ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

FORECASTING ELECTRICITY PRICES IN TURKEY: A COMPARISON OF CLASSICAL ECONOMETRICS AND MACHINE LEARNING TECHNIQUES

M.A. THESIS

Yunus Emre KARAGÜLLE

Department of Economics

Economics M.A. Program

JUNE 2018

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

FORECASTING ELECTRICITY PRICES IN TURKEY: A COMPARISON OF CLASSICAL ECONOMETRICS AND MACHINE LEARNING TECHNIQUES

M.A. THESIS

Yunus Emre KARAGÜLLE
(412161023)

Department of Economics

Economics M.A. Program

Thesis Advisor: Prof. Dr. Bülent GÜLOĞLU

JUNE 2018

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ SOSYAL BİLİMLER ENSTİTÜSÜ

TÜRKİYE'DE GÜN ÖNCESİ ELEKTRİK FİYATLARI TAHMİNİ: KLASİK EKONOMETRİ VE MAKİNE ÖĞRENME TEKNİKLERİNİN KARŞILAŞTIRMASI

YÜKSEK LİSANS TEZİ

Yunus Emre KARAGÜLLE
(412161023)

İktisat Anabilim Dalı

İktisat Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Bülent GÜLOĞLU

HAZİRAN 2018

Yunus Emre KARAGÜLLE, a M.A. student of ITU Graduate School of Arts and Social Sciences student ID 412161023 , successfully defended the thesis/dissertation entitled "**FORECASTING ELECTRICITY PRICES IN TURKEY: A COMPARISON OF CLASSICAL ECONOMETRICS AND MACHINE LEARNING TECHNIQUES**", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Bülent GÜLOĞLU** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Sencer ECER** Istanbul Technical University	....
	Prof. Dr. Nilgün ÇİL Istanbul University	....

Date of Submission : 2 May 2018
Date of Defense : 8 June 2018

To my family,

FOREWORD

First of all, I would like to thank my advisor Prof. Dr. Bülent Güloğlu for his contributions to my work.

I am also greateful to my friend Pınar Şenol who supported and guided me for the last two years.

I would like to thank my director Assoc. Prof. Dr. Nihat Alayoğlu whose provided convenience to me throughout the thesis stage.

Lastly, I am very thankful to my family for supporting and encouraging me to go after my goals.

MAY 2018 Yunus Emre KARAGÜLLE

TABLE OF CONTENTS

ABBREVIATIONS

MAE	: Mean Absolute Error
AIC	: Akaike Information Criteria
ANN	: Artificial Neural Network
APE	: Absolute Percentage Error
App	: Appendix
AR	: Autoregressive
ARX	: Autoregressive with exogenous
ARMA	: Autoregressive Moving Average
ARMAX	: Autoregressive Moving Average with Exogenous
DAM	: Day-Ahead Market
EXIST	: Energy Exchange Istanbul
MA	: Moving Average
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
NARX	: Non-linear Autoregressive Network with Exogenous Variables
RMSE	: Root Mean Square Error
TEM	: Turkish Electricity Market
TETAS	: Turkish Electricity Trading and Contracting
WMAE	: Weakly Mean Absolute Error

SYMBOLS

$\boldsymbol{\alpha}$	: Coefficient of AR process
$\boldsymbol{\beta}$	: Coefficient of MA process
$\mathbf{y_t}$	: Electricity Market Clearing Price
$\boldsymbol{\rho}$	: Coefficient of Exogenous Variables
$\boldsymbol{y_i}$	: Actual Price
$\boldsymbol{f_i}$	: Forecasted Price
$\mathbf{T}$	: Number of Forecasted Period

LIST OF TABLES

Page

LIST OF FIGURES

Page

FORECASTING ELECTRICITY PRICES IN TURKEY: A COMPARISON OF CLASSICAL ECONOMETRICS AND MACHINE LEARNING TECHNIQUES

SUMMARY

Liberalization of electricity markets trends has been widening in the world since 1990s and the process is still continuing in many countries. As a result of liberalization, the private retailer and distribution companies role has increased in the markets. Like other countries, government enterprises was operating the Turkish Electricity market before 2000s. However, Turkish Government enacted a electricity market law for deregulation of the market in 2001. Accordingly, Turkish Electricity Trading and Contracting (TETAS) was established in 2001 with the aim of sustaining the energy trade and reach a competitive electricity market. Afterwards, Day-Ahead market and Intra-Day markets are established in the following years. As a result of the liberalization process, forecasting electricity prices has become an important requirement in energy companies' decision-making process to maximize profits by offering competitive bid and ask rates.

On the contrary to other time series, electiricity prices have complex structures and dynamic determinants which makes difficult to forecast prices. For examle, one of the determinants of the electricity is demand and it is known that demand is highly related with temperatures. In summer, electricity demand expeced to increase due to cooling needs. However, it may change accordingly with tempetures. Futhermore, electiricty demand is also changing according to hour of the day or day of the week and so on. Another thing is that, generation source and type and the amount may also affect the prices. Consequently, electiricty prices are generally representing complex non-linear structures.

There are many methods for forecasting electricity prices. However, artificial neural networks are known to be good forecasting method for covering non-linear effect of the prices. On the other hand, time series methods are easy to forecast and interpret the relationships. Lastly, quantile regressions are capturing the different conditional distributions and may be useful for capturing spkies. Therefore, artificial neural networks, time series and quantile regression is used.

First, non-linear autoregressive network with exogenous inputs (NARX) is used with Levenberg-Marquardt learning algorithm. It is a two-layer feedforward network with sigmoid activation function. It is known that load is one of the most important factor affecting the prices, thus it is taken as inputs. Furthermore, to take seasonality in account, hour of the day, day of the week, season of the year, previous prices and other inputs are considered. The next day's prices are forecasted in out-of-sample data with multi-step ahead method and rolling window mechanism.

Second, autoressive moving average (ARMA) models with different AR(p) and MA(q) orders are applied. To estimate a model and forecasts series in the model, series should be stationary and stable. Ng-Perron test is applied for testing whether series is stationary. Price series is found to be stationary. Next day's prices are forecasted in

out-of-sample data with multi-step ahead method and rolling window mechanism. Furhermore, Optimal (p,q) are chosen after trial and error proces.

Thirdly, ARMA models with exogenous variables are applied to see whether exogenous variables are affacting the performance of the autoregressive model. As in the ARMA model, series should be stationary. As in the NARX model, load is taken as inputs. Besides, hour of the day, day of the week and other variables are considered. Next day's prices are forecasted in out-of-sample data with multi-step ahead method and rolling window mechanism.

Finally, quantile regression is applied. Different comibations of exogenous variables applied for different quantiles. Next day's prices are forecasted separetly for each quantile with multi-step ahead forecast methods and rolling windows mechanism.

All methods with different combinations of the models are evaluted with mean absolute percentage error(MAPE) in out-of-sample data. According to findings, the NARX method outperformed the other models and forecast the electricity prices in TEM with lowest errors.

TÜRKİYE'DE GÜN ÖNCESİ ELEKTRİK FİYATLARI TAHMİNİ: KLASİK EKONOMETRİ VE MAKİNE ÖĞRENME TEKNİKLERİNİN KARŞILAŞTIRILMASI

ÖZET

Elektrik piyasalarının serbestleşmesi 1990'lı yıllardan bu yana dünyada yayılmaktadır ve bu süreç birçok ülkede devam etmektedir. Ayrıca serbestleşmenin bir sonucu olarak, piyasalarda perakendeci ve dağıtım şirketlerinin rolü her geçen gün artmaktadır. Diğer ülkelerde olduğu gibi, 2000'lerden önce Türkiye Elektrik Piyasası devlet iktisadi teşebbüsleri tarafından işletilmekteydi. Ancak, Türkiye Cumhuriyeti 2001 yılında piyasaların serbestleştirilmesi için bir elektrik piyasası kanunu çıkarmıştır. Bu doğrultuda, 2001 yılında enerji ticaretini sürdürmek ve rekabetçi bir elektrik piyasasına ulaşmak amacıyla Türkiye Elektrik Ticareti ve Taahhüt A.Ş. (TETAS) kurulmuştur. Daha sonraki yıllarda, Gün Öncesi Piyasa ve Gün İçi Piyasa hayata geçirilmiştir. Serbestleşme sürecinin sonucu olarak, elektrik fiyatları tahmini, rekabetçi alış ve satış teklifleri vererek karlarını maksimize etmek için enerji şirketlerinin karar verme sürecinde önemli bir gereklilik haline gelmiştir.

Diğer zaman serilerinin aksine, elektirik fiyatları tahmin edilebilmesi güç, karmaşık ve dinamik bir yapıya sahiptirler. Örneğin, elektriğin en önemli belirleyicilerinden biri taleptir. Talebin de sıcaklıklar ile doğrudan ilişkili olduğu bilinmektedir. Yaz aylarında, elektrik ihtiyacının soğutma ihtiyaçları nedeniyle artması ve kış aylarında ısınma ve aydınlatmaya olan ihtiyaçtan dolayı artması beklenmektedir. Fakat bu talep artışı sıcaklığa ve zamana bağlı olarak değişebilmektedir, sıcaklık her yaz döneminde farklı olabilmektedir. Bunun yanı sıra, elektrik üretiminde kullanılan kaynaklar ve üretim maliyetleri de elektrik fiyatlarını etkileyebilmektedir. Türkiye'de üretilen elektriğin büyük bir kısmı doğalgaz, ithal kömür ve hidrolik kaynaklıdır. Özellikle doğalgaz ve ithal kömür, üretim maliyetini arttırıcı bir unsura sahiptir. Sonuç olarak, elektrik fiyatları genellikle karmaşık ve doğrusal olmayan bir yapıya sahiptir.

Elektrik fiyatlarının tahmini için birçok yöntem bulunmaktadır. Bu çalışmada, yapay sinir ağları, zaman serileri ve quantile regresyon yöntemleri kullanılmıştır. Yapay sinir ağlarının, fiyatların doğrusal olmayan etkisini kavrayabilmek için iyi bir tahmin yöntemi olduğu bilinmektedir. Öte yandan, zaman serisi yöntemleri ile fiyatları tahmin etmek ve yorumlamak kolaydır. Son olarak, quantile regresyon koşullu farklı dağılımları yakalamak için yararlı olabilmektedir.

