P.ÖZPALA

STATISTICAL MODELING OF

HOURLY ELECTRICITY LOAD SERIES IN TURKEY

PINAR ÖZPALA

FEBRUARY 2013

METU 2013

STATISTICAL MODELING OF
HOURLY ELECTRICITY LOAD SERIES IN TURKEY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

PINAR ÖZPALA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ECONOMICS

FEBRUARY 2013

Approval of the Graduate School of Social Sciences

Prof. Dr. Meliha ALTUNIŞIK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science/Arts / Doctor of Philosophy.

Prof. Dr. Erdal ÖZMEN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Esma Gaygısız
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Esma Gaygısız (METU,ECON)

Prof. Dr. Erdal Özmen (METU,ECON)

Assoc. Prof. Dr. Serdal BAHÇE (Ankara Uni.,ECON)

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

Name, Last Name: PINAR, ÖZPALA

Signature:

ABSTRACT

STATISTICAL MODELING OF

HOURLY ELECTRICITY LOAD SERIES IN TURKEY

Özpala, Pınar

Msc., Department of Economics

Supervisor: Assoc. Prof. Dr. Esma Gaygısız

February 2013, 60 pages

This study investigates Heterogeneous Double Seasonal ARIMA model, Heterogeneous Periodic AR model and a nonlinear model of Multiple Regime Logistic STAR model by application of hourly electricity load data between years 2006-2007 in Turkey. Forecast results suggest that Heterogeneous Double Seasonal ARIMA model is the best among three methods forecasting up to one day.

Keywords: Smooth Transitional Autoregressive Model, Seasonal Autoregressive Moving Average, Fourier Transformation, Forecasting

ÖZ

TÜRKİYE'DE SAATLİK ELEKTRİK YÜKÜ

SERİLERİNİN İSTATİSTİKSEL MODELLEMESİ

Özpala, Pınar

Yüksek Lisans, İktisat Bölümü

Tez Yöneticisi: Doç. Dr. Esma Gaygısız

Şubat 2013, 60 sayfa

Bu çalışma, Heterojen Mevsimsel Otoregresif Tamamlanmış Hareketli Ortalama Modeli, Heterojen Periodik Otoregresif Model ve Çok Rejimli Lojistik Yumuşak Geçişli Bağışım Modeli ile Türkiye'de 2006-2007 yılları arasında talep edilen saatlik elektrik yükü serilerinin ekonometrik uygulamasını incelemektedir. Öngörü sonuçları, Mevsimsel Otoregresif Tamamlanmış Hareketli Ortalama Modelinin bir günlük öngörü aralığı için bu üç metot arasında en iyi sonucu verdiğini göstermektedir.

Anahtar Kelimeler: Yumuşak Geçişli Bağışım, Mevsimsel Otoregresif Hareketli Ortalama, Fourier Dönüşümü, Öngörü

To My Family

ACKNOWLEDGMENTS

I am thankful to my thesis supervisor Assoc. Prof. Dr. Esma Gaygısız for her support throughout this study. I also tender my thanks to examining committee members, Prof. Dr. Erdal Özmen and Assoc. Prof. Dr. Serdal Bahçe.

I would like to express my gratitude to my mother for her endless support, to my brother for his valuable critics and inferences, and to my father for providing me motivation and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

ABBREVIATIONS

ACF	Autocorrelation Function
ANN	Artificial Neural Networks
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average Model
ARMA	Autoregressive Moving Average
BOT	Build-Operate-Transfer
BO	Build-Operate
EDAŞ	Electricity Distribution Company
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedasticity
EGARCH-M	Exponential Generalized Autoregressive Conditional Heteroskedasticity in Mean
EPDK	Energy Market Regulatory Authority
ETKB	Ministry of Energy and Natural Resources
EÜAŞ	Electricity Generation Company
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
H-PAR	Heterogeneous Periodic Autoregressive
H-SARIMA	Heterogeneous Seasonal Autoregressive Integrated Moving Average
LM	Lagrange Multiplier
MA	Moving Average

MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimation
NLDC	National Load Dispatch Center
PACF	Partial Autocorrelation Function
SARIMA	Seasonal Autoregressive Integrated Moving Average
SBIC	Schwarz Bayesian Information Criteria
TEAŞ	Turkish Electricity Company
TEDAŞ	Turkish Electricity Distribution Company
TEİAŞ	Turkish Electricity Transmission Company
TEK	Turkish Electricity Corporation
TETAŞ	Turkish Electricity Trading Company
TOR	Transfer of Operating Rights

CHAPTER 1

INTRODUCTION

Hourly electricity load series is one of the major interests of both academic and operational environment. Academicians' point of view, load series is worth to examine due to its specific time series properties such as seasonality, non-stationarity, effect of explanatory variable, some nonlinearities. Applied sights' point of view, accurate short time forecasting of hourly load demand up to one week ahead is critical for nations in order to balance electricity produced and consumed at any time in the day. Due to non-durable property, balancing of electricity load is crucial for welfare. Bunn and Farmer (1985) computed for a year that an increase of only 1% in the forecast error caused an increase of 10 million pounds in operating costs per year for one electric utility in the United Kingdom. Their study shows the importance of accurate forecasting in the sense of national wealth.

At National Load Dispatch Center, the unit under the body of TEİAŞ which is in charge of real-time balancing of electricity demand and supply in Turkey, hourly consumption estimates published daily and is used to guide scheduling activities for the following day. The whole system day ahead balancing is planned according to forecasts of demand side. The errors in forecasts cause unbalance of the system in real-time and the unbalance costs of the system is afforded by the side which causes the unbalance. In order to minimize these costs, it is required to minimize the hourly forecast errors. Proper forecasting and real time balancing also strengthen the competitiveness of the trading companies by reducing the costs in liberalized system.

In this study, our target is to achieve accurate short term load forecasting required for controlling and scheduling of power systems. Thus we focus on univariate methods for prediction up to one day ahead. Although, the meteorological variables have significant influence on electricity demands, univariate methods are sufficient for short term load forecasting up to one day because the meteorological variable change in smooth fashion which can be captured in the demand series itself. We have considered three time series approach to model the load demand. First one is the Double Seasonal ARIMA model described in the study of Taylor (2003) but with allowing to heteroskedasticity, Periodic Autoregressive Model proposed by Taylor (2006) with allowing to heteroskedasticity as well, and a nonlinear model of Multiple Regime LSTAR model described in Teräsvirta (1994). We have found that in forecasting of load demand up to one day, Heterogeneous Double Seasonal ARIMA model gives the best results which MAPE is considered as the evaluation criteria, and the second best one is Heterogeneous PAR model in which Fourier series is used to specify seasonalities. Although we rejected the linearity against nonlinear LSTAR model, we cannot obtain good results with this model. Estimation and forecast results of the MR-LSTAR model provided by programming in Matlab, and the results for other two models are provided by using econometrics programme of E-views.

Modeling of Turkish hourly load data with all three models and adding ARCH-GARCH effect into the models has not been studied so far.

This study is organized as follows; Chapter 1 briefly introduces the study, Chapter 2 explains the historical development of electricity sector in Turkey, Chapter 3 reviews electricity load forecasting methods used in the literature, Chapter 4 analyses the data, describes the methods used and displays the estimation and forecast results of the models, Chapter 5 concludes the study and gives some perspectives about future work.

CHAPTER 2

HISTORY OF TURKISH ELECTRICITY SECTOR

The first electricity generation in Turkey began with a 2 KW power water mill which was established by Swiss and Italian corporate group in Tarsus. Electricity firstly distributed to İstanbul in 11 February 1910 from İstanbul Silahtarağa Power Plant which was the first organized power plant.

In the first years of the Turkish republic electricity generation was in the power of privileged partnerships which started before the Republic. The situation was the result of liberal politics of economy started to be implemented after İzmir Economy Congress. In this period private enterprises had obtained privileges. These privileges used by German MAN and AEG, Italian Marelli, Hungarian Ganz and Belgium corporations. Private national capital also entered this sector and Kayseri and Nearby Electricity Turkish Incorporated Company was founded with the contract signed in 1926.

In the period of 1930-50 the statist economy politics was started to implement in Turkey with the effect of adoption of these politics in the world after the world depression 1929. Insufficient power of private enterprises was also a factor that caused to implement statist politics in economy. The state's central role in economy reflected to energy sector. The Municipal Law enacted in 1933 gave municipals authorization on building, operating and transferring electricity power plant. Foreign capital and privileged corporations except Kayseri and Neighborhood Turkish Electricity Incorporated Company were nationalized. In

1935 Etibank, General Directorate of Mineral Research and Exploration, and General Directorate of Electrical Power Resources Survey and Development Administration were founded.

After 1950 private enterprises role in economy became crucial again, government allowed privileged electricity companies and so private national capital encouraged to enter into sector. Privileged electricity operation was the advice of World Bank. However, none of all four companies that were founded at this period in order to operate electricity was belong to foreign capital. The founded companies between 1952 and 1956 were Northwest Anatolia Electricity Turkish Company, Çukurova Electricity Company, Kepez Electrical Power Plant Trading Company. The first two ended operation in 1971 and the last two maintained by 2003 but then government took control.

In the period of 1960-80 was implemented planned development model. In 1963 the Ministry of Energy and Natural Resources and in 1970 Turkish Electricity Corporation were founded. With the foundation of TEK, the politics of privileged private companies were ended.

After 1980 free market economy model was adopted as a result of the economic stability programme referred as 24 January Decisions dated 24 January 1984 and the liberal economy politics that the government implemented. Thereby private sector gained importance and naturally the free market rules of liberal politics were reflected to the electricity sector. Thus the legal regulations were started in order to make private sector enter this sector. This politics laid a foundation in the symposium named as Energy and Oil Problem organized by İstanbul Chamber of Commerce in 1979 with opening speech of Turgut Özal.

One of the main reasons of privatization requirement in the electricity market was the growth of electricity demand and the limited capacity of government at these times. The rapid growth of population, urbanization and the growth of industry were among the reasons of stronger electricity demand. In order to catch a high rate of economic growth and compensate the demand of growing population, new investments that enlarges the electricity generation capacity was required. However, the government had difficulty in financing this big technologic investment because of the economic conditions of Turkey.

Besides, the monopoly of TEK obstructed to be formed an efficient, competitive, and profitable electricity industry. The cost raises in TEK was reflected to the price charged to last consumer. Electricity generated and consumed more expensive than the required. Being under expose of politic factors and central government interventions, low efficiency of old technology and over employment led to inefficient use of resources which caused to cut the research and development funds.

The process of privatization regulated by Laws Number 3096, 3291, and 4283. The Law Number 3096 enacted in 1984 regards on authorization and regulation of domestic and foreign enterprises subject to the provisions of private law other than the TEK to produce, transmit, distribute and trade electricity led to private sector to serve in electricity sector. This law is also legal foundation for privatization and regulation actions in electricity sector. This law includes privatization by authorization and transferring of operational rights but not includes transfer of ownership. Producer models started to be applied in the sector following by this law.

Law Number 3291 enacted in 1986 includes the regulations on transfer of ownership. Build-Operate-Transfer, Transfer of Operational Rights and Auto-

This law regards on privatization by transfer of ownership of TEK's current and reconstructed enterprises. However, Law BOT model was firstly mentioned in Law Number 3996 dated 1994. This law constructs legal foundation for BOT model.

Thus in the electricity sector new legal regulations was required. In 1994 with the cabinet decree, TEK was rebuilt as Turkey Electricity Generation and Transmission Company and Turkey Electricity Distribution Company. Electricity Energy Fund was founded with Law Number 3613 to give financial support to projects of private sector and to stabilize the electricity prices. Law number 3974 dated 1994 includes the principals of energy facilities.

Law Number 4283 regards on Establishing and Operating Electric Power Plants and Sale of Energy through the Build-Operate Model was enacted in 1997 authorizes companies on building and operating thermal plants with Build-Operate model in accordance with the national energy politics and regulates the principals of energy selling. This model gives rights to private sector to build and operate power plants owned by investors except hydroelectric power plants, geothermal and nuclear power plants, as well as all other power plants running on renewable energy sources.

With the Law Number 4047, projects regards on generation, transmit and distribute electricity removed from Law Number 3996 and added to Law Number 3096. Also Law Number 4047 provides advantages for finance of BOT projects. Government ensuring of BOT projects and tax reduction are among the advantages.

Electricity market privatization activities were hindered because of the problems through the application of BOT, BO, and TOR models. The purchase and payment guarantee given for long term, the over pricing of electricity contradicted to the construction of competitive market structure stated in Seventh Development Plan for Five Years. The overpriced rates of purchase and payment guarantee appeared to be a big burden for TEAŞ and Treasury in long term.

Electricity Market Law Number 4628 enacted in 2001. This Law was required to be enacted because of opposition of World Bank and IMF to such contracts that causes big burden on government and these organizations wanted this law to be enacted as a provision of supporting the maintenance of Economic Stability Programme. It was also required through the European Union adoption process.

As Electricity Market Law came into force, electricity sector began to revise. Electricity Market Regulatory Authority was founded with this law in order to regulate the electricity market. However, the authority name was changed to Energy Market Regulatory Authority by arrangements in EPK in 18 April 2001 after legislation of Natural Gas Market Law. The authority became responsible of regulation and supervision of electricity, natural gas, and oil markets with the Oil Law that was enacted in 2003. The most important structural change with the law is the separation of TEAŞ into three as Electricity Generation Company, Turkish Electricity Transmission Company, and Turkish Electricity Trading Company.

