ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

NEW MODEL FOR FORECASTING FINANCIAL DATA

M.Sc. THESIS

Özgün SEYMEN UZUN

Department of Mathematic Engineering

Mathematical Engineering Programme

JUNE 2019

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

NEW MODEL FOR FORECASTING FINANCIAL DATA

M.Sc. THESIS

Özgün SEYMEN UZUN
(509151214)

Department of Mathematical Engineering

Mathematical Engineering Programme

Thesis Advisor: Assoc. Prof. Dr. Burcu TUNGA

JUNE 2019

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

FİNANSAL VERİLERİN ÖNGÖRÜSÜ İÇİN YENİ BİR MODEL

YÜKSEK LİSANS TEZİ

Özgün SEYMEN UZUN
(509151214)

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Burcu TUNGA

HAZİRAN 2019

Özgün Seymen Uzun, a M.Sc. student of ITU Graduate School of Science Engineering and Technology 509151214, successfully defended the thesis entitled “NEW MODEL FOR FORECASTING FINANCIAL DATA”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Dr. Burcu Tunga** İstanbul Technical University	....
Jury Members :	**Assoc. Prof. Dr. Ali Ercengiz** Istanbul Technical University	....
	Prof. Dr. Ayşe Hümeyra Bilge Kadir Has University	....

Date of Submission : 24 April 2019
Date of Defense : 14 June 2019

To my family,

FOREWORD

I would like to thank to my advisors Dr. Eti Mizrahi and Assoc. Prof. Dr. Burcu Tunga. They spared their time to guide and support me to write this thesis. Additionally, I would like to thank to my husband Onur Uzun and my mother Naciye Seymen and my father Ogün Seymen to support me any time when I studied on this thesis.

June 2019

Özgün SEYMEN UZUN
(Mathematic Engineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

KAP	: Kamuoyu Aydınlatma Platformu
BIST	: Borsa İstanbul
ISE	: İstanbul Stock Exchange
EMAVAR	: Exponential Moving Average Vector Autoregressive
AR	: Autoregressive
MACD	: Moving Average Convergence Divergence
EMA	: Exponential Moving Average
SMA	: Simple Moving Average
VAR	: Vector Autoregressive
MA	: Moving Average
ADF	: Augmented Dickey-Fuller Test
RSI	: Relative Strength Index
VECM	: Vector Error Correction Model
RW	: Random Walk
MAPFE	: Mean Absolute Percentage Error
VHTS	: Virtual Historical Trader Software
FDI	: Foreign Direct Investment
RMSFE	: Root Mean Square Forecast Error
C#	: C sharp
PP	: Philips-Perron
KPSS	: Kwiatkowski, Phillips, Schmidt and Shi
ADF-GLS	: Augmented Dickey-Fuller Enhanced Least Square
NGP	: Ng-Perron
ACF	: Autocorrelation Function
PACF	: Partial Autocorrelation Function
AIC	: Akaike Information Criteria
BIC	: Bayesian Information Criteria

SYMBOLS

$\mathbf{H_0}$	: Null Hypothesis
$\mathbf{H_1}$	: Alternative Hypothesis
$\mathbf{R^2}$	: R-square
Adj-$\mathbf{R^2}$	: Adjusted R-square
$\boldsymbol{\mu}$	: Mean
$\boldsymbol{\sigma^2}$	:Variance

LIST OF TABLES

Page

LIST OF FIGURES

Page

NEW MODEL FOR FORECASTING FINANCIAL DATA

SUMMARY

Forecasting is a significant process of the time series. It is the process of making predictions of the future by using the past and present data and generally by analysis of trends. Forecasting of stock price index and its movement are one of the most common research field about the financial time series. The financial time series are complicated and non-linear. By this cause, forecast financial time series is difficult. Additionally, the stock price index is affected by many factors like economic, social, political events in complicated manner. Although these issues, the forecast of stock index is one of most remarkable researches for many industrial experts and scholars.

In this thesis, the new model EMAVAR is explained. The EMAVAR model is combined MACD and VAR model.

Technical analysis is the process of examining a recent price to determine possible future prices. The moving averages is one of the oldest and most popular technical analysis tools. A moving average is the average of prices at a given time. There are five popular types of moving average: simple, exponential, triangular, variable and weighted. The important difference between these MAs is weight to the most recent data. MACD is the most common technical indicator in technical analysis. It uses Exponential Moving Average (EMA) while calculating. 26-days EMA and 12-days EMA of the data is calculated. Then, their subtraction of these EMAs gives the MACD line. The signal line is obtained from 9-days EMA of MACD. When MACD is applied, the experts comment to decide buy-sell (bearish-bullish) behavior.

VAR is the most familiar forecasting model in the literature for multivariate time series. In VAR process, the endogenous (dependent variable) value is modeled with the exogenous (independent variable) values to estimate. VAR is used for multivariate time series. The purpose of thesis is to construct a new model to improve the accurancy estimation.

The EMAVAR model uses different exogeneous variables which are occurred by the result of MACD process. Afterwards, exogeneous variable series is used in VAR process for anticipate the next value.

The details of EMAVAR model is explained like as follows. First of all, obtained data from KAP which is the daily (5-days a week) closing stock prices of BIST-100 and the banking sector between 2012-2018, is rearranged to make healthy forecast. Missing data (by occurred because of public holidays) is eliminated with linear interpolation and process have been continued. Additionally, 6 years full data set is divided 1 year and 2 years parts with the purpose of the independence of historical data. After that, MACD process is applied on different data arrengement seperately. The stationarity of all obtained series from MACD is recognized with Unit Root Test. Finally, the VAR model is applied on this different series.

In addition, one of other forecasting of time series model is Autoregression (AR) model. The autoregression can be used to forecast an arbitrary number of periods into

the future. AR is used on the same data sets to compare and show the good estimation of EMAVAR model.

To sum up, EMAVAR model is compared to AR model. According to the result of researches and experiments, EMAVAR model is better than the other is in 2 years.

FİNANSAL VERİLERİN ÖNGÖRÜSÜ İÇİN YENİ BİR MODEL

ÖZET

Bir zaman serisi, bir değerin zaman içerisinde sıralanmış ölçümlerinin bir kümesi olarak tanımlanabilir. Zaman serileri trend, mevsimsellik, çevresel, düzensizlik olmak üzere dört bileşenden oluşur. Zaman serilerinin önemli bir süreçlerinden birisi tahmin yapmadır. Geçmişteki ve şimdiki verileri kullanarak ve genellikle trendlerin analizi ile geleceğe yönelik öngörülerde bulunma sürecidir. Finansal zaman serileri doğrusal değildir ve karmaşıktır. Bu nedenle, finansal zaman serilerinin tahmin edilmesi zordur. Borsa fiyat endeksinin tahmini ve endeksin hareketi, finansal zaman serileri ile ilgili en yaygın araştırma alanlarından biridir. Ayrıca, hisse senedi fiyat endeksi ekonomik, sosyal, politik olaylar gibi birçok değişik faktörden etkilenmektedir. Bu sorunlara rağmen, hisse senedi endeksi tahmini son zamanlarda uzmanlar ve akademisyenler için en popüler sorunlardan biri olmuştur.

Teknik analiz, muhtemel gelecekteki fiyatları belirlemek için son fiyatları inceleme olarak tanımlanabilir. Hareketli ortalamalar metodu ise en eski ve en popüler teknik analiz araçlarından biridir. Hareketli bir ortalama, belirli bir zamanda olan fiyatların ortalamasıdır. Beş popüler hareketli ortalama türü vardır: basit, üstel, üçgen, değişken ve ağırlıklı. Bu hareketli ortalamalar arasındaki en önemli fark, son zamanlara yakın verilere verilen ağırlıktır. Hareketli ortalama sürecinde, her bir gecikmeli hata terimi onun şimdiki değerini etkilemektedir. Hareketli ortalama yakınsak ıraksak indikatörü (Moving Average Convergence Divergence) teknik analizde en yaygın kullanılan teknik göstergedir. Hesaplanırken üstel hareketli ortalama kullanılır. Verilerin 26 günlük ve 12 günlük üstel hareketli ortalama değerleri hesaplanır. Hesaplanan 26 ve 12 günlük hareketli ortalamaların farkı MACD çizgisini verir. Sinyal çizgisi ise, MACD 'nin 9 günlük üstel hareketli ortalamasından elde edilir. Uzmanlar, MACD teknik analiz yöntemini kullanarak alım-satım (ayı-boğa) davranışına karar vermek için yorumda bulunurlar.

Vektör Otoregresif model (VAR), çok değişkenli zaman serileri için literatürdeki en bilinen tahmin etme modelidir. Çok değişkenli zaman serilerinde en önemli amaç serinin bileşenleri arasında öngörülebilir yani durağan lineer birleşimin araştırılmasıdır. Başka bir deyişle, serinin bileşenleri arasındaki kointegrasyon ilişkisinin belirlenmesidir. Böyle bir ilişki belirlendiği zaman, serinin durağan lineer birleşimleri üzerinden istatistiki sonuç çıkarımlar yapılabilir. Vektör Otoregresif modelinde, endojen (bağımlı değişken) değer, tahmin edilecek eksojen (bağımsız değişken) değerler ile modellenir. İçsel değişkenlerin birlikte ele alındığı eş anlı denklem modeli olan Vektör otoregresif model, Sims tarafından 1980 yılında geliştirilmiştir. VAR modelinde değişkenler içsel ve dışsal olarak ayrılmadığından belirlenme sorunu söz konusu değildir. Bu modelde, bütün değişkenler içsel kabul edilir ve bir değişken kendisinin ve modelde yer alan diğer değişkenlerin gecikmeli değerlerinin fonksiyonu olarak tanımlanır.

Bu tezde, finansal verilerin öngörüsü için oluşturulmuş yeni bir model olan Exponential Moving Average Vector Autoregressive Model (EMAVAR) açıklanmıştır. EMAVAR modeli, Moving Average Convergence Divergence teknik analiz indikatörü (MACD) ve Vector Autoregressive (VAR) modeli kullanılarak oluşturulmuştur.

EMAVAR modelinde, MACD işleminin sonucu olarak farklı seriler oluşturulur. Daha sonra, bu serilerden iki tanesi bir sonraki değeri tahmin etmek için VAR modelinde bağımsız değişken olarak kullanılır.

EMAVAR modelinin detayları şu şekilde açıklanabilir. Öncelikle Kamuoyu Aydınlanma Platformundan elde edilen 2012 – 2018 yılları arasındaki veriler (5-çalışma günü) alınmıştır. Elde edilen bu verilerde tatil günleri nedeniyle oluşan eksik değerler olduğu tespit edilmiştir. Daha sağlıklı tahminler elde edilmesi için veriler doğrusal enterpolasyon uygulanarak ortadan kaldırılmıştır ve bu süreç devam ettirilmiştir. Ayrıca, alınan 6 yıllık tam veri seti, tarihsel verilerin bağımsızlığını göstermek amacıyla 1 yıl ve 2 yıl bölümlerine ayrılmıştır. Daha sonra, MACD işlemi ayrı ayrı bu ayrılan veri bolümlerine uygulanmıştır. MACD 'den elde edilen tüm serilerin durağanlığı (ilgili serilerde birim kökün var olup olmadığı), Augmented Dickey – Fuller Birim Kök Testi ile tespit edilmiştir. Zaman serilerinin durağan olması olarak ifade edilen durum, zaman içinde varyansın ve ortalamanın sabit olması ve gecikmeli iki zaman periodundaki değişkenlerin ko-varyansının değişkenler arasındaki gecikmeye bağlı olup zamana bağlı olmaması olarak tanımlanabilir. MACD işleminden sonra elede edilen macd ve signal serileri bağımsız değişken olarak Vektör Otoregresif modelde kullanılmıştır. Bu uygulama farklı yıllardan oluşan serilere ayrı ayrı olarak yapılmıştır.

Ek olarak, zaman serisi modellemelerinde kullanılan diğer modellerden biri Autoregression (AR) modelidir. Otoregresyon, gelecekteki keyfi bir dönem sayısını tahmin etmek için kullanılabilir. Bir AR modelinde, bağımlı değişken geçmişteki değerinin bir fonksiyonudur. Birçok zaman serisi de bu süreci içermektedir. Bu tezde AR modeli, EMAVAR modeli ile daha iyi tahmin verdiğini göstermek için aynı veri setlerinde kullanılmıştır.

Tezin ilk ve ikinci bölümünde finansal zaman serilerinin öngörüsü için yapılan çalışmalar incelenmiş ve analiz edilmiştir. Daha sonra tezde kullanılacak olan modellerin matematiksel araştırmalarına ve destekleyici literatür bilgilerine Bölüm 3 ‘te yer verilmiştir. Korelasyon fonksiyonları, durağanlık tanımı ve tespit etme testleri, modellemenin tanımı, MACD teknik analiz indikatörü, AR modelinin detaylarına yine aynı bölümde yer verilmiştir.

Tezin 4. Bölümünde finansal verilerin öngörüsü için oluşturulan yeni model EMAVAR ’ın detaylı açıklamaları mevcuttur. Farklı yıllardan oluşturduğumuz veri setlerine EMAVAR modelinin uygulanmasının detaylı adımları Bölüm 5’te yer almaktadır. Karşılaştırma amacıyla kullanıdığımuız AR modeli uygulamalarının detayları da yine bu bölümdedir.

Son olarak, Bölüm 6 ’da EMAVAR ve AR modelinin karşılaştırmalı sonuç tabloları bulunmaktadır.

Bu tezin amacı, finansal serilerin öngörüsünde doğruluk tahminini iyileştirmek için yeni bir model oluşturmaktır.

Özetle, bu tezde Exponential Moving Average Vector Autoregressive (EMAVAR) modeli, AR modeli ile karşılaştırılmıştır. Yapılan araştırmalar ve deneylerden elde edilen sonuçlara göre EMAVAR modeli 2 yıllık veri üzerinde diğerlerinden daha bulunmuştur.