Çalışmada kullanılan ilk yöntem, dışsal girdilerle doğrusal olmayan otoregresif ağ (NARX), Levenberg-Marquardt öğrenme algoritması ile kullanılmıştır. NARX Sigmoid Aktivasyon Fonksiyonu ile iki gizli katmanlı ve bir sonuç katmanlı ileri besleme ağıdır. Elektrik yükünün fiyatları etkileyen en önemli faktörlerden biri olduğu bilinmektedir, bu sebeple yük miktarı dışsal girdi olarak modele eklenmiştir. Ayrıca mevsimselliği dikkate almak için günün saati, haftanın günü, yılın mevsimi, tarife kukla değişkenleri ve geçmiş fiyatlar dikkate alınmıştır. Bir sonraki günün fiyatları, çoklu adım yöntemi ve rolling window mekanizmasıyla örneklem dışı verilerde

tahmin edilmiştir. Öyle ki, iki yıllık verininin son haftası modele dahil edilmemiştir. Daha sonrasında, bir sonraki günün fiyatları, yani saati temsilen 24 farklı fiyat tek bir model ile tahmin edilmiştir. Daha sonrasında rolling window sistemi uygulanarak, yani örneklem 24 saat arttırılarak bir sonraki günün fiyatları tahmin edilmiştir.

Çalışmanın ikinci yöntemi, otoregresif haraketli ortalamalar modeli (ARMA) farklı AR(p) and MA(q) dereceleriyle uygulanmıştır. Zaman serilerinde tahminleme yapabilmek için, serilerin durağan olması gerekmektedir. Bu sebeple, elektrik fiyatlarının durağan olup olmadığını ölçmek için Ng-Perron durağanlık testi uygulanmış ve serinin durağan olduğu sonucuna ulaşılmıştır. Seri düzeyde durağan olduğu için ARMA modeli ile devam edilmiştir. Bir sonraki günün fiyatları, çoklu adım yöntemi ve rolling window mekanizmasıyla örneklem dışı verilerde tahmin edilmiştir.

Üçüncü yöntemde ise, dışsal değişkenlerin otoregresif modelin performansını etkileyip etkilemediğini görmek için dışsal değişkenlere sahip ARMA modelleri (ARMAX) uygulanmıştır. ARMA modelinde olduğu gibi, ARMAX modelinde de durağanlık koşulunun sağlanması gerekmektedir. Elektrik fiyatları serisinin Ng-Perron düzeyde durağan olduğu sonucuna ulaşılmıştır. NARX modelinde olduğu gibi, elektrik yükü dışşal değişken olarak modele eklenmiştir. Ayrıca mevsimselliği dikkate almak için günün saati, haftanın günü, geçmiş fiyatlar ve benzeri değişkenler modellerde kullanmıştır. Bir sonraki günün fiyatları, çoklu adım yöntemi ve rolling window mekanizmasıyla örneklem dışı verilerde tahmin edilmiştir.

Son olarak quantile regresyon yöntemi uygulanmıştır. Daha önceki modellerde olduğu gibi elektrik yükü dışsal değişken olarak alınmıştır. Ayrıca mevsimsellik etkisini dikkate almak için günün saati, çalışma günü, haftasonu, geçmiş fiyatlar ve benzeri değişkenler modele eklenmiştir. Model farklı yüzdelik dilimler için uygulanmıştır. Diğer yöntemlerde olduğu gibi, bir sonraki günü fiyatları çoklu adım yöntemi ve rolling window mekanizması ile örneklem dışı verilerde tahmin edilmiştir.

Farklı kombinasyonları denenen modeller, örneklem dışı veride ortalama mutlak yüzde hata (MAPE) ile karşılaştırılmış ve değerlendirilmiştir. Elde edilen sonuçlara göre, NARX yönetmi ile kurulan Model A, %9.65 MAPE hata oranı ile diğer yöntemlerden daha iyi bir sonuç vermiştir.

1. INTRODUCTION

Deregulations of the electricity markets have been extending since 1990s. Many developed and developing countries have established exchange markets to deregulate the market. As can be seen in Figure 1, first attempts for liberalization of the Turkish Electricity Market (TEM) was started in 2000's. However, the Day-Ahead Market (DAM) system was established in 2011. Now, Day-Ahead Market and Intra-Day Markets are operated by Energy Exchange Istanbul (EXIST). In consequence of liberalization process, electricity price forecasting has become an essential for energy companies' decision-making process. A company should forecast Day-Ahead prices and give competitive bid rates so that it can maximize its profit.

Figure 1.1: Liberalization process of Turkish Electricity Market.

However, electricity prices have very complex structures. Specifically, prices are depending on many components and one of them is electricity demand which changes dramatically from time to time. For example, electricity prices expected to increase when the weather is too cold or too hot, that is either in Winter or in Summer term.

Furthermore, when temperatures increasing, demand for electricity increases. On the other hand, production cost decreases with rains in spring, thus price of electricity decreases till the beginning of summer. Thus, electricity prices generally exhibit daily, weekly and sometimes monthly seasonality. On the other hand, electricity is non-storable which attract retailer companies to forecast future demand and meet customer demand without losing profit. Accordingly, electricity prices displaying high spikes due to complex structures of the electricity market.

1.1 Turkish Electricity Market

The electricity market was operated by public economic enterprises before 2000. However, in the beginning of the 2001, Turkey has joined liberalization trend and established Turkish Electricity Trading and Contracting (TETAS). TETAS's mission is sustaining electrical energy trade, reach competitive electricity market and being a transparent company (TETAS, Annual Report, 2016). According to TETAS Annual Report (2015), TETAS was selling and buying 80% of the electricity produced and consumed in Turkey, however it was around 40% in 2015. Table 1 shows the process of liberalization process in the TEM. Public share in the market has been decreasing since 2001. The role of private companies increasing in the TEM in each day.

Table 1: TETAS' share in electricity market.

Year	Generation (%)	Consumption (%)
2002	82	77
2003	81	77
2004	80	78
2005	80	78
2006	69	68
2007	47	46
2008	44	43
2009	43	43
2010	41	41
2011	36	35
2012	35	34
2013	55	53
2014	49	47
2015	43	42
2016	38	38

Source: TETAS Annual Report, 2016

In the TEM, total available production capacity exceeds the total consumption. According to the generation sources of the electricity, as shown in Figure 1.2, natural gas, hydro dam and imported coils are generating most of the electricity.

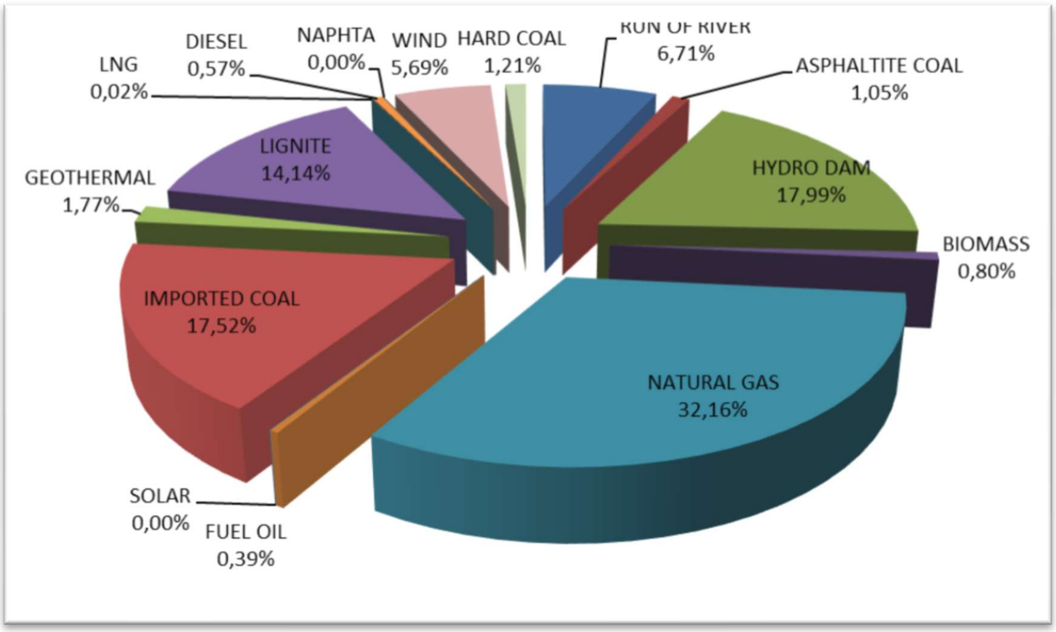

Figure 1.2: Proportionate division of electricity generation by sources in 2016.
Source: EPDK Electricity Market Development Report, 2016

1.2 Importance of The Study

TEM has been liberalized over the last 17 years and the process is continuing. In such a market where share of private companies increases, ambition of market players to maximize their profit is also increasing. As it mentioned before, companies seek to forecast future prices to maximize their profit. There are few studies on TEM, however, those are generally focusing on single forecasting methods and lack of capturing special characteristics of TEM.

In this study, electricity prices are forecasted by using different methods and results are compared to obtain the best forecasting method for TEM. Furthermore, special attention is given to capture TEM's characteristics, so that Day-Ahead electricity prices can be forecasted accurately.

2. LITERATURE REVIEW

Weron (2014) works on electricity price forecasting literature. According to this study, electricity price forecasting has become an important topic in the literature after 2000s because of the deregulation trend in the world. Besides, among different studies in Scopus and Web of Science are analyzed and most of them used either only neural networks, or only time series, or both. Electricity models are divided into five categories; multi-agent models, fundamental models, reduced-form models, statistical models, and computational intelligence models. Furthermore, time series models are classified in statistical models and neural networks are classified in computational intelligence models.

Even though statistical methods may not be good at modelling non-linear behavior of the electricity prices and performs poorly in the spiky periods, they can be useful in the non-spiky periods (Weron, 2014). Furthermore, computational intelligence models, like neural networks, are better to capture complexity and non-linearity of the electricity prices (Weron, 2014).

According to Weron (2014), the evaluation technics like absolute percentage error (APE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean absolute scaled error (MSE), and root mean square error (RMSE) have been used in the forecast evaluations, however, MAPE is the most common error measurement.

Singhal and Swarup (2011) conduct a multi-layer artificial neural networks (ANN) with back propagation algorithm by using half hourly prices and made multi-step ahead forecast, i.e., 48 hours. The ANN, which has one output layer and two hidden layers, where first layer includes 10 neurons and second layer include 5 neurons, is used. To take seasonality in account, previous prices of the electricity prices are used. Furthermore, the model includes historical demands, forecasted demands and time indices as inputs. Furthermore, day of the week, hour of the day, demand, marginal change in demand, price of yesterday, price of one week ago, price of two weeks ago, price of three weeks ago and price of three weeks ago are used as inputs in the model. Six months of the data is used for training the networks and different days for different months are tested. Market clearing prices are forecasted in the in-sample data and compared by using MAE and RMSE. The results are analyzed under three circumstances which are price with normal trend, small spike and large spike. According to findings, the model is useful in the normal trend regimes. However, it is not good small spike or large spike regimes.