According to the law the generation activities of electricity shall be performed by EÜAŞ, other public companies and their affiliate companies, private sector generation companies, autoproducers and autoproducer groups. TEİAŞ shall conduct the electricity transmission activities and shall be responsible for taking over all transmission facilities owned by the public, developing transmission investment plans for the proposed new transmission facilities and building and operating new transmission facilities. The transmission network was entirely owned by TEİAŞ. The distribution activities are separated into 21 zones and a company is authorized for each zone. The companies got through the privatization process and all of them has completed process.

Wholesale activities are conducted by TETAŞ and private sector wholesale companies in accordance with the provisions of this Law. These companies buy the electricity from the electricity generation companies and transmit through TEİAŞ facilities and finally sell it to the eligible consumers. TETAŞ is a public entity that conducts the trade activities on behalf of public. TETAŞ is primarily responsible of the trade of electricity generated in EÜAŞ's power plants. Legal entities that are engaged to retail sale activities have to obtain retail sale licenses. Retail sale companies are entitled to engage in retail sale or retail sale service activities without any limitation regarding regions. The distribution companies holding retail sale license are also entitled to retail sale. However, with the arrangement made in the Law, distribution companies engaged in generation and retail sale of electricity had to separate those activities on 1 January 2013. According to Electricity Market Law with the approval of EPDK electricity import and export activities are performed by TETAŞ, private sector wholesale companies, retail sale companies and distribution companies that is obtained retail sail license with the accordance of the provisions of the Law and the relevant regulations.

Electricity Market Law published with a purpose of ensuring the development of a financially sound and transparent electricity market operating in a competitive environment and the delivery of sufficient, good quality, low cost and environment-friendly electricity to consumers.

As a short and midterm program in the liberalization process of electricity market in order to reach the purpose in time, Electricity Energy Sector Reform and Privatization Strategy Paper was publicized on 17 March 2004. Following of the Strategy Paper, implementation of transitional Balancing and Reconciliation and price equalization mechanism started with Electricity Market Balancing and Settlement Regulation, which was published on 3 November 2004, aimed to stabilize supply and demand of electricity. With this law new free electricity market formed where the legal entities would take their place in the spot market, bids and the prices would be set hourly by a merit system like the one in England.

Market Finance Reconciliation Center was constructed as a unit within the body of TEİAŞ and is responsible for running the financial settlement system by calculating amounts payable or receivable by participants in the market. National Load Dispatch Center is the unit under the body of TEİAŞ in charge of real-time balancing of electricity demand and supply. Bid and offer prices submitted to NLDC by balancing mechanism participants.

CHAPTER 3

LITERATURE REVIEW

Many methods have been used in literature for modeling and forecasting short term load demand. Multiple regression, exponential smoothing, Box-Jenkins models and Kalman filters are among the classical approaches to short term load forecasting, and neural network, fuzzy logic are among the new approaches.

Classical approaches employed in the literature are Al-Hamadi and Soliman (2004); Amjady (2001); Bunn and Farmer (1985a), Cancelo et al. (2008), Dordonnat et al. (2008), Huang (2003), Huang et al. (2005), Nowicka et al. (2002), Taylor (2008), Taylor et al. (2006), Weron (2006).

New approaches employed are Badri (2012), da Silva et al. (2008), Darbellay and Slama (2000), Hippert et al. (2005), Hippert et al. (2001), Khotanzad et al. (1998), Metaxiotis et al. (2003), Reis and Alves da Silva (2005).

In literature electricity demand often modeled in terms of weather variables and past load series (e.g. Hor et al., 2005; Cancelo et al., 2008) or in terms of only past load series (Taylor 2003, Taylor 2010, Soares and Medeiros 2008). Weather-based online systems require default procedures in order to ensure robustness (Bunn, 1982). Of course, such methods are the only option when forecasting load in locations where weather forecasts are either unavailable or too costly (Soares and Medeiros, 2008). However, since load demand series can capture smooth change of weather variables in short term, univariate methods in load forecasting up to one day can be considered sufficient.

Taylor (2008a) in his empirical study showed that a univariate method that outperform a multivariate method up to about four hours ahead and a combination of forecasts from the two methods was found to be the most accurate approach up to a day ahead. This shows that univariate methods have a valuable role to play in short-term load forecasting.

Some authors as Fiebig et al. (1991), Peirson and Henley (1994), Ramanathan et al. (1997), Cottet and Smith (2003), and Soares and Souza (2006), Amaral (2008) treats each hour as a separate time series, such that 24 different models are estimated.

Multiple regression analysis for load forecasting uses the technique of weighted least-squares estimation. Based on this analysis, the statistical relationship between total load and weather conditions as well as the day type influences can be calculated. The regression coefficients are computed by an equally or exponentially weighted least-squares estimation using the defined amount of historical data.

Moghram and Rahman (1989) used multiple regression method to model hourly load in terms of explanatory variables such as weather and non-weather variables which are identified on the basis of correlation analysis between independent and dependent variables. In the paper it is compared with other models for each 24 hour load forecast. Barakat (1990) used the regression model by adjusting time series data for the effects of seasonality.

Haida and Muto (1994) used multiple regression model to forecast daily peak load where the explanatory variables were chosen based on the correlation analysis. Latest actual observation data before the forecast day is used in order to estimate coefficients of forecasting model. Coefficient estimations were done by historically exponential weighted least squares method, where weights were updated every day. In order to reduce the forecasting errors in transitional seasons they proposed a transitional technique. They presented a method which uses a regression model to forecast the nominal load and a learning method to forecast the residual load. Haida and Muto (1998) extended their method with trend techniques where the trend cancellation removes annual load growth by means of division or subtraction processes with morning load on the forecasting day and the trend estimation technique estimates the trend between the forecasting year's load and the past year's load by using the variable transformation techniques.

Hyde and Hodnett (1997) made linear regression analysis to estimate model coefficients where load forecasts identified as weather sensitive and weather insensitive load components. The approach of Ramanathan et al. (1997) is a multiple regression model included calendar and weather effects, one for each hour of the day, with a dynamic error structure as well as adaptive adjustments to correct for forecast errors of previous hours.

Charytoniuk et al. (1998) presented non-parametric regression. Non-parametric approach is data driven and does not assume any kind of distribution.

Cottet and Smith (2003) used multi-equation model each for 48 half hour in order to capture the intraday pattern, and developed forecast models within a Bayesian framework; however, with an assumption of a diagonal vector autoregressive structure for the error terms.

Exponential Smoothing assigns exponentially decreasing weights as the observation gets older.

Winters (1960) introduced well-known Holt-Winters exponential smoothing method in order to forecast time series models with seasonal cycles and trend.

Christiaanse (1971) has adapted the generalized exponential smoothing method of Brown (1965) in order to forecast hourly integrated loads with vector of fitting functions which were expanded with Fourier series.

Park (1991) decomposed his load model into three such as nominal, type and residual load. Nominal load was modeled by using Kalman filter where parameters estimated with exponentially weighted recursive least squares method. The type load was allocated for weekend load prediction which is updated exponential smoothing method. The residual load was predicted by autoregressive method estimated with recursive least squares method. The load modeling method of him is the mixture of autoregressive, exponential smoothing and general exponential smoothing method.

Taylor (2003) adapted the Holt-Winters exponential smoothing formulation for double seasonality. Moreover, they corrected for residual autocorrelation using a simple autoregressive model. They compare the hourly load forecasts produced by the new double seasonal Holt-Winters method with traditional Holt-Winters and from a multiplicative double seasonal ARIMA model. They concluded that the new double seasonal exponential smoothing method outperformed the others.

Gould et al. (2008) criticized the double seasonal HWT method of Taylor (2003) that it assumes same intra-day cycle for all days of the week and the weight updates based upon recent information are the same for each day of the week. He developed the proposed double seasonal model in Taylor (2003) by adding sub cycle of intra-day cycle into weekly cycle. They used common intraday cycle for days of the week that exhibit similar patterns of demand.

Taylor (2010) termed the method Gould et al. proposed as intraday cycle exponential smoothing due to its focus on intraday cycle and improved it with the inclusion of the adjustment for first order residual autocorrelation that was used in the double seasonal HWT method. Moreover Taylor (2010a) extended double seasonal exponential smoothing model into triple one by adding third cycle in order to describe daily cycle.

Taylor and McSharry (2007) considered five forecasting methods includes seasonal ARMA, periodic AR, double seasonal HWT, IC exponential smoothing, principal component analysis among double seasonal Holt-Winters performed best.

Time Series Methods is the use of a model based on previously observed values. Seasonal ARMA method is used for stochastic time series data with seasonal cycles. Seasonal ARMA model can be extended to multiple cycles (Box et al. (1994, p. 333)). The model that includes one cycle named as single seasonal, includes two cycle named as double and includes three cycle named as triple seasonal ARMA. Triple Seasonal ARMA can capture the daily, weekly and yearly cycle.

Hagan and Behr (1987) made the nonlinear extension of Box-Jenkins transfer function model and compare with seasonal ARIMA model. Moghram and

Rahman (1989) considered seasonal ARIMA with four other forecasting methods.

Soares and Medeiros (2008) proposed a model Two Level Seasonal Autoregressive Model by treating 24 hours as a different time series such that 24 separate models estimated, as in study of Ramanathan et al. (1997), in which seasonality captured by Fourier decomposition.

Cancelo (2008) modeled electricity consumption for each hour of the day separately in order to avoid intraday cycle by applying weather sensitive double seasonal ARIMA method which incorpates intraweek and intrayear cycles.

Taylor (2010) considered that for very short terms, modeling each hour separately can cause missing the level of the load. He also used triple seasonal ARMA model in his study as a natural extension of the two double seasonal ARMA models. Some authors added AR component to their basic models (Taylor (2003), Park (1991)).

State Space Methods is the use of a model which the number of inputs, outputs and states, the variables are expressed as vectors.

Harvey and Koopman (1993) developed a model formulated by unobserved components with time-varying splines to capture the intraday seasonal patterns of hourly electricity loads.

Dordonnat et al. (2008) presented multiple-equation linear time-varying regression model for the French national hourly electricity load, with one equation for each hour which all were estimated simultaneously. They allowed the cross correlation between the stochastic terms of the equations for different hours. They use Kalman filtering and associated algorithms to estimate and forecast their multivariate linear Gaussian state space model.

Soft Computing Methods such as Artificial Neural Network and Fuzzy Logic gets great interest of reseachers and practioners in last decades for load forecasting because of their capabilities for the nonlinear modeling of large multivariate data sets.

Charytoniuk and Chen (2000) compare forecasts from several differently structured ANNs. Hippert et al. (2010b) assert that defining the appropriate level of model complexity, and choosing the input variables are the challenges of NN modeling. They evaluated NN modeling within a Bayesian framework, in which input selection and model selection defined by Bayesian techniques.

Badri (2012) investigates the application of artificial neural networks and fuzzy logic as forecasting tools for predicting the load demand in short term category.

CHAPTER 4

DATA, METHODOLOGY AND EMPIRICAL RESULTS

4.1 Data

We have examined a dataset of hourly loads in Turkey from January 1 2006 to 31 December 2007. The last 168 hours is allocated for forecast evaluation.

We have employed Phillips-Perron test to the level with the spectral estimation method of Barlett-Kernel and bandwidth of Newey-West and null of nonstationarity (unit-roots) is strongly rejected. We also have employed Augmented Dickey-Fuller according to Schwarz Information Criterion with the maximum lags of 43 and we have rejected the non-stationarity again (see Table 1). That indicates we do not need differentiate series or add trend component to the model.

Table 1: Philips-Perron Stationarity Test Statistics for Hourly Load Series

	t-stat	P-value
Philips-Perron test stat	-29.977	0.0000
A. Dickey-Fuller test stat	-7.738	0.0000

Figure 1 displaying the load demand for two years reveals that the demand series shows seasonal pattern. Figure 2 points out that each hour has similar daily structure and weekly structure. We observe that generally in every day after 7 am, load demand increases and after 18 pm demand reduces. That reveals the structure of the load series are changing in a day similarly. Although each day has similar structure, the similarity of the week days also points out that there is weekly cycle. Especially the difference of structure between weekends and work days draws the attention. Figure 2 also reveals that there exist different patterns of load series for four seasons. For instance, while around 12 pm is the peak hour of the day in summer, around 18 pm is the peak hour in the other seasons. That's because 12 pm is generally the hottest hour of a summer day. Also we notice that in spring and autumn which are the transition seasons, the load demand does not reach 24.000 MW. However, we have not considered yearly cycle that reflects changes of the seasons. Taylor (2003) suggested that the yearly cycle resulted from temperature changes was not significant in short term load forecasting up to one day and load demand series can capture smooth change of weather variables in short term. We have also interpreted ACFs and PACFs such that series show daily and weekly periodicity (see Appendix A.1 Table 16).