1. INTRODUCTION

Forecasting of stock price index and its movement are one of the most remarkable applications of prediction of financial time series. The high-risk and high-efficiency are together in the stock market. The financial time series are complicated, dynamic, non-linear and chaotic in natural behavior (Abu-Mostafa & Atiya, 1996). These features cause that there is no exactly truly information that could be estimated from the past behavior. In addition, the stock market index is important for traders to make investment policies. However, the stock price index is effected by many factors like economic, social, political events in complicated manner. Although these issues, the forecast of stock index is one of most remarkable researches for many industrial experts and scholars.

There are a lot of researches to forecast stock price and its relationship between other macroeconomic values. The experts and scholars have been studied for this aim with different methods. Başçı and Karaca (2013) researched the relationship between ISE index 100 and set of four macroeconomic variables by using Vector Autoregressive model. They used Exchange, Gold, Import, Export as independent variables, ISE index 100 as dependent variables with the period from January, 1996 to October, 2011. The seasonal adjustment were made because all variable had seasonality. The first difference took all series to eliminate unit roots. The result of the study showed that it was explained 31% by share indices. In addition, Bator (2018) et al., examined the relationship the percentage of impact of endogeneous shocks and the direction of causality between some macroeconomic variables. These variables are foreign direct investment, unemployment rate and real gross domestic product exchange rate in Ghana. The Johansen cointegration test via VAR Model and VECM was used to determine long or short dynamic connections between the chosen macroeconomic variables with period from 1992 to 2016. Granger Causality Tests, Forecast Error Variance Decompositions and Impulse Response Functions were studied for the dynamic interactions. For determining stationarity, ADF Test were applied on data and all of variables of first differences were useful. Findings of this research, the effect of FDI variables were positive on the selected macroeconomic variables.

Moving Average Convergence Divergence (MACD) is the most known technical indicator utilized by investors. When MACD is applied, it makes recommendation for traders to buy-sell decision. MACD is computing with 3 Exponential Moving Average (EMA) values. 26-days EMA and 12-days EMA of the data is calculated. Then, their subtraction of these EMAs gives the MACD line. The signal line is obtained from 9-days EMA of MACD. The details of MACD is given in Section 3.7, the experts comment to decide buy-sell (bearish-bullish) behavior.

Vector Autoregression (VAR) Model is the most used forecasting model in the academic studies. The endogenous (dependent variable) value is modeled with the exogenous (independent variable) values to estimate. VAR is used for multivariate time series. The structure of VAR model is that each variable is a linear function of past lags of itself and past lags of the order variables.

In this project, the data used for modeling and forecasting is the daily (5-days a week) closing stock prices of BIST-100 and the banking sector between 2012-2018. The data is provided by KAP (KAP, 2019). Two different models are used for modeling and forecasting. A new model labeled as EMAVAR (Exponential Moving Average Vector Autoregressive), which is the main subject of this thesis, is compared with the AR model. A comparison has been done between the forecasted values obtained using these two different models. The aim of the comparison is showing that the new model give a better forecast.

The new model called as EMAVAR is inspired by the VAR model and the MACD indicator. In EMAVAR, the data is not modeled by means of different exogenous variables (as the VAR model), but by the different EMAs of the data itself. The EMAs are 9-12-26 days. These numbers are inspired by the formula of MACD indicator used in technical analysis.

This study was performed to produce a new model to improve the accuracy of estimation. This thesis is structured as follows: a general information about financial time series and literature review are given in the Section 2. 3rd Section is allocated to theoretical knowledge about basic statistics, modeling and forecasting used in this study such as; Autoregression (AR), Moving Average (MA), Vector Autoregression

(VAR), ADF, Unit Root Test, the information criterias for deciding model of goodness of fit. The EMAVAR model is explained in details in the 4^{th} section. Then, the applications of the EMAVAR and AR models on the data are given in the 5^{th} section. Finally, comparison, results and conclusions are given respectively in the 6^{th}, 7^{th} and 8^{th} section.

2. LITERATURE REVIEW

In recent years, many studies have been made to anticipate the financial time series or determine its movements by using different methods.

Technical indicators help traders for decision of making investment. For this purpose, Rosillo et al. (2013) compared the RSI, MACD, Momentum and Stochastic rules which known as technical indicators, in Spanish market companies with their studies. The computer tools was created to help the decision making in the stock exchange for small investors and remove the ambiguity caused by different indicators. In conclusion, it is achieved with the tools were wrote in C#, .NET software languages.

Technical analysis is very important to make investment for traders. To decide to buy or sale behavior was shown by using the technical indicators. Yazdi and Lashkari (2013) focused on the popular indicator MACD for 4 currencies: EURUSD, GBPUSD, USDCHF, USDJPY. The 4 currencies were compared with each other by using Virtual Historical Trader Software (VHTS) and found the better one.

Chong and Ng (2008) were used two technical indicators MACD and RSI to see profitability of these indicators. The 60-years data of the Stock Exchange FT30 Index in London were used in the experiment. The findings obtained from this study, in most cases, the RSI as well as MACD rules can produce good results than the buy and hold strategy.

Some principles are beneficial for economic researches: make model simple, use all data you have, use theory as a guide to chosen fundamental variables. However, theory does not give enough information about the dynamics. The researchers did not get successful result because they did not give importance to dynamic structure. The Vector Autoregressive (VAR) approach developed in the 1980s in first. It resolved this issue in the simple way by moving towards dynamics and away from collecting many causal variables. The VAR approach also resolved the problem of how to make long term forecasts where the fundamental variables themselves must be forecast (Allen & Fildes, 2001). There are a lot of studies about Vector Autoregression.

Gerçeker (n.d) studied the effects of foreign direct investments on economic growth. Economic growth of Turkey and the foreign direct capital investment was analyzed by using the period of January, 1995 and September, 2007 in monthly. First of all, Unit Root Test was used for determining the stationarity. The result of this test, take the first differences of variables to eliminate unit roots. Afterwards, VAR model was applied with eight lags. In conclusion, the relationship between these variables were determined positively and they acted each other in long term.

Özsoy (2009) aimed with this research to detect the direction and magnitude of the relationship between different levels of education and economic growth in Turkey. With this purpose, 1923-2005 Gross Domestic Product and the number of students in different class in Turkey were used for testing with VAR model. First of all, the Augmented Dickey-Fuller (ADF) Unit Root Test was used for determining the stationarity of data; secondly, the existence of cointegration was tested with Johansen method; thirdly, causality testing was determined with Granger Causality Test. Finally, conclusion of all of these tests, Vector Error Correction Model (VECM) was applied on series and these results were comment with Impulse Response Functions and Variance Decomposition Process. The findings obtained from this study, the relationship between education and economic growth in Turkey is positive. According to this conclusion, Turkey should pay special attention to education for sustainable economic growth.

VAR model was used different data of forecast. For example, Song and Stephen (n.d) studied to forecast tourism flows to Macau from 8 major countries which was selected over the period 2003 to 2008. He used this model because this model able to produce accurate medium to long term forecasts and does not require discrete forecasts of the explanatory variables.

Uysal et al. (2008) investigated shaping economic policy with basic objectives of growth and high inflation were intended to determine whether they constitute an obstacle to the realization of these goals. The annual data for the period of 1950 - 2006 were used, econometric analysis was performed with the help of VAR technique. Information was given about the theoretical framework of the subject and the studies in the literature. After the information on econometric methods and results, the relationship between the general level of prices and the growth rate, which are still highly sensitive to economic and political developments.

Forecast of exchange rate is very attractive issue from early time. It is affected from a lot of factors like economical, psychological or political conditions. Mida (2013) used four fundamental economic variables. They can be listed as inflation rate, interest rate, unemployment rate and industrial production index and studied their effect on fluctuations of exchange rates by using the Vector Autoregressive Model. The Vector Autoregressive model is a model to examine multivariate time series and is used for forecasting and describing dynamic performance of economic and financial series. The VAR model and Random Walk model were compared and the result of this research the RW model is better than VAR model for period one year. On the other hand, the VAR model was better than RW in the period 3 months.

The relationship between two or more variables can be determined by correlation and regression. If the the variables are multiple and the dependent variable has explanatory variable, a VAR model is used to define this relationship. In this research, the structural relationship and determining forecast performance of a VAR model and Time Series Regression with Lagged Explanatory Variables were focused on the data of some Nigerian economic series. The data was examined and the Root Mean Square Forecast Error (RMSFE) and Mean Absolute Percentage Forecast Error (MAPFE) were used as the measurement of criteria. According to findings of this research, the VAR model was better than Time series regression with Lagged Explanatory Variables model as shown by Meta diagnostic tools (A.I. & T.O., 2013).

3. THEORETICAL KNOWLEDGE

3.1 Hypothesis Testing

Hypothesis testing is the methodical process to test statements or ideas about a group or population. It is used for learning more about the characteristic in given population by using selected samples. The aim of hypothesis testing is to define the likelihood that a given population parameter is likely to be true.

The hypothesis testing consists of four steps, respectively: define the hypothesis, adjust the criteria for a decision, compute the test statistic and make a decision. These steps are explained as follows.

The first step is to state the value of population mean in null hypothesis. Null hypothesis is a declaration about a population parameter, it presumed to be true. The null hypothesis is symbolized as H_0. On the other hand, wrong thought about the null hypothesis is called as an alternative hypothesis. The alternative hypothesis is symbolized as H_1. The alternative hypothesis is defined as it is directly contradiction with the null hypothesis. The second one is setting criteria for a decision. In this step, the level of significance is used. Level of significance is defined that the criterion of determination. This determination is about the value stated in null hypothesis. The level of significance is generally accepted 5% in emprical science. The probability of obtaining a sample mean is less than 5%, when the null hypothesis is true. The stated value in the null hypothesis is rejected, the null hypothesis is true. Next step is computing in the mathematical test statistic. It is used for making a decision about the null hypothesis. The last and forth step is making decision. It is based on probability of finding sample mean. To sum up, there are two decisions about the null hypothesis. First one is to reject the H_0 that means the probability of obtaining sample mean is less than 5%, when the null hypothesis is true. Other decision is to retain the H_0 that means that the probability of obtaining sample mean is greater than %5, when the null hypothesis is true. The last definition about hypothesis testing is p-value. p-value is positive number between 0 and 1. It is the probability of obtaining sample mean result. It is used to compare to the level of significance (Introduction to Hypothesis Testing).

3.2 Confidence Interval

One type of interval estimates is confidence interval (CI). It is computed with help of the statistics of the observed data. It may contain the true value of an unknown population parameter in the calculation. The confidence interval is an inferential statistical solution tool, which is a kind of interval estimation for a population parameter in statistical science. It is possible to estimate the value of a parameter with a single number, and there is a range of two (lower and upper limit) numbers that can cover this parameter value. Thus, confidence intervals show how reliable an estimate is (Wikipedia, n.d.).

It is calculated with this formula:

$$CI = \mu \mp Z_{\frac{a}{2}} * \frac{\sigma}{\sqrt{n}} \tag{3.1}$$

where μ is mean, a is confidence level, σ is standard deviation and n is sample size.

3.3 Correlation Functions

Correlation functions are format as a variation of statistical tests, used to define the relationship between two or more sets of data.

The autocorrelation function can used for the following two purposes:

- To specify non-randomness in data set.
- If the data are not random, specifying an appropriate time series model.

AR(p) process lag number p determine with PACF, in the same way, MA(q) process lag number q determine with ADF.

3.3.1 Autocorrelation Functions

r_t represents the values of time series at time t. Autocorrelation functions of the series show correlations between r_t and r_{t-l} for $(l = 1,2,3 \ldots)$.

The autocorrelation between r_t and r_{t-l} equals to below equation:

$$\frac{Covariance(r_t,\ r_{t-l})}{Std.dev.(r_t)Std.dev.(r_{t-l})} = \frac{Covariance(r_t,\ r_{t-l})}{Variance(r_t)} \tag{3.2}$$

The denominator in the second side appears as variance because the standard deviation of a stationary series is the same at all times (Applied Time Series Analysis, 2018).

The theoretical value of an autocorrelation of particular lag is the same through the whole series in the weakly stationary series.

3.3.2 Partial autocorrelation functions

A partial autocorrelation is seen as conditional autocorrelation in general.

The partial autocorrelation formally described as:

$$\frac{Covariance(y, r_3 \mid r_1, r_2)}{\sqrt{Variance(y \mid r_1, r_2)Variance(r_3 \mid r_1, r_2)}} \tag{3.3}$$

In this formula y = response variable and r_1, r_2, and r_3 are predictor variables (Applied Time Series Analysis, 2018).

3.4 Stationarity

In general, stationary data is preferred for goodness of fit and forecast.

Stationary processes have the feature which the mean, the variance and autocorrelation structure are stable which means do not change over time. In details, the series has no trend, constant variance and constant autocorrelation and no seasonality.

Stationary process does not change over time. A time series $\{r_t\}$ is named to be strictly stationary if the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \ldots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers. In other words that is strict stationarity requires that the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is invariant under time shift. It can be said as it is a very vigorous condition that is hard to accept in empirical studies (Tsay, 2002).

A weaker version of stationarity is often accepted in literature. A time series $\{r_t\}$ is weakly stationary if both the mean of r_t and the covariance between r_t and r_{t-l} $(l = 1,2,3 \ldots)$ are time-invariant, where l is an arbitrary integer. In particular, if the follows are true, $\{r_t\}$ is weakly stationary

(a) $(r_t) = \mu$, which is constant

(b) $v(r_t, r_{t-l}) = \gamma_l$, which only depends on l.

An asset return series is weakly stationary which is commonly assumed in the finance literature. A sufficient number of historical returns are available for checking this assumption empirically. For instance, the data can be divided into subsamples and be checked the consistency of the results obtained (Tsay, 2002).

The example of stationary and non-stationary series can be shown in Figure 3.1 and 3.2.

3.4.1 Stationarity test

A unit root has a mathematical background about its name. This process can be explained as a series of monomials basically. Each monomial corresponds to a root. It is named as unit root, when one of these roots is equal to 1.