Keles et al. (2016) examines the extended forecast methods in the European Power Exchange Market. The hourly data contains 2 years from 2011 to 2013. The study is based on the ANN with back propagation algorithm. However, the models are developed by two different perspectives. First, there is no certain fact whether deseasonalizing and detrending the time series in ANN models (Keles at al, 2016). Thus, the ANN networks are modelled and compared with and without trend and seasonality. Second, hourly loads, power generation from renewable energy, daily available capacities by fuel type, fuel prices and public holidays, day of the week, hour of the day, month of the year and previous prices are specified as fundamental factors determining electricity prices. Among many variables, k-Nearest Neighbor, a filter-based method is used to specify most relevant inputs in the model. The structure of the ANN model is obtained by trial and error method. Three layers feedforward ANN with one output neuron is applied to make multi-step ahead forecast. Besides, ARIMA (1,1,1) is used as a benchmark models to compare with the ANN models. The accuracy of the model is evaluated by comparing RMSE in the out-of-sample data. According to results, the ANN models with selected inputs performed better than the seasonal ARIMA.

Panapakidis and Dagoumas (2016) measures the pure ANN models and hybrid models with using no pre-processed data in Southern Italy Electricity Market and covers the period from 2012 to 2015. The data is clustered in small homogenous groups to see whether it is increasing performance of the simple ANN models. One-step ahead forecast method is used to forecast prices. Therefore, same forecast is applied 24 times for different hours of the day. In all models, the most relative inputs are determined after trial and error method. The sole ANN models include day of the week, hour of the day, holiday dummy, hourly prices of yesterday, hourly prices of last week and so on. These variables are also used in the same model with renewable energy resources (RES) generation and natural gas prices, previous prices of loads and RES generation, predicted loads and RES generation. On the other hand, a hybrid model consist of unsupervised and supervised machine learning stages is applied to increase performance of the simple ANN by clustering similar patterns in small training sets. The method is applied at two stages. First the data is clustered to k different clusters. Second, the normal ANN process is applied to the clustered training sets and output is obtained. The VAR, ARMA, and GARCH models are also used and compared with the proposed model. Models are evaluated by APE, MAPE, MAE and Theil U statistic. The results show that even though the ANN is good at capturing non-linear structure of the electricity prices, performance with no pre-processed data is lower than the pre-processed data. However, it performs better than the other models.

Gareta et al. (2006) considers electricity price forecasting by feedforward neural networks. According to the study, electricity prices highly depend on demand and electricity demand represents seasonality. Therefore, day of the week, month of the year indicators are taken as inputs. Furthermore, it is stated that electricity prices are determined by different situations presented by the market. Hence, maximum, medium and minimum prices with some previous prices are taken into the model. Next day's prices are forecasted with multi-step ahead method i.e., 24 outputs in the output neuron. To avoid overfitting problem, number of hidden neurons are chosen after trial and error process. The results are evaluated by comparing the difference between actual and the forecasted prices in-sample and out-of-sample data. It is stated that neural networks have ability to cover complex non-linear structure of electricity prices and can be used to make accurate forecasts.

Gökgöz and Filiz (2016) forecasts the electricity prices in the TEM by feedforward artificial neural networks with Sigmoid Function. Gradient Descent, Gradient Descent Momentum, Levenberg-Marquardt and BFGS learning algorithms are used. The data is taken from EXIST and covers periods from 2012 to end of the 2014. In the ANN model, temperature is used as inputs which is weighted by the electricity consumption amount of each city. Besides, the natural gas prices are added to the model, however, since lack of gas market and the data in Turkey, the US Henry Hub Natural Gas Spot prices are used. Other variables in the models are; hour of the day, day of the week, load, previous prices and previous loads. Different combinations of the neurons are tested. TEM's prices are forecasted with multi-step ahead forecasting method. Model performances are evaluated in-sample data by MAPE. Levenberg-Marquardt and

BFGS learning algorithms are found to performed better than other algorithms with MAPE score of 9.76%.

Marcjasz et al. (2018) measures the non-linear autoregressive network with exogenous inputs (NARX) with long term seasonal component. The data includes two set, Global Energy Forecasting Competition 2014 (GEFCom2014) from 2011 to 2013 and the Nordic Pool from 2013 to 2015. The long term seasonal component has been ignored in the electricity price forecasting literature until recently (Weron, 2014). Therefore, the seasonal component autoregressive framework with different methods are performed. Prices are forecasted simultaneously with one step-ahead forecasting method and 24 different predictions for each hour. The seasonal component ARX, ANN, seasonal component ANN models are applied. In the ARX model similar day technics is used. According to this technic, the forecast for hour h on Monday is very similar to last Monday's prices and this same for the Saturdays and Sundays. Besides, price of Friday is very similar to Thursdays and same is applied for Tuesdays, Wednesdays, Thursdays. MATLAB's NARX structure with one hidden layer and one output layer and the Levenberg-Marquardt algorithm with Sigmoid Transfer Function is used. The performances are tested with weakly mean absolute error (WMAE) loss function and models are compared with Diebold-Mariano tests. According to findings, performance of the NARX is improved with seasonal components to the model.

Mirakyan et al. (2017) establishes composite models for forecasting electricity prices. The ANN, support vector regression and rigid regression are used. The daily data covers the period starting from 2012 to 2015 in European Power Energy Exchange. The purpose of the study is to test whether composite forecasts are performing better than the individual models. The composite forecast is used to choose either an individual model or combine forecast from the models. Results are evaluated in out-of-sample data with RMSE, MAPE and median APE. Consequently, it is argued that the composite forecast improves the performance of individual models.

Nogales et al. (2002) evaluate two different time series models, dynamic regression and transfer functions to forecast electricity prices for Spanish Market and California Electricity Market. According to paper, electricity prices are high frequency data which contains multiple seasonality, calendar effect, high volatility and huge spikes. Therefore, time series models are modeled with exogenous variables to capture these effects. The demand is chosen as inputs because it is believed that demand can partially explain the characteristics of the price. In the dynamic regression previous values of prices and demands are included in the model. However, in the transfer function, which has same structure with autoregressive moving average with exogenous variables (ARMAX), past prices and demand are taken as inputs and both series are assumed to be stationary as in the ARMA model. The parameter estimation is based on maximum likelihood function. Next day's prices are forecasted with multi-step ahead forecast and the results are evaluated by mean week error and forecast mean square error. It is concluded that transfer functions are performed better than dynamic regression.

Girish (2016) forecasts electricity prices in Indian electricity market using autoregressive- GARCH models. The hourly data covers periods from 2010 to 2013. One-step ahead forecast is applied with rolling window method to forecast prices in the out-of-sample data. It is mentioned that demand is more than the supply in the Indian market which makes it different from other electricity markets. The pre-processed data is used to develop ARIMA, ARIMA-GARCH, ARIMA-EGARCH and ARIMA-PARCH models. Seasonality is taken in account by using hourly, weekly and daily dummies. The models are evaluated MAE, MAPE, RMSE and Theil Inequality coefficient technics. ARIMA-GARCH, ARIMA-EGARCH and ARIMA-PARCH are the best models for west region, southern region and north/north-eastern regions respectively.

Ziel and Weron (2018) compares the univariate and multivariate modelling performance of electricity price forecasting. According to the study, multivariate frameworks forecasting electricity prices by applying 24 different models with one-step ahead forecast. However, univariate framework is forecasting next day's prices with one single model and 24 multi-steps ahead forecast. In the univariate model, multi-step ahead forecasting is applied with recursive scheme, that is, after forecasting price of y_t, it is taken as inputs to forecast next hour's prices y_{t+1}. The AR, VAR, Lasso models are applied both in univariate and multivariate framework and exogeneous variables are not included in the models. However, seasonal dummies i.e., hour of the day, day of the week, hour of the week and weakly mean of hourly frequency, the daily mean of hourly frequency and the weekly mean of daily frequency is taken as inputs. Models are evaluated with MAE and mean percentage deviation from the best model.

According to findings, multivariate model performance is not outperformed the univariate model performance.

3. DATA

Data set is taken from EXIST transparency platform. It contains 2 years, starting from 18 December 2015 to 31 December 2017. The data contains market clearing price (MCP), load for 17880 hours. MCP for 2 years can be seen in Figure 3.

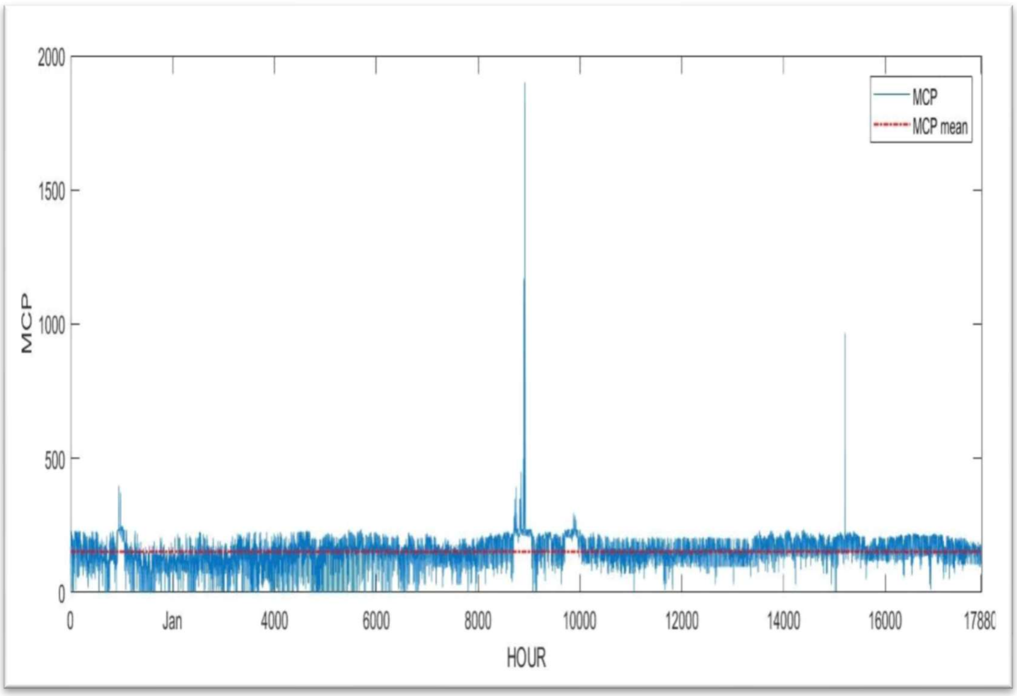

Figure 3: Market clearing price and average market clearing price.