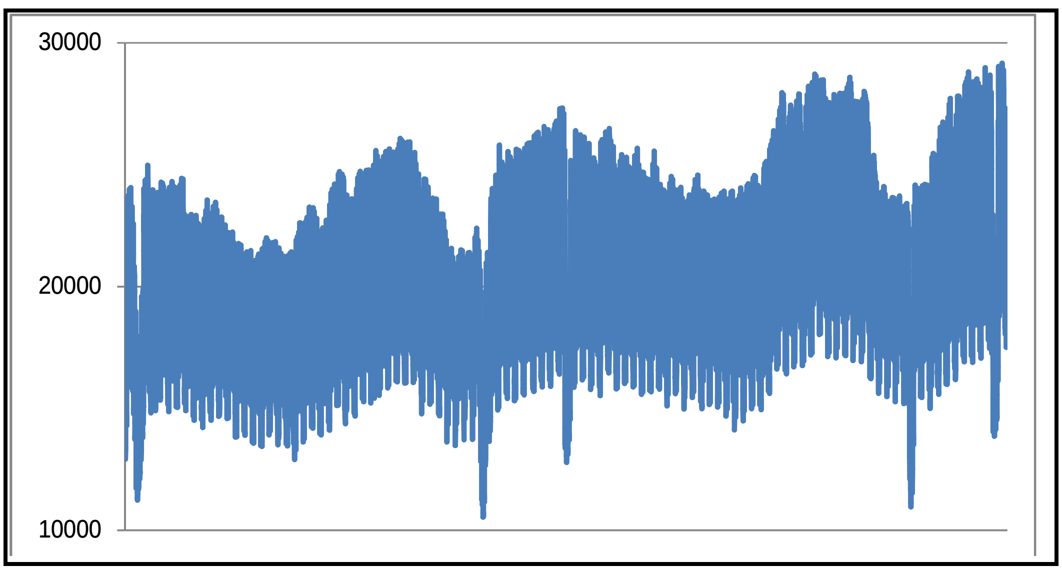
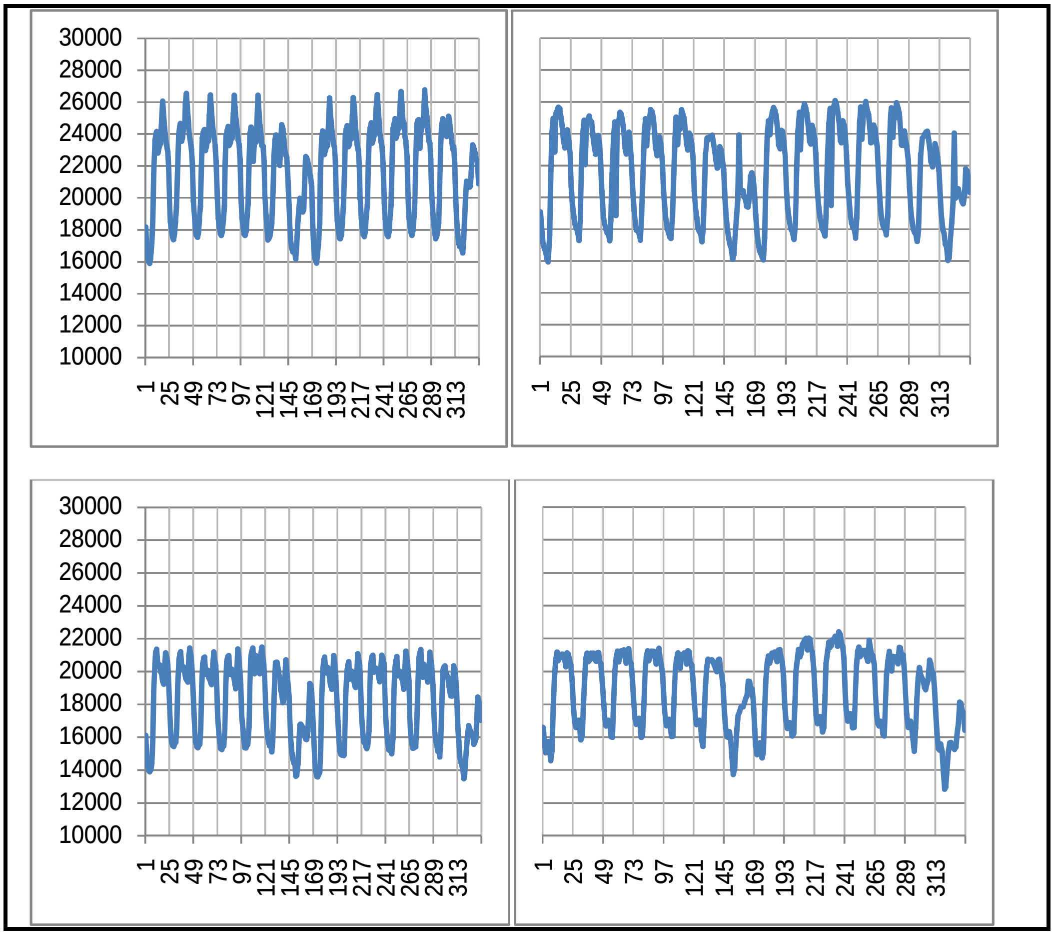


Figure 1: Hourly Load Series in MW for the Period between 1 January 2006 and 31 December 2007



Notes: Upper part is for December and August, respectively and lower part is for April and September, respectively. All of the series includes one week and start from Monday and end with Sunday. Load demands are in MW.

Figure 2: Hourly Load Series for Two Week of Different Seasons in 2006

4.2 Methodology and Empirical Results

In this paper we compare time series models in the literature based solely on statistical arguments. We have modeled load series for three methods and compare the forecast results up to one day horizon. The models discussed here are Heterogeneous Double Seasonal Autoregressive Moving Average Model, Heterogeneous Periodic Autoregressive Model and Multiple Regime Logistic Smooth Transition Autoregressive Model. Estimation and forecast results of the first two models are provided by using econometrics programme of E-views and results of the latter one provided by programming in Matlab.

4.2.1 Heterogeneous Double Seasonal ARIMA Model

Box et al. (1994, p. 333) wrote that the standard ARMA model for single seasonality can be extended for multiple seasonal cycles. Further, Taylor (2003) expressed extended seasonal ARIMA model capturing both the intraday and the intraweek seasonal cycles as

$$\phi_p(L)\Phi_{P_1}(L^{s_1})\Phi_{P_2}(L^{s_2})\nabla^d\nabla_{s_1}^{D_1}\nabla_{s_2}^{D_2}(y_t - a - bt) = \theta_q(L)\Theta_{Q_1}(L^{s_1})\psi_{Q_2}(L^{s_2})\varepsilon_t \qquad (1)$$

where y_t is the load in period t; a is a constant term; b is the coefficient of a linear deterministic trend term; ε_t is a white noise error term; L is a lag operator; ∇ is the difference operator, ∇_{s_1} and ∇_{s_2} are seasonal difference operators, s_1 and s_2 are intraday and intraweek cycles; and ϕ_P, Φ_{P_1}, $\Phi_{P_2}, \Theta_{Q_1}, \psi_{Q_2}$ are polynomial functions of p, P_1, P_2, Q_1 and Q_2 respectively.

This model (1) can be expressed as ARIMA (p, d, q)x(P_1, D_1, Q_1)x(P_2, D_2, Q_2). Darbellay and Slama (2000) apply the model for hourly load data of Czech Republic by setting s_1=24 to model within-day cycle of 24 hours, and s_2=168 to model the within-week cycle of 168 hours.

Taylor (2003) indicates that although the model can be extended to triple seasonal model to model yearly seasonal cycle, it is not significant in the context of forecasting up to a day. This leads us to use the Double Seasonal ARIMA model used as in the study of Taylor (2003). However our study differs from the study by Taylor (2003) in that we take into account heteroskedastic nature of the residuals and ARCH-GARCH models in the estimation.

The autocorrelation function and partial autocorrelation function of the time series are observed to select the order of the model (see Appendix A.1 Table 15 and Table 17). The selected model is estimated by non linear least square technique based on the fact that Nonlinear Least Squares estimates are asymptotically equivalent to Maximum Likelihood estimates and are asymptotically efficient under the assumption of $\varepsilon_t \sim N(0, \sigma^2)$ (Amaral, 2008). Coefficients are estimated simultaneously by Marquardt nonlinear least squares algorithm. We compared the Schwarz Bayesian Information Criteria and MAPEs for different ARIMA models. We used the logarithmic transformation of series which provides improvement in SBIC (see Appendix A.1 Table 15).

Double Seasonal ARMA model we estimated is expressed as:

$$(1-\phi_1 L)\left(1-\phi_{s_1} L^{s_1}\right)\left(1-\phi_{s_2} L^{s_2}\right)(lny-c) = (1+\rho_1 L)\left(1+\rho_{s_1} L^{s_1}\right)\left(1+\rho_{s_2} L^{s_2}\right)\varepsilon_t \quad (2)$$

where L is the lag operator and s_1 is the daily seasonal cycle and s_2 is the weekly cycle for hourly load series.

After the model (2) is estimated, it is subjected to heteroskedasticity test. We applied the Lagrange Multiplier test for conditional heteroskedasticity proposed by Engle (1982). The null hypothesis of the test is no ARCH effect in the model with the assumption of normal distribution of errors. GARCH, EGARCH, EGARCH-in-Mean effect in the ARMA series are evaluated.

The family of ARCH (p, q) model is estimated by maximum likelihood method with Marquardt-Levenberg algorithm. ARCH lags are added to the model until heteroskedasticity in the residuals is eliminated. We have estimated SARIMA model for each member of ARCH family to model residuals and chosen ARCH model which gives the best MAPE result (see Appendix A.1 Table 17).

Following the introduction of the GARCH effect into the model, Ljung-Box Q test (Godfrey, 1988) is applied for no remaining serial correlation in the residuals. It is a statistical test of whether any group of autocorrelations of the time series is zero.

The most appropriate model is SARIMA (1,0,1)x$(1,0,1)_{24}$ x$(1,0,1)_{168}$ model (2) in which $s_1 = 24$ and $s_2 = 168$ (see Table 14, Table 16). The LM test statistics in the Table 2 suggests that the estimated residuals from the model (2) are heterogeneous and that there is ARCH effect in the model. By considering MAPEs EGARCH-M model is chosen (see Appendix A.1 Table 17).

Table 2: ARCH LM Test Statistics for Double Seasonal ARMA Model

Prob. Chi-square (at lag 3)	0.0000
Variable	Coefficient
c	0.0002 (0.000)***
ε_{t-1}^2	0.3878 (0.008)***
ε_{t-2}^2	-0.0740 (0.008)***
ε_{t-3}^2	0.0090 (0.008)

Notes: Values in brackets are standard deviations of estimated parameters of the LM test equation. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

We estimated heterogeneous SARIMA model expressed as:

$$(1-\phi_1 L)\left(1-\phi_{s_1}L^{s_1}\right)\left(1-\phi_{s_2}L^{s_2}\right)(lny-c-\vartheta sqrt(h_t))$$
$$= (1+\rho_1 L)\left(1+\rho_{s_1}L^{s_1}\right)\left(1+\rho_{s_2}L^{s_2}\right)\varepsilon_t \qquad (3)$$

where

$$\ln(h_t) = c_1 + c_2 * |\varepsilon_{t-1}/sqrt(h_{t-1})| + c_3 * \varepsilon_{t-1}/sqrt(h_{t-1}) + c_4 * \ln(h_{t-1})$$

Table 3: Estimation Results of SARIMA Model with EGARCH-M (1, 1)

Variable	Coefficient
$sqrt(h)$	0.002 (0.001)**
c	11.168 (0.331)***
$ar(1)$	0.967 (0.001)***
$sar(24)$	0.052 (0.001)***
$sar(168)$	0.938 (0.001)***
$ma(1)$	-0.093 (0.006)***
$sma(24)$	0.243 (0.003)***
$sma(168)$	-0.260 (0.003)***
$c(1)$	-3.215 (0.045)***
$c(2)$	0.552 (0.006)***
$c(3)$	0.042 (0.004)***
$c(4)$	0.672 (0.005)***

Notes: Values in brackets are standard deviations of estimated parameters. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

According to estimation results of the model (3) displayed in Table 3, we conclude that one hour ago, one day ago and one week ago load demand affect significantly load demand in positive manner as we expected. Significance of seasonal terms in the model reveals that load series shows daily and weekly pattern.

The coefficient of determination of the estimated model (3) is computed as 0.99 and it shows that the regression line approximates the real data points very good. The comparison of actual and estimated data is displayed in the Figure 3 and illustrates that the estimated load demand is very close to real load demand.

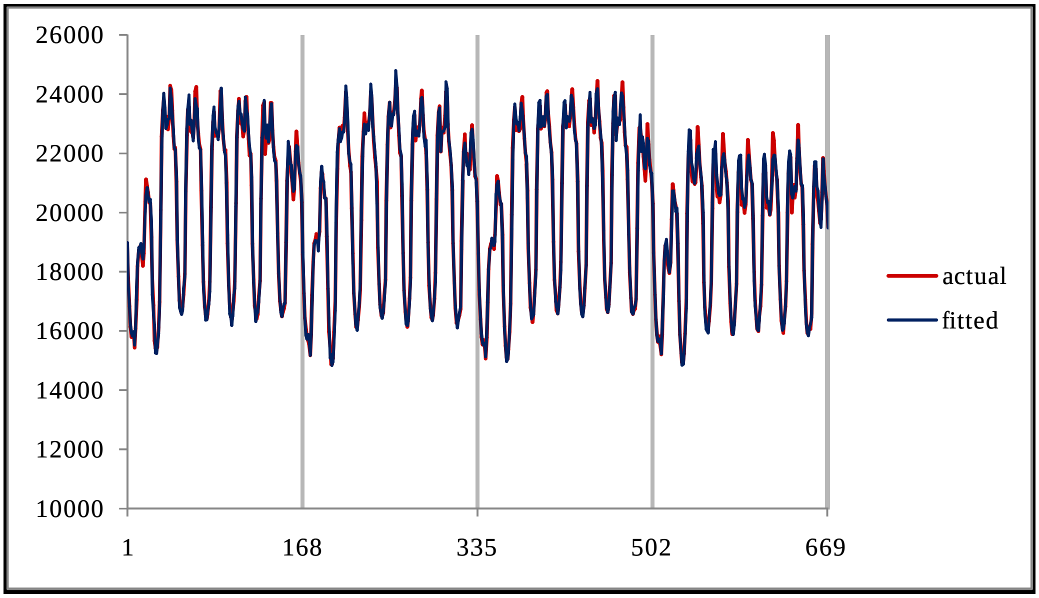


Figure 3: Actual and Fitted Values of H-SARIMA Model for One Month

We have employed misspecification test on the estimated model. Ljung-Box Test is introduced to detect serial correlation. The p-values of the Q statistics indicate that residuals of the model have serial correlation problem (see Table 4). We cannot avoid correlation problem.