Unit root tests are used for determining stationarity of the time series. Unit roots cause non-stationarity. In the literature, there are many tests for this aim. Dickey and Fuller's DF-test and ADF test (Dickey and Fuller, 1979) is the first two of most popular tests. Phillips-Perron test (Phillips and Perron, 1988) and KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992) can be given as examples in the same way. ADF-GLS test (Elliot, Rothenberg and Stock, 1996) and NGP test (Ng and Perron, 1995 and 2001) are used also less frequently for this aim.

The Dickey-Fuller test is the more known unit root test in the literature. It is based on linear regression. If serial correlation appears in the data, Augmented Dickey-Fuller test is preferred. The ADF handles bigger, more complex models.

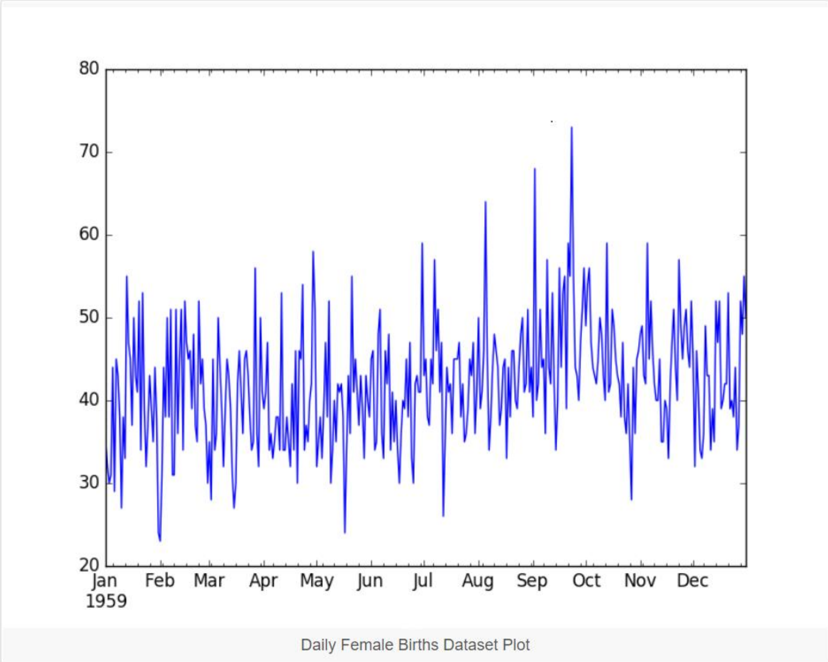

Figure 3.1 : The example of stationary data graphic (Mastery, 2019).

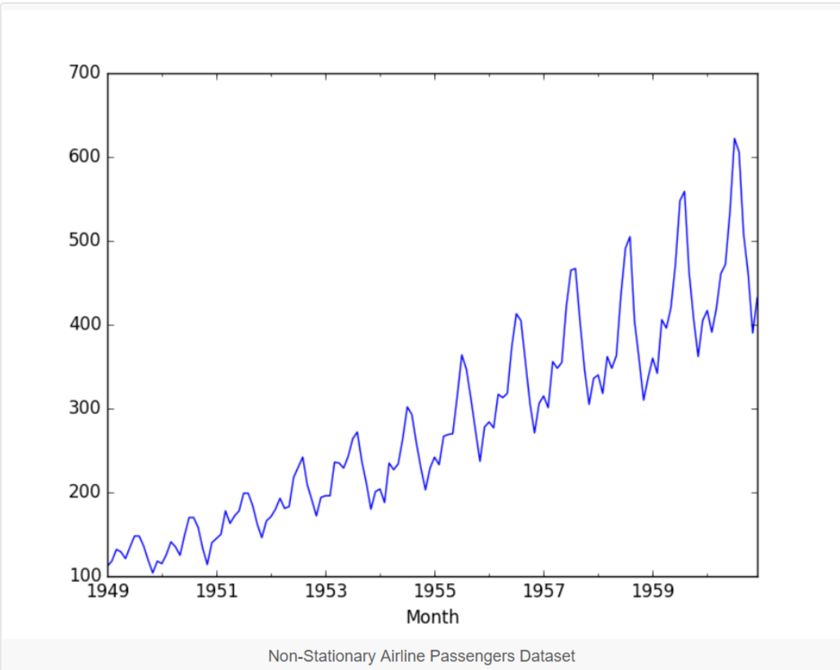

Figure 3.2 : The example of non-stationary data graphic (Mastery, 2019).

3.4.2 Granger Causality

Granger causality test is one of the tests used for determining the causal route of the relationship between two related data.

The transformation of non-stationary data into a stationary one may cause lack of information, that's why we use the Granger Causality Test to have the opportunity to use the non-stationary data for modeling.

The operational definition of Granger is based on these hypothesis: Let two related time series be X and Y, respectively. Firstly, the future cannot be the reason of the past. The cause always occurs before the conclusion. If the past causes the present or future, definitive causality occurs only. This needs a lagged between the reason and the conclusion. Secondly, causality can only be determined for stochastic processes, not two deterministic processes (Işığıçok, 1994). Thirdly, Y is named as the Granger cause of X, if the past values of X are used. Additinoally, X can be more successful than when the past values of X are not used.

If X is effected by both lagged values of X and lagged values of Y, the situation is called Y Granger causes X. In the other way, If Y is effected by both lagged values of Y and lagged values of X, it is called X Granger causes Y. However, if X is Granger cause Y and Y is Granger cause X, the causality relationship between X and Y is called as bi-directional causality. If only one exists, then it is called as uni-directional causality. If both do not exist, then the variables X and Y are independent from each other (ekolar.com, 2018).

3.4.3 Cointegration

To model multivariate time-series is complicated because of non-stationarity, in particular economic data. Engle and Granger studied about cointegration in 1980s. The stated that the linear combination of two or more non-stationary series can be stationary. If the linear combination of this type series exists, the non-stationary time series can called cointegrated. As a simple example, X_t and Y_l are both non-stationary series, the linear combination of these time series is $X_t - kY_l$ is stationary. In result, X_t and Y_l are cointegrated (Eviews, 2018).

There are a lot of tests to determine the cointegration relation as Engle-Granger, Johansen, Phillips–Ouliaris etc. The simple one is Engle-Granger test based on ADF unit root test. For cointegration, Johansen test is another test that allows for more than one cointegrating relationship, unlike the Engle–Granger method.

3.5 Modeling

3.5.1 Least square estimation

The least square is the method about forecasting paramaters by minimizing the squared inconsistencies between observed data and their expected values. If Y is response value and X is covariables, $Y = f(X) + noise$. The function f is regression function. f is estimated from sampling n covariables and their response (Van De Geer, 2005).

3.5.2 Goodness of fit

Goodness of fit test is used to control if sample data fits a distribution from a certain population. In other words, it explains you if the sample data demonstrates the data which would anticipate to find in the actual population.

A widely used statistic method is R-square which is used for measuring goodness of fit of a stationary model. It is defined as:

$$R^2 = 1 - \frac{Residual\ sum\ of\ squares}{Total\ sum\ of\ squares} \tag{3.4}$$

Adjustment R^2 is defined as:

$$Adj\_R^2 = 1 - \frac{Variance\ of\ residuals}{Variance\ of\ r_t} \tag{3.5}$$

where σ_r^2 is the sample variance of r_t (Tsay, 2002).

3.5.2.1 Akaike information criteria

There are several information criteria, which are used as a model selection guide, and the model should selected with the smallest information criterion. They are based on likelihood.

Akaike Information Criteria is one of the most known information criteria. To choose the model that minimizes the negaive likelihood penalised bu the number of parameter is the concept of AIC. The equation is shown as:

$$AIC = \frac{-2}{T}\ln(likelihood) + \frac{2}{T} \times (number\ of\ paramaters) \tag{3.6}$$

where likelihood function is evaluated at the maximum likelihood estimates and T is sample size.

3.5.2.2 Bayesian information criteria

The other commonly used information criteria is the Bayesian Information Criteria. BIC is based on a Bayesian Framework as an estimation of Bayes factor. The BIC is differ from AIC with second term in the equation.

According to Bayesian viewpoint, BIC is intended to explore the most potential model given data (Acquah, 2010).

The information criteria is used as a model selection guide and the model should selected with the smallest information criterion.

3.6 AR, MA, VAR Models

In this section, some beneficial simple econometric models for examining financial time is given such as AR, MA, VAR models.

White Noise:

If a time series $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance, it is called as a white noise series. More specifically, if the time series is normally distributed with mean zero and variance σ^2, this series is named as a Gaussian white noise (Tsay, 2002).

For the white noise series, all the ACFs are zero. If all sample ACFs are close to zero in the practice, the series are white noise series.

3.6.1 Autoregressive model (AR(p))

A simple model that makes utilization of predictive power is

$$r_t = \phi_0 + \phi_1 r_{t-1} + \alpha_t \tag{3.7}$$

where $\{\alpha_t\}$ is white noise series $((\alpha_t \approx N(0, \sigma_\alpha^2)$, the errors are independently distributed with a normal distribution that has mean 0 and constant variance. This model is in the same as simple linear regression model in where r_t is the dependent variable and r_{t-1} is explanatory variable. Actually, this equation is represent to an autoregressive model of order one which is showed as AR(1) model.

Properties of AR(1):

The mean of r_t is $\mu = \frac{\phi_0}{1-\phi_1}$

The variance of is $Var(r_t) = \frac{\sigma_\alpha^2}{1-\phi_1^2}$

The correlation between observations h time periods apart is $\rho_h = \phi_1^h$.

A simple generalization of the AR(1) model is the autoregressive model order of p which represents as AR(p) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \alpha_t \tag{3.8}$$

where p is a non-negative integer and $\{\alpha_t\}$ is white noise series with mena zero and variance σ_α^2. The AR(p) model can be explained as in the same as multiple linear regression model with lagged values helping as the explanatory variables (Tsay, 2002).

3.6.2 Moving Average model (MA(q))

Moving average model is another useful model in modeling return series in financial literature.

Let's $\alpha_t \approx N(0, \sigma_\alpha^2)$, the errors are independently distributed with a normal distribution that has mean 0 and constant variance.

The 1st order MA model, denoted by MA(1) is

$$r_t = c_0 + \alpha_t - \theta_1 \alpha_{t-1} \tag{3.9}$$

where c_0 is constant and $\{\alpha_t\}$ is white noise series.

The qth order MA model, denoted by MA(q) is

$$r_t = c_0 + \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_q \alpha_{t-q} \tag{3.10}$$

where $q > 0$.

Properties of MA(1):

The mean : $E(r_t) = c_0$

The variance of is $Var(r_t) = \sigma_\alpha^2(1 + \theta_1^2)$

The ACF is $\rho_1 = \frac{\theta_1}{1+\theta_1^2}$, and $\rho_h = 0$ for $h \geq 2$.

Note that all autocorrelations are 0 in the theoretical ACF, only for lag 1 is non-zero value.

3.6.3 Vector autoregressive model (VAR(p))

The VAR model is simple model for modeling asset returns in the literature. r_t is a multivariate time series and the VAR model order of 1 (represent as VAR(1)) follows as

$$r_t = \phi_0 + \Phi r_{t-1} + \alpha_t \tag{3.11}$$

where ϕ_0 is a k-dimensional vector, Φ is a $k \times k$ matrix, and $\{\alpha_t\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix. The covariance matrix is required to be positive definite in the literature; otherwise, the dimension of r_t may be reduced.

The following two equations include in the VAR(1) model like as:

$$r_{1t} = \phi_{10} + \Phi_{11} r_{1,t-1} + \Phi_{12} r_{2,t-1} + \alpha_{1t} \tag{3.12}$$

$$r_{2t} = \phi_{20} + \Phi_{21} r_{1,t-1} + \Phi_{22} r_{2,t-1} + \alpha_{2t} \tag{3.13}$$

where Φ_{ij} is the (i,j)$^{\text{th}}$ element of Φ and ϕ_{i0} is the i$^{\text{th}}$ element of ϕ_0.

The following equation shows VAR(p) model:

$$r_t = c + \Phi_1 r_{t-1} + \Phi_2 r_{t-2} + \cdots + \Phi_p r_{t-p} + \alpha_t \tag{3.14}$$

where r_{t-i} is the i$^{\text{th}}$ lag of r, *c* is a *k*-vector of constants, Φ_i is a time-invariant ($k \times k$)-matrix and α_t is a *k*-vector of error terms.

3.7 Moving Average Convergence Divergence (MACD)

Technical analysis is the process of examining a recent prices to determine possible future prices. This issue is done by comparing the current price and recent price to anticipate a suitable outcome.

In technical analysis, MACD is the Moving Average Convergence Divergence indicator. The definition and types of moving averages is given below before defining MACD.

One of the oldest and most popular technical analysis tools is moving average. A moving average is the average of prices at a given time. There are five widespread types of moving average: simple, exponential, triangular, variable and weighted. The important difference between these MAs is weight to the most recent data. Simple MA perform equal weight to the prices. Exponential MA performs more weight to recent data.

A simple moving average is calculated by adding the prices n time periods and then dividing by n. For example, add the most recent 10 days prices and then divide by 10. The result is the arithmetic average price of the last 10 days.

$$Simple\ MA\ = \frac{\sum_{1}^{n} closing\ price}{n} \tag{3.15}$$

An exponential moving average reduces the lag by applying more weight to recent data. The EMA is calculated by calculating a percentage of today's closing price to yesterday's moving average value. A given day's EMA calculation depends on the EMA calculations for all the days prior to that day. First, calculation initial EMA value with the simple moving average is needed. A simple moving average is used to calculate first value of exponential moving average; it is used as the previous period's EMA in the first calculation. Then, the weighting multiplier is calculated. Final step is calculating the exponential moving average for each day between the initial EMA value and today.

$$Initial\ SMA\ = \frac{\sum_{1}^{n} closing\ price}{n} \tag{3.16}$$

$$Multiplier\ = 2/(time\ period\ + 1) \tag{3.17}$$

$$EMA\ = \{EMA(today) \times multiplier\} + EMA(previous\ day) \times (1 - Multiplier) \tag{3.18}$$

Investors generally buy when the price rises above its moving average and sell when the price falls below its moving average.