Figure 3 shows that MCP were generally between 100-200 TL/MWh. There are a few unusual prices. The maximum price observed in 2016 is observed in 23 December at 2 pm with value 1899.99 TL/MWh, however the maximum prices in 2017 is observed in 9 November at 2pm. On the other hand, minimum price observed in the period is 0/MWh. MCP has mean 152.30 TL/MWh, drawn red in the figure. Besides, standard deviation of 57.16 TL/MWh for 2 years period.

4. METHODOLOGY

There are plenty of methods to forecast electricity prices. However, there is no unique way for forecasting prices. Electricity markets are distinctive from each other and needs specific methods for different markets. In the electricity price forecasting literature, neural networks and time series are the most used methods. The neural networks are generally performing better for capturing non-linear effects of the series. On the other hand, time series are easy to conduct and implement the results.

4.1 Non-linear Autoregressive Network with Exogenous Inputs

Non-linear autoregressive network with exogenous inputs (NARX) is one of type of the recurrent neural networks. A typical NARX is two-layer feedforward network with sigmoid activation functions. The network has dynamic structures because when the network completes calculations and obtain output, it feed back to inputs of the networks and modified the errors. Figure 4.1 shows the process of NARX network.

$$y(t) = f\begin{pmatrix} y(t-1), y(t-2), \dots y\left(t-n_y\right), x(t-1), \\ x(t-2), \dots, x(t-n_x) \end{pmatrix} \tag{4.1}$$

Equation 4.1 represents the mathematical form of Figure 4.1. In the equation, dependent variable y(t) at time t regressed on passed values of exogenous variables denoted by x(t). When the output is calculated, network feed back to inputs to minimize the errors.

There are no stylized facts about the configuration of the model for a specific electricity market. Weron (2014) mentioned that there is no rule for choosing minimum set of most effective input variables, however, one global rule can be found for specific markets. Accordingly, there are two models left after trial and error process.

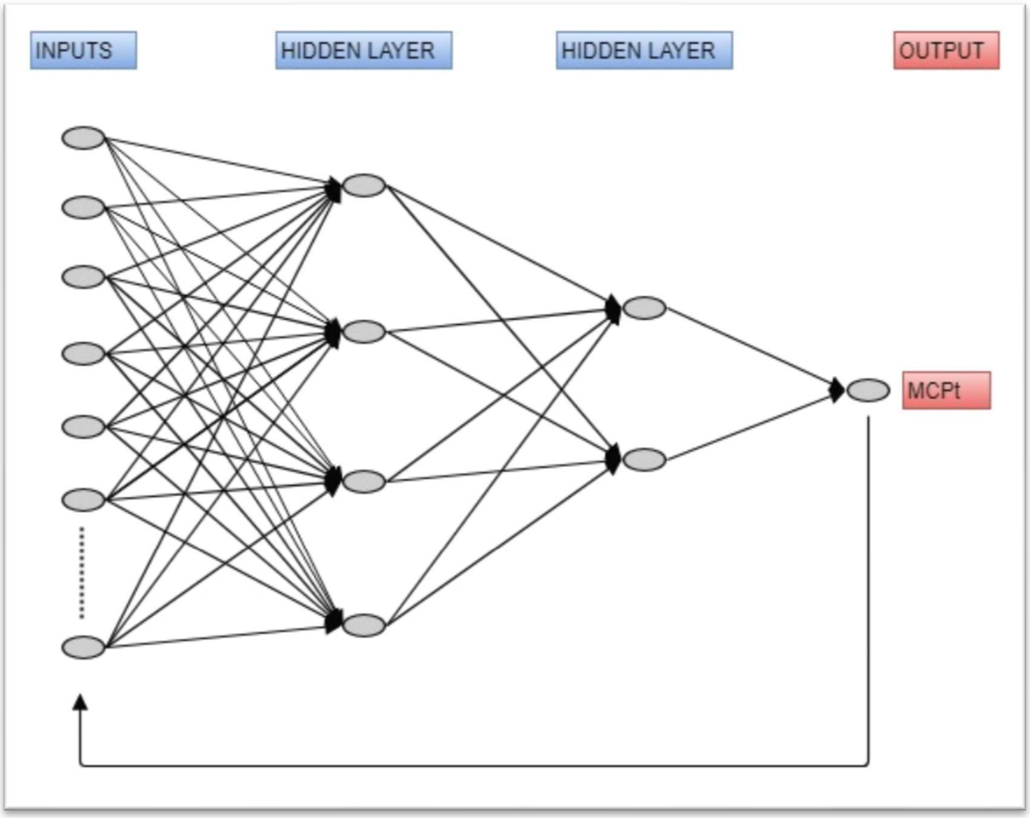

Figure 4.1: NARX network structure.

4.1.1 Model A

The model has two-layer feedforward neural networks with 3 neurons in the hidden layers. Besides, Levenberg-Marquardt learning algorithm with Sigmoid Function is used.

Electricity is non-storable goods, therefore amount of electricity produced for a given period must be consumed. If it is not the case, campanies needs to offer for loading or non-loading in the market. Besides, companies have to supply electricity to costumer as much as they demanded. Consequently electricity load is a very important exogenous variable used in the model.

It is well known that electricity prices generally presents seasonalities. The optimum lag is determined by checking partial autocorrelation and autocorrelation of the prices. Accordingly, number of delay is chosen 24 in the model.

There are three different tariff periods with regards to consumption levels and companies sells electricity to customer at different levels in these tariff periods if they are not eligible customer. These are, Day,Peak and Night hours tariffs. Even though the prices are same for all costumers whose buying electricity from energy companies, these tariff hours are showing the general demand structure of the market. Therefore, night hours-tariff 3 and daytime hours-tariff 1 added as a dummy variable. On the other hand, increasing needs for heating and lighting in winter, increases demand and increase the average prices in winter and it is known that highest prices occurred in december in the data. Thus, dummy for november, december and january is taken as winter dummy. Moreover, day of the week, hour of the day, public holidays are taken as inputs. However, since forecasted period includes last day of the year and it presents different consumption behaviors than other holidays and casual days, dummy for 31 December is also taken. All variables are Figure 4.2 is given to summarize and visiualize the network structure.

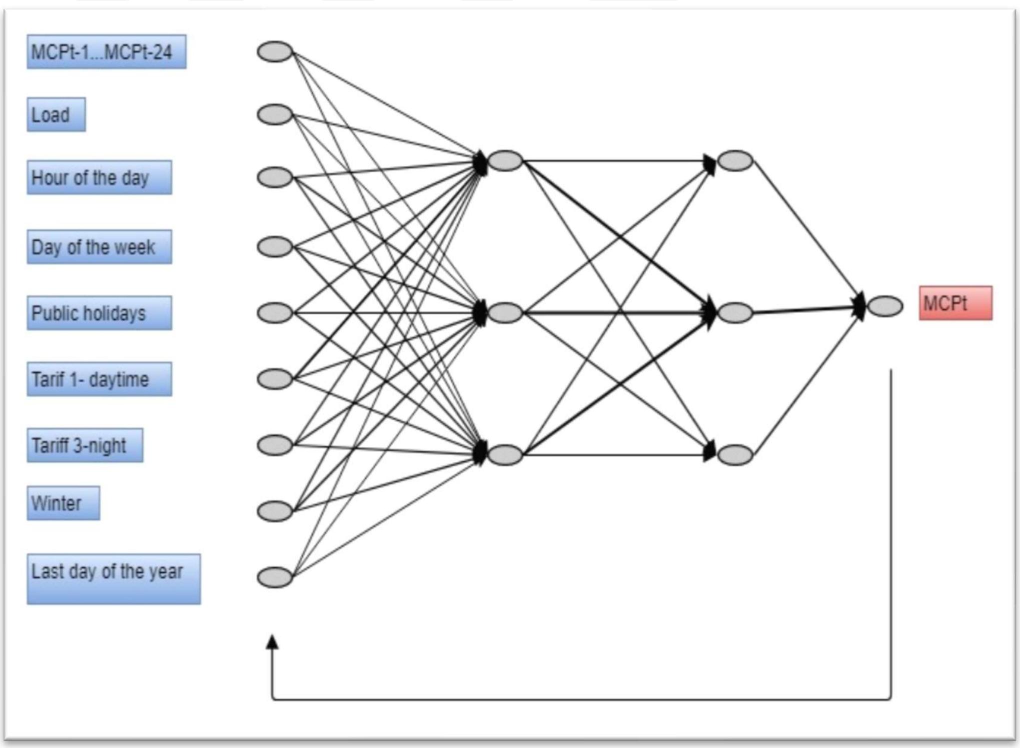

Figure 4.2: Model A.

4.1.2 Model B

The model B and the model A has few different variables, and model B is simple version of Model A. Like in the model A, electricity load is taken as exogenous variable and the number of delay is 24. Additionally, day of the week, hour of the day and public holidays are also taken.

However, in this model last day of the year is not included as a dummy in the model and instead of using night hours tariff 3, dummy for peak hours-tariff 2 is added. All variables are given in NARX structure in Figure 4.3.