Table 4: Ljung-Box Test Statistics for H-SARIMA Model

	Test for qth order autocorrelation			
q	1	12	24	168
Q-stat				

Notes: Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

In Figure 4 comparison of actual and forecasted values for one work day of winter are displayed. The illustration implies that the model gives accurate forecasts.

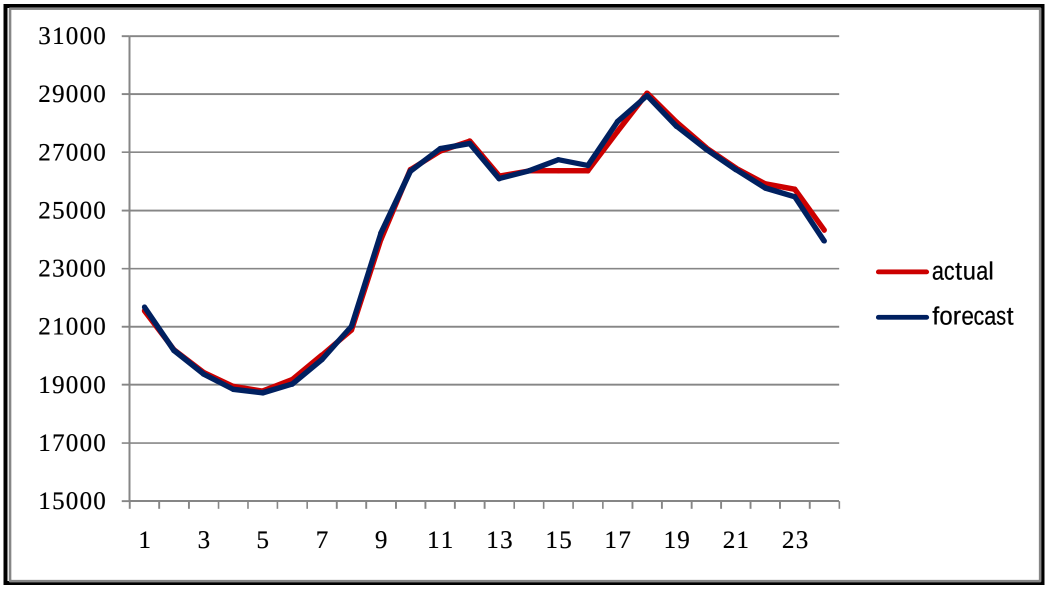


Figure 4: Actual and Forecasted Values of H-SARIMA Model up to One Day

Forecast errors are evaluated according to Mean Absolute Percentage Error. MAPE of the proposed Double Seasonal ARMA model with EGARCH-M (1, 1) effect up to one day forecast horizon is computed as 0.57 which is shown in Table 5.

Table 5: MAPE results for H-SARIMA Model

Hour	1	2	3	4	5	6	7	8	9	10	11	12
MAPE	0.61	0.36	0.33	0.37	0.36	0.43	0.48	0.49	0.53	0.50	0.48	0.47
Hour	13	14	15	16	17	18	19	20	21	22	23	24
MAPE	0.46	0.43	0.50	0.51	0.55	0.54	0.54	0.52	0.50	0.51	0.53	0.57

4.2.2 Heterogeneous Periodic AR Model

Taylor (2006) proposed the Periodic AR model to model electricity load data in which the parameters change with the periods. The coefficients of the model are time-variant. The study shows AR model are sufficient and MA terms are unnecessary.

Soares and Medeiros (2008) proposed Two Level Seasonal AR model where two models are estimated separately; one of them modeling potential load with Fourier transition component and dummy variables, second of them modeling irregular load allowing heteroskedasticity in errors. Amaral (2008) used Periodic AR model like in the study of Taylor (2006) for his linear model.

Our model is Periodic AR model used in the study of Taylor (2006). However unlike Taylor (2006), we model heteroskedasticity structure of the residuals. We multiplied coefficients with Fourier series to specify daily and weekly periodicity.

The model estimated is expressed as:

$$lny = -c + \alpha_i lny_t L^i + \phi_{s_{1i}} lny_t L^{s_1} + \phi_{s_{2i}} lny_t L^{s_2} + \varepsilon_t$$

$$\phi_{s_1} = \alpha_{s_1} + \sum_{j=1}^{h_1} \left(\gamma_{1j} \cos 2j\pi \frac{D(t)}{s_1} + \gamma_{2j} \sin 2j\pi \frac{D(t)}{s_1} \right) + \sum_{j=1}^{h2} \left(\tau_{1j} \cos 2j\pi \frac{W(t)}{s_2} + \tau_{2j} \sin 2j\pi \frac{W(t)}{s_2} \right)$$

$$\phi_{s_2} = \alpha_{s_2} + \sum_{j=1}^{h_1} \left(\varphi_{1j} \cos 2j\pi \frac{D(t)}{s_1} + \varphi_{2j} \sin 2j\pi \frac{D(t)}{s_1} \right) + \sum_{j=1}^{h_2} \left(\psi_{1j} \cos 2j\pi \frac{W(t)}{s_2} + \psi_{2j} \sin 2j\pi \frac{W(t)}{s_2} \right) \quad (4)$$

We apply an LM test for detect heteroskedasticity. In order to detect serial correlation in residual series, Ljung-Box Q test is used.

The lags are decided 1, 24 and 168 and two periods are chosen according to ACF and PACF, one of them cycles at every 24 hour to specify daily periodicity and the other one cycles at every 168 hour to specify weekly periodicity.

The number of harmonics is decided by minimizing SBIC. We evaluated harmonics up to 5 and the model with 5 harmonics has minimum SBIC (see Appendix A.2 Table 17). LM test results for the model (4) shown in Table 6 reveal that the residuals of the model have heteroskedastic structure.

Table 6: ARCH LM Test Results

Prob. Chi-square (at lag 3)	0.0000
Variable	Coefficient
c	0.0003 (0.000)***
ε_{t-1}^2	0.5227 (0.008)***
ε_{t-2}^2	-0.1750 (0.008)***
ε_{t-3}^2	0.1262 (0.008)***

Notes: Values in brackets are standard deviations of estimated parameters of the LM test equation. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

However after LM Test is applied and ARCH models are introduced into the model, some harmonics turn to be insignificant. Thus, we eliminate the insignificant harmonics from the model and estimate again. We have chosen 4 harmonics for daily periodicity and 3 harmonics for weekly periodicity. The determination of heteroskedasticity structure of the model is made by comparing MAPE results (Appendix A.2 Table 18) and EGARCH (1, 1) model is chosen to model heteroskedasticity in the residuals .

According to estimation results in Table 7 there are 3 significant harmonics to describe daily periodicity and 4 significant harmonics to describe weekly periodicity. The coefficients are highly significant. The series have statistically significant daily and weekly periodic structure.

Table 7: Estimation Output of PAR Model with EGARCH (1, 1)

α_0	-0.134 (0.011)***
α_1	0.250 (0.002)***
α_{s1}	0.488 (0.002)***
λ_{11}	0.107 (0.002)***
λ_{21}	-0.025 (0.002)***
τ_{11}	-0.132 (0.003)***
τ_{21}	-0.222 (0.002)***
λ_{12}	0.008 (0.002)**
λ_{22}	0.022 (0.002)***
τ_{12}	0.080 (0.002)***
τ_{22}	-0.056 (0.002)***
λ_{13}	-0.019 (0.002)***
λ_{23}	-0.018 (0.002)***
τ_{13}	-0.021 (0.002)***
τ_{23}	0.069 (0.002)***
λ_{14}	0.009 (0.002)***
λ_{24}	-0.010 (0.002)***
αs_2	0.275 (0.002)***
φ_{11}	-0.108 (0.002)***
φ_{21}	-0.025 (0.002)***
ψ_{11}	0.131 (0.002)***
ψ_{21}	0.222 (0.002)***
φ_{12}	-0.008 (0.002)**
φ_{22}	-0.022 (0.002)***
ψ_{12}	-0.081 (0.002)***
ψ_{22}	0.056 (0.002)***
φ_{13}	0.019 (0.002)***
φ_{23}	0.018 (0.002)***
ψ_{13}	0.021 (0.002)***
ψ_{23}	-0.070 (0.002)***
φ_{14}	-0.010 (0.002)***
φ_{24}	0.010 (0.002)***
c(1)	0.000 (0.000)***
c(2)	0.707 (0.000)***
c(3)	0.217 (0.010)***

Notes: Values in brackets are standard deviations of estimated parameters of the LM test equation. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

In Figure 5, the actual and fitted values of H-PAR model are compared for one month. The coefficient of determination equals to 0.99 that indicates the regression line fit the data very well.

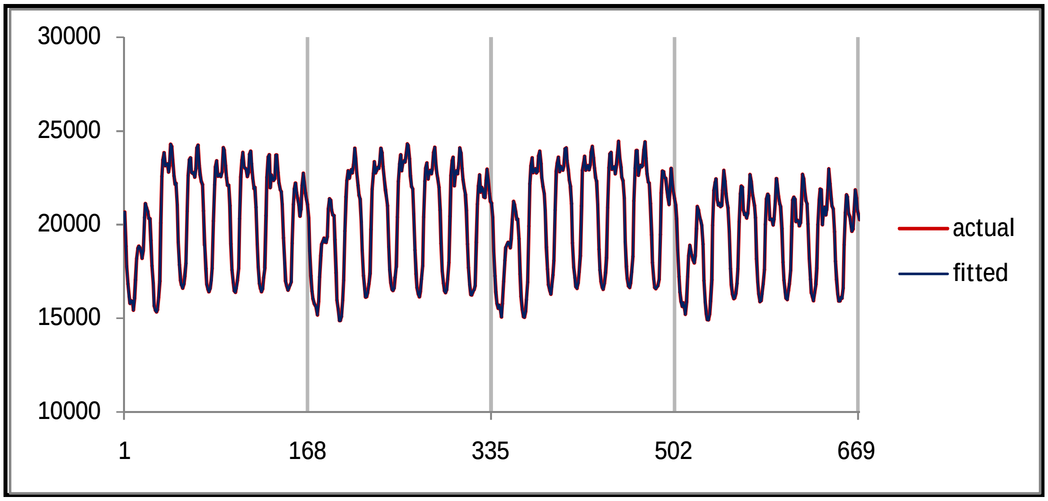


Figure 5: Actual and Fitted Values of H-PAR Model

Then, Ljung-Box Test is introduced to detect serial correlation. The p-values of the Q statistics in Table 8 indicate that residuals of the model have serial correlation problem. We cannot avoid serial correlation problem in this model either.

Table 8: Ljung-Box Test Statistics for H-PAR Model

	Test for qth order autocorrelation			
q	1	12	24	168
Q-stat	(4616.3)***	(24151.)***	(33010.)***	(48099.)***

Notes: Values in brackets are standard deviations of estimated parameters of the test statistics. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

In Figure 6 comparison of actual and forecasted load demand data up to one day horizon is displayed. The illustration shows that the difference between actual and forecasted demand between about 3 am and 7 am where the load demand is lowest is highest. Thus the model is worse between the hours that the load demand is lowest.

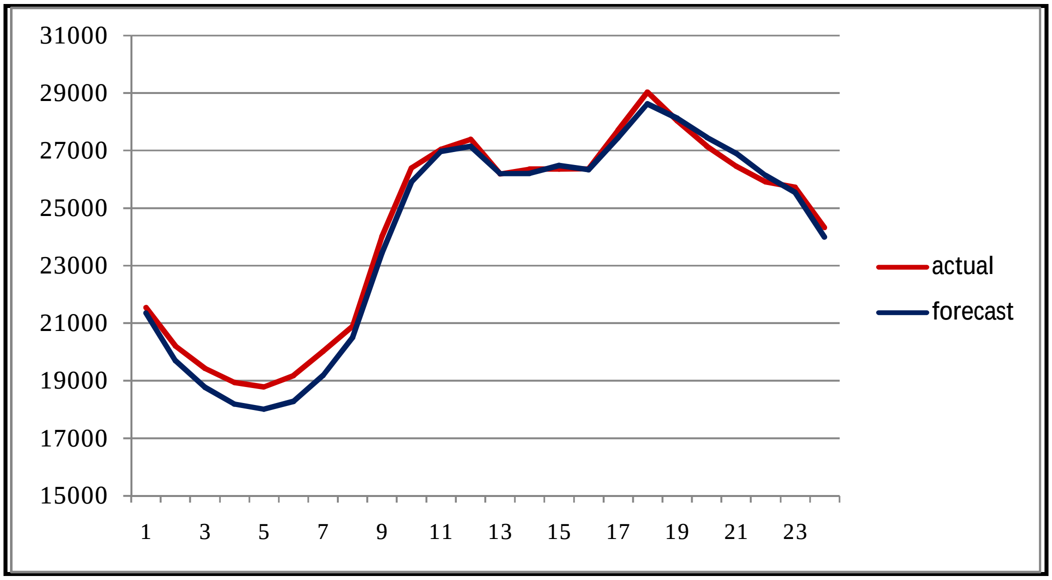


Figure 6: Actual and Forecasted Values of H-PAR Model up to One Day

In Table 9 MAPE results up to one day forecast horizon is displayed. MAPE increases at first hours, then decreases gradually. MAPEs approves that the model has highest errors where the demand is lowest.

