusive hence they presented that there was no need to add I and MA parts also R-square values are high enough. For the price model with the effect of best possible ARIMA(24,1,3) parameters, final hybrid multiple regression model gave R-square: %87,4 and MAPE %52,66. As a result, the comparison of actual and observed data is studied and the power of model is tested with illustrating on various regression tests. The reason behind the high value of mean absolute percentage error is the standard volatility of the price. Because, price formation is a complex process outcome such as; transmission network constraints (Day-Ahead Planning processes), open capacities of power plants that give hourly prices, difference between the demand expected by the producer and the actual consumer demand, unexpected hourly deviations factors can create high percentage of errors due to the multiplicity of these independent variables which are not integrated into the model. It will be useful to collect data about the mentioned independent variables that are not integrated, in order to improve the price model and to minimize percentage errors. Also, in price data major difficulty has been obtaining the necessary price data for the analysis, some of the values were missing in the data base of EPIAS. Hence in order to attain statistical sufficiency, the study is to be performed without lack of price data in the future .Moreover, in the results section there are two days of two seasons as spring and summer, in future studies randomly 4 days can be picked to observe different seasons effect in load and price forecasting including 4 seasons in the year. Consequently, still the results have shown that proposed models especially load models gave very low percent of errors with extremely accurate day ahead forecasts considering Turkeys electricity load and price profile.

References

[1] C. Hamzaçebi, "Forecasting of turkey's net electricity energy consumption on sectoral bases," *Energy policy*, vol. 35, no. 3, pp. 2009–2016, 2007.

[2] M. Ungor, M. K. Kalafatcilar *et al.*, "On the role of productivity and demographic factors for growth: The case of turkey, 2004-2012," Research and Monetary Policy Department, Central Bank of the Republic of Turkey, Tech. Rep., 2013.

[3] "Electrical energy sector report in 2016, tetas," http://http://www.tetas.gov.tr/tr-TR/Sektor-Raporlari, accessed: 2017-04-17.

[4] G. Heinemann, D. Nordmian, and E. Plant, "The relationship between summer weather and summer loads-a regression analysis," *IEEE Transactions on Power Apparatus and Systems*, no. 11, pp. 1144–1154, 1966.

[5] D. Lijesen and J. Rosing, "Adaptive forecasting of hourly loads based on load measurements and weather information," *IEEE Transactions on Power Apparatus and Systems*, no. 4, pp. 1757–1767, 1971.

[6] M. T. Hagan and S. M. Behr, "The time series approach to short term load forecasting," *IEEE Transactions on Power Systems*, vol. 2, no. 3, pp. 785–791, 1987.

[7] D. C. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," *IEEE transactions on Power Systems*, vol. 6, no. 2, pp. 442–449, 1991.

[8] D. Srinivasan, S. S. Tan, C. Cheng, and E. K. Chan, “Parallel neural network-fuzzy expert system strategy for short-term load forecasting: System implementation and performance evaluation,” *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1100–1106, 1999.

[9] I. Erkmen and A. Ozdogan, “Short term load forecasting using genetically optimized neural network cascaded with a modified kohonen clustering process,” in *Intelligent Control, 1997. Proceedings of the 1997 IEEE International Symposium on.* IEEE, 1997, pp. 107–112.

[10] A. K. Topalli and I. Erkmen, “A hybrid learning for neural networks applied to short term load forecasting,” *Neurocomputing*, vol. 51, pp. 495–500, 2003.

[11] A. K. Topalli, I. Erkmen, and I. Topalli, “Intelligent short-term load forecasting in turkey,” *International Journal of Electrical Power & Energy Systems*, vol. 28, no. 7, pp. 437–447, 2006.

[12] C. Hamzaçebi and F. Kutay, “Yapay sinir ağlari ile türkiye elektrik enerjisi tüketiminin 2010 yilina kadar tahmini,” *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 19, no. 3, 2004.

[13] G. Ceylan and A. Demirören, “Yapay sinir ağları ile gölbaşı bölgesini kısa dönem yük tahmini,” 2004.

[14] T. YALÇINÖZ, S. HERDEM, and U. EMİNOĞLU, “Yapay sinir ağlari ile niğde bölgesinin elektrik yük tahmini,” 2002.

[15] M. Natrella, “Nist/sematech e-handbook of statistical methods,” 2010.

[16] G. E. Box and G. M. Jenkins, “Some recent advances in forecasting and control,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 17, no. 2, pp. 91–109, 1968.

[17] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[18] M. de Smith, "Statsref: Statistical analysis handbook-a web-based statistics," 2015.

[19] "Electricity transmission company data set," https://www.teias.gov.tr.

[20] "Turkish energy markets management company data set," https://seffaflik.epias.com.tr/.

[21] "Hourly temperature data set in years 2011-2012 for istanbul, turkey," https://www.wunderground.com/weather/tr/istanbul/.

[22] D. Thevenard, "Methods for estimating heating and cooling degree-days to any base temperature." *ASHRAE Transactions*, vol. 117, no. 1, 2011.