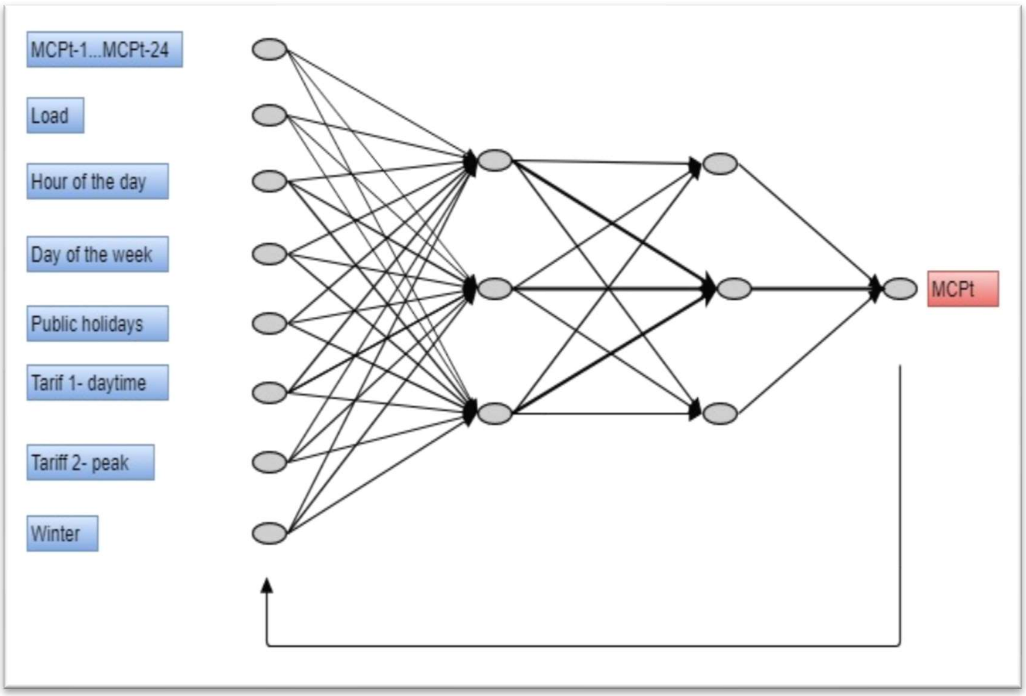

Figure 4.3: Model B.

4.2 Autoregressive Moving Average Model

The autoregressive moving average (ARMA) is composed of two parts, that is autoregressive (AR) and moving average (MA). ARMA considers the random structure and time correlation of the phenomenon (Weron,2014). It is a very common method for forecasting time series. The general form of ARMA with AR of order p and MA of order q is written in equation 4.2.

$$y(t) = \left(\delta + a_1 y_{t-1} + \cdots + a_p y_{t-p} + u_t - \boldsymbol{\beta}_1 u_{t-1} - \cdots - \boldsymbol{\beta}_q u_{t-q}\right) \quad \text{and } u_t \sim WN(o, \sigma_u^2) \tag{4.2}$$

Where u_t is independent white noise series, besides, α's, β's represents AR and MA parameters respectively. If p=0 then the process is reduced to MA and if q=0 than the process is reduced to AR process.

To estimate and forecast ARMA model, series should be stationary and invertible. Therefore, stationarity is tested with Ng-Perron test presented in Table 4.1.

Table 4.1: Ng-Perron test for stationarity.

Null Hypothesis: MCP has unit root
Exogenous: Constant
Bandwidth: 53(Newer-West automatic)
Sample:12/18/2015 - 12/31/2017
Included Observation:17880

	MZa	MZt	MSB	MPT
Ng-Perron Statistics	-4021.400	-44.837	0.011	0.007
Asymptotic Critical Values	-13.800***	-2.580***	0.174***	1.780***
	- 8.100**	-1.980**	0.233**	3.170**
	- 5.700*	-1.620*	0.275*	4.450*
HAC corrected Variance				1606.650

*** 1 %, ** 5 percent and * 10 significance level. Asymtotic critical values from Ng-Perron (2001, Table 1).

In the Ng-Perron test, the null hypothesis implies that series has unit root, i.e. non-stationary. MZa, MZt, MSB and MTB has different distributions. According to Table 4.1, the null hypothesis is rejected for all significance level and different distributions. It implies that the MCP is stationary. Therefore, ARMA model can be used safely.

Stability conditions for AR part are provided. That is, the characteristic equation,

$$\lambda^P - \alpha_1 \lambda^{p-1} - \cdots - \alpha p = 0 \tag{4.3}$$

only has roots with absolute values smaller than one, i.e. roots are inside the unit circle.

On the other hand, stability is related with MA part and it is invertible if the characteristic equation,

$$\lambda^q - \alpha_1 \lambda^{q-1} - \cdots - \alpha q = 0 \quad (4.4)$$

only has roots with absolute values smaller than one, i.e. roots are inside the unit circle.

Steps for ARMA model construction and forecasting are summarized in Figure 4.4.

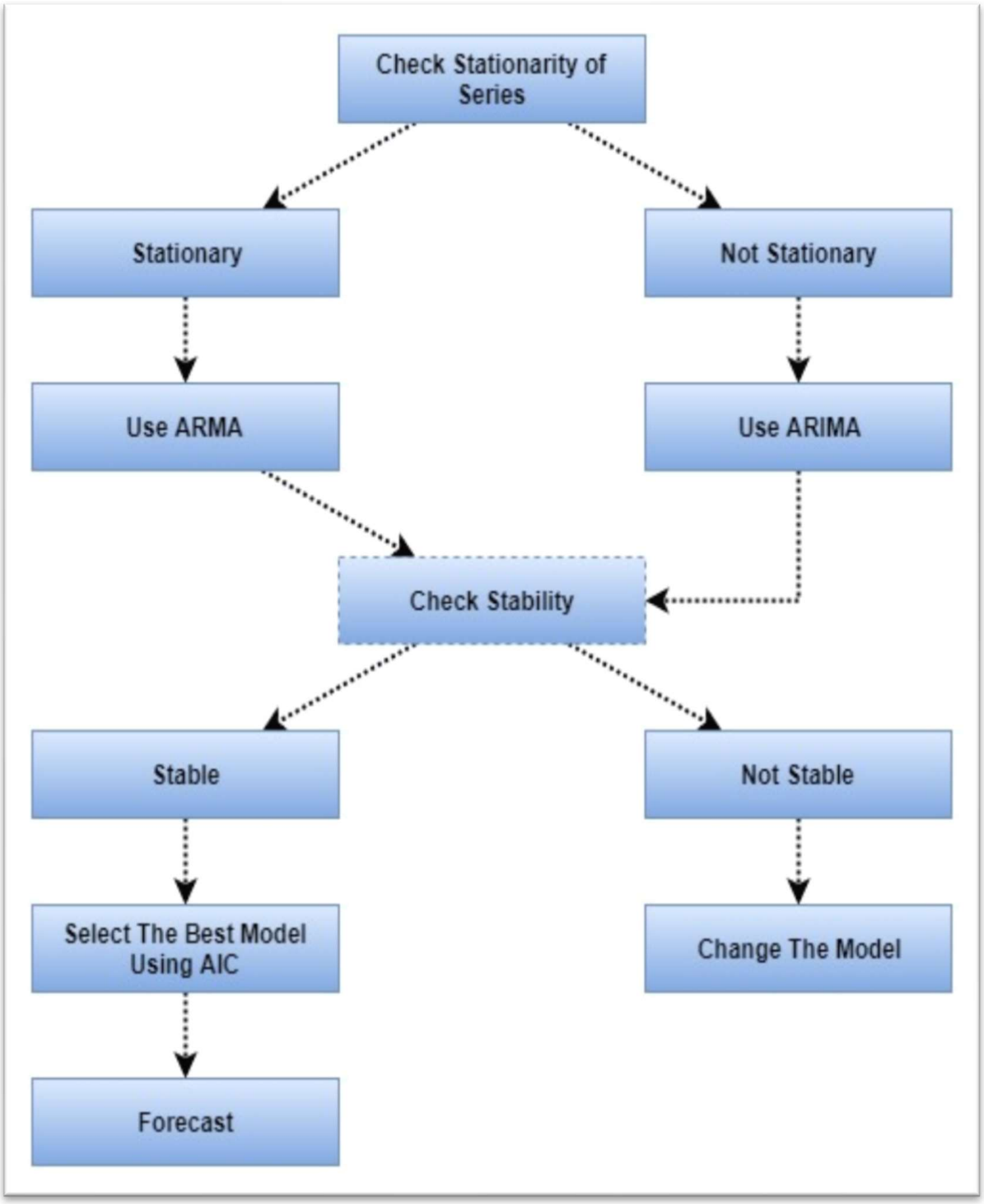

Figure 4.4: ARMA model construction.

Lastly, in the ARMA process, natural logarithm of market prices is used. In the Besides, optimal ARMA (p, q) is obtained by using autocorrelation functions, partial autocorrelation functions and Akaike Information Criterion (AIC) is taken account in model selection period. Lastly, to take care of daily seasonality, same hour yesterday (y_{t-24}) and same hour two days ago (y_{t-48}) are added into the ARMA models.

4.3 Autoregressive Moving Average Model with Exogenous Covariates

To see whether exogenous variables are affacting forecast performance of simple autoregressive models and whether it is performed better than other models, autoreggressive moving average with exogenous variables (ARMAX) is used. The ARMAX model enhance the simple ARMA model by adding exogenous variables. The ARMAX equation can be given as,

$$y(t) = \left(\delta + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \rho_{1k} x_{tk-1} + \cdots + \rho_{tk} x_{tk-k} + u_t - \beta_1 u_{t-1} - \cdots - \beta_q u_{t-q}\right) \quad (4.5)$$

$$and\ u_t \sim WN(o, \sigma_u^2)$$

and can be reduced to,

$$y(t) = \delta + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{k=1}^{k} \rho_k x_{tk} + u_t + \sum_{j=1}^{q} \beta_j u_{t-j} \quad (4.6)$$

$$\text{and } \mathrm{u_t \sim WN(o, \sigma_u^2)}$$

where $\boldsymbol{u_t}$ is independent white noise series, besides, $\boldsymbol{a_p}$, $\boldsymbol{\beta_q}$ repsesents AR and MA parameters respectively same with Equation 4.2. However, in these equations, $\boldsymbol{x_t}$ represents k different number of exogenous variable at time t, and $\boldsymbol{\rho_{tk}}$ represents the coefficients of k-th exogenous variable at time t.

As in the ARMA model, series should be stationary and invertible. It is known that the MCP is stationary. The process in Figure 4.4 will be followed to make a forecast.

Two ARMAX (p, q) model has been applied with different p and q. The structure of both model is same as in Equation 4.5 however, they have different exogenous variables which is shown in Table 4.2. For example, to take daily seasonality in account other than hourly seasonality, market clearing prices of yesterdays and two days ago are also added into the model. Futhermore, the autocorrelation functions and partial autocorrelations functions are taken into account and AIC is used to choose best model among them.

Table 4.2: ARMAX (p, q) models.