Table 9: MAPE results for H-PAR Model

Hour	1	2	3	4	5	6	7	8	9	10	11	12
MAPE	0.86	1.03	1.29	1.52	1.66	1.84	1.93	1.84	1.93	1.84	1.82	1.74
Hour	13	14	15	16	17	18	19	20	21	22	23	24
MAPE	1.61	1.52	1.42	1.34	1.32	1.26	1.25	1.23	1.18	1.16	1.15	1.13

4.2.3 Multiple Regime Logistic Smooth Transition Autoregressive Model

We considered modeling nonlinear relationships by Multiple Regime Smooth Transition Autoregressive model. The model is a type of STAR model which uses a logistic function as a transition function.

The logistic STAR (LSTAR) model was first introduced to the time series literature by Chan and Tong (1986) and further developed by Teräsvirta (1994). Amaral (2008) proposed a new formula by introducing periodic behavior into a STAR model. Medeiros and McAleer (2008) studied Multiple Regime Heterogeneous LSTAR model for electricity price data. In this study we use the methodology of Teräsvirta (1994) in estimating Multiple Regime LSTAR model. We assume normality of the residuals and we do not allow heteroskedasticity.

Our model is in the form as follows:

$$y_t = G(w_t, s_t;\ \psi) + \varepsilon_t = \beta_0' w_t + \sum_{m=1}^{M} \beta_m' w_t f(s_t;\ \gamma_m, c_m) + \varepsilon_t \qquad (5)$$

where $G(w_t, z_t;\ \psi)$ is a nonlinear function of the variables w_t and s_t, w_t lagged values of dependent variable y_t, s_t is a transition variable represented by lagged value of y_t, $\psi \in R^{(M+1)(p+1)+2M}$, $f(s_t;\ \gamma_m, c_m)$ is the logistic function given by

$$f(s_t;\ \gamma_m, c_m) = \frac{1}{1+e^{-\gamma m(s_t - c_m)}} \qquad (6)$$

and ε_t is a random noise.

The parameter c in the equation (5) can be interpreted as the threshold of transition one regime to another, in the sense that the logistic function changes monotonically from 0 to 1 as transition variable s_t increases. Logistic function (6) is approaching from 0 to 0.5 as s_t increases while $s_t < c$, 1 to 0.5 as s_t decreases while $s_t < c$ and equals to 0.5 when $s_t = c$. The parameter γ determines the smoothness of the change in the value of the logistic function and thus smoothness of the transition from one regime to the other. As γ is very large, the change of the logistic function becomes almost instantaneous at $s_t = c$. Thus, the equation turns into TAR model as $\gamma \rightarrow \infty$. When $\gamma \rightarrow 0$, the logistic function become becomes equal to a constant and when $\gamma = 0$, the model reduces to the linear model.

Linear autoregressive model specified of order p using a model selection criterion SBIC. However it is renewed by comparing the MAPEs. Linearity is tested against STAR model. The transition variable of the alternative model is represented by one lagged value of the dependent variable, thus load demand of one hour ago from the estimated hour. Linearity test is repeated until most appropriate transition variable is determined. Linearity tests serve two purposes. Firstly the test reveals that a linear model is suitable for modeling or not since the nonlinear model is used if only the linear model is rejected. Secondly, if linearity is rejected for at least one of the transition variable, it allows choosing the model with the transition variable which makes the strongest rejection to the null hypothesis and gives the smallest p-value between candidates. However if there are several p-values which are close to each other, the model specification will be postponed to forecasting stage of procedure.

In this study LM type linearity test described by Teräsvirta (1994) is used. F-version LM test statistics under null hypothesis has an approximate F-distribution with n and $T-(k+p+1)-n$ degrees of freedom. It is recommended to use for small and moderate samples (Teräsvirta pg.70). Thus we use χ^2-version LM test statistics. Although the test statistics constructed under the assumption of normality we introduced heteroskedasticity robust version of LM test against unspecified forms of heteroskedasticty in the error process $\{\varepsilon_t\}$ following Davidson and MacKinnon (1985) and Wooldridge (1990) (see Appendix A.3 for more detail).

The selection of starting values is important for the optimization procedure. Nonlinear models are very sensitive to the initial values because of the iterative estimation technique. We used two dimensional grid search technique for initialization. When γ and c are fixed, model is linear in the parameters. Thus constructing a grid in the parameters of transition function and allows estimating parameters conditionally on the parameters included in the grids by minimizing sum of squared residuals.

The parameters can be estimated by conditionally maximum likelihood and nonlinear least squares. Two methods are equivalent when $\varepsilon_t \sim NID(0, \sigma^2)$. We have estimated model parameters by nonlinear least squares (7) technique. The vector of parameters ψ is estimated as:

$$\begin{aligned} \hat{\psi} &= \underset{\psi}{\operatorname{argmin}}\, \Theta_T(\psi) \\ &= \underset{\psi}{\operatorname{argmin}}\, \sum_{t=1}^{T} (y_t - G(w_t, s_t, \psi))^2 \end{aligned} \tag{7}$$

Determining the number of regimes of the nonlinear model is based on the test of remaining nonlinearities. The test works by introducing additional regime to the model and testing the significance of the additional regime (see Appendix A.3.2 for details).

First we have chosen lags for linear model 1, 2, 24, 168 by minimizing SBIC which equals to 15.90. However we eliminated lag 2 from the linear model, thus it is renewed to the model with lags 1, 24, 168 which gives SBIC 16.14 since it gives the better MAPE. After determining the linear autoregressive model, we introduced robust LM-test (see Appendix A.3.1), compute χ^2 distributed LM statistic and reject linearity against STAR model with the significance level of 1%. We tried until 3 lagged value of the dependent variable as the transition variable and we significantly reject linearity of the model against alternative with each of them. We determined the transition variable by comparing the MAPEs (see Appendix A.3 Table 19). As a result we have chosen lag one of the dependent variable as the transition variable. The transition variable is normalized by dividing to its sample standard deviation before estimation in order to normalize smoothing parameters. Remaining nonlinearity test is rejected when the nonlinear regime number has reached to 4 (see Table 10). That means, there are 5 statistically significant regimes in the model.

Table 10: Linearity and Remaining Non-Linearity Test Statistics for MR-LSTAR model

	Linearity	Remaining Non-Linearity			
χ^2	270.922	264.542	105.147	33.740	3.688
p-value	0.000	0.000	0.000	0.000	0.988

Estimated model is expressed explicitly as:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-24} + \alpha_3 y_{t-168} + \beta dummy_t - (\lambda_{m0} + \lambda_{m1} y_{t-1} + \lambda_{m2} y_{t-24} + \lambda_{m3} y_{t-168})(1 + exp\{-\gamma_m (y_{t-1} - c_m)\})^{-1} + \varepsilon_t \quad (8)$$

where m is the number of nonlinear regime which is m=1, 2, 3,4, and dummy is to model holidays differently which is one for holidays and zero for other ones.

γ_m and c_m are smoothing parameter and threshold variable of the nonlinear regimes, respectively. y_{t-1} is the one lagged of dependent variable and is chosen as the transition variable.

Table 11: Estimation Output of MR-LSTAR Model

Variable	Coefficient
α_0	605.429 (543.976)
α_1	0.768 (0.029)***
α_2	0.028 (0.020)*
α_3	0.173 (0.015)***
β	1288.925 (41.798)***
λ_{10}	-1473.932 (278.088)***
λ_{11}	-0.612 (0.182)***
λ_{12}	0.399 (0.128)***
λ_{13}	0.303 (0.078)***
λ_{20}	799.766 (258.202)***
λ_{21}	0.101 (0.147)
λ_{22}	-0.189 (0.111)**
λ_{23}	0.032 (0.055)
λ_{30}	-188.328 (436.224)
λ_{31}	0.375 (0.065)***
λ_{32}	-0.059 (0.027)**
λ_{33}	-0.306 (0.039)***
λ_{40}	-283.073 (569.217)
λ_{41}	-0.041 (0.030)*
λ_{42}	-0.123 (0.014)***
λ_{43}	0.172 (0.024)***
γ_1	3.475 (0.578)***
γ_2	10.254 (4.784)**
γ_3	6.373 (1.276)***
γ_4	10.256 (3.933)***
c_1	5.296 (0.112)***
c_2	5.498 (0.041)***
c_3	6.488 (0.040)***
c_4	7.594 (0.044)***

Notes: Values in brackets are standard deviations of estimated parameters of the test statistics. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

The estimation results of the model (8) are shown in Table 11. Non-normalized threshold parameters are computed as

$$(c_1, c_2, c_3, c_4) = (17187.64, 17843.21, 21056.16, 24645.57)$$

Thus the transition from first regime to the second realizes when the load demand of the one hour ago equals to 17187.64 and to third one at 17843.21, to fourth one at 21056.16 and to fifth one at 24645.57. When we examine daily graph of the load series, we realizes that there are five different periods within the day. Load demand of lower than first threshold variable 17187.64 approximately corresponds to night times, between first threshold and second threshold variable 17187.64 and 17843.21 corresponds to early morning times, between 17843.21 and 21056.16 corresponds to early hours of working day, between 21056.16 and 24645.57 corresponds to between hours of before lunch time and end of the day on spring and autumn months, greater than 24645.57 corresponds to between hours of before lunch time and end of the day on winter months and summer months.

In a typical work day the peak load demand is at about 6 pm, it diminishes until about 6 am and it rises until 6 pm with slight decreases at lunch time. The loads greater than 24.645 demanded at the peak hours of about 18.00 at winter days. For summer days the peak loads are at around 12 pm, thus load greater than 24.645 demanded at around 12 pm. By the end of the working day demand is diminishing until reaching to 17.183 at the lowest demand hour of about 6 am at the starting hour of the working day. Daily cycle of load demand at weekends resembles to weekdays but the levels are lower.

The comparison of actual and fitted values is displayed for one month in the figure 7. The coefficient of determination of the model is 0.93.

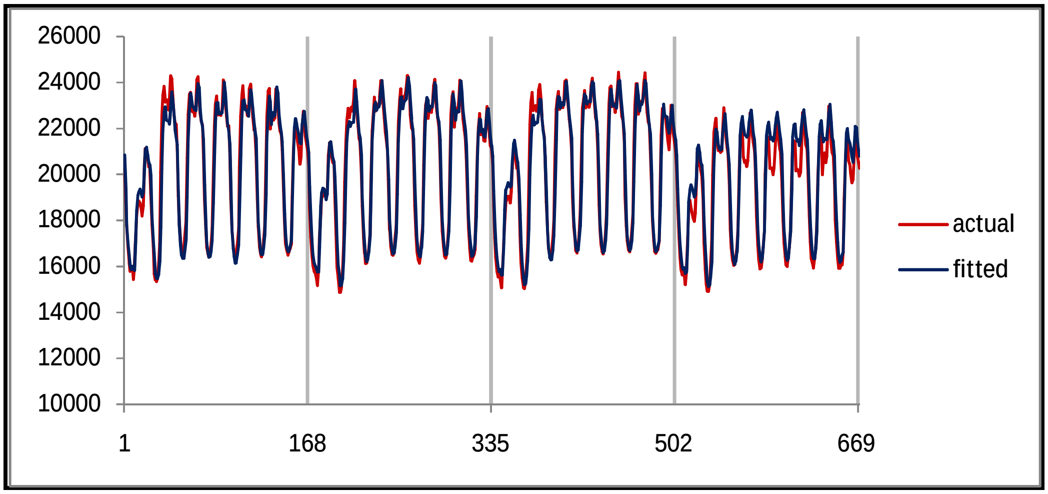


Figure 7: Actual and Fitted Values of MR-LSTAR Model for One Month

Ljung-Box Test is introduced to detect serial correlation. The p-values of the Q statistics indicate that residuals of the model has serial correlation problem (see Table 12). Although we have modeled the series as nonlinear, we cannot avoid serial correlation problem as well.

Table 12: Ljung-Box Test Statistics for MR-LSTAR Model

	Test for qth order autocorrelation			
q	1	12	24	168
Q-stat	(8295.5)***	(24514.45)***	(38846.5)***	(95899.3)***

Notes: Values in brackets are standard deviations of estimated parameters of the test statistics. Significance levels are denoted as *, **, ***, for %1, %5 and %10 significance levels, respectively.

We have not obtained good results from forecasting of load demand with MR-LSTAR model up to one day. However, Figure 8 and Table 13 displays that this model also gives good results up to 6 hour forecast horizon.