The MACD is one of the most popular trend-following momentum indicator that shows the relationship between two moving averages of prices. The MACD was generated by Gerald Appel in 1970s (Yazdi & Lashkari, 2013).

The MACD is calculated by subtracting a 12-days MA and 26-days MA.

$$MACD = EMA\,(12 - day\ closing\ prices) - EMA\,(26 - day\ closing\ prices) \tag{3.19}$$

$$Signal = EMA\,(9 - day\ MACD) \tag{3.20}$$

This indicator swings above and below zero. MACD is above the zero, if 12-days EMA is higher than 26-days EMA. This situation is named as bullish, it implies that shift in the supply or demand lines. MACD is below the zero, if the 12-days EMA is smaller than 26-days EMA. This situation is named as bearish; it implies a bearish or downward, shift in the supply or demand lines. A 9-days MA of the MACD is named as the signal line, which predicts the convergence of two moving averages.

Another indicator is histogram, is difference between MACD and signal by solid block. Analysis of MACD is easy for traders. When MACD is above its 9-days EMA, histogram is positive. When the MACD is below its 9-days EMA, histogram is negative. The histogram was developed in 1986 by Thomas Aspray (Huang & Kim, 2006). The example of MACD is shown in Figure 3.3.

Figure 3.3 : The example of MACD (.dash, 2019).

The MACD is most effective in wide-swinging trading markets. There are three methods to use the MACD. These can be listed as follows: crossovers, overbought-oversold conditions, and divergences.

Crossovers: The basic rule of MACD is to sell action when the MACD falls below its signal line. Identically, when the MACD rises above its signal line, buy action consists.

Overbought-Oversold Conditions: When the shorter MA move back form the longer EMA, it means that the price is exceed its limit and will soon turn back to levels that are more acceptable.

Divergences: A bearish divergence is shown when the MACD is making new lows while prices fail to reach new lows. A bullish divergence is shown when the MACD is making new highs while prices fail to reach new highs (Achelis, 1995).

3.8 Forecasting

Forecasting is a significant process of time series. Forecasting is the process of making predictions of the future by using the past and present data and generally by analysis of trends.

3.8.1 Forecasting AR model

One of the most important of time series analysis is forecasting. In Equation 3.8, AR(p) general formaula is given. The autoregression can be used to forecast an arbitrary number of periods into the future. Firstly, time index h and interested in forecasting r_{h+l} where $l \geq 1$. Time index h is named as forecast origin and l named as forecst horizon. $\hat{r}_h(l)$ is the forecast of r_{h+l}usssing the minimum squared error loss function and F_h is the collection of information avaliable at forecast origin h. Then, the forecast $\hat{r}_k(l)$ is chosen such that

$$E\{[r_{h+l} - \hat{r}_h(l)]^2 \mid F_h\} \leq min_g \, E\,[r_{h+l} - g\,]^2 \mid F_h] \tag{3.21}$$

where g is a function of the information available at time h, that is a function of F_h (Tsay, 2002).

3.8.2 Forecasting VAR model

Constructing VAR model includes three steps. First one of all, use the test statistic or information criterion to determine the order. Second one is guess the specified model by using least square estimation. Last one is the statistic of residuals to check adequacy of a fitted model. If the fitted model is suitable, then it can be used to make forecasts and make outcome respecting the dynamic relationship between variables (Tsay, 2002).

4. NEW MODEL EMAVAR

The new model called EMAVAR (Exponential Moving Average Vector Autoregressive) is inspired by the VAR model and the MACD indicator. MACD is one of the most common technical indicator in technical analysis. It gives the direction of movement with using recent data. In addition, it provides more meaningful data by calculation exponential moving averages. For this reason, MACD is used in this thesis and more details about MACD is given in Section 3.7. VAR model is one of most successful, flexible and easy to use model in the literature about analysis of multivariate time series. The VAR model has proven to be useful. Because, it describes the dynamic behavior of multivariate time series. It can be made conditional on the potential future paths of specified variable in the model so it is flexible (Vector Autoregressive Models for Multivariate Time Series). In the result of these thoughts, The new model EMAVAR is occurred. In EMAVAR, the data is not modeled by means of different exogenous variables (as the VAR model), but by the different EMAs of the data itself. The EMAs obtain by MACD and used in VAR process for forecasting.

5. EMAVAR and AR MODELS APPLICATION

5.1 Data

The data used in this project are BIST-100 closing index and BIST-100 banking sector index between 2012 and 2018. The data are obtained from KAP website as five-days work week. Missing values appear by cause of public holidays. Modeling the data with missing values may not give a good forecast. That's why the missing values are completed by linear interpolation. Details of the linear interpolation is given in Section 5.1.1.

5.1.1 Linear interpolation

Interpolations techniques are used to anticipate the missing values in data set. The linear interpolation is the simplest for of these techniques. In the linear interpolation, two data points are connected each other with the straight line (Noor, Yahaya, Ramli, & Mustafa Al bakri, 2014). The equations of the linear interpolation function can be written as following:

$$f_1(x) = b_0 + b_1 (x - x_0) \tag{5.1}$$

where *x* is the independent variable, x_0 is the inital value of independent variable *x* and $f_1(x)$ is the value of the dependent variable for *x*. Then we get following equations:

$$b_0 = f(x_0) \tag{5.2}$$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{5.3}$$

5.2 Data arrengement for modeling

The data of XU100 and XBank get from KAP are from January, 2012 to June, 2018. First of all, MACD is applied on the 6 years. After then, with the purpose of seeing other effects, to use the part of 1 year and 2 years of this data set is decided. 1 year includes from May, 2017 to June, 2018, 2 years include May, 2016 to June, 2018 and 6 years include all of them. Linear interpolation is used to eliminate missing data and data set is converted as full 5-weekdays. Firstly, the linear interpolation is applied on all data. In addition, this process is applied on 1 year and 2 years, seperately. However, the result does not change. Therefore, the all data which is applied linear interpolation in the first step for eliminating missing days and it is used in the other steps by divided wanted date arrangement.

The data sets with missing values is shown in Figure 5.1 and 5.3, and completed data sets in Figure 5.2 and 5.4.

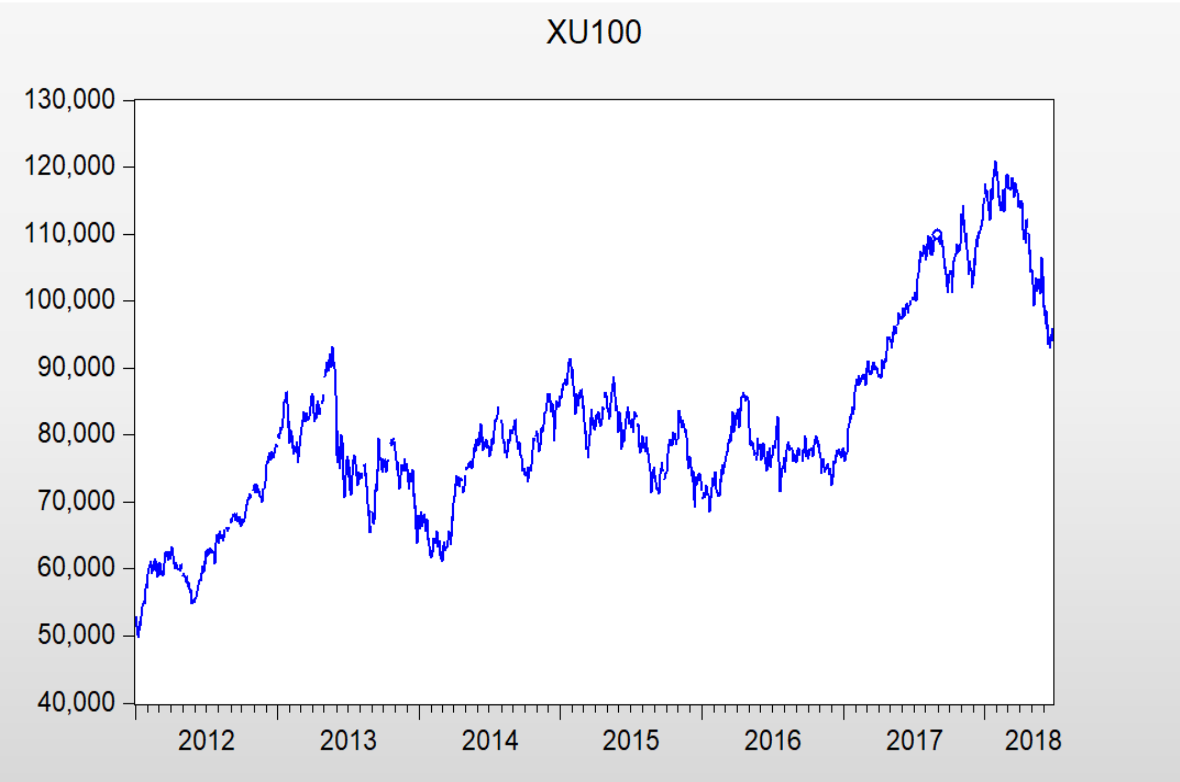

Figure 5.1 : Full data of XU100 with missing values (6 years).

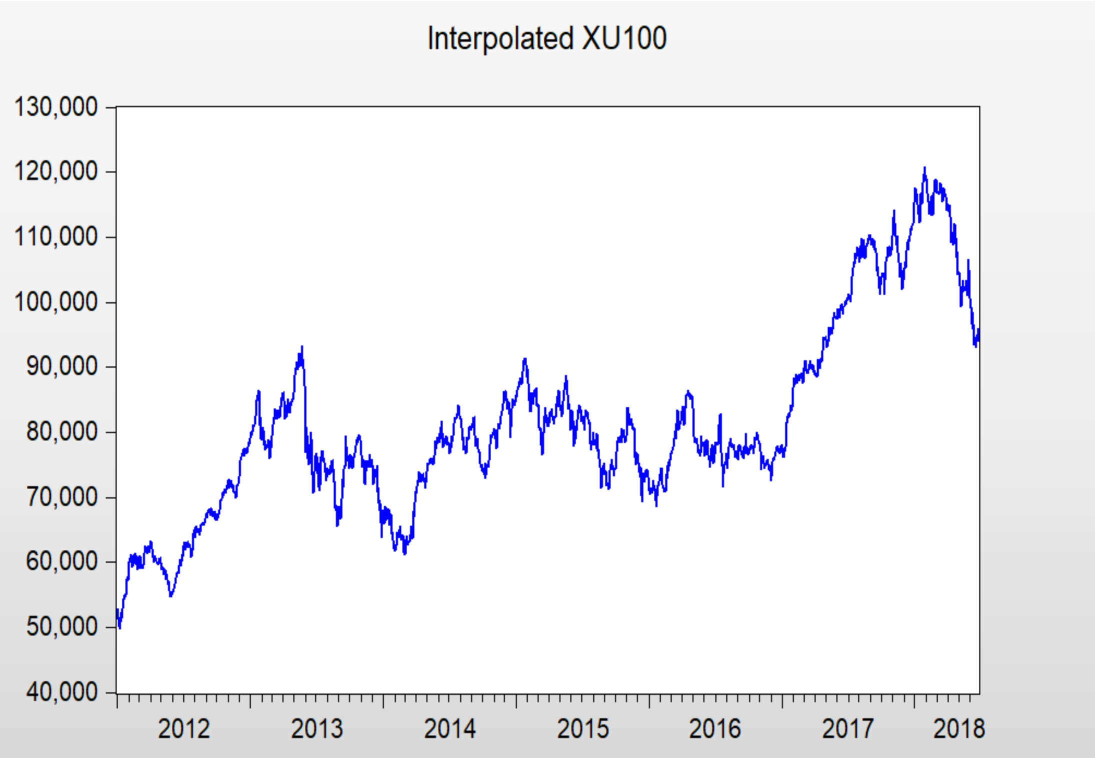

Figure 5.2 : Completed data of XU100 by linear interpolation (6 years).

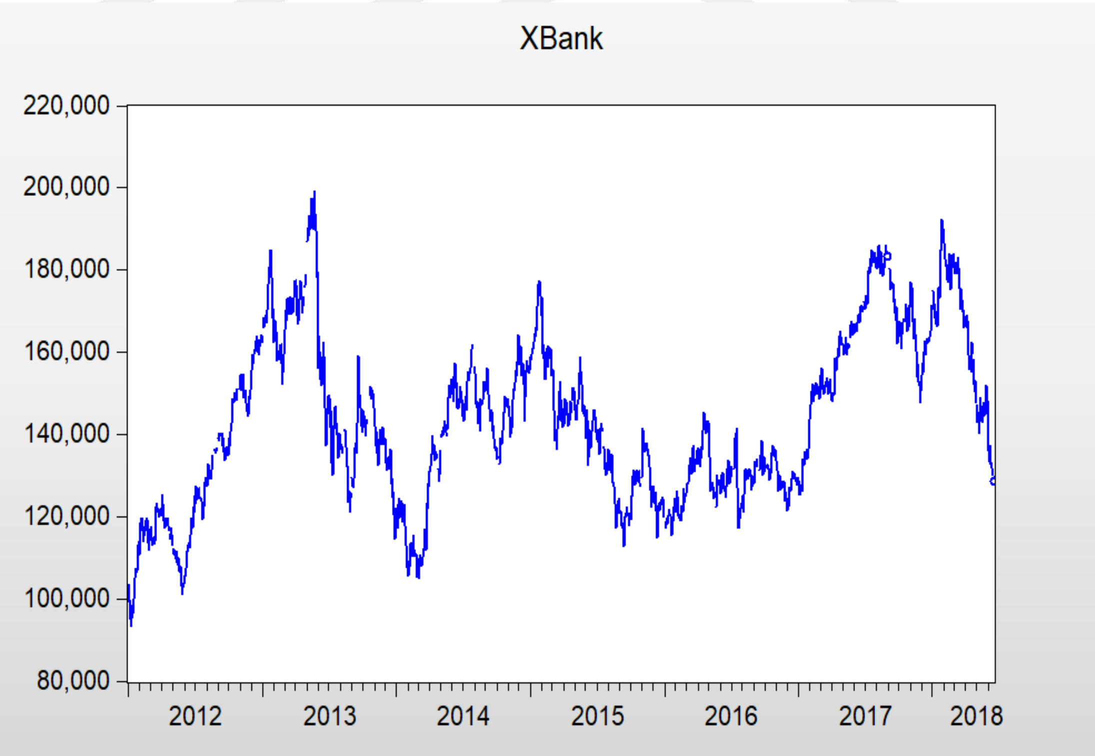

Figure 5.3 : Full data of XBank with missing values (6 years).