Variables	Model C (p, q)	Model D (p, q)
Load	X	X
Hour of Day	X	
Day of Week	X	
Working Days		X
Weekends		
Holidays		
Tariff 1 - Daytime	X	
Tarif 2 - Peak	X	
Tarif 3 - Night		X
Winter	X	X
y_{t-24}	X	X
y_{t-48}	X	X

4.4 Quantile Regression

The next model is quantile regression. In the OLS regression estimates the regression coefficients at the mean and coefficients are same for all distribution. However, Koenkerr and Basett (1978) introduced quantile regression which enables to capture conditional distribution of the dependent variable given its explanatory variables (Nowotarski and Weron, 2018). The quantile equation can be written as,

$$Q_N(\beta q) = \sum_{i:y_i \geq x_i'\beta}^{N} q|y_i - x_i'\beta_q| + \sum_{i:y_i \leq x_i'\beta}^{N} (1-q)|y_i - x_i'\beta_q|) \tag{4.7}$$

Where $Q_N(\beta_q)$ represents the conditional q-th quantile of MCP, x_i is vector explanatory variables and βq is vector parameters for quantile q (Nowotarski and Weron, 2018). Quantile estimator minimizes the equation over β_q and different choices of q estimate different values of β_q. Table 4.3 shows variables used in q-th quantile regression.

Table 4.3: Quantile regression model.

Variables	Qth quantile
Load	X
Hour of Day	
Day of Week	
Working Days	X
Weekends	
Tariff 1 - Daytime	X
Tarif 2 - Peak	X
Tarif 3 - Night	
Winter	
y_{t-21}	X
y_{t-22}	X
y_{t-23}	X
y_{t-24}	X
y_{t-25}	X
y_{t-26}	X
y_{t-27}	X
y_{t-48}	X
y_{t-72}	X

5. RESULTS

5.1 Forecast Performance Evaluation

In the electricty price forecasting literature, different performance evaluations are used. Mean absolute percentage error (MAPE) is common and useful method of measuring errors. MAPE may be misleading when the price is zero. However, in the forecasting period,prices are positive.

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\frac{|\, y_i - f_i|}{y_i} \times 100 \tag{5.1}$$

Where T is number of forecasted period, $\boldsymbol{y_i}$ and $\boldsymbol{f_i}$ represent actual and forecasted values of electricity prices. MAPE is calculated in out-of-sample forecast.

5.2 Model Comparison

In this study, several methods with different settings are constructed for forecasting Day-Ahead electricity market clearing prices and the best models are selected. The data start from 18 December 2015 to 31 December 2017. The last week of the data is dropped from the models and the next 24 hours prices has been forecasted by the model in the out-of-sample data.

First, graph of the best five models are given. Figure 5.1 shows MCP with Model A forecast, Figure 5.2 represents MCP with Model B forecast, Figure 5.3 represents MCP with ARMA (4,4), Figure 5.4 represents MCP 40th quantile and Figure 5.5 represent MCP with 50th quantile and Figure 5.6 represents the best five models together.

According to Figure 5.1 Model A captures the general structure and the direction of movements in the MCP except the last day of the year. As mentioned before electricity consumption changes dramatically in the 31 December, even though the dummy for last day of the year was added, forecast for 31 December is relatively worse than the other days.

Model B is in the Figure 5.2. This model has nearly the same with Model B. The model is generally well capturing the directions of the prices and performance is relatively worse in December 31.

ARMA (4,4) forecast is not good fitting the actual MCP. Eventhough, forecasted result are mostly following the right direction, there are big differences between forecasted and actual values of prices.

Additionally, 40th quantile forecast performs well for the first five days. It is generally moving with the actual prices and very good at capturing the spikes. Unlike the other models, it is generally performing best in the spiky hours.

Quantile 50th forecast performs well for the first three days and good at following sudden increases in these days. However, it performs poorly for the rest of the week.

In the Figure 5.6, all models are given together. As discussed before, Model A seems to be the best model forecasting electricity prices in TEM.

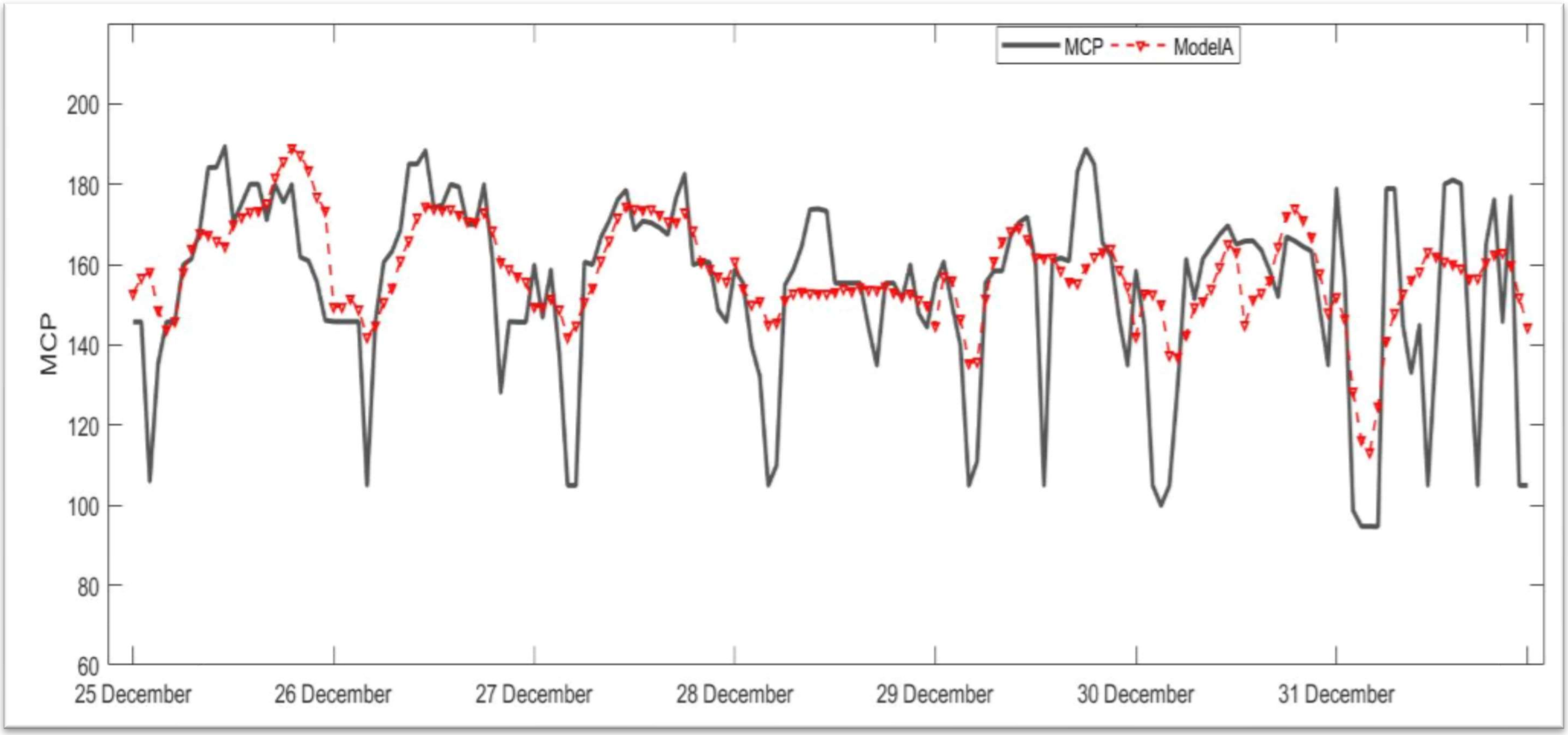

Figure 5.1:MCP and Model A.

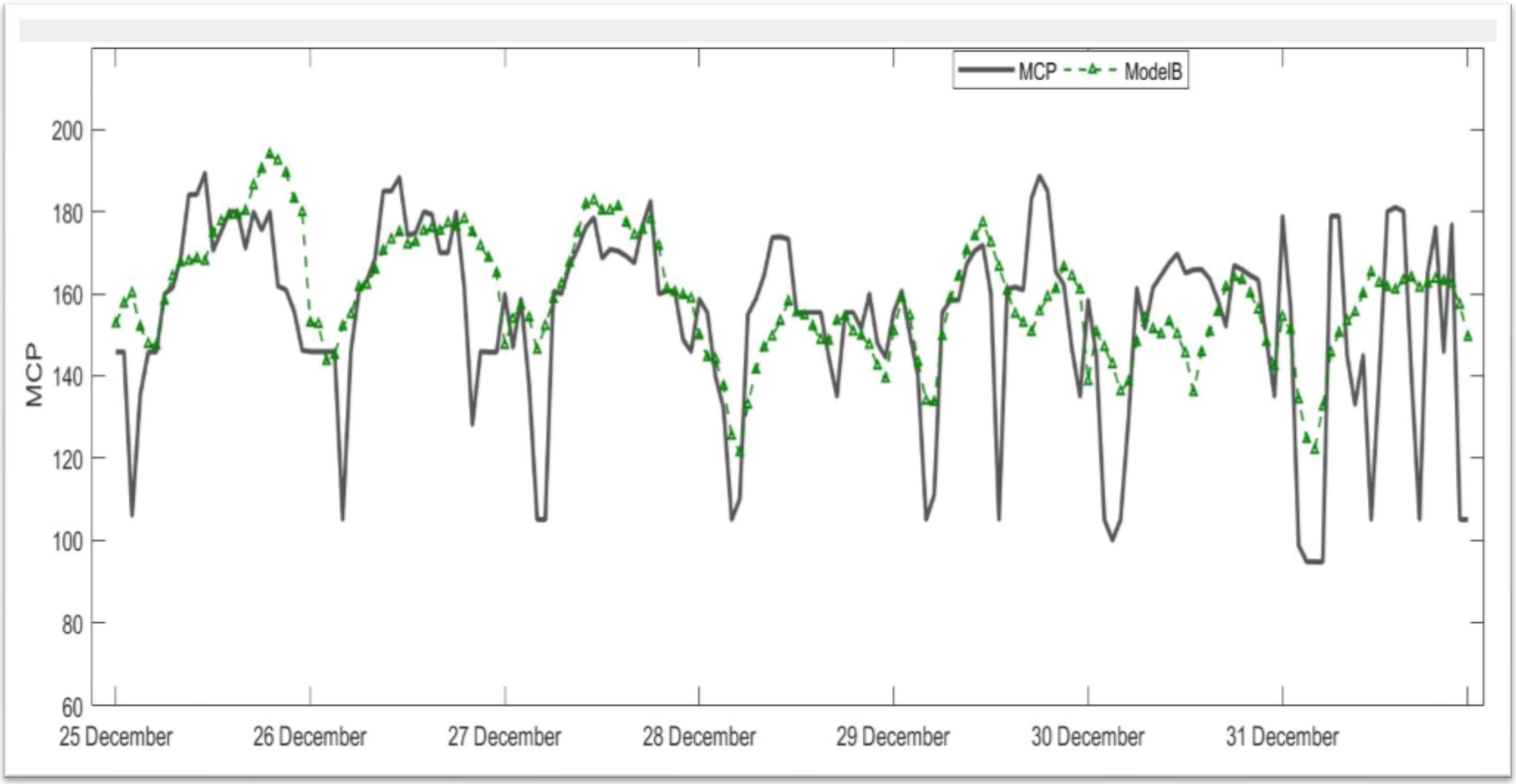

Figure 5.2: MCP and Model B.