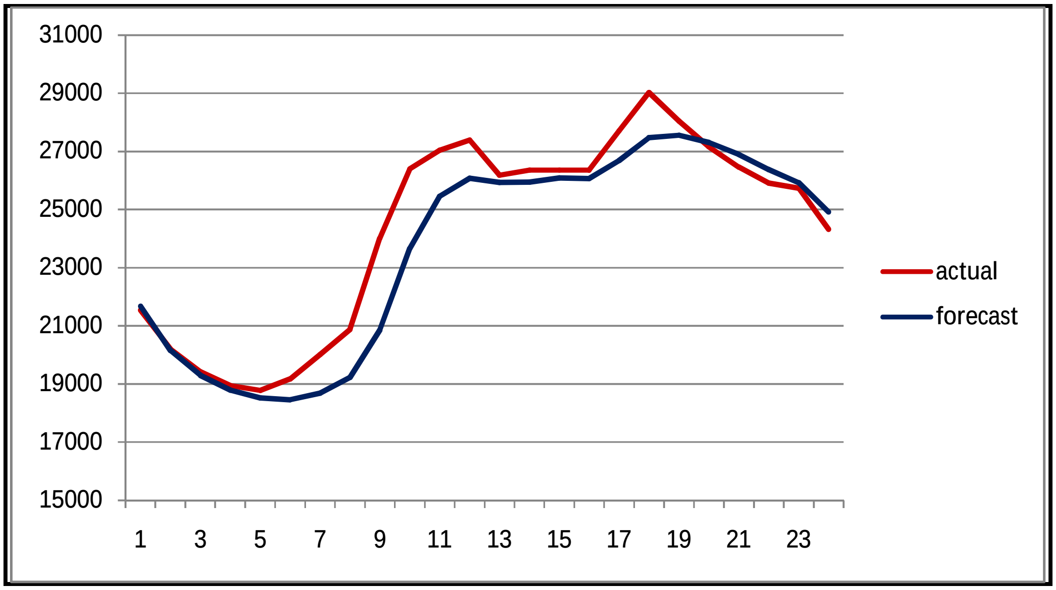


Figure 8: Actual and Forecasted Values of MR-LSTAR Model up to One Day

Table 13: MAPE results for MR-LSTAR Model

Hour	1	2	3	4	5	6	7	8	9	10	11	12
MAPE	0.63	0.43	0.51	0.59	0.74	1.24	2.01	2.75	3.91	4.55	4.67	4.68
Hour	13	14	15	16	17	18	19	20	21	22	23	24
MAPE	4.40	4.19	3.98	3.80	3.80	3.88	3.77	3.61	3.52	3.44	3.32	3.29

4.3 Comparison of Models

We have compared the MAPE results of the three models and resulted that the Heterogeneous Double Seasonal ARIMA model gives the best MAPE results in forecasting up to 24 hours (see Figure 9). Figure 9 shows that the MAPE results of H-PAR model gets closer to the H-SARIMA model as time horizon extends.

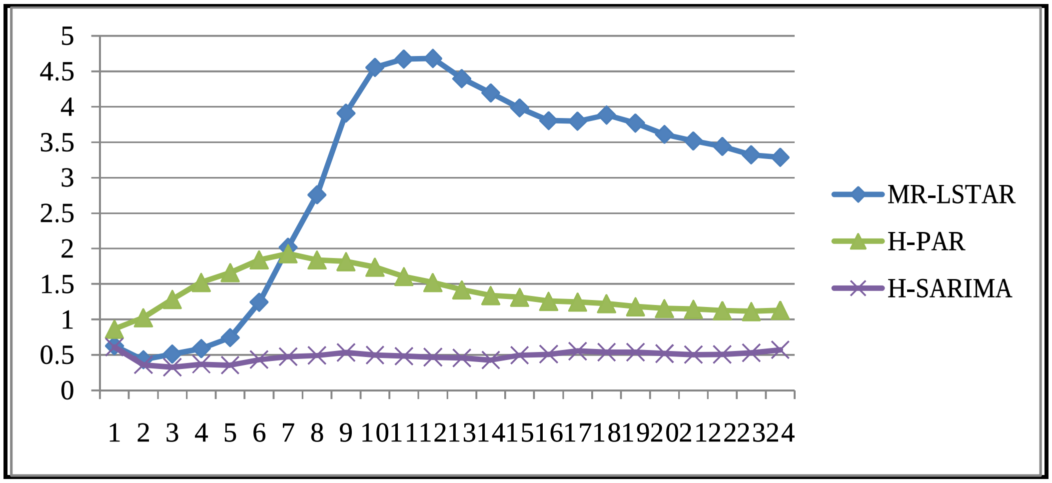


Figure 9: Comparisons of MAPEs between Models Forecasted up to 24 Hours Horizon

Heterogeneous Periodic AR model gives generally good results, however the results are worsen when the demand is lowest. The model is just failed to forecast load demand at the lowest points. On the other hand, from the H-PAR model we have obtained smaller forecast errors for advanced hours up to two weeks (see Figure 10).

MRLSTAR model gives the worse forecast results but it surpasses the forecast success of Heterogeneous Periodic AR model for the first 6 hours. The reason is the MRLSTAR model is better to forecast at lowest points than the H-PAR model. Figure 10 reveals that the difference between forecasted and the actual values of the H-PAR model for the second week is smaller than the other two models.

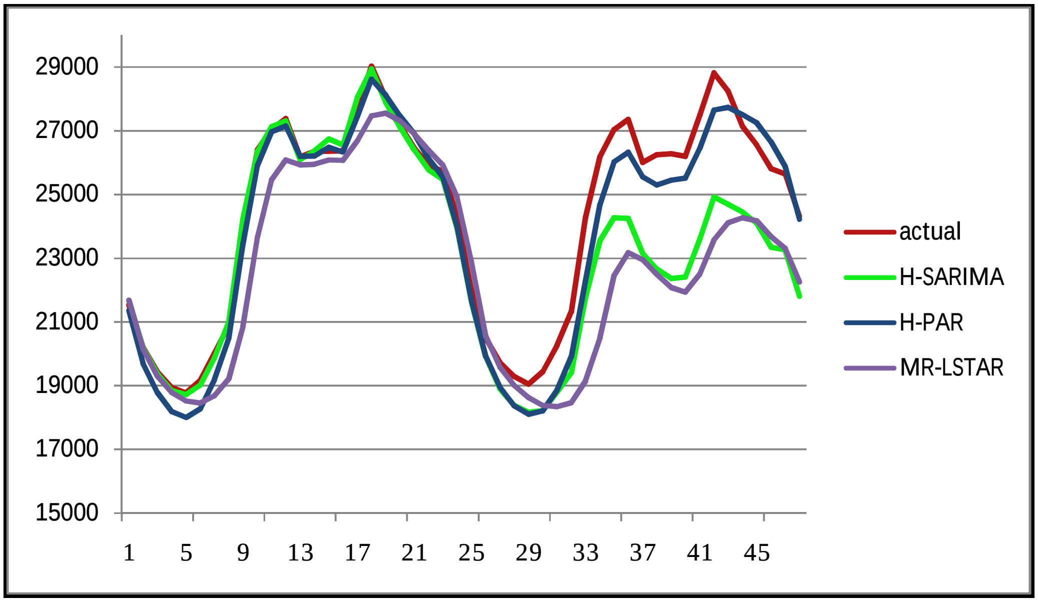


Figure 10: Comparison of Forecasted Models and Actual Data for Two Weeks

CHAPTER 5

CONCLUSION

In this study we have evaluated three univariate time series models in the context of short term forecasting up to one day in application of Turkish hourly electricity load data between the years 2006 and 2007. We have MAPE for H-SARIMA model of 0.57, for H-PAR model 1.13 and for MR-LSTAR model 3.29. Soares and Medeiros (2008) gets mean of MAPEs at 3.76 for the one day forecast horizon at hour 1 am and Taylor and McSharry (2007) gets mean of MAPEs for PAR model at 2.5 for the one day forecast horizon. We had best results with Heterogeneous SARIMA model up to one day. However SARIMA results get worse than Heterogeneous PAR model when extending the forecast horizon up to 48 hours. Although linearity is significantly rejected against MRLSTAR model, we have not obtained good results from nonlinear model longer than 6 hours horizon. The lack of nonlinear model may be excluding of heterogeneity from the model.

As a future research, it is suggested to considering the heterogeneity in MRLSTAR model, and evaluating the model with temperature variable as the transition variable and also it is an option to consider specifying periodicity of MR-LSTAR model with Fourier series.

REFERENCES

Al-Hamadi, H. and Soliman, S. (2004). "Short-Term Electric Load Forecasting Based on Kalman Filtering Algorithm with Moving Window Weather and Load Model". *Electric Power Systems Research*, 68, 47−59.

Amaral, L.F., Souza R.C., and Stevenson M. (2008). "A Smooth Transition Periodic Autoregressive (STPAR) Model for Short-Term Load Forecasting". *International Journal of Fore*casting, 24, 603–615.

Amjady, N. (2001). "Short-Term Hourly Load Forecasting Using Time Series Modeling with Peak Load Estimation Capability". *IEEE Transactions on Power Systems*, 16, 498−504.

Badri, A., Ameli, Z., and Birjandi, A.M. (2012). "Application of Artificial Neural Networks and Fuzzy Logic Methods for Short Term Load Forecasting". *Energy Procedia*, 14, 1883 – 1888

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994). Time Series Analysis: Forecasting and Control, third ed. Prentice Hall, New Jersey.

Bunn, D.W. (2000). "Forecasting Loads and Prices in Competitive Power Markets". *Proceedings of the IEEE*, 88, 163–169.

Bunn, D.W. (1982). "Short-Term Forecasting: A Review of Procedures in The Electricity Supply Industry". *Journal of the Operational Research Society*. 33, 533–545.

Bunn, D. and Farmer, E. (1985). Comparative Models for Electrical Load Forecasting. New York: Wiley.

Cancelo, J.R., Espasa, A., and Grafe, R. (2008). "Forecasting From One Day to One Week Ahead for The Spanish System Operator". *International Journal of Forecasting* 24, 588–602.

Charytoniuk, W. and Chen, M. (2000). "Very Short-Term Load Forecasting Using Artificial Neural Networks", *IEEE Transactions on Power Systems*, 15.

Christiaanse, W.R. (1971). "Short-Term Load Forecasting Using General Exponential Smoothing". *IEEE Transactions on Power Apparatus and Systems* PAS-90, 900–910.

Cottet, R. and Smith, M. (2003). "Bayesian Modeling and Forecasting Of İntraday Electricity Load". Journal of the American Statistical Association, 98, 839−849.

da Silva, A., Ferreira, V., and Velasquez, R. (2008). "Input Space to Neural Network Based Load Forecasting". *International Journal of Forecasting*, 24, 616−629 (this issue).

Darbellay, G.A. and Slama, M. (2000). "Forecasting the Short-Term Demand for Electricity – Do Neural Networks Stand a Better Chance? ". *International Journal of Forecasting*, 16, 71–83.

Dordonnat, V., Koopman, S., Ooms, M., Dessertaine, A., and Collet, J. (2008). "An Hourly Periodic State Space Model for Modelling French National Electricity Load". *International Journal of Forecasting*, 24, 566−587

Engle, R. F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of The Variance of UK Inflation". Econometrica, 50, 987−1007.

Fiebig, D., Bartels, R., and Aigner, D. (1991). "A Random Coefficient Approach to The Estimation of Residual End-Use Load Profiles". *Journal of Econometrics*, 50, 297−327.

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

Hagan, M.T. and Behr, S.M. (1987). "The Time Series Approach to Short-Term Load Forecasting". *IEEE Transactions on Power Systems* PWRS-2, 785–791.

Haida, T. and Muto, S. (1994). "Regression Based Peak Load Forecasting Using a Transformatıon Technique". *IEEE Transactions on Power Systems*, 9.

Haida, T., Muto, S., Takahashi, Y., and Ishi Y. (1998). "Peak Load Forecasting Using Multiple-Year Data with Trend Data Processing Techniques". *Electric Engineering in Japan*, 124.

Harvey, A. and Koopman, S. (1993). "Forecasting Hourly Electricity Demand Using Time-Varying Splines". *Journal of the American Statistical Association*, 88, 1228–1237.

Hippert, H., Bunn, D., and Souza, R. (2005). "Large Neural Networks of Electricity Load Forecasting: Are They Overfitted?". *International Journal of Forecasting*, 21, 425–434.

Hippert, H., Pedreira, C.E., and Souza, R.C. (2001). "Neural Networks for Short-Term Load Forecasting: A Review and Evaluation". *IEEE Transactions on Power Systems* 16, 44–55.

Hyde, O. and Hodnett, P.F. (1997). "An Adaptable Automated Procedure for Short-Term Electricity Load Forecasting". *IEEE Transactions on Power Systems*, 12, 84-93.

Hor, C.-L., Watson, S. J., and Majithia, S. (2005). "Analyzing the Impact Of Weather Variables on Monthly Electricity Demand". *IEEE Transactions on Power Systems*, 20, 2078–2085.

Huang, S.-J. (2003). "Short-Term Load Forecasting via ARMA Model Identification Including Non-Gaussian Process Considerations". *IEEE Transactions on Power Systems*, 18, 673–679.

Huang, C.M., Huang, C.-J., and Wang, M. L. (2005). "A Particle Swarm Optimization to Identify the ARMAX Model for Shortterm Load Forecasting". *IEEE Transactions on Power Systems*, 20, 1126–1133.