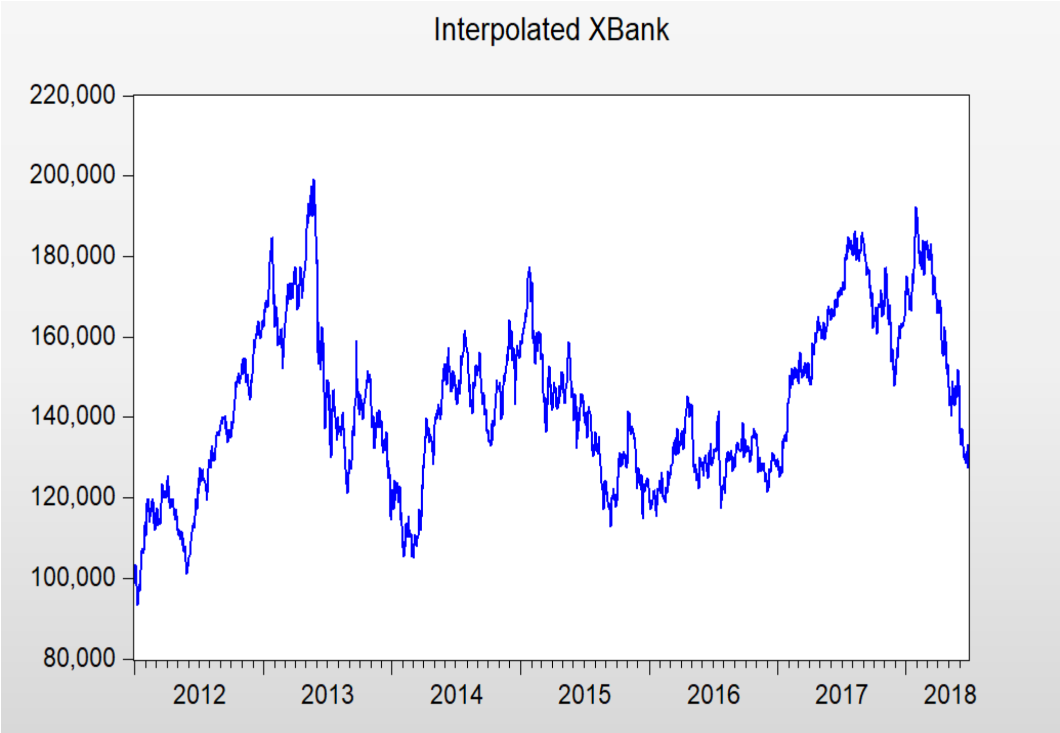

Figure 5.4 : Completed data of XBank by linear interpolation (6 years).

5.3 AR and EMAVAR Models For XU100

In this section, forecasting of selected data set about XU100 with AR(1) and EMAVAR models are told. The details about of data arrengement is given in Section 5.2. To compare the models, AR(1) model is applied on the different data arrangement of XU100 for forecasting next-one day. After that, thought with more well model, EMAVAR, is applied on the same data. The details about of AR and EMAVAR models are given in Section 3.6.1 and 4, respectively. After this section, interpolated XU100 is named as XU100 shortly and interpolated XBank is named as XBank shortly in the same way.

1 year and 2 years are given in Figure 5.5 and Figure 5.6, respectively. 6 years of XU100 is given in Figure 5.2.

For EMAVAR, these data are used in MACD process. By the result of MACD process, macd and signal series are obtained. It is shown as below in Figure 5.7, 5.8, 5.9, respectively for 1 year, 2 years and 6 years.

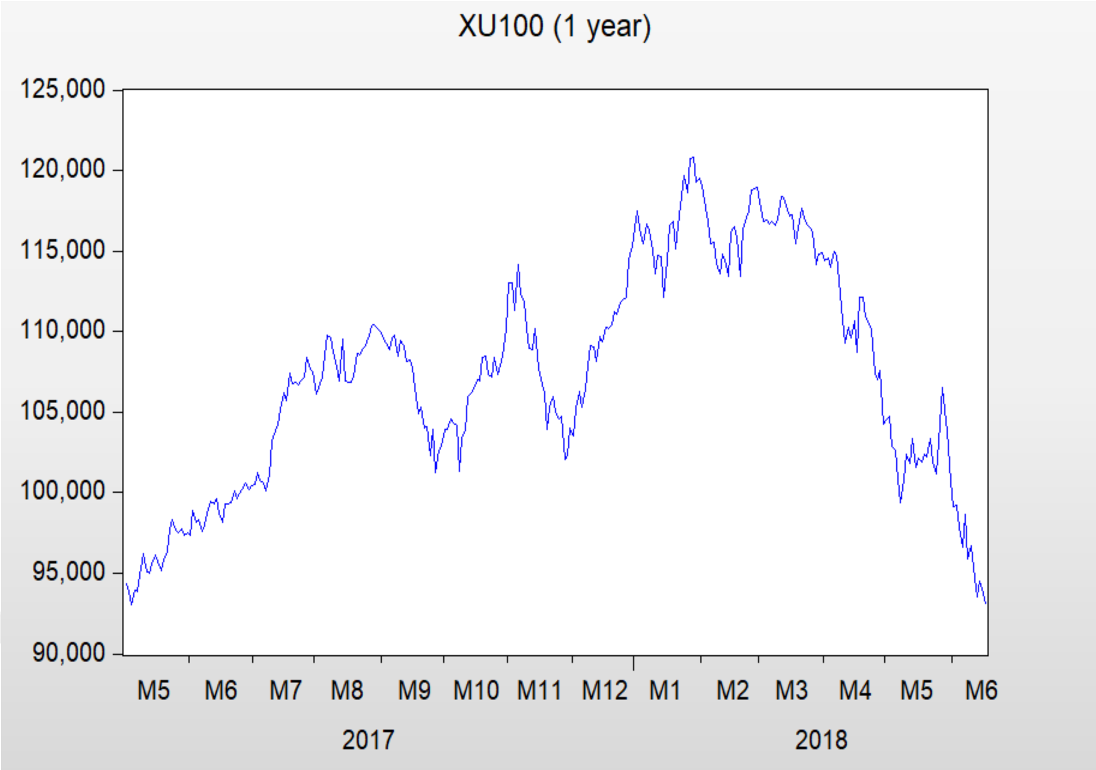

Figure 5.5 : 1 year of XU100.

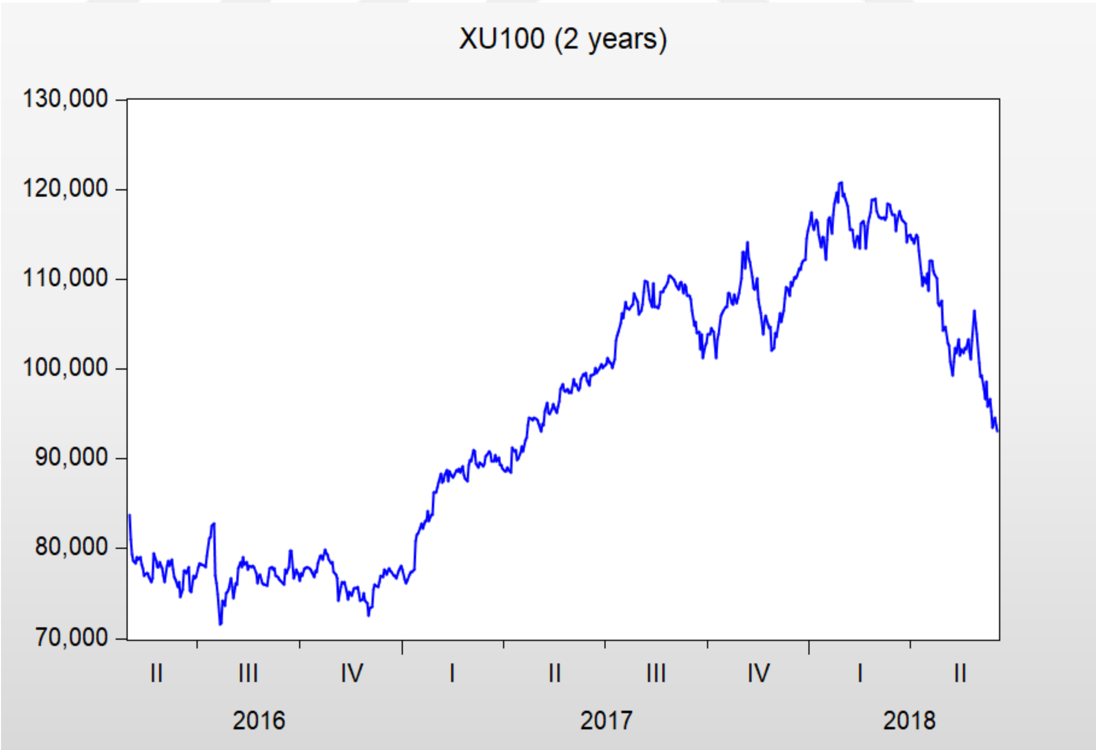

Figure 5.6 : 2 years of XU100.

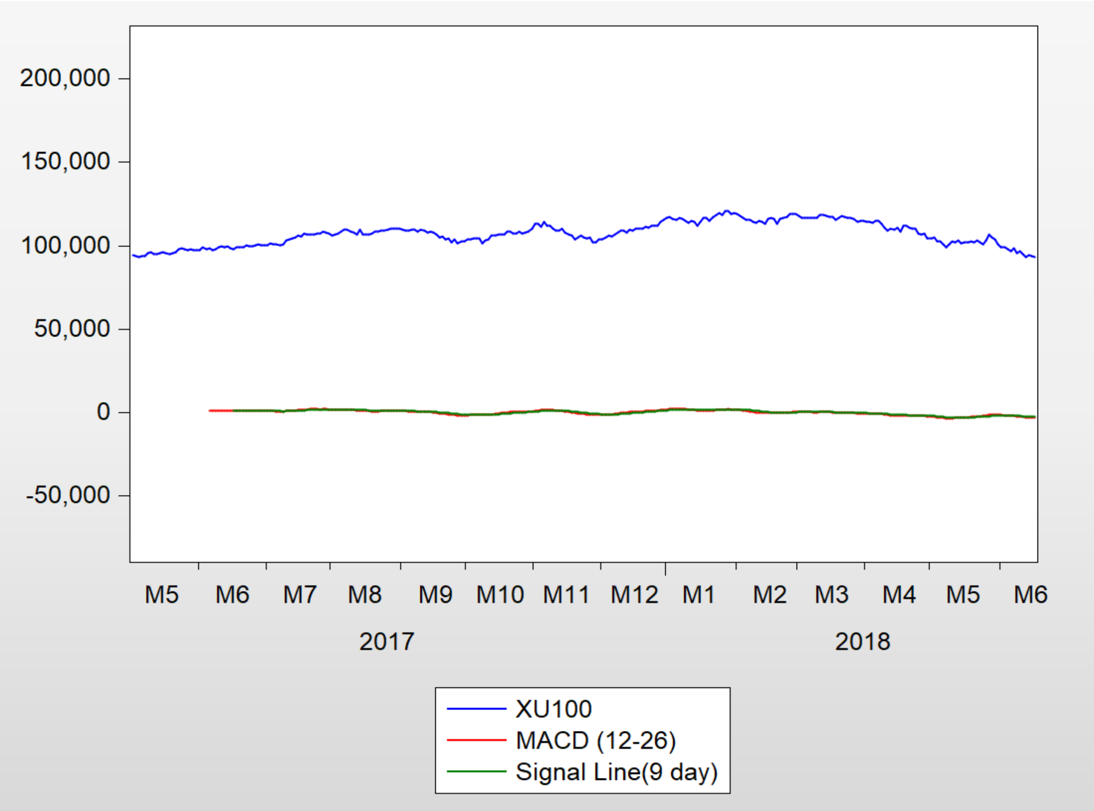

Figure 5.7 : The obtained series from MACD of XU100 (1 year).