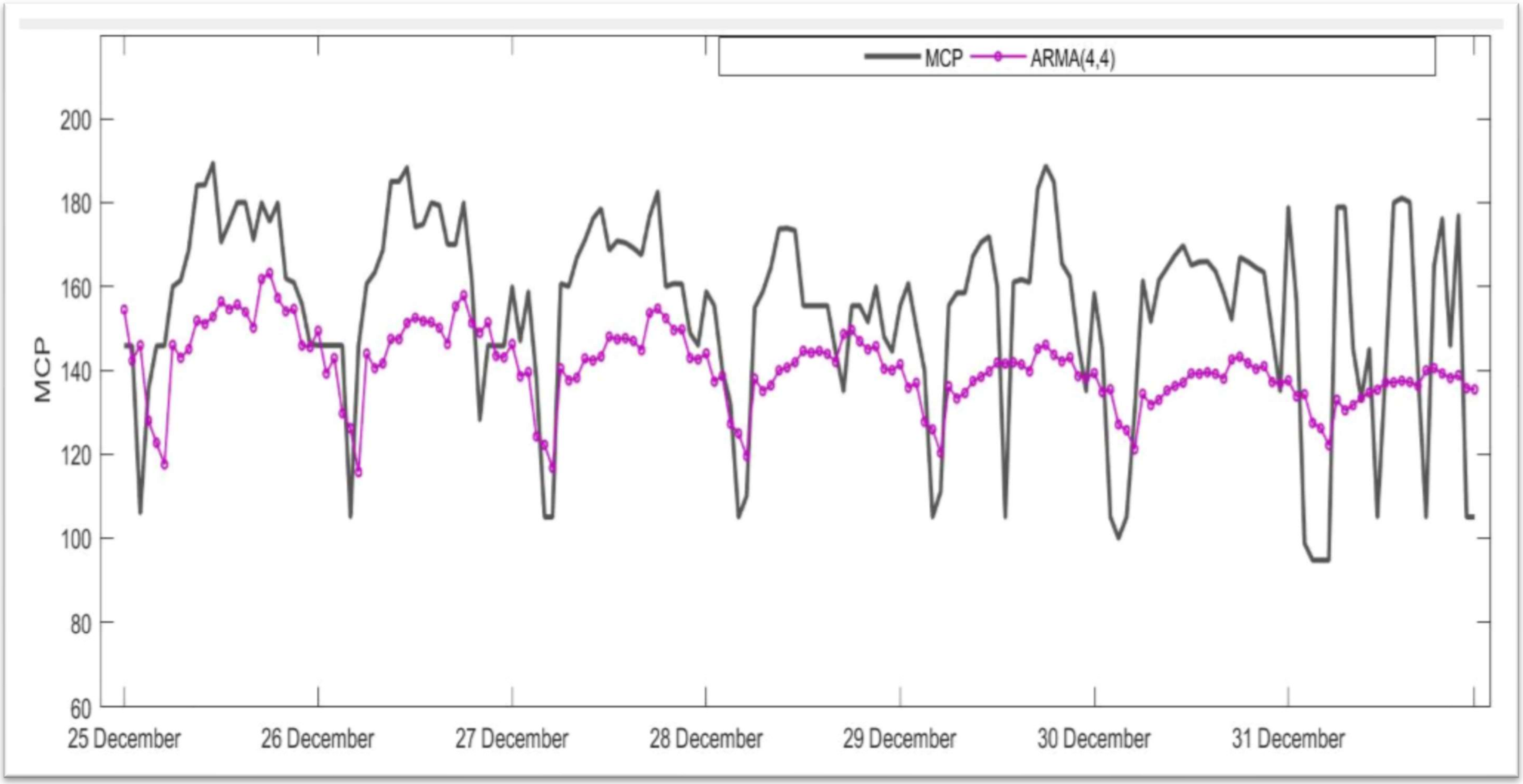

Figure 5.3: MCP and ARMA (4,4).

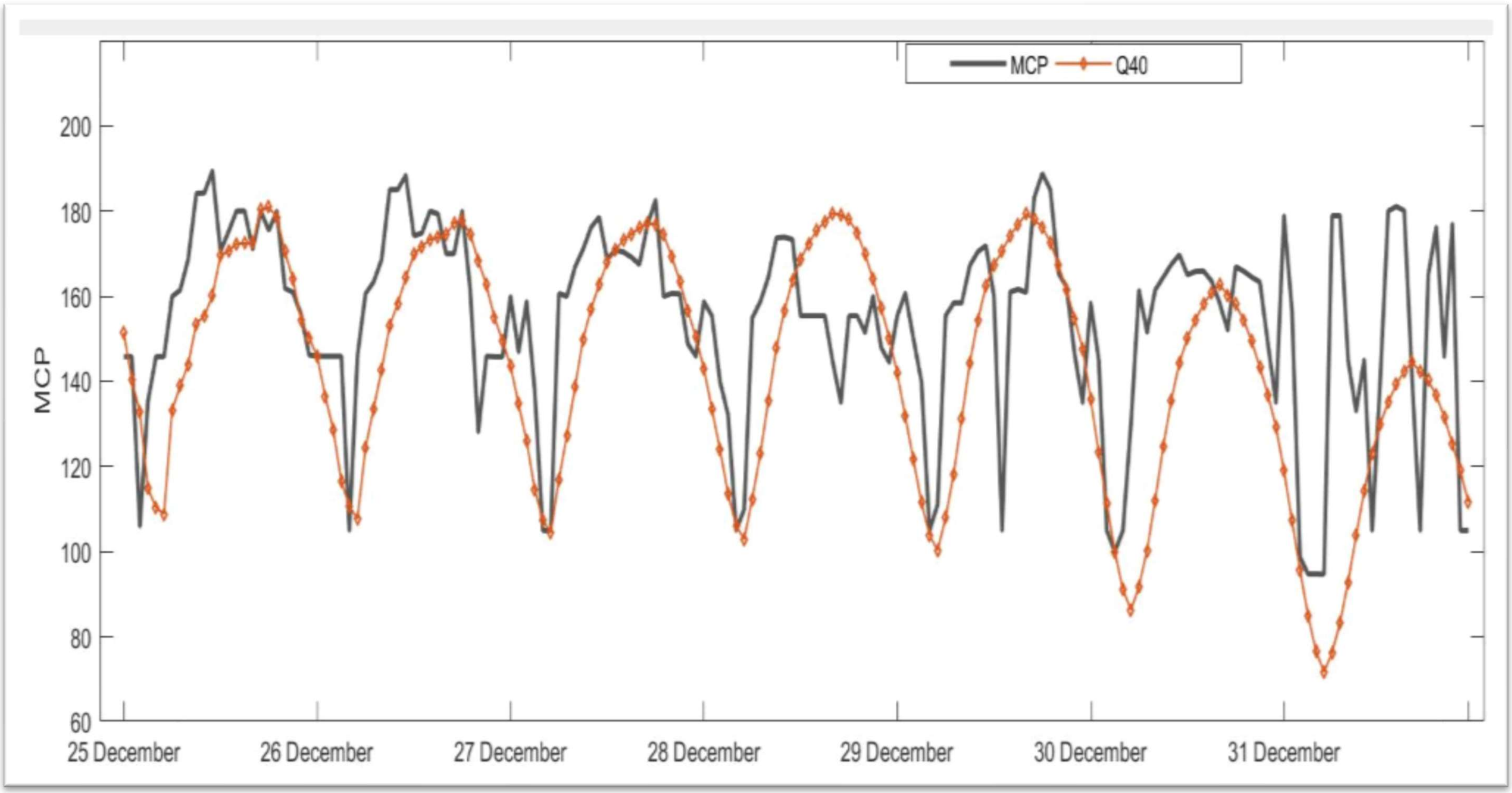

Figure 5.4: MCP and 40th quantile.

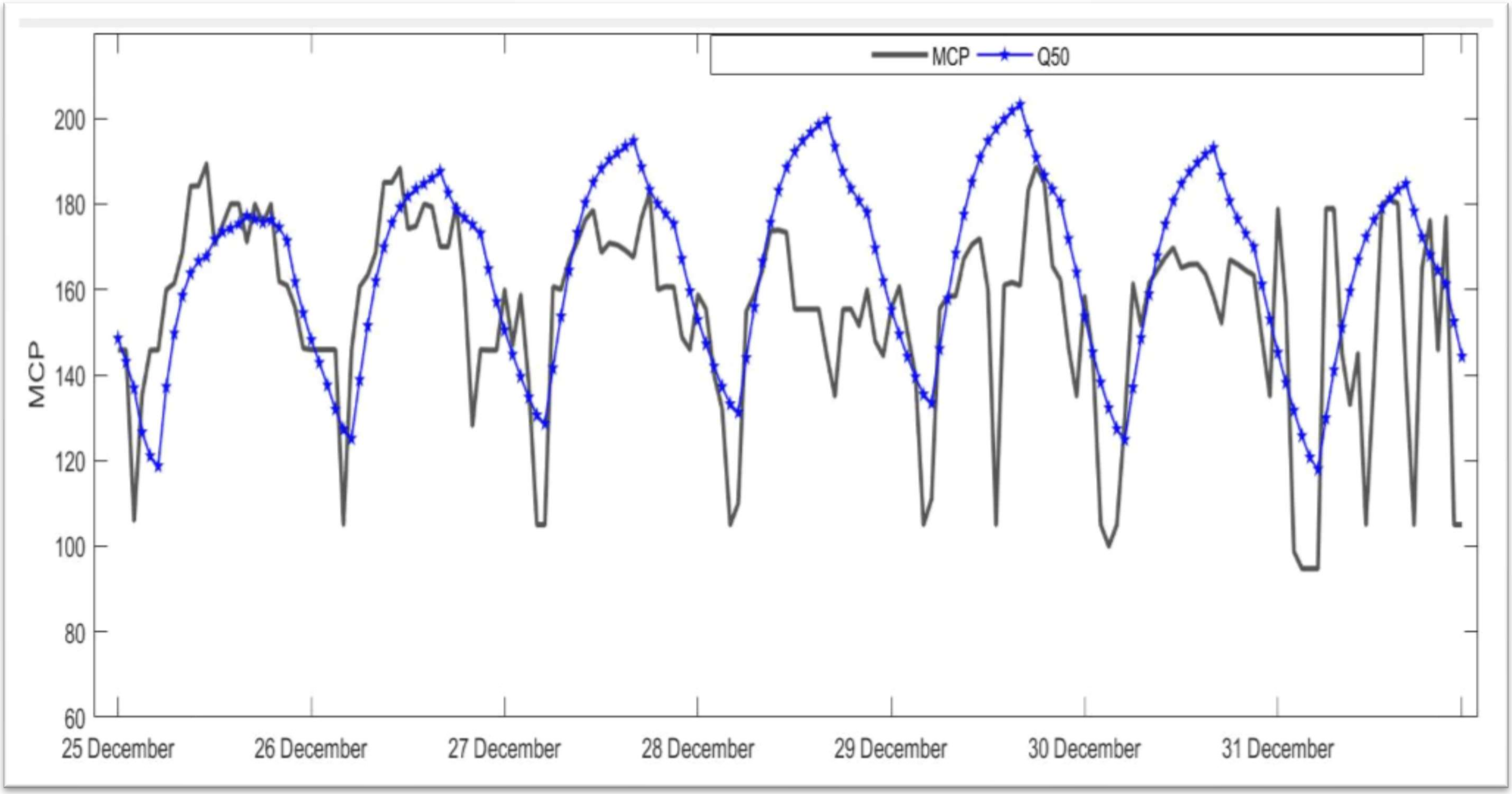

Figure 5.5: MCP and 50th quantile.

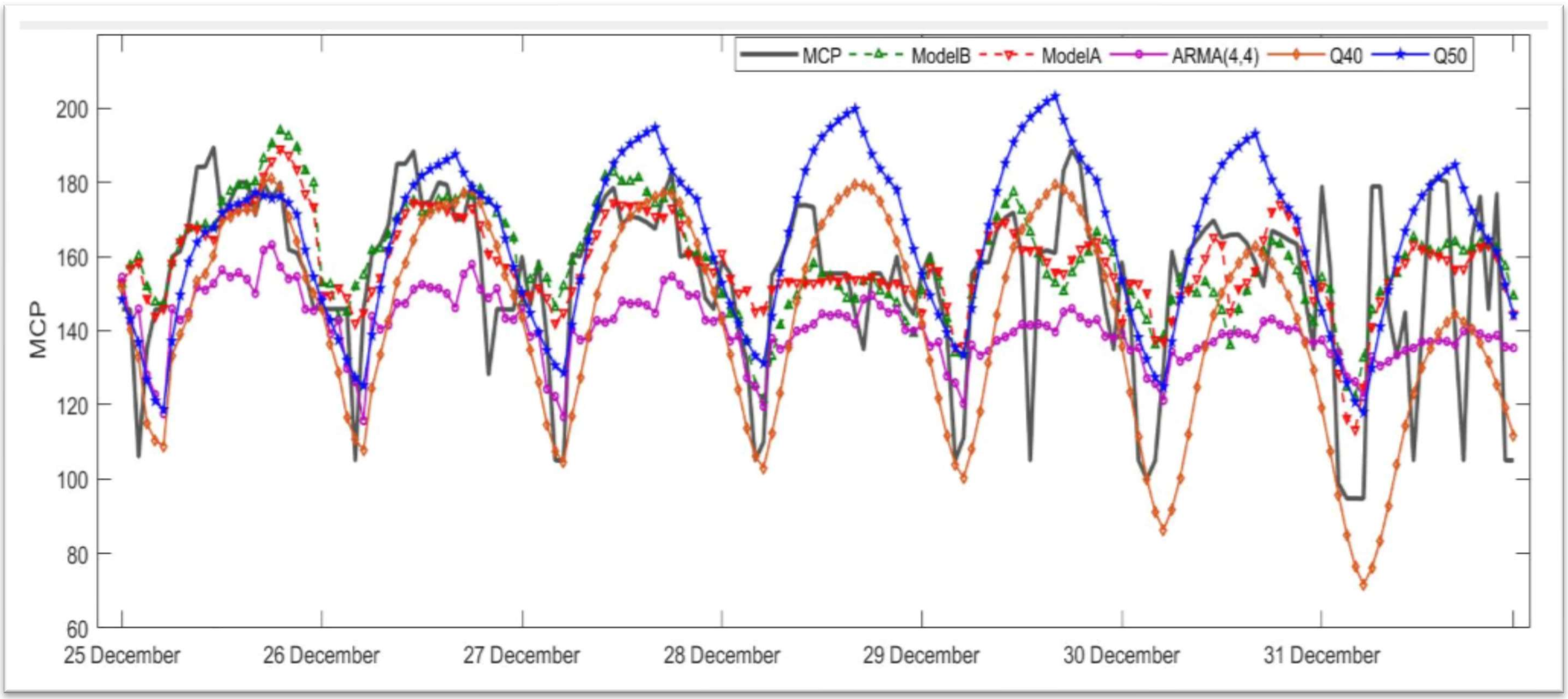

Figure 5.6: MCP and the best five forecast.

Furthermore, models are compared by using MAPE performance evaluation method in the out-of-sample data and the selected models with their scores are presented in Table 5. In the table, measurement errors are given as daily average.

Table 5:Forecast performance evaluation.

Year	MAPE (%)
Model A	9.65
Model B	10.46
ARMA (4, 4)	13.57
ARMA (5,4)	15.97
Model C (4,4)	16.83
Model D (4,4)	17.44
40 th Quantile	12.77
50 th Quantile	14.61

In the Table 5, error measurements are calculated and given as daily error average MAPE. To have best forecast, the errors should be minimized. Among other indicators, MAPE is relatively easy to interpret and comprehend the errors rate. Since there are no points that includes zero in the forecasting period, MAPE can be used to interpret the models.

According to the Table 5 and the figures discussed, Model A which is constructed by NARX network has lowest MAPE. Model A has an MAPE of 9.65% and it is a reasonable error rate for forecasting in such a spiky market. On the other hand, other models based on time series structure performed relatively worse.

5.3 Conclusion

Beginning with the liberalization of the TEM, electricity price forecasting has become a substantial way to maximize energy companies' profit. The liberalization process continues and have not been reached maturity. Besides, there are few studies on electricity price forecasting in the TEM. Therefore, in this study, electricity market clearing prices estimated with NARX, ARMA, ARMAX and quantile regression methods and different combinations of these models are constructed by using data starting from December 2015 to December 2017. Forecasting performances are compared with MAPE. According to empirical results, NARX method with Model A outperformed the other methods and the models. It has reasonably low error rates showing that Model A captures the characteristic of the TEM's clearing prices.

Eventhough NARX model is performing best in the TEM with lowest daily average error rates, quantile regression is performing better than NARX in the spkiy hours. In the following studies, combination of these models can lead to a better forecasting performance.

REFERENCES

EPDK. (2016). Electricity market development report. ANKARA. Retrieved from https://www.epdk.org.tr/Detay/Icerik/3-0-24/elektrikyillik-sektor-raporu

Gareta R., Romeo L. M., Gil A. (2006). Forecasting of electricity prices with neural networks. Energy Conversion & Management, 47, 1770-1778.

Girish, G. (2016). Spot electricity price forecasting in Indian electricity market using autoregressive-GARCH models. Energy Strategy Reviews, 11-12, 52-57.

Gökgöz F., Filiz F. (2016). Electricity price forecasting in Turkey with artificial neural network models. Investment Management and Financial Innovations, 13(3), 150-158.

Keles D., Scelle J., Paraschiv F., Ficthner W. (2016). Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Applied Energy, 162, 218-230.

Koenker R., Basett Jr. G. (1978). Regression Quatiles. Econometrica, 33-50.

Marcjasz G., Uniejewski B., Weron R. (2018). On the importance of the long-term seasonal component in the day-ahead electricity price forecasting with NARX neural networks. International Journal of Forecasting.

Mirakyan A., Meyer-Renschhausen M., Koch A. (2017). Composite forecasting approach for next-day electricity price forecasting. Energy Economics, 66, 228-237.

Nogales F. J., Contreras J., Conejo A. J., Espinola R. (2002). Forecasting next day electricity prices by time series models. IEEE Transactions on Power Systems, 17, 342-348.

Nowotarski J., Weron R. (2018). Recent advances in electricity price forecasting: A reviews of probabilistic forecasting. Renewable and Sustainable Energy Reviews, 81, 1548-1568.

Panapakidis I. P., Dagoumas A. S. (2016). Day-ahead electricity price forecasting via the application of neural network based models. Applied Energy, 172, 132-151.

Singhal D., Swarup K. S. (2011). Electricity price forecasting using artificial neural networks. Electrical Power and Energy Systems, 33, 550-555.

TETAS. (2015). Annual Report. Ankara. Retrieved from http://www.tetas.gov.tr/File/?path=ROOT%2f1%2fDocuments%2fFaaliyet+Raporu%2fTETA%C5%9E+2015+YILI+FAAL%C4%B0YET+RAPORU.pdf

TETAS. (2016). Annual Report. ANKARA. Retrieved from http://www.tetas.gov.tr/File/?path=ROOT%2f1%2fDocuments%2fFaaliyet+Raporu%2f2016+Y%C4%B1l%C4%B1+Faaliyet+Raporu+v2.pdf

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30, 1030-1081.

Ziel F., Weron R. (2018). Day-ahead electricity price forecastin with high-dimensional structures: Univariate vs. multivariate framworks. Energy Economics, 70, 396-420.

CURRICULUM VITAE

Name Surname : Yunus Emre Karagülle

Place and Date of Birth : Istanbul- December 17, 1992

E-Mail : karagullemre@gmail.com

EDUCATION

- **B.Sc.** : 2016, Anadolu University, Economics
- **M.A.** : 2018, Istanbul Technical University, Economics

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Karagülle Y.E., Güloğlu B.** (2018). Forecasting Day-Ahead Electricity Prices with Machine Learning and Time Series Techniques. ENSCON International Congress of Energy, Economy and Security, (*p.9).* Istanbul.