Hyde, O. and Taylor, J. W. (2010) “An Evaluation of Bayesian Techniques for Controlling Model Complexity and Selecting Inputs in a Neural Network for Short-Term Load Forecasting”. *Neural Networks*, 23, 386-395.

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). "Forecasting with Exponential Smoothing: The State Space Approach". *Springer-Verlag*, Berlin, Heidelberg.

Khotanzad, A., Afkhami-Rohani, R. and Maratukulam, D. (1998). "ANNSTLF - Artificial Neural Network Short-Term Load Forecaster - Generation Three". *IEEE Transactions on Power Systems*, 13, 1413−1422.

Mbamalu, G.A.N. and El-Hawary, M.E. (1993). "Load Forecasting via Suboptimal Seasonal Autoregressive Models and Iteratively Reweighted Least Squares Estimation". *IEEE Transactions on Power Systems,* 8, 343–348.

Metaxiotis, K., Kagiannas, A., Askounis, D., and Psarras, J. (2003). "Artificial Intelligence in Short Term Electric Load Forecasting: A State-Of-Art Survey for The Researcher". *Energy Conversion and Management*, 44, 1525−1534.

Moghram, I. and Rahman, S. (1989). "Analysis and Evaluation Of Five Short-Term Load Forecasting Techniques". *IEEE Transactions on Power Systems,* 4, 1484–1491.

Nowicka-Zagrajek, J. and Weron, R. (2002). "Modeling Electricity Loads in California: ARMA Models with Hyperbolic Noise". *Signal Processing*, 82, 1903−1915.

Pai, P.-F. and Hong, W.-C. (2005). "Support Vector Machines with Simulated Annealing Algorithms in Electricity Load Forecasting". *Energy Conversion and Management, 46, 2669–2688.*

Park, J.H., Park, Y.M., and Lee, K.Y. (1991). "Composite Modeling For adaptive Short-Term Load Forecasting". *IEEE Transactions on Power Systems 6, 450–457.*

Peirson, J. And Henley, A. (1994). "Electricity Load and Temperature". *Energy Economics*, 16, 235−243.

Ramanathan, R., Engle, R. F., Granger, C., Vahid-Arahi, F. and Brace, C. (1997). "Short-Run Forecasts Of Electricity Loads and Peaks". *International Journal of Forecasting*, 13, 161–174.

Reis, A. and Silva, A. (2005). "Feature Extraction via Multiresolution Analysis for Short-Term Load Forecasting". *IEEE Transactions on Power Systems, 20, 189–198.*

Soares, L.J. and Medeiros, M.C. (2008). "Modeling And Forecasting Short-Term Electricity Load: A Comparison of Methods with an Application to Brazilian Data". *International Journal of Forecasting* 24, 630–644.

Soares, L. and Souza, L. (2006). "Forecasting Electricity Demand Using Generalized Long Memory". *International Journal of Forecasting*, 22, 17–28.

Taylor, J.W. (2008). "An Evaluation of Methods for Very Short-Term Load Forecasting Using Minute-by-Minute British Data". *International Journal of Forecasting*, 24, 645–658.

Taylor, J.W. (2003). "Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing". *Journal of the Operational Research Society*, 54, 799–805.

Taylor, J.W. (2006). "Density Forecasting for The Efficient Balancing of the Generation and Consumption of Electricity". *International Journal of Forecasting*, 22, 707–724.

Taylor, J.W. (2010). "Triple Seasonal Methods for Short-Term Electricity Demand Forecasting". *European Journal of Operational Research*, 204, 139–152.

Taylor, J.W., deMenezes, L.M., and McSharry, P. (2006). "A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day Ahead". *International Journal of Forecasting*, 22, 1–16.

Taylor, J.W. and McSharry, P.E. (2007). "Short-Term Load Forecasting Methods: An Evaluation Based on European Data". *IEEE Transactions on Power Systems*, 22, 2213-2219.

Teräsvirta, T. (1994). "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models". *Journal of the American Statistical Association,* 89, 208–218.

Teräsvirta, T., Tjostheim, D., and Granger, C.W.J., (2010). "Modeling and Forecasting with Nonlinear Economic Time Series". Oxford University Press, Newyork.

Weron, R. (2006). "Modelling and Forecasting Electric Loads and Prices". Chichester: Wiley.

APPENDICES

APPENDIX A

TEST RESULTS

A.1 Test Results for Heterogeneous Double Seasonal ARIMA Model

Table 14: ACF and PACF of Hourly Load Series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|*******	|*******	1	0.949	0.949	15769.	0.000
|******|	****| |	2	0.839	-0.614	28090.	0.000
|***** |	| |	3	0.702	0.008	36725.	0.000
|**** |	| |	4	0.560	-0.014	42219.	0.000
|*** |	| |	5	0.426	0.009	45406.	0.000
|** |	*| |	6	0.301	-0.137	46992.	0.000
|* |	| |	7	0.187	0.018	47606.	0.000
|* |	|* |	8	0.098	0.151	47776.	0.000
| |	|* |	9	0.043	0.094	47809.	0.000
| |	| |	10	0.018	-0.036	47814.	0.000
| |	| |	11	0.013	-0.018	47817.	0.000

Notes: It is noticeable that PACF displays a sharp cutoff at lag 2, while the ACF decays slowly. Thus, the series are first modeled as AR(2) and ARMA(1,1) and the new form of ACF and PACF are observed.

Table 15: Comparison of SBIC of ARMA (1, 1) models

	Pure	Transformed
SBIC	-3.051	16.359

Notes: Pure indicates ARMA (1, 1) model with original hourly load data is the dependent variable, Log Transformed indicates ARMA (1, 1) model with logarithmic transformation of hourly load data is the dependent variable

Table 16: ACF and PACF of Hourly Load Series $\mathbf{SARIMA(1,0,1)x(1,0,1)_{24}}\boldsymbol{x}$ $\mathbf{(1,0,1)_{168}}$

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	1	-0.001	-0.001	0.0296	
| |	| |	2	0.021	0.021	7.3544	
| |	| |	3	-0.022	-0.022	15.997	
| |	| |	4	-0.052	-0.053	63.135	
| |	| |	5	-0.029	-0.029	77.929	
| |	| |	6	0.008	0.010	79.083	
| |	| |	7	-0.004	-0.005	79.316	0.000
| |	| |	8	-0.008	-0.013	80.515	0.000
| |	| |	9	-0.023	-0.025	89.451	0.000
| |	| |	10	-0.016	-0.016	93.670	0.000
| |	| |	11	-0.015	-0.014	97.377	0.000
| |	| |	12	0.021	0.019	104.66	0.000
| |	| |	13	0.025	0.022	115.73	0.000
| |	| |	14	-0.027	-0.031	128.04	0.000
| |	| |	15	-0.032	-0.034	145.13	0.000
| |	| |	16	-0.016	-0.013	149.81	0.000
| |	| |	17	0.001	0.004	149.82	0.000
| |	| |	18	0.026	0.022	161.33	0.000
| |	| |	19	0.027	0.020	174.19	0.000
| |	| |	20	0.015	0.011	178.30	0.000
| |	| |	21	-0.010	-0.009	179.86	0.000
| |	| |	22	0.002	0.006	179.92	0.000
| |	| |	23	0.054	0.059	230.73	0.000
| |	| |	24	0.057	0.059	287.34	0.000
| |	| |	25	0.043	0.039	319.50	0.000
| |	| |	26	-0.003	-0.004	319.70	0.000
| |	| |	27	-0.020	-0.010	326.85	0.000
| |	| |	28	-0.002	0.013	326.91	0.000
| |	| |	29	0.002	0.012	326.99	0.000
| |	| |	30	-0.000	-0.000	326.99	0.000
| |	| |	31	-0.001	-0.005	327.01	0.000
| |	| |	32	-0.008	-0.005	328.04	0.000
| |	| |	33	-0.015	-0.007	331.94	0.000
| |	| |	34	-0.024	-0.015	341.51	0.000
| |	| |	35	0.035	0.037	363.02	0.000
| |	| |	36	-0.002	-0.006	363.06	0.000
| |	| |	37	-0.002	-0.009	363.12	0.000
| |	| |	38	-0.002	-0.001	363.21	0.000
| |	| |	39	-0.028	-0.018	376.77	0.000
| |	| |	40	-0.027	-0.022	389.41	0.000
| |	| |	41	-0.006	-0.011	390.11	0.000
| |	| |	42	-0.008	-0.016	391.11	0.000

Notes: Both ACF and PACF of model AR(2) have strong spikes at multiples of 24 and multiples of 168). Thus the model renewed by adding sar(24), sar(168), sma(24) and sma(168)components. ACF and PACF have still spikes at multiples of 24 but not strong as before. We then try several combinations of SARIMA model with periodicity of 24 and 168 lags.

Table 16 (Continued)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	43	0.003	-0.008	391.30	0.000
| |	| |	44	0.002	-0.005	391.34	0.000
| |	| |	45	0.004	0.002	391.65	0.000
| |	| |	46	0.010	0.007	393.21	0.000
| |	| |	47	0.042	0.033	423.99	0.000
|** |	|* |	48	0.214	0.206	1213.9	0.000
| |	| |	49	0.049	0.052	1255.2	0.000
| |	| |	50	-0.016	-0.020	1259.7	0.000
| |	| |	51	-0.021	-0.010	1267.6	0.000
| |	| |	52	-0.018	0.012	1273.4	0.000
| |	| |	53	-0.018	-0.001	1279.2	0.000
| |	| |	54	0.011	0.007	1281.2	0.000
| |	| |	55	-0.007	-0.010	1282.1	0.000
| |	| |	56	-0.001	0.002	1282.1	0.000
| |	| |	57	-0.019	-0.008	1288.0	0.000
| |	| |	58	-0.014	-0.005	1291.5	0.000
| |	| |	59	-0.005	0.003	1291.9	0.000
| |	| |	60	0.010	0.002	1293.7	0.000
| |	| |	61	0.011	0.001	1295.9	0.000
| |	| |	62	-0.000	0.012	1295.9	0.000
| |	| |	63	-0.006	0.014	1296.5	0.000
| |	| |	64	-0.013	-0.001	1299.4	0.000
| |	| |	65	-0.015	-0.015	1303.1	0.000
| |	| |	66	-0.009	-0.025	1304.5	0.000
| |	| |	67	0.001	-0.017	1304.5	0.000
| |	| |	68	0.004	-0.008	1304.7	0.000
| |	| |	69	-0.019	-0.022	1311.2	0.000
| |	| |	70	-0.005	-0.015	1311.6	0.000
| |	| |	71	0.007	-0.021	1312.3	0.000
|* |	|* |	72	0.140	0.117	1650.7	0.000
| |	| |	73	-0.002	-0.018	1650.8	0.000
| |	| |	74	-0.008	-0.015	1651.8	0.000
| |	| |	75	-0.035	-0.023	1673.3	0.000
| |	| |	76	-0.024	-0.010	1682.9	0.000
| |	| |	77	0.001	0.010	1682.9	0.000
| |	| |	78	-0.006	-0.008	1683.5	0.000
| |	| |	79	0.006	0.001	1684.0	0.000
| |	| |	80	0.003	0.000	1684.2	0.000
| |	| |	81	0.001	0.008	1684.2	0.000
| |	| |	82	-0.017	-0.008	1689.3	0.000
| |	| |	83	-0.007	-0.016	1690.1	0.000
| |	| |	84	-0.000	-0.007	1690.1	0.000
| |	| |	85	0.023	0.022	1699.1	0.000
| |	| |	86	0.014	0.024	1702.7	0.000

Notes: Both ACF and PACF of model AR(2) have strong spikes at multiples of 24 and multiples of 168). Thus the model renewed by adding sar(24), sar(168), sma(24) and sma(168)components. ACF and PACF have still spikes at multiples of 24 but not strong as before. We then try several combinations of SARIMA model with periodicity of 24 and 168 lags.