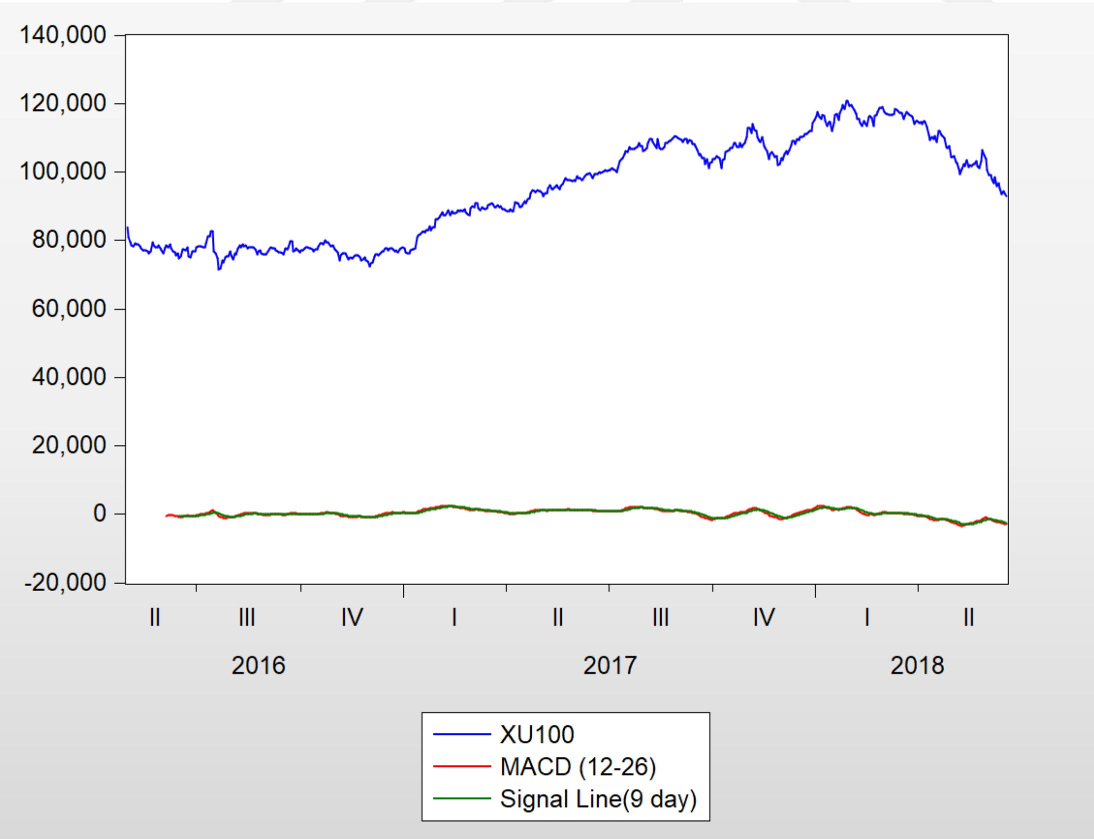

Figure 5.8 : The obtained series from MACD of XU100 (2 years).

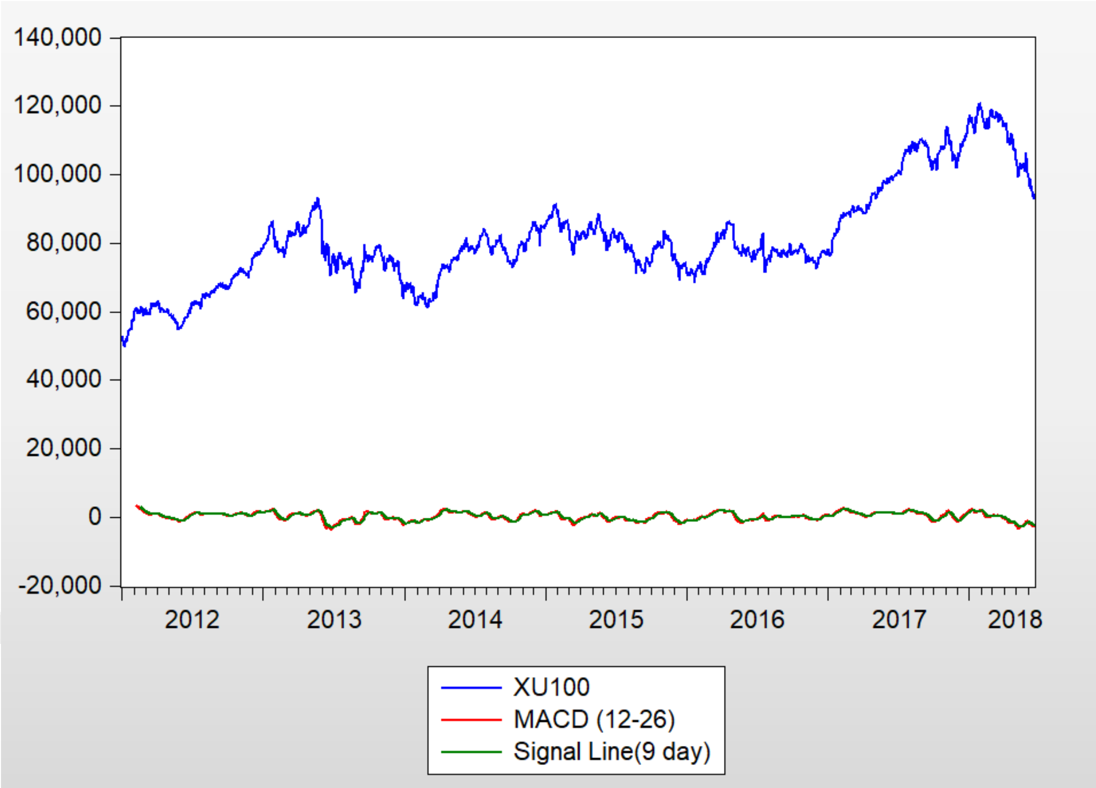

Figure 5.9 : The obtained series from MACD of XU100 (6 years).

After EMA process in MACD appliance, the data set size has been reduced because the VAR process will be applied. It is shown in graphics. In the result, all of them occur from June and June. The aim of this investigation, which is to apply the model on the different data parts, to observe the effect of historical data and past recent data on forecasting values. Macd and signal series is used as exogeneous variables in VAR process.

In VAR process, macd and signal series are stationary, and XU100 serie is non-stationary in 6 years. Therefore, the cointegration and VECM are investigated.

While the same process is applying 1 year and 2 years data, non-stationarity of macd and signal series are detected with Unit Root Test. The process has been continued with utilizing first differences and anticipate one-next day. First differences of series are used to eliminate non-stationarity.

The result of Unit Root Test is decided through null hypothesis. An example of macd series can show in Figure 5.10. Null hypothesis is not rejected because the probability value is greater than 0,05. Thus, macd series is non-stationary. In the other Figure 5.11, the stationarity for first difference of example series. Null hypothesis is rejected because probability value is smaller than 0,05.

The findings obtained from this test, it is understood to use first differences of series. These first differences is used in VAR model and controlled the correlogram of the first differences series. An example of correlogram presentation is shown in Figure 5.12. Then forecast is done with first differences.

Augmented Dickey-Fuller Unit Root Test on MACD__12_26_

Null Hypothesis: MACD__12_26_ has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.837377	0.3619
Test critical values:	1% level	-3.455387	
	5% level	-2.872455	
	10% level	-2.572660	

Figure 5.10 : An example for Unit Root Test macd series (1 year).

Augmented Dickey-Fuller Unit Root Test on D(MACD__12_26_)

Null Hypothesis: D(MACD__12_26_) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.933477	0.0000
Test critical values:	1% level	-3.455387	
	5% level	-2.872455	
	10% level	-2.572660	

Figure 5.11 : An example for Unit Root Test 1st difference of macd series (1 year).

.

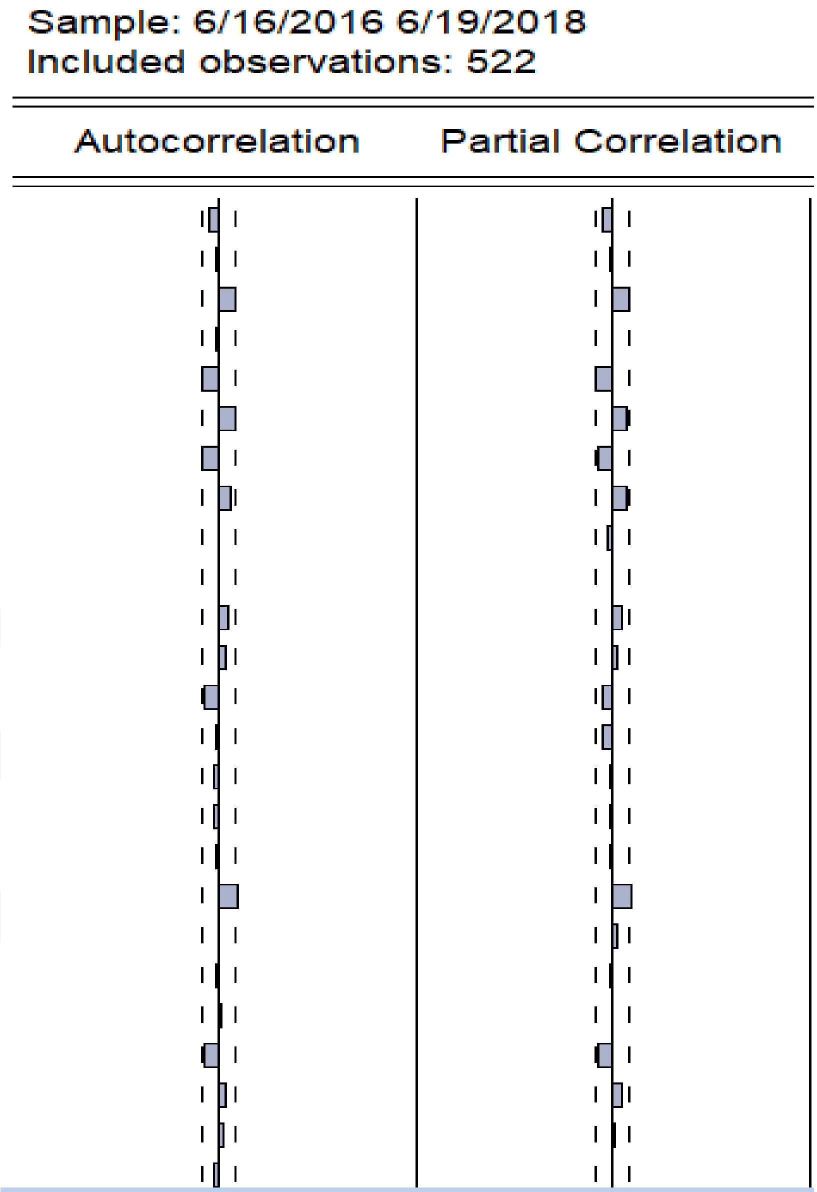

Figure 5.12 : An example of correlogram of firs difference of close (2 years).

5.4 AR and EMAVAR Models For XBank

Afterwards, the same process is applied on XBank data with the same data arrengement. The aim is to show that the model can be applied in other financial data like XBank.

6 years of XBank is given in Figure 5.4. In Figure 5.13 and 5.14, XBank data set is given, respectively.

Macd and signal series are used as exogeneous variable in EMAVAR model with XBank as seen in Figure 5.15, Figure 5.16 and Figure 5.17.

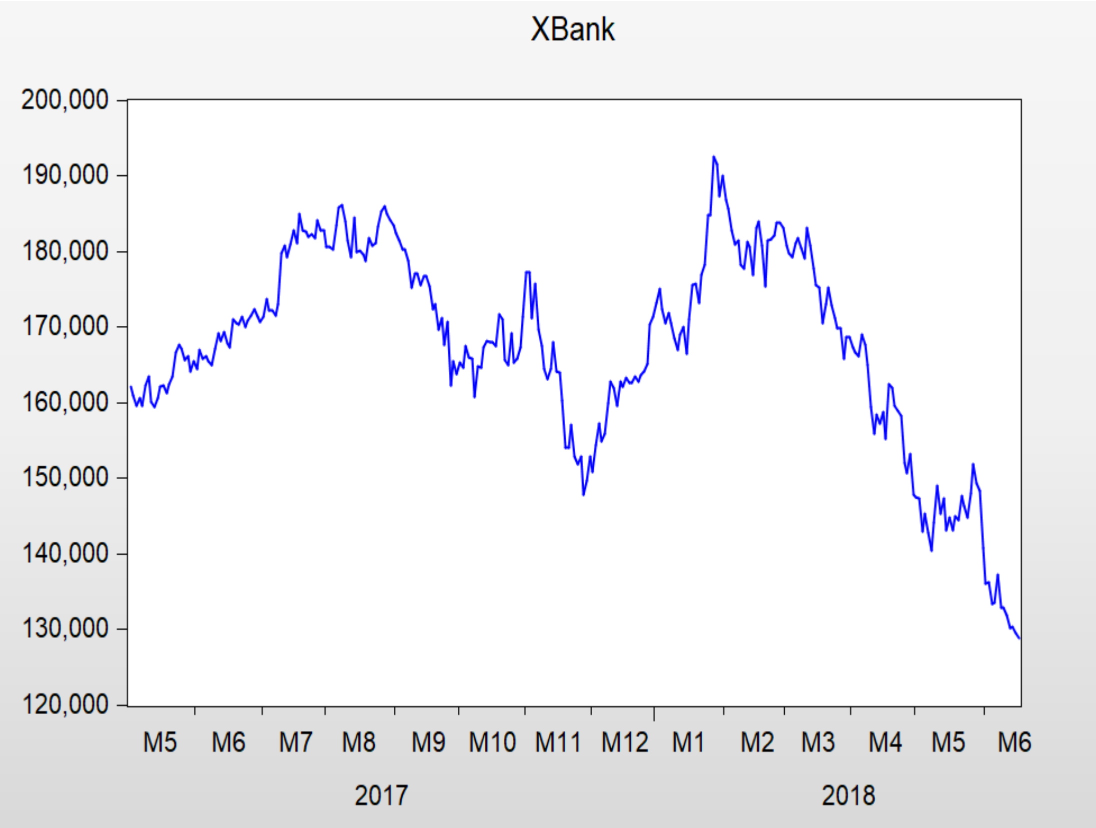

Figure 5.13 : 1 year of XBank.

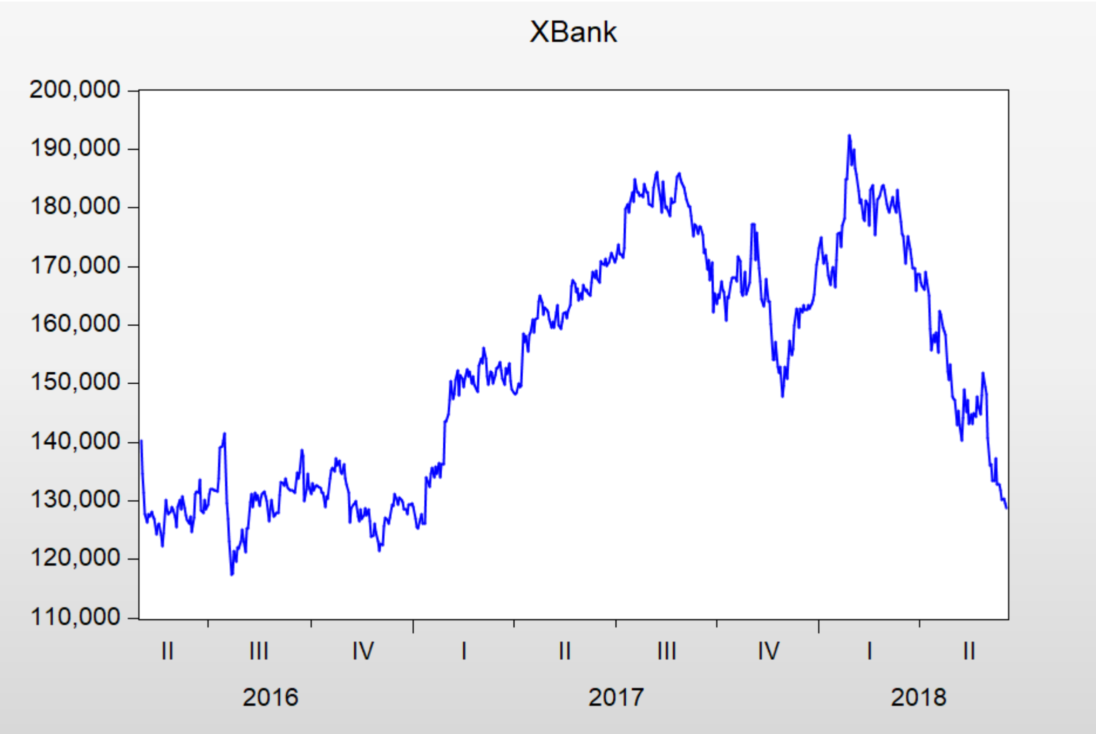

Figure 5.14 : 2 year of XBank.

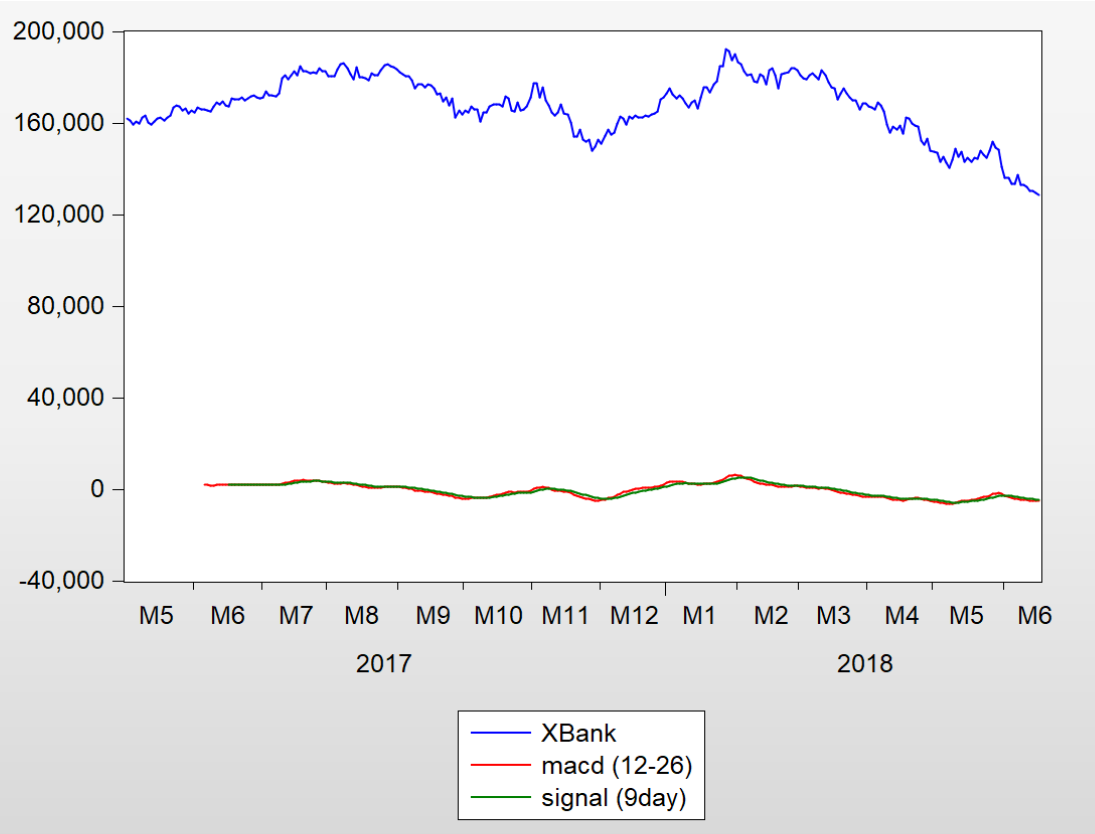

Figure 5.15 : The obtained series from MACD of XBank (1 year).

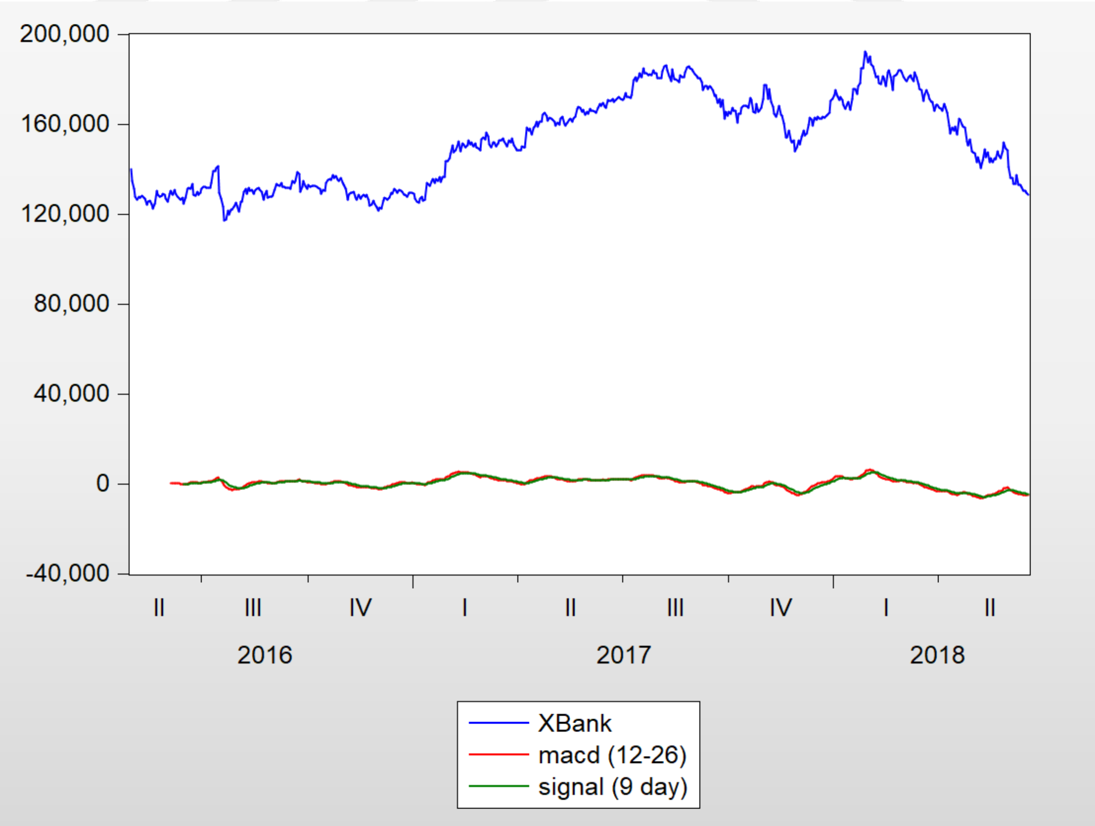

Figure 5.16 : The obtained series from MACD of XBank (2 years).

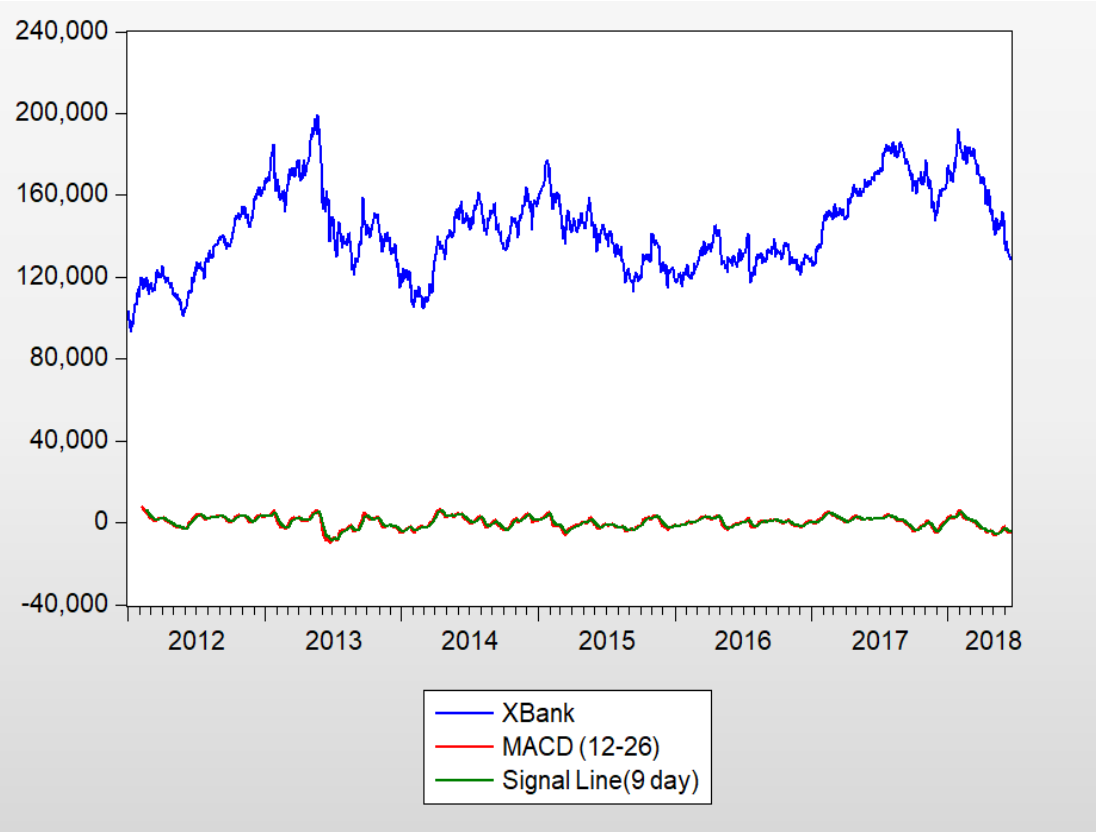

Figure 5.17 : The obtained series from MACD of XBank (6 years).

Obtained results are presented as a table in Section 6.

6. FORECASTING APPLICATION

After applying respectively AR and EMAVAR models on XU100 and XBank data, a forecast has been done for each model for 6 years, 2 years and 1 year data. Here below, in Table 6.1 and Table 6.2, one can find the actual, forecasted values and error percentage for AR model and all the same things for EMAVAR model in Table 6.3 and 6.4.

Table 6.1 : The result of AR model for XU100.

Period	Real Value	Forecasted Values	Error(%)
1 year	94436,87	93776,51	0,69926
2 year	94436,87	93776,55	0,69922
6 year	94436,87	93775,09	0,70076

Table 6.2 : The result of AR model for XBank.

Period	Real Value	Forecasted Value	Error(%)
1 year	131048,81	129521,49	1,16
2 year	131048,81	129517,54	1,16
6 year	131048,81	129511,70	1,17

After these results, forecasts again have been done for 1, 3 and 5 days, in different time intervals to be sure about the error percentage. Here below, all the results with the related time intervals have been given in Table 6.5, 6.6, 6.7 and 6.8.

Table 6.3 : The result of EMAVAR model for XU100.

Period	Real Value	Forecasted Value	Error(%)
1 year	94436,87	93787,30	0,68
2 year	94436,87	93973,61	0,49
6 year	94436,87	93090,08	1,42

Table 6.4 : The result of EMAVAR model for XBank

Period	Real Value	Forecasted Value	Error(%)
1 year	131048,81	128686,21	1,8
2 year	131048,81	129788,85	0,96
6 year	131048,81	128856,76	1,6

.

Table 6.5 : The result of AR model for multiple day forecast for XU100

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
03/02/2014 – 02/02/2015	1 day	3.02.2015	87508,52	89706,65	2,51
		3.02.2015	87508,52	89706,65	
	3 days	4.02.2015	85910,14	89700,52	3,82
		5.02.2015	85789,31	89694,38	
		3.02.2015	87508,52	89706,65	
		4.02.2015	85910,14	89700,52	
	5 days	5.02.2015	85789,31	89694,38	4,60
		6.02.2015	84987,42	89688,25	
		9.02.2015	84574,27	89682,11	
04/02/2013 – 02/02/2015	1 day	3.02.2015	87508,52	89616,66	2,41
		3.02.2015	87508,52	89616,66	
	3 days	4.02.2015	85910,14	89584,38	3,68
		5.02.2015	85789,31	89545,52	
		3.02.2015	87508,52	89616,66	
		4.02.2015	85910,14	89584,38	
	5 days	5.02.2015	85789,31	89545,52	4,21
		6.02.2015	84987,42	89505,97	
		9.02.2015	84574,27	89465,78	

Table 6.5 (continued) : The result of AR model for multiple day for XU100

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
15/06/2017 – 18/06/2018	1 day	19.06.2018	94436,87	93035,57	1,48
	3 days	19.06.2018	94436,87	93035,57	1,75
		20.06.2018	94551,86	93027,85	
		21.06.2018	95057,23	93020,14	
	5 days	19.06.2018	94436,87	93035,57	1,86
		20.06.2018	94551,86	93027,85	
		21.06.2018	95057,23	93020,14	
		22.06.2018	95852,11	93012,43	
		25.06.2018	94008,29	93004,71	
15/06/2016 – 18/06/2018	1 day	19.06.2018	94436,87	93035,59	1,48
	3 days	19.06.2018	94436,87	93035,59	1,75
		20.06.2018	94551,86	93027,89	
		21.06.2018	95057,23	93020,20	
	5 days	19.06.2018	94436,87	93035,59	1,86
		20.06.2018	94551,86	93027,89	
		21.06.2018	95057,23	93020,20	
		22.06.2018	95852,11	93012,51	
		25.06.2018	94008,29	93004,82	

Table 6.6 : The result of AR model for multiple day for XBank

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
03/02/2014 – 02/02/2015	1 day	3.02.2015	166210,94	173544,04	4,41
	3 days	3.02.2015	166210,94	173544,04	6,76
		4.02.2015	159693,00	173517,22	
		5.02.2015	161689,02	173490,40	
	5 days	3.02.2015	166210,94	173544,04	7,95
		4.02.2015	159693,00	173517,22	
		5.02.2015	161689,02	173490,40	
		6.02.2015	158380,73	173463,59	
		9.02.2015	157586,32	173436,78	
4/02/2013 – 02/02/2015	1 day	3.02.2015	166210,94	173542,52	4,41
	3 days	3.02.2015	166210,94	173542,52	6,76
		4.02.2015	159693,00	173514,17	
		5.02.2015	161689,02	173485,82	
	5 days	3.02.2015	166210,94	173542,52	7,46
		4.02.2015	159693,00	173514,17	
		5.02.2015	161689,02	173485,82	
		6.02.2015	158380,73	173457,48	
		9.02.2015	157586,32	173429,15	

Table 6.6 (continued): The result of AR model for multiple day for XBank

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
15/06/2017 – 18/06/2018	1 day	19.06.2018	131048,81	128737,14	1,76
	3 days	19.06.2018	131048,81	128737,14	1,98
		20.06.2018	131428,71	128715,48	
		21.06.2018	131481,48	128693,84	
	5 days	19.06.2018	131048,81	128737,14	1,67
		20.06.2018	131428,71	128715,48	
		21.06.2018	131481,48	128693,84	
		22.06.2018	133084,27	128672,19	
		25.06.2018	127352,51	128650,55	
15/06/2016 – 18/06/2018	1 day	19.06.2018	131048,81	128735,17	1,77
	3 days	19.06.2018	131048,81	128735,17	1,99
		20.06.2018	131428,71	128711,56	
		21.06.2018	131481,48	128687,96	
	5 days	19.06.2018	131048,81	128735,17	1,67
		20.06.2018	131428,71	128711,56	
		21.06.2018	131481,48	128687,96	
		22.06.2018	133084,27	128664,35	
		25.06.2018	127352,51	128640,75	

Table 6.7 : The result of EMAVAR model for multiple day for XU100

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
03/02/2014 – 02/02/2015	1 day	3.02.2015	87508,52	89678,66	2,48
		3.02.2015	87508,52	89678,66	
	3 days	4.02.2015	85910,14	89645,46	3,75
		5.02.2015	85789,31	89613,26	
		3.02.2015	87508,52	89678,66	
		4.02.2015	85910,14	89645,46	
	5 days	5.02.2015	85789,31	89613,26	4,50
		6.02.2015	84987,42	89582,08	
		9.02.2015	84574,27	89551,95	
4/02/2013 – 02/02/2015	1 day	3.02.2015	87508,52	89616,66	2,41
		3.02.2015	87508,52	89616,66	
	3 days	4.02.2015	85910,14	89584,38	3,68
		5.02.2015	85789,31	89545,52	
		3.02.2015	87508,52	89616,66	
		4.02.2015	85910,14	89584,38	
	5 days	5.02.2015	85789,31	89545,52	4,18
		6.02.2015	84987,42	89505,97	
		9.02.2015	84574,27	89465,78	

Table 6.7 (continued): The result of EMAVAR model for multiple day for XU100

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
15/06/2017 – 18/06/2018	1 day	19.06.2018	94436,87	93071,61	1,45
	3 days	19.06.2018	94436,87	93071,61	1,75
		20.06.2018	94551,86	93020,06	
		21.06.2018	95057,23	92991,18	
	5 days	19.06.2018	94436,87	93071,61	1,88
		20.06.2018	94551,86	93020,06	
		21.06.2018	95057,23	92991,18	
		22.06.2018	95852,11	92968,90	
		25.06.2018	94008,29	92952,36	
15/06/2016 – 18/06/2018	1 day	19.06.2018	94436,87	93152,33	1,36
	3 days	19.06.2018	94436,87	93152,33	1,55
		20.06.2018	94551,86	93213,51	
		21.06.2018	95057,23	93282,41	
	5 days	19.06.2018	94436,87	93152,33	1,58
		20.06.2018	94551,86	93213,51	
		21.06.2018	95057,23	93282,41	
		22.06.2018	95852,11	93351,68	
		25.06.2018	94008,29	93419,28	

Table 6.8 : The result of EMAVAR model for multiple day forecast for XBank

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
03/02/2014 – 02/02/2015	1 day	3.02.2015	166210,94	173357,4685	4,30
	3 days	3.02.2015	166210,94	173357,4685	6,53
		4.02.2015	159693,00	173149,9153	
		5.02.2015	161689,02	172948,3145	
	5 days	3.02.2015	166210,94	173357,4685	7,62
		4.02.2015	159693,00	173149,9153	
		5.02.2015	161689,02	172948,3145	
		6.02.2015	158380,73	172752,7347	
		9.02.2015	157586,32	172563,2094	
4/02/2013 – 02/02/2015	1 day	3.02.2015	166210,94	173280,9733	4,25
	3 days	3.02.2015	166210,94	173280,9733	6,43
		4.02.2015	159693,00	172986,5952	
		5.02.2015	161689,02	172687,5889	
	5 days	3.02.2015	166210,94	173280,9733	7,08
		4.02.2015	159693,00	172986,5952	
		5.02.2015	161689,02	172687,5889	
		6.02.2015	158380,73	172383,8799	
		9.02.2015	157586,32	172075,4644	

Table 6.8 (continued): The result of AR model for multiple day for XBank

Sampling Data	Period	Forecasted Date	Real Value	Forecasted Value	Error
15/06/2017 – 18/06/2018	1 day	19.06.2018	131048,81	128686,2134	1,80
	3 days	19.06.2018	131048,81	128686,2134	2,13
		20.06.2018	131428,71	128523,3296	
		21.06.2018	131481,48	128368,8241	
	5 days	19.06.2018	131048,81	128686,2134	1,92
		20.06.2018	131428,71	128523,3296	
		21.06.2018	131481,48	128368,8241	
		22.06.2018	133084,27	128211,3522	
		25.06.2018	127352,51	128052,6743	
15/06/2016 – 18/06/2018	1 day	19.06.2018	131048,81	128916,186	1,63
	3 days	19.06.2018	131048,81	128916,186	1,79
		20.06.2018	131428,71	128966,2258	
		21.06.2018	131481,48	129016,7729	
	5 days	19.06.2018	131048,81	128916,186	1,43
		20.06.2018	131428,71	128966,2258	
		21.06.2018	131481,48	129016,7729	
		22.06.2018	133084,27	129056,5118	
		25.06.2018	127352,51	129087,0413	

7. RESULTS AND DISCUSSION

BIST-100 closing index (5 days a week), and BIST-100 banking sector index (5 days a week) between 2012-2018 are used in this project. AR model and the new EMAVAR has been applied respectively to 6, 2 and 1 year data.

Results are shown in Tables 6.1, 6.2, 6.3 and 6.4. One can notice that there is no a major difference between 6-2 and 1 year forecasted data using AR model. But it is also noticed that the 2 years data forecasts have less percentage error in both models. The percentage errors for 6-2 and 1 year are different in EMAVAR model. This fact can be explained as the EMAVAR model is more sensitive. EMAVAR model gives 0,49 percentage error for 2 years XU100 data and 0,96 percentage error for 2 years XBank data. While AR model gives 0,69 and 1,16 percentage error, respectively. After these results, forecasts have been done for 1, 3 and 5 days, in different time intervals to be sure about the error percentage. In same way, the 2 years forecast is more good outcome again. This difference can be commented as Exponential Moving Average gives weight to recent data, this fact helps to catch the pattern. Other comments if any are left to the financers.

Finally, It can be said that EMAVAR is a more efficient model than AR for the selected data.

8. CONCLUSION

In this thesis, AR and EMAVAR model are compared. The aim is to compare the efficency of these two models in forecasting. The new model EMAVAR is inspired by technical indicator MACD and the most common financial model VAR. Two different financial data in 3 different time periods are used to analyze the behaviour of the models.

BIST-100 closing index (5 days a week), and BIST-100 banking sector index (5 days a week) data between 2012-2018 are used for modeling and forecasting. Firstly, the raw data is interpolated because of missing data in holidays. AR and EMAVAR models are applied respectively to the interpolated data.

The new model EMAVAR gives the better forecasting result in 2 year for XU100 and XBank data.

REFERENCES

.dash. (2019). Retrieved from Investpoedia: https://www.investopedia.com/terms/m/macd.asp

A.I., T., & T.O., O. (2013). Measuring Forecasting Performance of Vector Autoregressive and Time Series regression Models. *AMERICAN JOURNAL OF SCIENTIFIC AND INDUSTRIAL RESEARCH*, 49-58. doi:10.5251/ajsir.2013.4.1.49.58

Abu-Mostafa, Y., & Atiya, A. (1996). Introduction to Financial Forecasting. *Applied Intelligence, 6*, 205-213.

Achelis, S. (1995). *Technical Analysis From A to Z.* USA: Mc Graw-Hill.

Acquah, H.-G. (2010, January). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics, 2*, 001-006. Retrieved from http://www.academicjournals.org/JDAE

Allen, P., & Fildes, R. (2001). Econometric Forecasting. *Principles of Forecasting: A Handbook for Researchers and Practitioners*.

***Applied Time Series Analysis*.** (2018). Retrieved from The Pennsylvania State University: https://newonlinecourses.science.psu.edu

Başçı, E. S., & Karaca, S. S. (2013). The Determinants of Stock Market Index: VAR Approach to Turkish Stock Market. *International Journal of Economics and Financial Issues, 3*(1), 163-171.

Chong, T.-L., & Ng, W.-K. (2008). Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. *Applied Economics Letters, 15*(14), 1111–1114. doi:10.1080/13504850600993598
ekolar.com. (2018).

***Eviews*.** (2018). Retrieved from http://www.eviews.com/help

Gerçeker, M. (n.d.). *VAR Model Yardımıyla Türkiye Ekonomisinde Doğrudan Yabancı Sermaye Yatırımlarının Ekonomik Büyüme Üzerindeki Etkilerinin İncelenmesi.*

Huang, B., & Kim, Y. (2006). A test of MACD Trading Strategy.

Institute, T. U. (n.d.). Retrieved from http://www.sci.utah.edu/

Introduction to Hypothesis Testing. (n.d.). In *Probability and The Foundations of Inferential Statictics.*

Işığıçok, E. (1994). *Zaman Serilerinde Nedensellik Çözümlemesi.*
KAP. (2019). Retrieved from Kamuoyu Aydınlatma Platformu: https://www.kap.org.tr/tr/Endeksler

Mastery, M. L. (2019). Retrieved from https://machinelearningmastery.com/time-series-data-stationary-python/

Mida, J. (2013). Forecasting Exchange Rates: A VAR Analysis.

Noor, M., Yahaya, A., Ramli, N., & Mustafa Al bakri, A. (2014). Filling Missing Data Using Interpolation Methods: Study on the Effect of Fitting Distribution. *Key Engineering Materials, 594-595,* 889-895. doi:10.4028/www.scientific.net/KEM.594-595.889

Özsoy, C. (2009). Türkiye'deki Eğitim ve İktisadi Büyüme Arasındaki İlişkinin VAR modeli ile Analizi . *The Journal of Knowledge Economy & Knowledge Management*, 71-83.

Rosillo, R., de la Fuente, D., & Brugos, J. (2013). Technical analysis and the Spanish stock exchange: testing the RSI, MACD, momentum and stochastic rules using Spanish market companies. *Applied Economics*, 1541-1550. doi:10.1080/00036846.2011.631894

Song, H., & Stephen, F. (n.d.). *Forecasting International Tourist Flows to Macau.*

Tetteh-Bator, E., Adjieteh, M., Jin, L., & Asenso, T. (2018). Vector Autoregressive Models for Multivariate Time Series Analysis; Macroeconomic Indicators in Ghana. *Mathematical Theory and Modeling, 8*(1), 34-48.

Tsay, R. (2002). *Analysis of Financial Time Series.* John Wiley & Sons, Inc.

Uysal, D., Mucuk, M., & Alptekin, V. (2008). Türkiye Ekonomisinde Vektör Otoregresif Model ile Enflasyon - Büyüme İlişkisinin Analizi. *ZKÜ Sosyal Bilimler Dergisi*, *4*(8), 55-71.

Van De Geer, S. A. (2005). Least Squares Estimation. *Encyclopedia of Statistics in Behavioral Science*, *2*, 1041-1045.

Vector Autoregressive Models for Multivariate Time Series. (n.d.).
Wikipedia. (n.d.).

Yazdi , S., & Lashkari, Z. (2013). Technical analysis of Forex by MACD Indicator. *International Journal of Humanities and Management Sciences (IJHMS), 1*(2), 159-165.

CURRICULUM VITAE

Name Surname : **Özgün SEYMEN UZUN**

Place and Date of Birth : **İzmir / 1992**

E-Mail : **ozgunseymen@gmail.com**

EDUCATION :

- **B.Sc.** : 2015, İstanbul Technical University, Science & Letter Faculty, Mathematical Engineering Department
- **M.Sc.** : 2019, İstanbul Technical University, graduate School of Science Engineering and Technology, Mathematical Engineering Department