Table 16 (Continued)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	87	-0.008	0.010	1703.7	0.000
| |	| |	88	-0.015	0.004	1707.9	0.000
| |	| |	89	-0.013	-0.005	1710.9	0.000
| |	| |	90	-0.015	-0.014	1714.7	0.000
| |	| |	91	-0.005	-0.014	1715.2	0.000
| |	| |	92	-0.003	-0.010	1715.3	0.000
| |	| |	93	-0.003	-0.005	1715.5	0.000
| |	| |	94	-0.022	-0.028	1723.6	0.000
| |	| |	95	0.044	0.019	1757.5	0.000
|* |	| |	96	0.098	0.051	1924.5	0.000
| |	| |	97	0.013	-0.014	1927.6	0.000
| |	| |	98	-0.014	-0.013	1930.8	0.000
| |	| |	99	-0.023	-0.008	1940.2	0.000
| |	| |	100	-0.039	-0.025	1966.8	0.000
| |	| |	101	-0.020	-0.010	1973.6	0.000
| |	| |	102	-0.005	-0.007	1974.0	0.000
| |	| |	103	0.002	-0.001	1974.1	0.000
| |	| |	104	0.016	0.012	1978.5	0.000
| |	| |	105	-0.004	-0.002	1978.8	0.000
| |	| |	106	-0.002	0.007	1978.8	0.000
| |	| |	107	0.002	-0.004	1978.9	0.000
| |	| |	108	0.003	-0.004	1979.1	0.000
| |	| |	109	0.009	0.002	1980.5	0.000
| |	| |	110	0.006	0.007	1981.1	0.000
| |	| |	111	-0.004	0.006	1981.5	0.000
| |	| |	112	-0.014	-0.000	1984.7	0.000
| |	| |	113	-0.019	-0.007	1990.7	0.000
| |	| |	114	-0.023	-0.018	2000.0	0.000
| |	| |	115	-0.001	-0.006	2000.1	0.000
| |	| |	116	0.015	0.011	2003.8	0.000
| |	| |	117	-0.005	0.002	2004.3	0.000
| |	| |	118	-0.016	-0.015	2008.7	0.000
| |	| |	119	0.022	0.015	2017.3	0.000
| |	| |	120	0.054	0.003	2067.3	0.000
| |	| |	121	-0.004	-0.015	2067.7	0.000
| |	| |	122	0.006	0.018	2068.4	0.000
| |	| |	123	-0.017	0.005	2073.1	0.000
| |	| |	124	-0.013	-0.001	2076.0	0.000
| |	| |	125	-0.012	-0.007	2078.6	0.000
| |	| |	126	-0.016	-0.016	2083.0	0.000
| |	| |	127	-0.013	-0.014	2085.7	0.000
| |	| |	128	0.007	-0.001	2086.5	0.000
| |	| |	129	0.003	0.002	2086.7	0.000
| |	| |	130	-0.010	-0.008	2088.6	0.000

Notes: Both ACF and PACF of model AR(2) have strong spikes at multiples of 24 and multiples of 168). Thus the model renewed by adding sar(24), sar(168), sma(24) and sma(168)components. ACF and PACF have still spikes at multiples of 24 but not strong as before. We then try several combinations of SARIMA model with periodicity of 24 and 168 lags.

Table 16 (Continued)

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>| |</td><td>| |</td><td>131</td><td>0.013</td><td>0.012</td><td>2091.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>132</td><td>0.020</td><td>0.022</td><td>2098.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>133</td><td>-0.012</td><td>-0.025</td><td>2100.9</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>134</td><td>0.023</td><td>0.011</td><td>2109.9</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>135</td><td>0.006</td><td>0.013</td><td>2110.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>136</td><td>-0.007</td><td>0.007</td><td>2111.3</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>137</td><td>-0.008</td><td>0.005</td><td>2112.4</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>138</td><td>-0.028</td><td>-0.017</td><td>2126.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>139</td><td>-0.011</td><td>-0.010</td><td>2128.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>140</td><td>-0.004</td><td>-0.009</td><td>2128.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>141</td><td>-0.010</td><td>-0.010</td><td>2130.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>142</td><td>-0.028</td><td>-0.028</td><td>2143.8</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>143</td><td>-0.033</td><td>-0.054</td><td>2162.4</td><td>0.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>144</td><td>0.108</td><td>0.062</td><td>2365.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>*| |</td><td>145</td><td>-0.064</td><td>-0.067</td><td>2437.0</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>146</td><td>-0.016</td><td>-0.011</td><td>2441.2</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>147</td><td>-0.023</td><td>-0.009</td><td>2450.0</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>148</td><td>-0.028</td><td>-0.008</td><td>2463.3</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>149</td><td>-0.010</td><td>-0.004</td><td>2465.2</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>150</td><td>-0.013</td><td>-0.017</td><td>2468.2</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>151</td><td>-0.000</td><td>-0.006</td><td>2468.3</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>152</td><td>0.013</td><td>0.001</td><td>2471.4</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>153</td><td>0.007</td><td>0.003</td><td>2472.3</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>154</td><td>0.017</td><td>0.013</td><td>2477.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>155</td><td>-0.000</td><td>0.000</td><td>2477.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>156</td><td>-0.003</td><td>-0.006</td><td>2477.2</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>157</td><td>0.009</td><td>-0.008</td><td>2478.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>158</td><td>-0.000</td><td>-0.004</td><td>2478.5</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>159</td><td>-0.009</td><td>-0.007</td><td>2479.8</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>160</td><td>-0.014</td><td>-0.007</td><td>2483.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>161</td><td>-0.017</td><td>0.002</td><td>2488.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>162</td><td>-0.017</td><td>0.002</td><td>2493.3</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>163</td><td>0.005</td><td>0.011</td><td>2493.8</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>164</td><td>0.017</td><td>0.015</td><td>2498.8</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>165</td><td>0.005</td><td>0.011</td><td>2499.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>166</td><td>-0.000</td><td>0.011</td><td>2499.1</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>167</td><td>-0.003</td><td>-0.025</td><td>2499.3</td><td>0.000</td></tr>
<tr><td>| |</td><td>| |</td><td>168</td><td>-0.019</td><td>-0.041</td><td>2505.3</td><td>0.000</td></tr>
</table>

Notes: Both ACF and PACF of model AR(2) have strong spikes at multiples of 24 and multiples of 168). Thus the model renewed by adding sar(24), sar(168), sma(24) and sma(168)components. ACF and PACF have still spikes at multiples of 24 but not strong as before. We then try several combinations of SARIMA model with periodicity of 24 and 168 lags.

Table 17: Comparison of Some SARIMA Models

SARIMA	SBIC				
	Pure	Heterogeneous			
		GARCH	EGARCH	GARCH-M	EGARCH-M
$(1,0,1)x(1,0,1)_{24}x(1,0,1)_{168}$	-5.472	-5.878	-5.859	-	-5.859
$(2,0,0)x(1,0,1)_{24}x(1,0,1)_{168}$	-5.475	-5.874	-5.869	-	-
$(2,0,1)x(1,0,1)_{24}x(1,0,1)_{168}$	-5.472	-5.885	-5.884	-	-5.884
$(2,0,0)x(1,0,2)_{24}x(1,0,1)_{168}$	-5.509	-5.918	-5.909	-	-
$(2,0,0)x(1,0,3)_{24}x(1,0,1)_{168}$	-5.533	-5.937	-5.931	-	-
$(2,0,0)x(1,0,4)_{24}x(1,0,1)_{168}$	-5.533	-5.954*	-5.952	-	-
$(2,0,0)x(2,0,1)_{24}x(1,0,1)_{168}$	-5.503	-5.897	-5.878	-	-
$(2,0,0)x(2,0,2)_{24}x(1,0,1)_{168}$	-5.515	-5.918	-5.908	-	-
$(2,0,0)x(1,0,2)_{24}x(1,0,2)_{168}$	-5.472	-5.875	-5.867	-	-5.867
$(0,1,1)x(1,0,1)_{24}x(1,0,1)_{168}$	-5.519	-5.936	-5.926	-5.955	-5.933
$(0,1,1)x(2,0,1)_{24}x(1,0,1)_{168}$	-5.541	-5.966	-5.956	5.954	-5.942
SARIMA	**MAPE**				
	Pure	Heterogeneous			
		GARCH	EGARCH	GARCH-M	EGARCH-M
$(1,0,1)x(1,0,1)_{24}x(1,0,1)_{168}$	0.076	0.106	0.069	-	0.057
$(2,0,0)x(1,0,1)_{24}x(1,0,1)_{168}$	0.077	0.091	0.114	-	-
$(2,0,0)x(1,0,1)_{24}x(1,0,1)_{168}$	0.075	0.097	0.137	-	0.155
$(2,0,0)x(1,0,2)_{24}x(1,0,1)_{168}$	0.077	0.093	0.110	-	-
$(2,0,0)x(1,0,3)_{24}x(1,0,1)_{168}$	0.079	0.082	0.103	-	-
$(2,0,0)x(1,0,4)_{24}x(1,0,1)_{168}$	0.079	0.077	0.095	-	-
$(2,0,0)x(2,0,1)_{24}x(1,0,1)_{168}$	0.172	0.198	0.210	-	-
$(2,0,0)x(2,0,2)_{24}x(1,0,1)_{168}$	0.119	0.162	0.172	-	-
$(2,0,0)x(1,0,2)_{24}x(1,0,2)_{168}$	0.065	0.183	0.142	-	0.189
$(0,1,1)x(1,0,1)_{24}x(1,0,1)_{168}$	0.108	0.349	0.363	0.196	0.218
$(0,1,1)x(2,0,1)_{24}x(1,0,1)_{168}$	0.105	0.232	0.599	0.709	0.582

Notes: MAPEs compared for log transformed load forecasts and for24 hours forecast horizon. The model with best MAPE has been chosen between the models that has close SBIC.

A.2 Test Results for Heterogeneous Periodic AR Model

TABLE 18: SBIC for Number of Harmonics in PAR Model

Harmonics	SBIC
$(h_1, h_2) = (0,0)$	-3.547
$(h_1, h_2) = (1,1)$	-3.903
$(h_1, h_2) = (2,2)$	-4.190
$(h_1, h_2) = (3,3)$	-4.385
$(h_1, h_2) = (4,4)$	-4.461
$(h_1, h_2) = (5,5)$	-4.524

Notes: Number of harmonics of Fourier transformation is selected by Schwarz Information Criteria. However, after the heteroskedasticity test, appropriate ARCH effect is tried on the model and the insignificant Fourier harmonics have been removed from the model until all the variables are significant at %5. Further the MAPEs are compared to determine the final model.

TABLE 19: Comparison of MAPEs for Different Modeling of Heterogeneity in PAR Model

	HETEROGENEITY			
	GARCH	GARCH-M	EGARCH	EGARCH-M
MAPE	0.1669	0.1667	0.1712	0.1686

Notes: MAPEs are compared for log transformed load forecasts and for 24 hours forecast horizon. Smallest MAPE is PAR model with GARCH-M (1,1)

A.3 Test Results for Multiple Regime Logistic STAR Model

TABLE 20: MAPE Comparisons of the Model with Different Transition Variables

d	1	2	3
MAPE	3.28	4.15	4.93

Notes: d denotes the lag order of transition variable.

A.3.1 LM Type Heteroskedasticity Robust Linearity Test

Teräsvirta approximated the transition function by a third order Taylor expansion around the null hypothesis of $\gamma = 0$. The final auxiliary regression after merging terms and parameterization the approximation, the basis for the test is as follows:

$$y_t = \beta_0' w_t + \sum_{j=1}^{3} \beta_j' w_t s_t^j + \varepsilon_t{}^*, t = 1, \dots, T$$

where $\varepsilon_t{}^* = \varepsilon_t + R_3(\gamma,\ c, s_t)\psi_j' z_t$, $R_3(\gamma,\ c, s_t)\psi_j' w_t$ is the remainder

The null hypothesis is $\beta_1 = \beta_2 = \beta_3 = 0$

$\varepsilon_t{}^* = \varepsilon_t$ is under the null hypothesis so that the, the remainder does not affect the asymptotic distribution theory if an LM type test is used.

In the testing procedure, firstly the estimation of linear model under H_0 is done and the residuals e_t and the residual of sum of squares SSR_0 are computed. Secondly, residuals are regressed on w and $h_t = (w_t' s_t, w_t' s_t^2, w_t' s_t^3)$ and the residuals and residual sum of squares SSR_1 are computed. Finally asymptotic test statistic is computed as:

$$LM_{\chi^2} = T\ \frac{SSR_0 - SSR_1}{SSR_0}$$

or the F-version

$$LM_F = T \frac{(SSR_0 - SSR_1)/n}{SSR_1/\{T - (k + p + 1) - n\}}$$

The difference between LM test and robust LM test is the computation of SSR_1. In the latter one, same as the first one, the estimation of linear model under H_0 is done and the residuals and the residual of sum of squares SSR_0 are computed. Then, h_t is regressed on w and the n-dimensional residual vectors $r_t, t = 1, \dots, T$ computed and then $SSR_0$is regressed on $e_t r_t$, residuals and residual sum of squares are computed SSR_1.

A.3.2 LM Type Remaining Nonlinearity Test

Consider a STAR model with M limiting regimes:

$$y_t = \beta_0' w_t + \sum_{m=1}^{M} \beta_m' w_t f(s_t; \gamma_m, c_m) + \varepsilon_t$$

Null Hypothesis: $\gamma_m = 0$

Alternative Hypothesis: $\gamma_m > 0$

The model is not identified under the null hypothesis. Thus we need to follow Teräsvirta (1994) and solve the identification problem as in linearity test introduced by approximating the transition function by a third order Taylor expansion around the null hypothesis of $\gamma_m = 0$ such as:

$$y_t = \breve{\beta}\, w_t + \sum_{m=1}^{M-1} \beta'_m\, w_t f(s_t;\ \gamma_m, c_m) = +\alpha'_1 w_t s_t + \alpha'_2 w_t s_t^2 + \alpha'_3 w_t s_t^3 + \varepsilon_t^*$$

where $\varepsilon_t^* = \varepsilon_t^* + R(s_t; \gamma_M, c_M)$ is the remainder.

APPENDIX B

TEZ FOTOKOPİSİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : ÖZPALA

Adı : PINAR

Bölümü : İKTİSAT

TEZİN ADI (İngilizce) : STATISTICAL MODELING OF HOURLY ELECTRICITY LOAD SERIES IN TURKEY

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☐

1. Tezimin tamamından kaynak gösterilmek şartıyla fotokopi alınabilir. ☐
2. Tezimin içindekiler sayfası, özet, indeks sayfalarından ve/veya bir bölümünden kaynak gösterilmek şartıyla fotokopi alınabilir. ☐
3. Tezimden bir bir (1) yıl süreyle fotokopi alınamaz. ☐

TEZİN KÜTÜPHANEYE TESLİM TARİHİ: