REPUBLIC OF TURKEY
VAN YÜZÜNCÜ YIL UNIVERSITY
INSTUTE OF NATURAL AND APPLIED SCIENCES
STATISTICS DEPARTMENT

# THE 2018 FORECASTING OF OPEC OIL PRICE BY USING TIME SERIES MODEL

MASTER'S THESIS

PREPARED BY: Hawre Asaad OMARBL
SUPERVISOR: Prof. Dr Fevzi ERDOĞAN

VAN-2017

REPUBLIC OF TURKEY
VAN YÜZÜNCÜ YIL UNIVERSITY
INSTUTE OF NATURAL AND APPLIED SCIENCES
STATISTICS DEPARTMENT

# THE 2018 FORECASTING OF OPEC OIL PRICE BY USING TIME SERIES MODEL

MASTER'S THESIS

PREPARED BY: Hawre Asaad OMARBL

VAN-2017

**THESIS ACCEPTANCE AND APPROVAL PAGE**

This thesis entitled "**The 2018 Forecasting of OPEC Oil Price by Using Time Series Model**" presented by Hawre Asaad OMARBL under the supervision of Prof. Dr. Fevzi ERDOĞAN in the department of Statistics has been accepted as a Master Thesis according to the rules of Higher Education Instruction of Republic of Turkey on 11/10/2017 with unanimity the member of jury.

Chair : Prof.Dr. Fevzi ERDOĞAN Signature

Member : Doç.Dr. Eray ÇELİK Signature

Member : Yard Doç.Dr. Hamit MİRTAGİOĞLU Signature

This thesis has been approved by the committee of institute of science on 20/ 10/ 2017 with decision number 2017/ 48 - I

Signature
Prof Dr. Sıtkı SENSOY
Director of Institute

## THESIS STATEMENT

I hereby declare that all information presented in this thesis were gathered and presented within the frame of ethical behaviors and academic rules, and all kinds of information that does not belong to me have been cited appropriately as stipulated by the thesis writing rules.

Hawre OMARBL

## ABSTRACT

## THE 2018 FORECASTING OF OPEC OIL PRICE BY USING TIME SERIES MODEL

Hawre Asaad OMARBL
M.Sc. Thesis., Statistics Science
Supervisor: Prof. Dr. Fevzi ERDOĞAN
Oct 2017, 77 pages

This thesis aims to study and analyse the monthly data of OPEC crude oil prices from January 2005 to February 2017 in order to find and apply a method that can forecast oil prices for the following years. This is important because oil prices play a significant role in the stationary power of money, which in turn affects the economic stabilities of both the developed and growing countries. In this study ARIMA models were used to forecasting the coming two years (the period from March 2017 till December 2018). The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) residual variances were used to select the appropriate ARIMA model for the dependent parameters investigated, in comparison with the oil prices before subsequent forecasting. The thesis attained that the monthly time series for the price is non-stationary and it has a trend, which is mostly due to the global inflation of money that took place after January 2005. Before any analysis was made, some modifications (differentiation) were applied to the series to turn them into stationary time series, after which the most competent models for the forecasting could be deduced. The thesis forecasted for the future monthly data for money supply using the model ARIMA.



**Keywords:** Akaike information criterion (AIC), Augmented Dickey-Fuller Test, Bayesian Information Criterion (BIC), Time series model ARIMA, Z-test, forecast point.

# ÖZET

## ZAMAN SERİLERİ MODELİ KULLANARAK OPEC PETROL FİYATININ 2018 YILI TAHMİNİ

OMARBL, Hawre
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Prof. Dr. Fevzi ERDOĞAN
Ekim 2017, 77 sayfa

Tez, Ocak 2005'den Şubat 2017'ye kadar OPEC Ülkesi için OPEC Ham Petrol Fiyatının aylık verilerinin doğruluğunu çalışmayı ve analiz etmeyi amaçlamıştır. Paranın istikrarında ve gelişmiş ve gelişen ülkelerin ekonomik istikrarında en önemli role sahip olduğu için bu çalışma yapılmıştır. Bu çalışmada önümüzdeki iki yılı tahmin etmek için ARIMA modelleri kullanılmıştır (Mart 2017'den Aralık 2018'e kadar olan dönem). Sonraki tahminden önce petrol fiyatıyla karşılaştırmalı olarak araştırılan bağımlı parametreler için ARIMA modellerinin seçilmesi için Akaike Bilgi kriteri (AIC) ve Bayes Bilgi Kriteri (BIC) artık varyansı kullanılmıştır. Tez, fiyat için aylık zaman serilerinin istikrarlı olmadığı ve Ocak 2005'den sonra paranın enflasyonu sebebiyle bir trendi olduğunu bulmuştur ve serileri sabit zaman serilerine değiştirmek gerekmiştir, böylece, gelecek dönem tahmini için en yetkin modelleri edinmiştir. Tez ARIMA modelini kullanarak para arzı için gelecek aylık verilerini tahmin etmiştir.

**Anahtar kelimeler:** Akaike bilgi kriteri (AIC), Bayesian Bilgi Kriteri (BIC), Genişletilmiş Dickey-Fuller Test, Zaman serisi model ARIMA, Z-testi tahmin noktası.

## ACKNOWLEDGEMENT

First and foremost, I would like to thank Allah, and wish to express my sincere and deepest appreciation to my supervisor Prof. Dr. Fevzi ERDOĞAN for the guidance, advice and idea during this thesis, my best regards also to all other members of the department of statistics, especially to Doç. Dr. H. Eray ÇELIK, Asst. Prof. Dr. Hatice TAŞKESEN, Asst. Prof. Dr. Mahmut KARA and my dear friend Sherzad Karim for his helps. I am also grateful to everyone whom I have learned a lot from. Thanks to my parents for their encouragement. I would also like to give special thanks to my father as he always had my back and he was supporting me and encouraging me for best goals.

2017

Hawre OMARBL

**CONTENTS**

**Page**

**Page**

## LIST OF TABLES

**Table** **Page**

## LIST OF FIGURES

**Figures** **Pages**

## SYMBOLS AND ABBREVIATIONS

Along with a description of some symbols and abbreviations used in this study are presented below.

| Symbols | Description |
|---|---|
| **N** | Number of original time series |
| **p** | Rank order non-seasonal autoregressive model |
| **P** | Rank order seasonal autoregressive model |
| $\emptyset_i$ | None seasonal autoregressive model coefficient |
| $\Phi_i$ | Seasonal autoregressive model coefficient |
| $\boldsymbol{\theta_i}$ | None seasonal moving average model coefficient |
| $\boldsymbol{\Theta_i}$ | Seasonal moving average model coefficient |
| **d** | None seasonal difference degree |
| **D** | Seasonal difference degree |
| **t** | Time |
| $\boldsymbol{a_i b_i}$ | Fourier series factor |
| **S** | Length of the season |
| **Q** | Rank of seasonal moving average model |
| **q** | Rank of none seasonal moving average model |
| **K** | Number of lags |
| $\boldsymbol{\rho_k}$ | Time series auto-correlation |
| $\boldsymbol{\mu}$ | Time series mean |
| $\boldsymbol{\sigma_z^2}$ | Time series variance |

| | |
|---|---|
| $r_k$ | Auto-covariance time series at offset k |
| $r_0$ | Time-series variation |
| $\rho_{xy}$ | Partial auto-correlation between series X and Y |
| $r_{xy}$ | Auto-covariance between series X and Y |
| $\beta$ | Back shift operator |
| $\sigma_\varepsilon$ | Variance of error |
| $\nabla$ | Differences factor |
| $Z_t$ | Original time series |
| $W_t$ | Stationary time series |

| **Abbreviations** | **Description** |
|---|---|
| **MSE** | Mean square error |
| **SE** | Standard error |
| **ACF** | Auto correlation function |
| **PACF** | Partial auto correlation function |
| **AR** | Auto regressive |
| **MA** | Moving average |
| **ARMA** | Auto regressive moving average |
| **ARIMA** | Auto regressive integrated moving average |
| **DF** | Dickey-Fuller |

## APPENDIX INDEX

**APPENDIX** **Page**

# 1. INTRODUCTION

Oil is the most important source of energy. According to 2013 British petroleum statistical report, oil accounted for 33.1% of the world primary energy consumption in 2012. Even though there are many other sources of energy (natural gas and coal), oil's transportation cost is negligible with respect to its price. Therefore, it can be exported to any desired destination and it has an international market.

The Organization of Petroleum Exporting Countries (OPEC) is the main source of crude oil. About 72.6% of the world oil reserves are under OPEC ownership. OPEC exercises enormous influence on the world prices of oil due to many factors, in particular to its spare oil capacity of roughly 4 million barrels per day (in 2014, based on an assumed 85% utilisation rate). Additionally, according to the International Energy Agency (IEA), non-OPEC countries had an excess demand of 35.5 million barrels per day in 2013. This gap is satisfied with the oil supplied by the 12 members of OPEC. In 2013, the non-OPEC share of global oil production was approximately 59.8% of the total world production of oil. Due to the non-OPEC, excess demand-supply, it is a common belief that non-OPEC producers behave as price takers and that OPEC may play a central role in the world oil market by adjusting its production and setting the prices of oil.

Crude oil is a key commodity for the economy. The effect of increasing price and its daily fluctuations affect not only the economies and financial markets but extend to reach individuals as well. This is because an increase in oil prices has a direct effect on petrol prices; in addition, it also affects the prices of other goods and services. Therefore, forecasting crude oil is a very important task to reduce the impact of price fluctuations, and help investors, hedgers, and individual to make informed decisions when dealing with energy markets.

There are ample studies addressing the accuracy of crude oil volatility modelling and forecasting. These include Autoregressive Conditional Heteroscedasticity, ARCH-type models (Fong and See, 2002; Giot and Laurent, 2003), Asymmetric threshold autoregressive (TAR) model (God by et al., 2000), and artificial based forecast methods (Fan et al., 2008a), Interval Least Square (ILS) (Xu et al., 2008),

Support Vector Machine (SVM) (Xie et al., 2006), Artificial Neural Networks (ANN) (Kulkarni and Haidar, 2009), Adaptive Network-based Fuzzy Inference System (ANFIS) (Ghaffari and Zare, 2009), Fuzzy Neural Network (Liu et al., 2007), Autoregressive Moving Average (ARMA) (Cabedo and Moya, 2003)...etc. However, the complexity of the model specification does not guarantee high performance on out-performed out-of-sample forecasts.

One of the models that have gained enormous popularity in many areas and forecasting research practice is Box-Jenkins method. However, despite the fact that the Box-Jenkins method is powerful and flexible, it is not able to handle the volatility that is present in most data series. To handle the volatility in the crude oil data, this thesis utilizes the forecasts obtained from the Box-Jenkins model by using AIC and BIC approaches to select best model and forecasting.

## 1.1. Objective of the Study

The specific objective of this thesis are as follows: to identify patterns of oil price from OPEC over the period of 2005 to 2017, then develop the suitable time series model and use it to estimate next year’s forecast.

## 2. LITERATURE REVIEW

Interest on examination of time series was initiated by one of the French athletes (Joseph Fourier, 1807) when he simplified the time series in the form of infinite limits, total Pocket (Sine) and (Cosine); and this approach was named the Fourier Series for his efforts.

$$Z_t = \sum_{i=1}^{\infty} a_i Cos \frac{2\pi_{ti}}{N} + \sum_{i=1}^{\infty} b_i Sin \frac{2\pi_{ti}}{N} \quad (1.1.1)$$

Where:

$a_i, b_i$: Fourier series factor

t: time

N: Number Views original time series

Stocks (1879), Schuster (1906) and Beveridge (1922) all applied Fourier string in data analysis for time series to find sessions or hidden periods. Yule (1926) found out that simplified string Fourier does not give a good prediction, and cannot be used to estimate the hidden courses for the series because the length and size of the real series are not consistent. He instead proposed the idea of "Autoregressive" approach. Walker (1931) expanded the method of Yule where he found the regression model of none seasonal degree (p).

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varepsilon_t \quad (1.1.2)$$

Where:

$p$: None seasonal autoregressive model rank.

$\emptyset_i (i = 1,2, \dots, p)$: none Seasonal autoregressive model coefficient.

$\varepsilon_t$: Random errors.

Statsky (1937) added Moving Average none seasonal MA(q):

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1.1.3)$$

Where:

q: rank of Non-Seasonal Moving Average model.

$\theta_i (i = 1,2,...,q)$: parameter of Non-Seasonal Moving Average model.

Wold (1954) innovated the idea of models (Autoregressive and moving average) called "mixed models" symbolized with ARMA (p,q). Researcher Durbin' study (1960) on the regression model parameters showed a way that used least squares. Then, Walker (1962) generalized findings of Durbin and mixed models of ARMA (p, q). The researchers Box & Cox (1964) made a research on mathematical transfers to the non-stationary time series around variance. The researchers Box & Jenkins conducted a study of models of stationary and non-stationary time series between 1962 and 1967. The researchers Box and Etal (1967) studied on Seasonal Multiplicative Model in the non-stationary time series which consists of two parts: seasonal and non-seasonal part, symbolized as ARIMA (p, d, q) $(P, D, Q)^S$

Where:

p: rank of none seasonal autoregressive model.

q: rank of none seasonal moving average model.

d: none seasonal difference degree.

S: length of the season.

P: rank of the seasonal autoregressive model.

Q: rank of the seasonal moving average model.

D: seasonal difference degree.

Watts, Box & Jenkins (1970-1980) suggested suitable models with which to analyze stationary and none stationary time series. The researchers Box & Jenkins (1989) identified the steps for making for time series (Identification, Estimation, Diagnostic, Forecasting). McLeod (1977) made little modifications on the method of Box & Jenkins in estimating the parameters of the models ARMA (p, q) and his results were found to be closer to the maximum possible estimates. Dent & Swanson (1988) developed a method suitable for forecasting in the field with rare information from ARMA (p, q) through testing in the field of transportation. Later, Hasza (1980) studied the behaviour of

autocorrelation function model ARMA (p,q) and found a function that relies on these models and offset intervals.

Wood Ward & Gray (1981) used partial autocorrelation function in determining mixed models ARMA (p, q) as a generalisation of the findings of Box & Jenkins in their study. Tiao & Hillmer (1980) suggested that the value of the time series for any time that can be represented in two ways: first as the aggregate effects of seasonal outcomes, general direction and randomness at the time of these outcomes, and the second method as the product of these three results.

Many researchers used methods of time series analysis in different fields of applications including: (Nizar Mustafa Al-Sarraf, 1981), (Saadoun al-Muhsin Muhammad, 1982), (Iftikhar Hameed Al-naqash, 1982), (Sardar Tahir Ali, 1983), (Kanaan Abdul-Latif Abdul-Razzaq, 1984) (Haitham Salim al-Zubaidi, 1985), (Abbas Fadel Abbas, 1985), (Mohammed Habib Alsharot, 1986), (Nevzat Mohamed Ahmed, 1986), (Kameran Fadhil Othman, 1987), (Karima Abdolkazem Khafaji, 1989), (Buthaina Abdagadir Abdulaziz, 1994,) (Nadia Omar Ahmed, 2000).

## 3. METHODOLOGY

### 1.2. Time Series

Time series is a set of observations generated sequentially in time. If the set is continuous then the time series is continuous, if the set is discrete then the time series is discrete. When the observations are completed at constant time interval and time is considered at a discrete breakdown, then time series is considered a discrete time series. Also, when observations are recorded with continuous intervals and time periods are completed at fixed time intervals then time series is a continuous time series.

### 1.3. Applications

The usage of time series models is twofold:

- Obtain an understanding of the underlying forces and structure that produced the observed data.
- Fit a model and proceed to forecast, monitoring or even feed-back and feed-forward control.

### 1.4. Time Series Analysis Goals

- Describe
- Explain
- Forecast
- Control
- The development of a short-term time series (data generation).

## 1.5. Time Series Analysis Usage

Time Series Analysis is used for many applications:

- Economic Forecasting
- Sales Forecasting
- Budgetary Analysis
- Stock Market Analysis
- Yield Projections
- Process and Quality Control
- Inventory Studies
- Workload Projections
- Utility Studies
- Census Analysis and, etc...

## 1.6. Methods used to model time series

The fitting of time series models can be an ambitious undertaking. There are many methods of model fitting including the following:

- Box-Jenkins ARIMA models.
- Box-Jenkins Multivariate Models.
- Holt-Winters Exponential smoothing (single, double, triple).

The user's application and preference will decide the selection of the appropriate technique.

## 1.7. Model Building Strategy Box and Jenkins (1976)

There are three main steps in the process, each of which may be several times

- Model specification (or identification).
- Model fitting.
- Model diagnostics.

#### 1.7.1.1. Model specification (or identification)

The classes of time series models are selected that may be appropriate for a given observed series. In this step, we look at the:

- Time plot of the series.
- Compute many different statistics from the data.
- Apply any knowledge of the subject matter in which the data arise, such as biology, business, or ecology.

#### 1.7.1.2. Model fitting

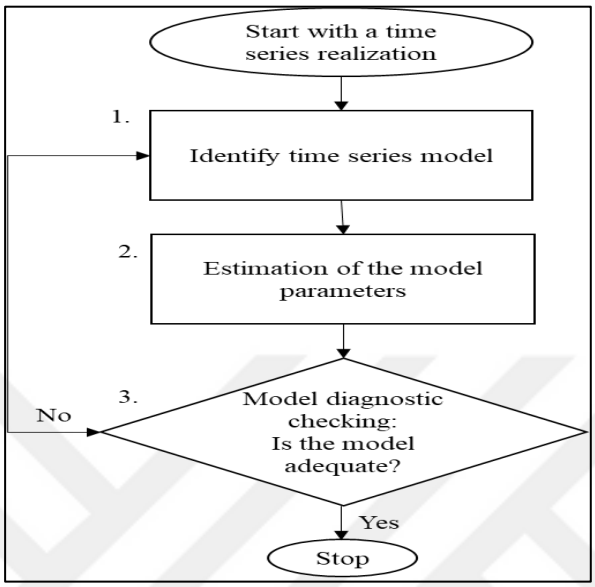
Model fitting consists of finding the best possible mates of those unknown parameters within a given model. In this thesis, criteria such as AIC, BIC, least squares and maximum likelihood for estimation were all utilized.

#### 1.7.1.3. Model diagnostics

The model diagnostic is concerned with assessing the quality of the specified and estimated model. How well does the model fit the data? Are the ones of the model reasonably well satisfied? If no inadequacies are found, the modelling may be assumed to be complete, and the model may be used, for example, to forecast future values. Otherwise, another model needs to be chosen in the light of the inadequacies found; that is, model specification step must be repeated.

## 1.8. Objectives of Analysis

The aims of analysis the time series are to Describe, Understand Forecast and Control.




## 1.9. Technique of Analysis

The graphical plot is used to determine the most suitable specification plotting the data over time and the auto-correlation function (ACF) and partial auto-correlation Function (PACF) are then plotted.

If the time series plot is stationary in the mean, it can be assessed, as there is no evidence of a change in the mean overtime. The time series plot is stationary in the variance if the plotted series illustrates no change in the variance over time. If the time series patterns are non-stationary, non-stationarity factors should be removed; this can be done with the method called “differencing” or “differentiation”.

The “different” series are defined as the change between each observation in the original series. Usually, the difference approach is used first anyway because it is very important to removing non-stationary factors from sets.

Seasonality is defined as a pattern that repeats itself over fixed intervals of time. If (AR) and (MA) models are used, the seasonal aspects of the data can be seen, and then the ACF and PACF can be tested at the seasonal lags.

## 1.10. Augmented Dickey-Fuller Test

In econometrics and statistics, an augmented Dickey–Fuller test (ADF) tests the null hypothesis of a unit root is present in a time series sample and the alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. Augmented Dickey-Fuller t-statistic uses to know time series data stationary or not; the result is a negative number, the more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

The unit root test is then carried out under the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$. Once a value for the test statistic

$$DF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

Is computed it can be compared to the relevant critical value for the Dickey–Fuller Test. If the test statistic is less (this test is non-symmetrical so we do not consider an absolute value) than the (larger negative) critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present.

**Hypotheses for the test**

The null hypothesis(H0) for this test is that there is a unit root.

The alternate hypothesis(H1) differs slightly according to which equation you're using. The basic alternate is that the time series is stationary (or trend-stationary).

The more negative the DF test statistic, the stronger the evidence for rejecting the null hypothesis of a unit root.

## 1.11. Akaike Information Criterion (AIC)

Akaike information criterion (AIC) is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, AIC will choose the best model from a set, it won't specify anything about absolute quality. In other words, if all of your models are poor, it will choose the best of a bad bunch. Therefore, once the best model is selected, running a hypothesis test to figure out the relationship between the variables in the model and the outcome of interest should be considered. Hence, AIC provides a means for model selection.

AIC is founded on information theory: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. In doing so, it deals with the trade-off between the goodness of fit of the model and the complexity of the model.

Suppose that we have a statistical model of some data. Let ($\hat{L}$) be the maximum value of the likelihood function for the model; let $k$ be the number of estimated parameters in the model. Then the AIC value of the model is the following.

$AIC = 2k - 2\ln(\hat{L})$

Where:

$\hat{L}$: the maximized value of the likelihood function of the model M, i.e. $\hat{L} = p(x|\hat{\theta}, M)$ where $\hat{\theta}$ are the parameter values that maximize the likelihood function.

x = the observed data;

k = the number of free parameters to be estimated.

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting because increasing the number of parameters in the model almost always improves the calculated goodness of the fit.

## 1.12. Bayesian Information Criterion (BIC)

Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC).

When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. Both BIC and AIC attempt to resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC.

$$BIC = \ln(n)\, k - 2\ln(\hat{L})$$

$\hat{L} =$ the maximized value of the likelihood function of the model $M$, i.e. $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function;
$x =$ the observed data;
$n =$ the number of data points in $x$, the number of observations, or equivalently, the sample size;
$k =$ the number of free parameters to be estimated. If the model under consideration is a linear regression, $k$ is the number of regressors, including the intercept.

**Properties**

It is independent of the prior or the prior is a constant.
It can measure the efficiency of the parameterized model in terms of predicting the data.
It penalises the complexity of the model where complexity refers to the number of parameters in the model.
It is closely related to other penalised likelihood criteria such as AIC.

## 1.13. Stationary

The time series will be stationary if there is no change in mean, variance and auto-correlation as time is changed.

There are two types of stationary series.

### 1.13.1. Strictly stationary

The series will be Strictly Stationary if the joint probably distribution of $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$ is equal to the joint probability distribution $(Z_{t_1+k}, Z_{t_2+k}, \dots, Z_{t_n+k})$, and the following conditions are met:

- Mean is constant every time:

$$\boldsymbol{E(Z_t) = \mu} \tag{1.13.1}$$

- Variance is discrete every time:

$$\boldsymbol{E(Z_t - \mu)^2 = \sigma_z^2} \tag{1.13.2}$$

- Auto-covariance function does not depend on time:

$$\boldsymbol{E(Z_t Z_{t-1}) = \gamma_k} \tag{1.13.3}$$

### 1.13.2. Weakly stationary

Time series are “Weak Stationaries” if just first two moments of it re time-invariant.

## 1.14. None Stationary

If the properties of series change with time, they are called are non-stationary series; and non-stationarity in the time series means organisational changes about the mean and the variance with a change of time. In such a case non-stationarity factors should be removed before time series analysis are conducted, because their presence will give misleading results in forecasting. There are two types of non-stationary series:

### 1.14.1. None stationary around the mean

Some time series fluctuate around a mean, these are called “series non-stationary around mean”. This non-stationarity can be removed by calculating suitable differences. The main method of taking differences are as follows:

### 1.14.2. Difference operator

In time series analysis, the first difference operator Δ is a special case of the lag polynomial.

$\Delta Z_t = Z_t - Z_{t-1}$

$\Delta Z_t = (1 - B)Z_t$

Similarly, the second difference operator works as follows:

$\Delta(\Delta Z_t) = \Delta Z_t - \Delta Z_{t-1}$

$\Delta^2 Z_t = (1 - B)\Delta Z_t$

$\Delta^2 Z_t = (1 - B)^2 Z_t$

The above approach generalises to the i[th] difference operator

$\Delta^i Z_t = (1 - B)^i Z_t$

B: backshift operator.

### 1.14.3. None stationary around variance

If the variability increases over time that means there is a trend in this time series and that it is non-stationary around variance. A time series with stationary around mean but not around variance should be made stationary using a suitable transform. There may be a time series that both in the mean and in the variance, is non-stationary in this case, its variance should be made stationary first.

Generally, for variance transforming we use power transformation of Box-Cox:

$$F(x_t) = x_t^{(\lambda)} = \frac{x_t^{(\lambda)} - 1}{\lambda}$$

$x_t$: Original series

$F$: Transformed series

$\lambda$: Power transformation

### 1.15. Auto-Correlation

This is the correlation between the elements of a series and others from the same series separated from them by a given interval.

$$\rho_k = \frac{E(Z_t - \mu)(Z_{t-k} - \mu)}{\sqrt{\sigma_{Z_t}^2 \sigma_{Z_{t-k}}^2}}$$

Since stationary series:

$$\sigma_Z^2 = \sigma_{Z_{t-k}}^2 = E(Z_t - \mu)^2$$

Then:

$$\rho_k = \frac{E(Z_t - \mu)(Z_{t-k} - \mu)}{E(Z_t - \mu)^2} = \frac{\gamma_k}{\gamma_0}$$

Because a stationary time series variance is constant and equal to all the periods. And the correlation properties:

$$\rho_0 = 1$$

$$\rho_k = \rho_{-k}$$

$$|\rho_k| \leq 1$$

## 1.16. Partial Auto-Correlation

Partial auto correlation is defined as the direct autocorrelation between and $y_t$ and $y_{t+k}$. The function also calculates the degree of relation between $y_t$ and $y_{t+k}$ since the impact of alternative time lags on $y$ is stable. Partial autocorrelation function shows the partial correlation of a time series with its native lagged values which qualify for the values of the time series at all shorter lags. That's correlation between two observations under the hypothesis that we know and held into account the values of some other set of variables. Suppose $(r_{kk}: k = 1,2,3 \ldots)$ a set of partial autocorrelations at various lags $y$ given by:

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1}, j^r k - 1}{1 - \sum_{j=1}^{k-1} r_{k-1}, j^r j}$$

Where:

$r_k, j = r_{k-1}, j - r_{kk} r_{k-1,k-j}, \quad j = 1,2,3, \ldots, k-1$

## 1.17. Time Series Models

### 1.17.1. Auto-regressive model AR(p)

The public version of the template is:

$$\boldsymbol{Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varepsilon_t} \qquad (1.17.1)$$

Where:

$\varepsilon_t \sim N(0, \sigma_\varepsilon)$

$\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ are parameters of AR(p), the equation (1.17.1) is called the regression model and symbolises her code AR(p) model just may be written using the Back-Shift Operator:

$\emptyset(B) Z_t = \varepsilon_t$

$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$

Where:

$B$: Backshift operator.

In order to verify the condition permanently, the template must be the roots of the characteristic equation ($\emptyset(B) = 0$ absolute value) is greater than one. To find auto-correlation we take into consideration the equation (1.17.1) and batting ($Z_{t-k}$) and taking the expectation, we get:

$$E(Z_t Z_{t-k}) = \phi_1 E(Z_{t-1} Z_{t-k}) + \phi_2 E(Z_{t-2} Z_{t-k}) + \ldots + \phi_p E(Z_{t-p} Z_{t-k}) + E(\varepsilon_t Z_{t-k}) \qquad (1.17.2)$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \ldots + \phi_p \gamma_{k-p} \qquad (1.17.3)$$

And dividing the equation (1.17.3) on ($\gamma_0$) Get the auto-correlation function:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p} \qquad (k \geq 1) \qquad (1.17.4)$$

It is decreasing exponentially. While partial auto-correlation function equal to zero after rank p They disrobed after a period (p).

If ($k = 1,2, \ldots, p$) and we substitute equation No. (1.17.1) get p of equations:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \ldots + \phi_p \rho_{p-1}$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_2 + \ldots + \phi_p \rho_{p-2}$$

: : : : :

$$\rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots + \phi_p$$

Which can estimate values ( $\rho_1, \rho_2, \ldots, \rho_p$ Equation solution above).

The regression model could be non-seasonal and seasonal time series. The formula is that rebooting fixed intervals if there is a template called Seasonal Auto-Regressive Model, and denoted as AR $(P)^S$ and worded as follows:

$$Z_t = \Phi_1 Z_{t-S} + \Phi_2 Z_{t-2S} + \ldots + \Phi_P Z_{t-PS} + \varepsilon_t \qquad (1.17.5)$$

Where:

P: The rank of seasonal auto-regression model.

$\Phi_i$ : Seasonal Autoregressive model coefficient.

S: The length of the season.

The template can be written by a ricochet back:

$$\Phi(B)^S Z_t = \varepsilon_t$$

And auto-correlation for this model is:

$$\rho_{kS} = \Phi^k \qquad k = 1,2,\ldots \qquad (1.17.6)$$

### 1.17.2. Moving-average model MA(q)

The public version of the template is:

$$Z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ... - \theta_q\varepsilon_{t-q} \quad (1.17.7)$$

Where:

$\theta_1, \theta_2, ..., \theta_q$ are parameters of MA(q), the equation (1.17.7) is called moving average model and symbolise MA (q) and can write the template using the backshift operator:

$$Z_t = \theta(B)\varepsilon_t$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$$

Where the variation form is:

$$\gamma_0 = Var(Z_t) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma_\varepsilon^2$$

And that auto-correlation function template is:

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + ... + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2} & \forall k = 1,2,...,q \\ 0 & \forall k \geq q+1 \end{cases} \quad (1.17.8)$$

Watch that auto-correlation function equal to zero after class $q$ it cuts out after the period (q). While partial auto-correlation function decreasing exponentially when it's seasonal, it is called Seasonal Moving-Average model and denoted MA $(Q)^S$:

$$Z_t = \varepsilon_t - \Theta_1\varepsilon_{t-S} - \Theta_2\varepsilon_{t-2S} - ... - \Theta_Q\varepsilon_{t-QS} \quad (1.17.9)$$

Where:

S: The length of the season

Q: Rank of seasonal moving average model

$\Theta_i$: Seasonal moving average coefficient

The template can be written by a backshift operator:

$$Z_t = \Theta(B)\varepsilon_t$$

$$\Theta(B) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - ... - \Theta_Q B^{QS}$$

### 1.17.3. Mixed (Auto-Regressive Moving-Average) model ARMA(p,q)

Many time series cannot be represented as a model to be only moving-average or only auto-regression, because these chains often contain characteristics of both of the models and therefore cannot be represented in a standard model. However, there is one model that gives these specifications, called the mixed model (autoregressive movingaverage) and abbreviated to ARMA(p,q); where (p) is regression rank (q) represents moving average rank while (p,q) represent the rank. General formula hybrid template is as follows:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \quad (1.17.10)$$

And abbreviated by a retracement back to:

$$\phi(B) Z_t = \theta(B) \varepsilon_t$$

This auto-correlation function and form are:

$$\rho_k = \phi_1 \rho_{k-1} + ... + \phi_p \rho_{k-p} \quad (1.17.11)$$

It is decreasing. The partial autocorrelation function behaviour the same.

Autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (integrated) can be applied one or more times to eliminate the non-stationarity.

## 4. APPLY PART

### 4.1. Data

In this study we analysed 12 years of crude oil price from Jan 2005 until Feb 2017 and concluded that oil price has a 12 months' seasonal cycle.

Lowest during Nov – Feb, highest during Jun – Sep.

Major uptrend during 2007 - 2008 and downtrend since 2014.

Data Source: opec.org

Time Frame: January 2005 ~ Apr 2017.

Table 4.1 Descriptive Summary of OPEC data

| | |
|---|---|
| Maximum value | 131.22 |
| Minimum value | 26.5 |
| 1st Quartile | 52.78 |
| Median | 72.59 |
| Mean | 76.18 |
| 3rd Quartile | 104.91 |
| Standard Division | 26.42 |

The oil prices from 2005 to 2017 based on the OPEC crude oil price data is presented in Table 4.1 above. The crude oil price differs, starting $26.5 to $131.22 for a barrel. This reveals that the price of crude oil possibly will go through enormous changes from period to period. The table also shows the mean value ($76.18) and the S.D ($26.42) of the crude oil price data.

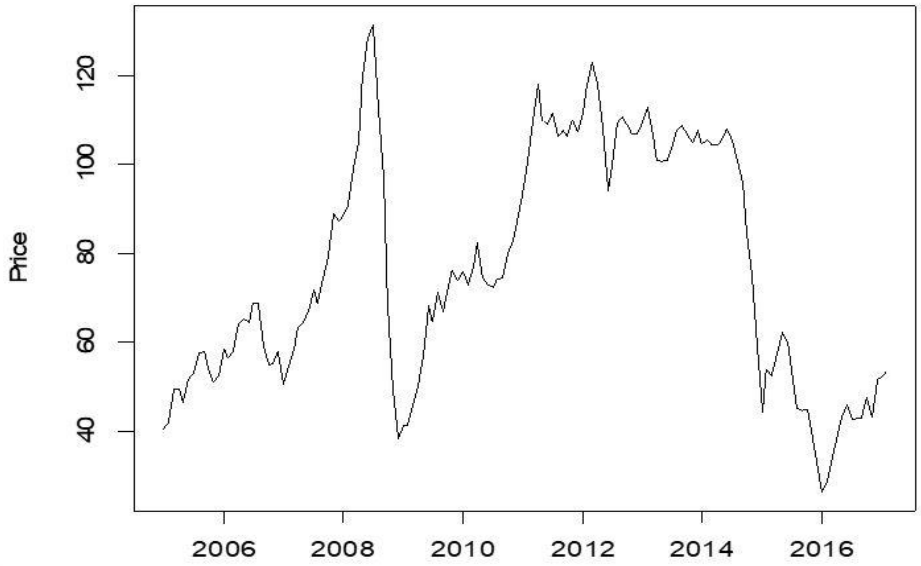



Figure 4.1. Price vs Time.

Figure 4.1 illustrates the graph of crude oil price against the time and how crude oil price gradually increased from 2005 until 2008, and from 2010 until 2014. Furthermore, the price saw major changes from time to time.

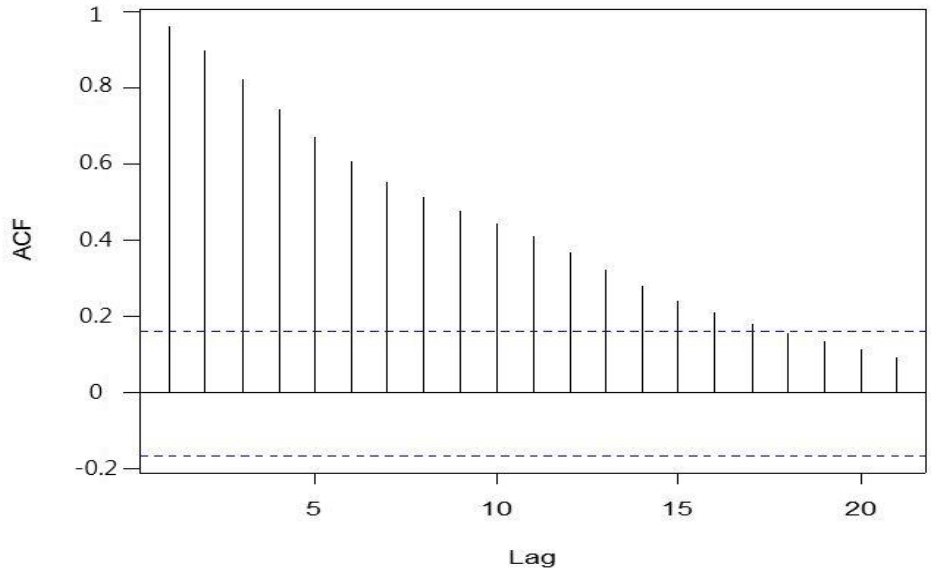



Figure 4.2 Sample ACF of opec price.

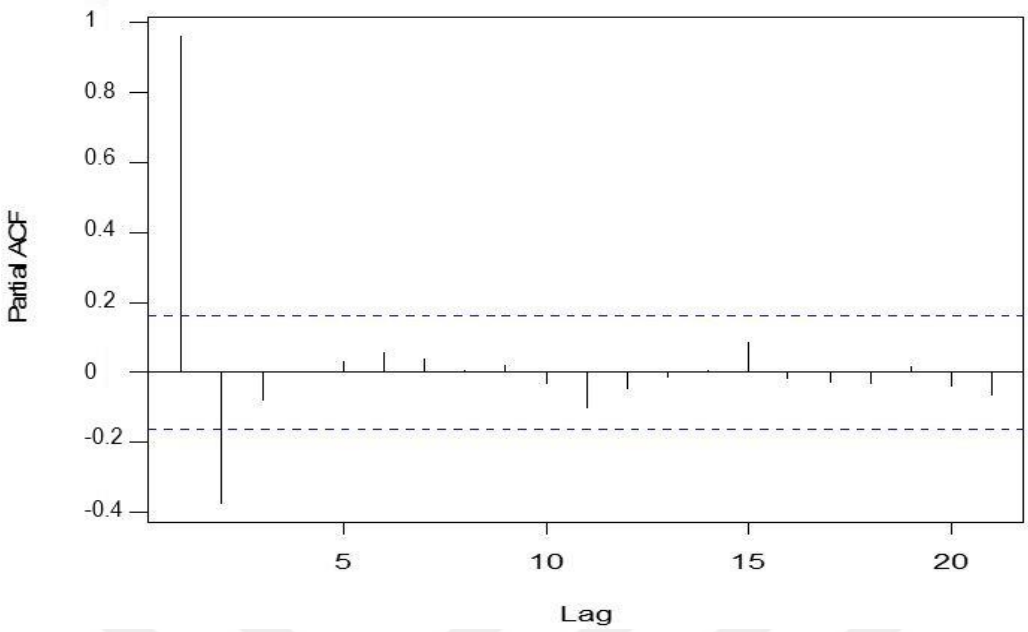



Figure 4.3. Sample PACF of OPEC price.

The easy way to realise the series are not stationary is by figure 4.2 and figure 4.3; these show that many ACF and PACF lags are significant, which means the series are not stationary.

## 4.2. Test for Stationary

The data stationarity must be confirmed for the application of the time series models (especially ARIMA model). As long as the price variation has not changed significantly (excluding the 2008 and 2014 values), the assumption of constant variance may be normal. However, it is incontestable that oil price data has a trend, so the data will probably not be stationary right away to achieve model of time series. Hence, the differences in the model should be modeled, rather than the actual values. In case if the differences are still not stationary, more differencing should be applied.

In the thesis, the first differencing on the data was performed, and the resultant series were tested using Augmented Dickey-Fuller Test. The approach is demonstrated in the following steps:

(ADF) Augmented Dickey Fuller Test

Lag order = 0,

DF = -1.533,

P-value = 0.7709

The alternative hypothesis is Stationary

P-value greater than alpha=0.05, that means time series is not stationary.

Non-seasonal difference and seasonal difference were used to find a number of seasonal and non-seasonal differences to obtain the stationarity of time series. The result needed to get the time series stationary to get first difference is displayed below:

(ADF) Augmented Dickey Fuller Test for first difference of OPIC price

Lag order = 0,

DF = -7.2492,

P-value = 0.01

The alternative hypothesis is Stationary

A p-value less than alpha=0.05 was found, which means that the time series was now stationary. In other words, the result shows that the time series need to the first differenced to get the stationary time series, and only for once.

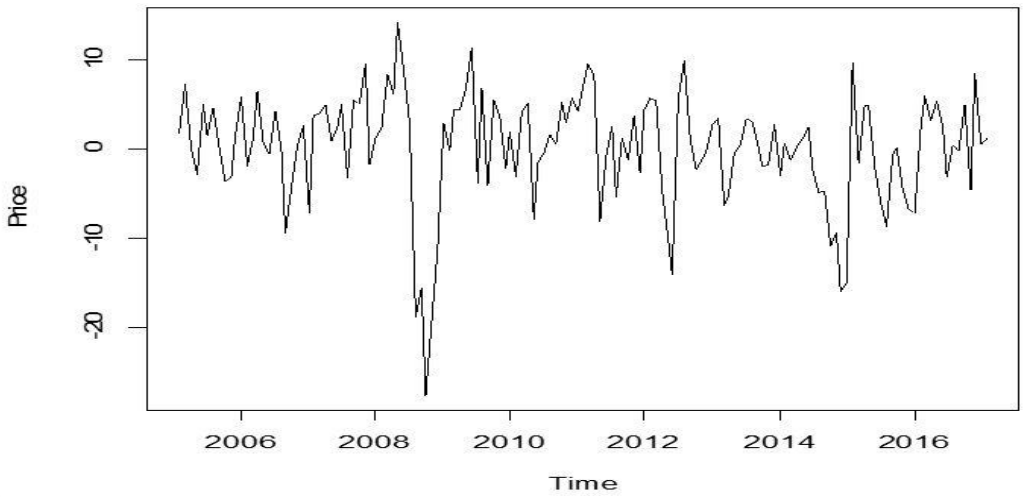



Figure 4.4. Difference time series plot of monthly OPEC price.

Figure 4.4 shows that after a first difference of the time series of OPEC prices, the time series plot has some variation but no trends can be seen. Meanwhile, the time series appear to be stationary.

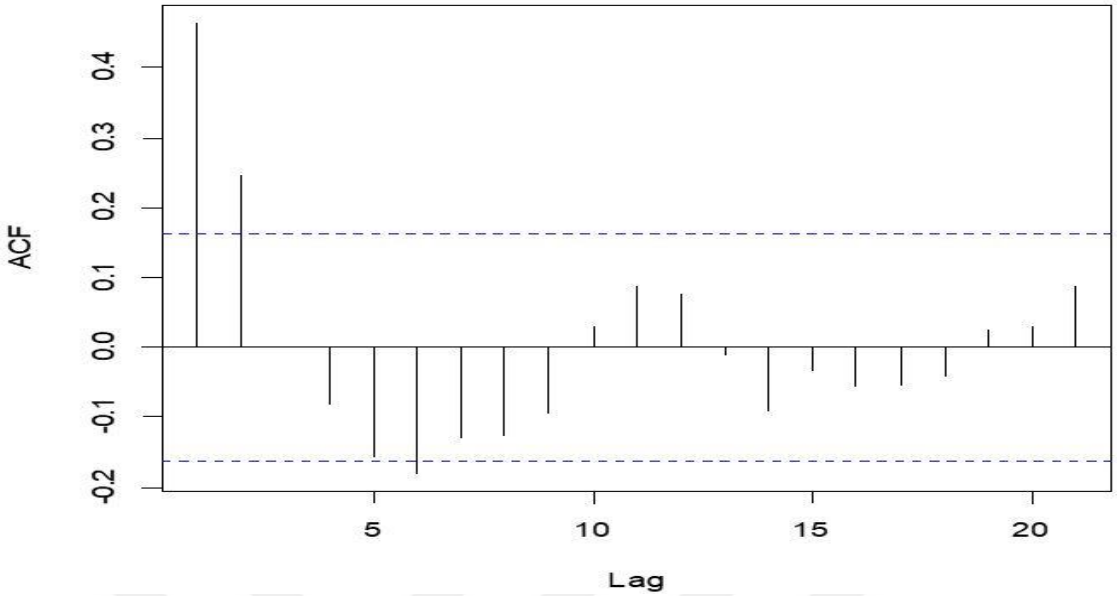



Figure 4.5. Sample ACF of price.

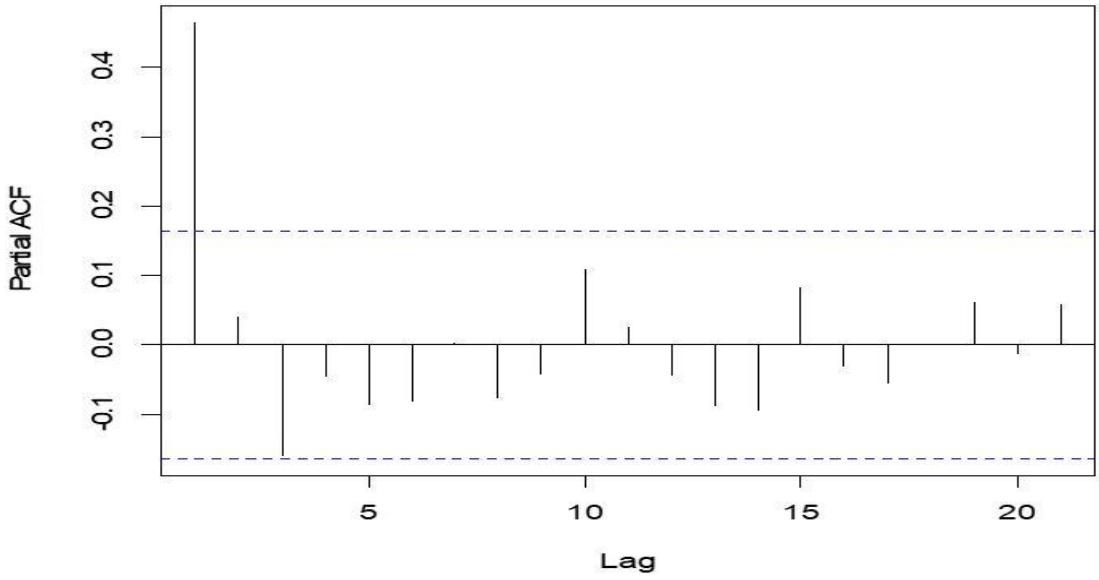



Figure 4.6. PACF of OPEC price.

The Figure 4.5 shows that the sample ACF is tailed off, but Figure 4.6 shows that the sample PACF value is only significant at lag 1, and it is clear that an ARIMA (1,1,0) model is suitable as a first consideration.

### 4.3. Modeling

A forward stepping calculation of each model with BIC and AIC values was completed to assist in evaluating and choosing the best and most precise model of parameters that fits the data. A basic model was assumed for the data. The next step was to use this model to verify the BIC and AIC output of the ARIMA model, and choose the best model which yielded the minimum BIC and AIC values.

Table 4.2 ARIMA parameters with BIC and AIC

| Models | AR | D | MA | BIC | AIC |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 932.060 | 926.107 |
| 2 | 0 | 1 | 2 | 925.127 | 918.196 |
| 3 | 0 | 1 | 3 | 929.997 | 918.090 |
| 4 | 0 | 1 | 4 | 934.631 | 919.748 |
| **5** | **1** | **1** | **0** | **922.968** | **917.015** |
| 6 | 1 | 1 | 1 | 927.810 | 918.880 |
| 7 | 1 | 1 | 2 | 929.944 | 918.037 |
| 8 | 1 | 1 | 3 | 934.909 | 920.026 |
| 9 | 2 | 1 | 0 | 927.715 | 918.784 |
| 10 | 2 | 1 | 1 | 931.419 | 919.512 |
| 11 | 2 | 1 | 2 | 934.838 | 919.955 |
| 12 | 2 | 1 | 3 | 939.799 | 921.939 |
| 13 | 3 | 1 | 0 | 928.889 | 918.982 |
| 14 | 3 | 1 | 1 | 932.152 | 917.268 |
| 15 | 3 | 1 | 2 | 936.696 | 918.836 |
| 16 | 1 | 0 | 0 | 969.951 | 961.000 |
| 17 | 2 | 0 | 2 | 943.908 | 926.006 |
| 18 | 1 | 2 | 0 | 948.950 | 943.010 |
| 19 | 1 | 2 | 1 | 926.348 | 917.439 |
| 20 | 2 | 2 | 0 | 953.228 | 944.318 |
| 21 | 2 | 2 | 1 | 931.002 | 919.123 |
| 22 | 3 | 2 | 0 | 953.551 | 941.672 |
| 23 | 3 | 2 | 1 | 932.488 | 917.639 |
| 24 | 4 | 2 | 0 | 956.399 | 941.550 |
| 25 | 4 | 2 | 1 | 937.259 | 919.440 |

According to Table 4.2, the lowest AIC and BIC value is on row 5 with AIC (917.015), BIC (922.9684) and parameter ARIMA (1,1,0).

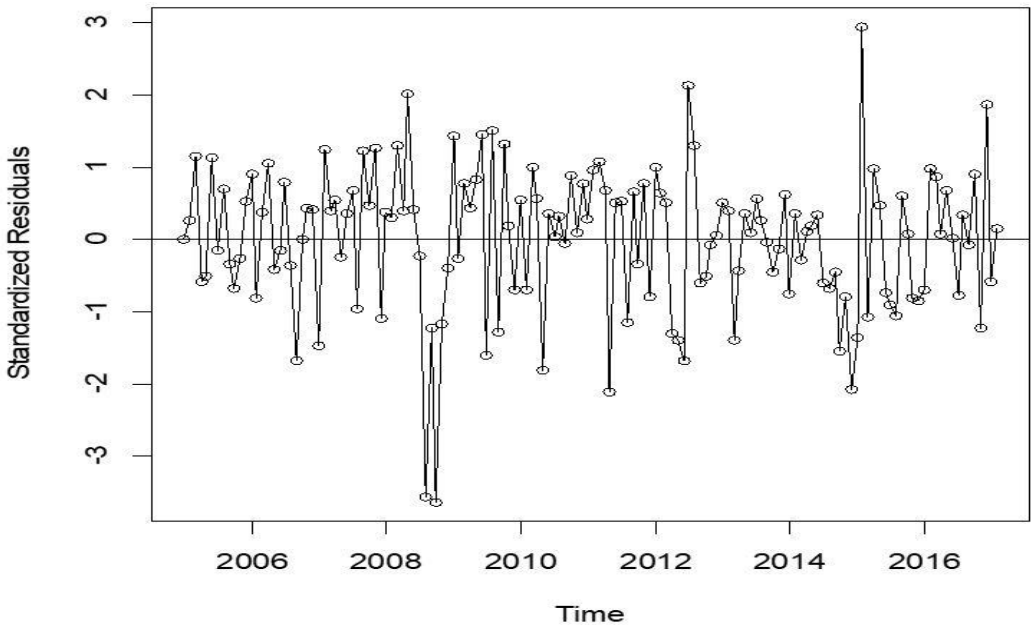



Figure 4.7. Residuals from ARIMA (1,1,0) model.

According to the residual plots, the first check of diagnostic was to review the residual plots over time. If the achieved model is adequate, it should give a plot suggests a rectangular scatter around the zero-horizontal level without any trends at all.

Figure 4.7 shows such a plot of standardized residues from the AR(1) model close-fitting to time series for OPEC prices. Standardization allows us to see residuals of rare sizes easily. This plot supports the model, as no trends exist.

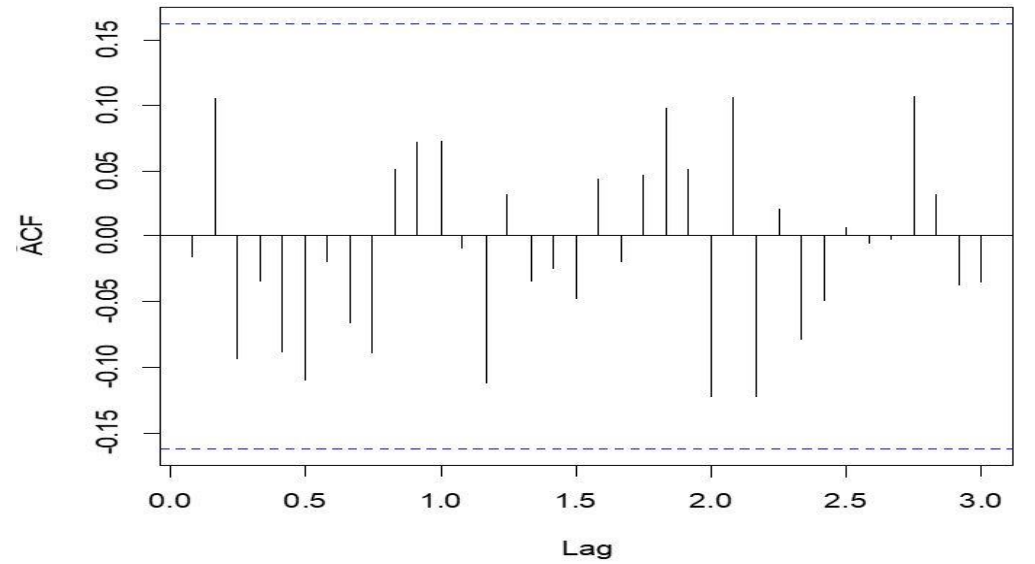



Figure 4.8. ACF of residual from ARIMA (1,1,0).

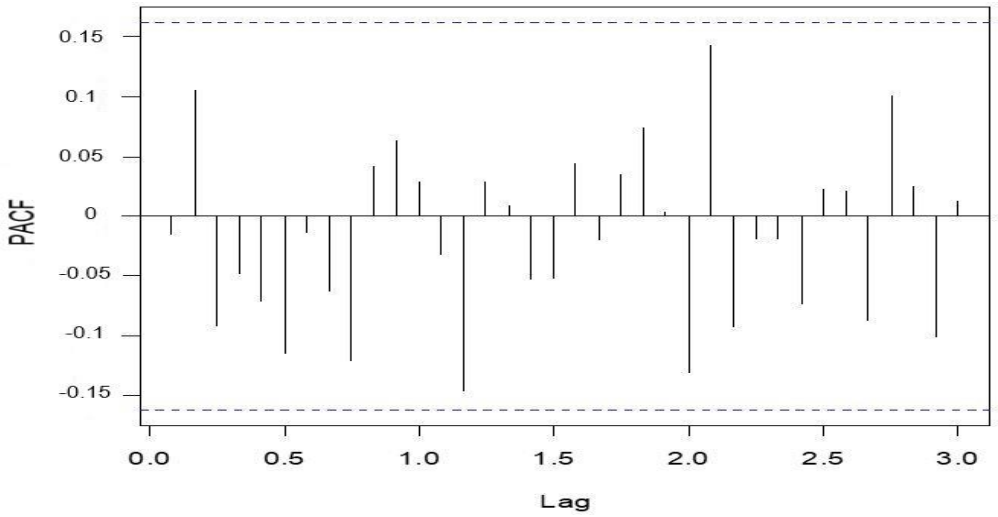



Figure 4.9. PACF of Residuals from the ARIMA (1,1,0) model.

Figure 4.8 and Figure 4.9 show the ACF and PACF residuals. Horizontal lines are based on intermittent plotting at large error ± standard delay. There is no indication of auto-correlation in the residuals of this model.

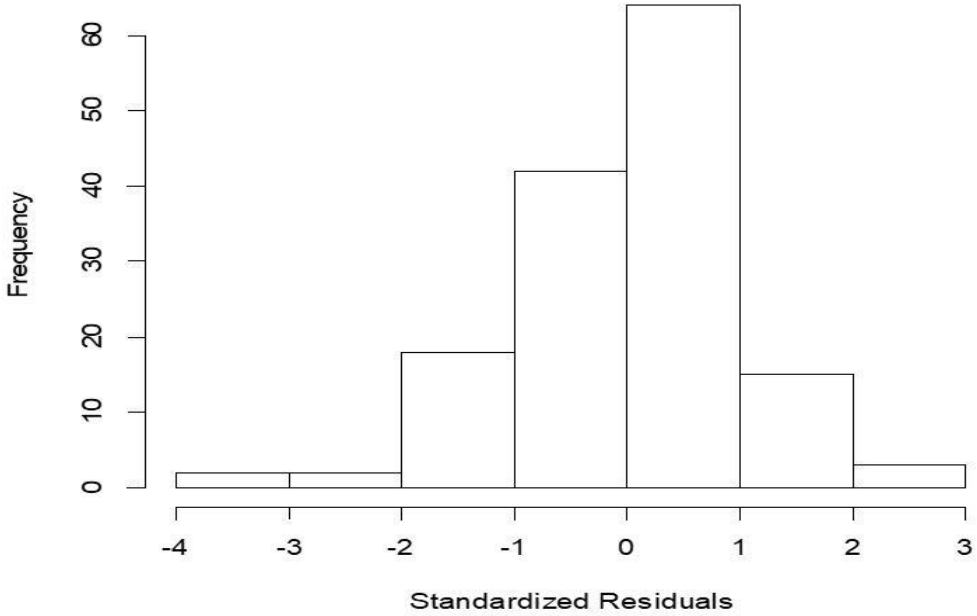



Figure 4.10. ARIMA (1,1,0) Residuals.

Figure 4.10 shows the residual histogram. Note that the overall shape is convex, but definitely not perfect. Quantitative quantification plot should give more details as to why. Quantile-quantile plots are an effective tool for assessing the normal states in such cases.

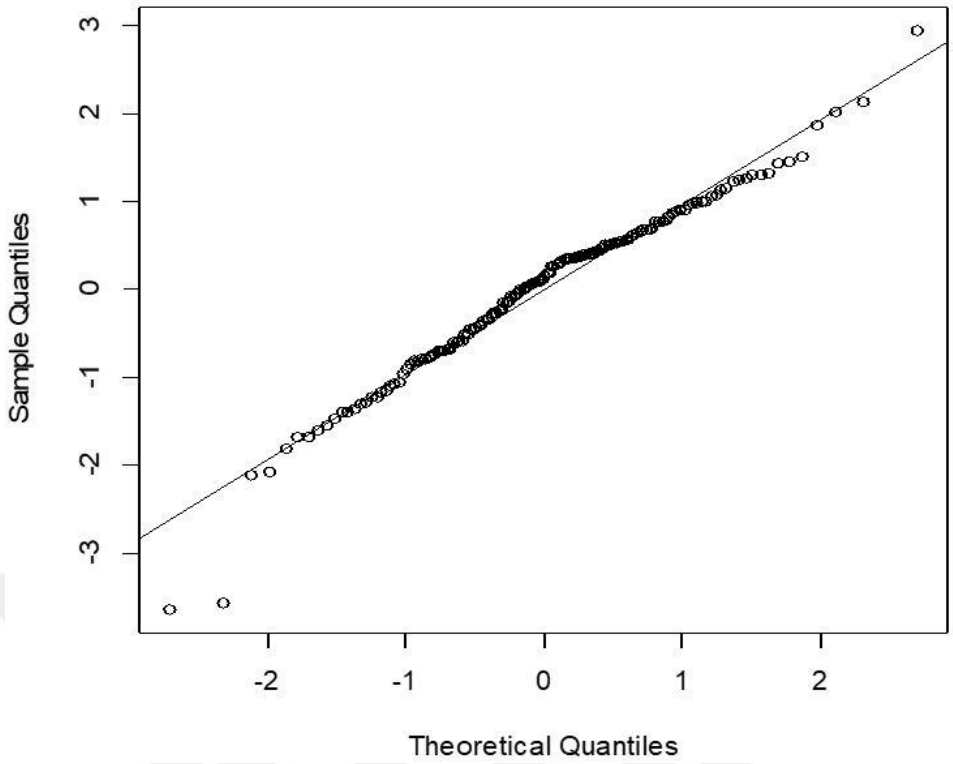



Figure 4.11. Residuals from the ARIMA (1,1,0) model normal Q-Q Plot.

Q-Q plot of residues from AR (1,1,0) model estimated for industrial color chain is given in figure 4.11 and the plot seems to follow a well enough straight line, especially in the extreme values.

Coefficients**:**

| ARIMA Model | Estimate | SE | Z-value | P value |
|---|---|---|---|---|
| AR 1 | 0.4603 | 0.073203 | 6.288 | 3.22E-10 |

Significant Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

Diagnostic Z-test was used to find out that the model is adequate.

The model is adequate because P-value is less than alpha=0.05, which means that the model can be used for forecasting.

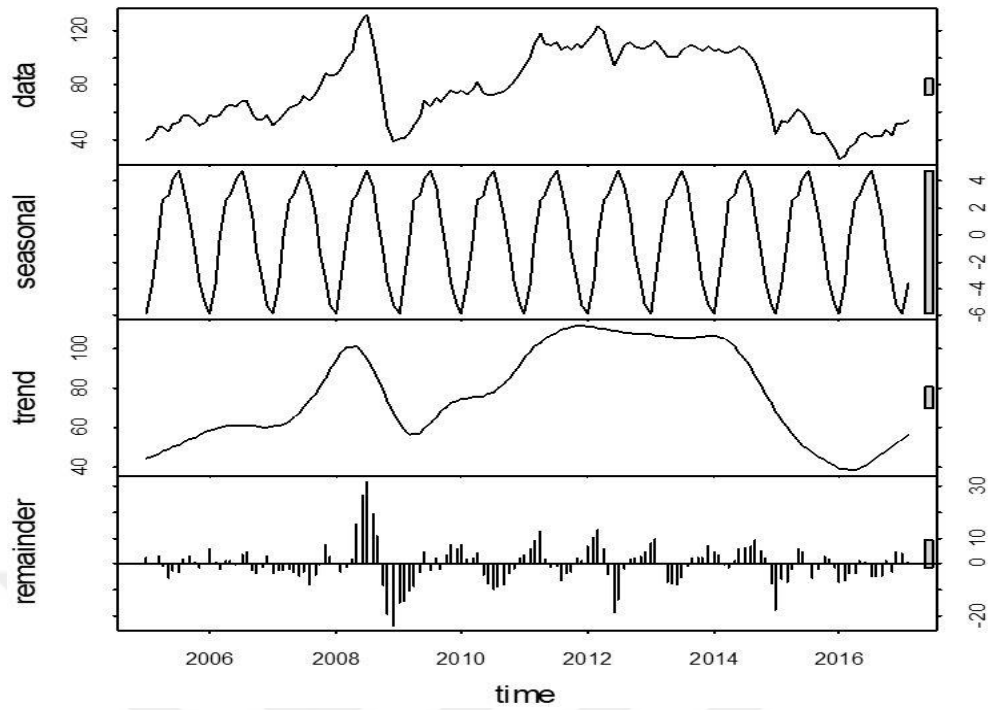



Figure 4.12. S TL plot of oil price.

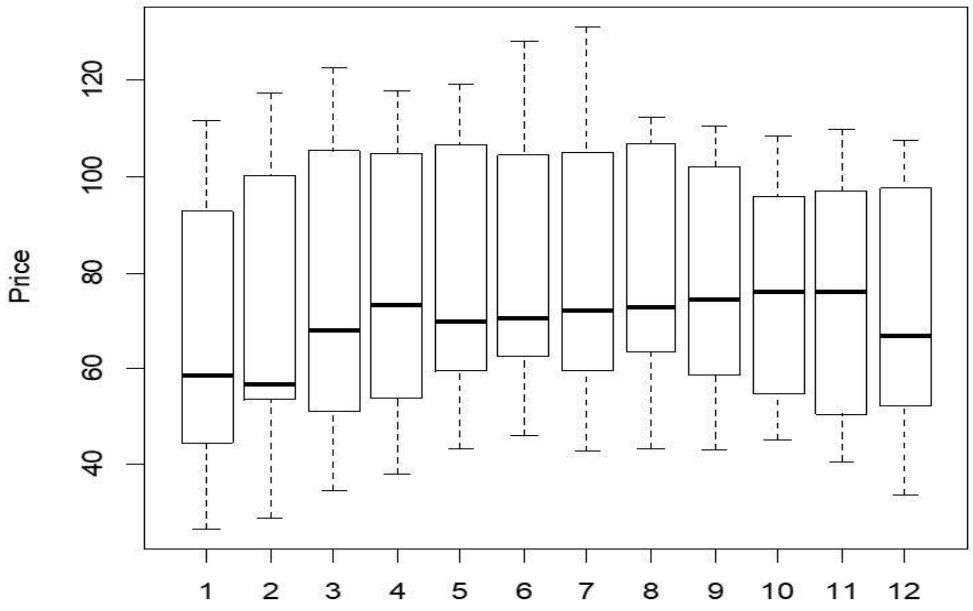



Figure 4.13. Boxplot of Price vs Month.

## 4.4. Forecasting

In this step, a forecast for 2 years (from Mar 2017 until Dec 2018), using the model derived, is presented. The forecast result below shows price within a year.

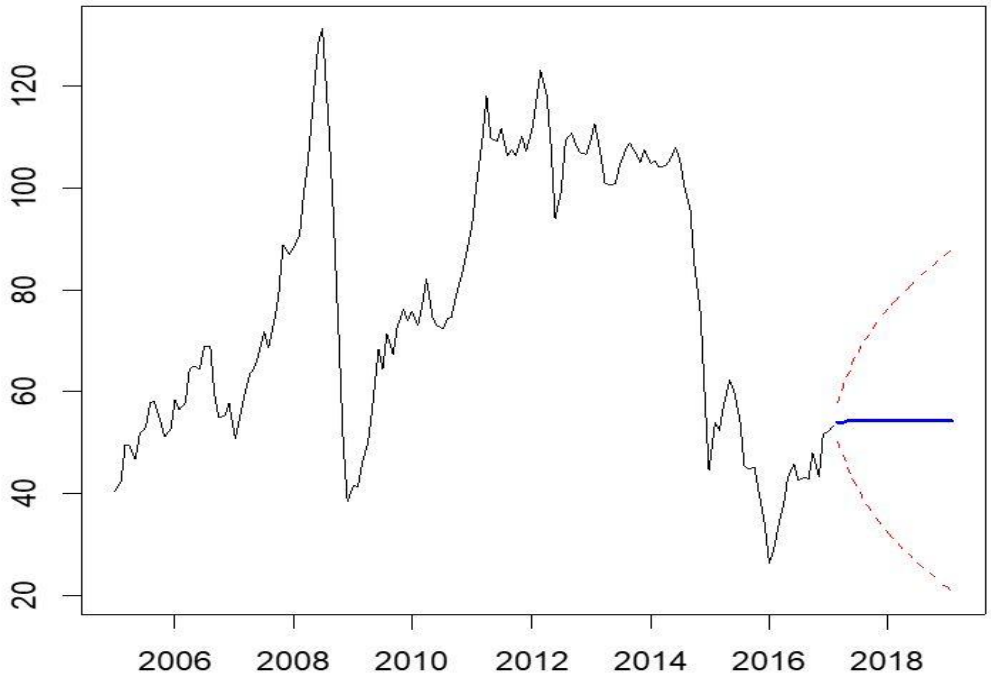



Figure 4.14. Forecasts from ARIMA (1,1,0).

Table 4.3 Forecast results Feb 2017 - Sep 2018 with drift

| Date | Point Forecast | Lower | Higher |
|---|---|---|---|
| Mar-17 | 53.89474 | 50.08245 | 57.70702 |
| Apr-17 | 54.13627 | 47.38899 | 60.88355 |
| May-17 | 54.24745 | 44.96501 | 63.52989 |
| Jun-17 | 54.29862 | 42.82344 | 65.77381 |
| Jul-17 | 54.32218 | 40.92313 | 67.72123 |
| Aug-17 | 54.33302 | 39.21729 | 69.44875 |
| Sep-17 | 54.33801 | 37.66634 | 71.00969 |
| Oct-17 | 54.34031 | 36.2395 | 72.44111 |
| Nov-17 | 54.34137 | 34.9135 | 73.76923 |
| Dec-17 | 54.34185 | 33.67075 | 75.01296 |
| Jan-18 | 54.34208 | 32.49785 | 76.18631 |
| Feb-18 | 54.34218 | 31.38447 | 77.29989 |
| Mar-18 | 54.34223 | 30.32249 | 78.36197 |
| Apr-18 | 54.34225 | 29.30545 | 79.37905 |
| May-18 | 54.34226 | 28.32812 | 80.3564 |
| Jun-18 | 54.34226 | 27.38618 | 81.29835 |
| Jul-18 | 54.34227 | 26.47606 | 82.20848 |
| Aug-18 | 54.34227 | 25.59473 | 83.0898 |
| Sep-18 | 54.34227 | 24.73963 | 83.94491 |
| Oct-18 | 54.34227 | 23.90855 | 84.77599 |
| Nov-18 | 54.34227 | 23.09956 | 85.58497 |
| Dec-18 | 54.34227 | 22.311 | 86.37353 |

## 5. CONCLUSION

The main objective of this thesis study was to find the expected price of oil for the year 2018 using a derived method based on the prices of previous years.

The sample ACF and PACF values were calculated from the original price series using the R-program model, and relevant graphs were drawn for each of them. These models were then used to find out the most suitable model for the forecast.

To achieve stationarity, the series were transformed by a difference of one. The plot of the difference series indicated that the series were evenly distributed around the mean. After the distribution (ACF and PACF) of the differentiated series, the ARIMA model was determined to be the most appropriate one for the forecast, in the form of ARIMA (1, 1, 0). The parameters of the supplied model were also estimated. The model was then subjected to a statistical diagnostic test using the ADF statistic test. The analysis proved that the model was statistically significant, adequate and accurate. The derived model was then used to predict OPEC oil price values for the following years. Expectations of the model are towards a good preservation of original data trends which do not increase or decrease significantly.

However, the price is forecasted to be higher by the ARIMA (1, 1, 0) model with the constant, compared to the previous model. The difference between the average prices suggested by the models was as high as $54 for 2018.

It has been forecasted that the prices do not increase and decrease year over year.

**REFERENCES**


Abraham, B. and Ledolter, J., 2009. ***Statistical Methods for Forecasting*, Vol. 234.** John Wiley & Sons.

Al-Saleh, Y., Upham, P. and Malik, K., 2008. Renewable energy scenarios for the KSA. ***Tyndall Centre for Climate Change Research Working Paper,*** 125.

Anderson, O.D. and Perryman, M.R., 1982. Applied time series analysis. ***Proceedings of The International Conference Held at Houston,*** Texas, August 1981. North Holland.

Anderson, T.W., 2011. ***The Statistical Analysis of Time Series,* Vol. 19.** John Wiley & Sons.

Arteche, J., 2002. Semiparametric robust tests on seasonal or cyclical long memory time series***. Journal of Time Series Analysis,* 23**(3): 251-285.

Box, G.E. and Jenkins, G.M., 1976. **Time Series Analysis, Control, and Forecasting. *San Francisco, CA: Holden Day,* 3226**(3228): 10.

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. ***Time Series Analysis. Forecasting and Control.*** John Wiley & Sons.

Brockwell, P.J. and Davis, R.A., 2016. ***Introduction to Time Series and Forecasting.*** springer.

Castillejos, A.M., 2006. ***Management of Time Series Data*** (Doctoral dissertation, University of Canberra).

Chatfield, C., 1980. ***The Analysis of Time Series.*** an introduction 2nd edition Chapman and Hall.

Chatfield, C., 2016. ***The Analysis of Time Series: An Introduction.*** CRC press. Hamilton, J., 2002. D. (1994), ***Time Series Analysis.***

Cochrane, J.H., 2005. Time series for macroeconomics and finance. ***Manuscript, University of Chicago.***

Dean, J. and Ghemawat, S., 2010. MapReduce: a flexible data processing tool. ***Communications of the ACM,* 53**(1): 72-77.

Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. ***Journal of the American Statistical Association.***

Eickhoff, M.I.R.K.O., McNickle, D. and Pawlikowski, 2006. Analysis of the time evolution of quantiles in simulation. ***International Journal of Simulation,* 7**(6).

Enders, W., 2004. Applied econometric time series, by walter. ***Techno Metrics,* 46**(2): 264.

Fuller, W.A., 1976. Introduction to statistical time series, New York. John Wiley. ***Fuller Introduction to Statistical Time Series 1976.***

Fuller, W.A., 2009. ***Introduction to Statistical Time Series (*Vol. 428).** John Wiley & Sons.

Frank, H. A., 1993. ***Budgetary Forecasting in Local Government. New Tools and Techniques.*** Praeger Pub Text.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2014. ***Bayesian Data Analysis* (Vol. 2).** Boca Raton, FL: CRC press.

Geweke, J., 1984. Inference and causality in economic time series models. ***Handbook of Econometrics,*** 2: 1101-1144.

Geweke, J. and Porter-Hudak, S., 1983. The estimation and application of long memory time series models. ***Journal of Time Series Analysis,* 4**(4): 221-238.

Jeremic, B., 2008. by: Boris Jeremic Department of Civil and Environmental Engineering University of California, Davis January 30, 2008.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J., 1998. ***Forecasting Methods and Applications,*** John Wiley & Sons. Inc, New York.

Montgomery, D.C., Jennings, C.L. and Kulahci, M., 2015. ***Introduction to Time Series Analysis and Forecasting.*** John Wiley & Sons.

Montgomery, D.C., Johnson, L.A. and Gardiner, J.S., 1990. ***Forecasting and Time Series Analysis.*** McGraw-Hill Companies.

O'Donovan, T.M., 1983. Short term forecasting: An introduction to the Box-Jenkins approach. ***John Wiley & Sons, Inc., 605 Third Ave., New York,*** 10158, USA.

Pankratz, A., 2009. ***Forecasting with Univariate Box-Jenkins Models. Concepts and Cases* (Vol. 224).** John Wiley & Sons.

Peden, E.A. and Freeland, M.S., 1998. Insurance effects on US medical spending (1960 -1993). ***Health Economics,* 7**(8): p.671-687.

Shiller, R.J. and Perron, P., 1985. Testing the random walk hypothesis: Power versus frequency of observation. ***Economics Letters,* 18**(4): 381-386.

Singhal, R.A., Tiwari, C.B., Sahai, H. and Misra,1989. Estimation and testing procedures for non-normal samples ***Communications in Statistics-Theory and Methods.***

Tarantola, A., 2005. ***Inverse Problem Theory and Methods for Model Parameter Estimation.*** Society for Industrial and Applied Mathematics.

Tuzunturk, S., 2009. The relationship between volatility and volume on the Istanbul stock exchange. ***International Journal of Sustainable Economy,* 1**(3): 289-304.

Ye, M., Zyren, J. and Shore, J., 2005. A monthly crude oil spot price forecasting model using relative inventories. ***International Journal of Forecasting,* 21**(3).

**APPENDIX 1.**

**R Codes**

```
library(tseries)
OPEC<-read.csv("Opec2005-2017.csv",sep=",")
opec.k<-ts(OPEC$price, start=c(2005,1),frequency = 12)
# plots of time series, ACVF, ACF & PACF for original data
ts.plot(opec.k,ylab="Price",main="Fig.1  Monthly Opec Price")
Month=c("J","F","M","A","M","J","J","A","S","O","N","D")
points(window(opec.k,start=c(2005,1)),pch=Month)
acf(OPEC$price,main=expression(Fig.2~~Sample~~ACF~~of~~opec~~price))
pacf(OPEC$price,main=expression(Fig.3~~Sample~~PACF~~of~~opec~~price))
#Fig.1 shows that the time series is not stationary because there are trend but there are
not seasonal.
#The Fig.2 also shows that the time series is not stationary because a lot of lags are
significant.
#But Fig.3 showa that there are only two lags are significant, we will consider
specifying AR(2).
#  ststionary test
adf.test(opec.k,k=0)
# p.value=0.7709 greater than alpha=0.05, that means the time series is not stationary.
# ndiffs() and nsdiffs()  in R were used to find number of difference and seasonal
non.seaso.<-ndiffs(opec.k)
seaso.<-nsdiffs(opec.k,frequency(opec.k))
rbind("number of difference"=non.seaso.,"number of seasonal difference"=seaso.)
# the result shows that the time series need to once difference to get stationary.
# one difference was used to get stationary
opec.diff<-diff(opec.k)
adf.test(opec.diff, alternative="stationary", k=0)
# p.value=0.01 less than alpha=0.05, that means the time series is stationary.
ts.plot(opec.diff,ylab="Price",main="Fig.4 Difference time series plot of Monthly Opec
Price")
opec.diff<-diff(OPEC$price)
acf(opec.diff,main=expression(Fig.5~~Sample~~ACF~~of~~opec~~price))
pacf(opec.diff,main=expression(Fig.6~~Sample~~PACF~~of~~opec~~price))
#after one difference the time series is stationary because there are trend and no
seasonal, it is showed in Fig.4.
#The Fig.5 shows that the sample ACF is tails off,but Fig.6 shows that the sample
PACF value only at lag 1 is significant, we see clearly an ARIMA(1,1,0) model worthly
is first consideration.
#Plots of the Residuals
#Our first diagnostic check is to inspect a plot of the residuals over time. If the model is
#adequate, we expect the plot to suggest a rectangular scatter around a zero horizontal
```

```
#level with no trends whatsoever.
library(TSA)
opec.model=Arima(opec.k,order= c(1,1,0))
plot(window(rstandard(opec.model)),ylab='Standardized Residuals',
type='o',main=expression(Fig.7~~Residuals~~from~~the~~ARIMA(list(1,1,0))~~Mode
l))
abline(h=0)
#Fig.7 shows such a plot for the standardized residuals from the AR(1) model
#fitted to the opec price time series. Standardization allows us to see residuals of
#unusual size much more easily. The parameters were estimated using maximum
likelihood.
#This plot supports the model, as no trends are present.
###
acf(window(rstandard(opec.model)),lag.max=36,main=expression(Fig.8~~ACF~~of~~
Residuals~~from~~the~~ARIMA(list(1,1,0))~~Model))
pacf(window(rstandard(opec.model)),lag.max=36,main=expression(Fig.9~~PACF~~of
~~Residuals~~from~~the~~ARIMA(list(1,1,0))~~Model))
#Graph of the sample ACF & PACF of these residuals are shown in Fig.8 & Fig.9 The
dashed
#horizontal lines plotted are based on the large lag standard error of ±. There are no
#evidence of autocorrelation in the residuals of this model.
hist(window(rstandard(opec.model)),xlab='Standardized Residuals',
main=expression(Fig.10~~Residuals~~from~~the~~ARIMA(list(0,1,1))~~Model))
#we investigate the question of normality of the error terms via the residuals.
#Fig.10 displays the histogram of the residuals. The shape is somewhat
#“bell-shaped” but certainly not ideal. Perhaps a quantile-quantile plot will tell us more.
qqnorm(window(rstandard(opec.model)),main=expression(Normal~~Q-Q~~Plot))
title(main=expression(Fig.11~~Residuals~~from~~the~~ARIMA(list(1,1,0))~~Model),
line=3)
qqline(window(rstandard(opec.model)))
#quantile-quantile plots are an effective tool for assessing normality.Here we apply
them to residuals.
#A quantile-quantile plot of the residuals from the AR(1) model estimated for the
#industrial color property series is shown in Fig.11 The points seem to follow the
#straight line fairly closely—especially the extreme values. This graph would not lead
us
#to reject normality of the error terms in this model.
# Select best model
# auto.arima() was used to choose a best model
auto.arima(opec.k)
# the result shows that ARIMA(1,1,0) is the best model according AICc AIC and BIC.
# Here AIC & BIC for 24 models were used for unsure the ARIMA(1,1,0) model is the
best model
param.iter<-                c(0,1,1)
param.iter<-rbind(param.iter, c(0,1,2))
param.iter<-rbind(param.iter, c(0,1,3))
param.iter<-rbind(param.iter, c(0,1,4))
param.iter<-rbind(param.iter, c(1,1,0))
```

```
param.iter<-rbind(param.iter, c(1,1,1))
param.iter<-rbind(param.iter, c(1,1,2))
param.iter<-rbind(param.iter, c(1,1,3))
param.iter<-rbind(param.iter, c(2,1,0))
param.iter<-rbind(param.iter, c(2,1,1))
param.iter<-rbind(param.iter, c(2,1,2))
param.iter<-rbind(param.iter, c(2,1,3))
param.iter<-rbind(param.iter, c(3,1,0))
param.iter<-rbind(param.iter, c(3,1,1))
param.iter<-rbind(param.iter, c(3,1,2))
param.iter<-rbind(param.iter, c(1,0,0))
param.iter<-rbind(param.iter, c(2,0,2))
param.iter<-rbind(param.iter, c(1,2,0))
param.iter<-rbind(param.iter, c(1,2,1))
param.iter<-rbind(param.iter, c(2,2,0))
param.iter<-rbind(param.iter, c(2,2,1))
param.iter<-rbind(param.iter, c(3,2,0))
param.iter<-rbind(param.iter, c(3,2,1))
param.iter<-rbind(param.iter, c(4,2,0))
param.iter<-rbind(param.iter, c(4,2,1))
colnames(param.iter)<-c("AR", "d", "MA")
arimaBIC<-function(ts, params=param.iter){
 aic<-apply(params,1,function(p) {
  mo.param<-p[1:3]
  Arima(ts, mo.param)$aic
})
 bic<-apply(params,1,function(p) {
  mo.param<-p[1:3]
  Arima(ts, mo.param)$bic
 })
 result<-cbind(params, bic,aic)
 colnames(result)<-c(colnames(params), "BIC","AIC")
 rownames(result)<-1:nrow(result)
 result
}
l<-opec.k
result<-arimaBIC(l)
result
h<-((result[,4]+result[,5])/2)
minBIC<-which(h==min(h))
list("Minimum value of AIC and BIC"=minBIC)
# the result shows that ARIMA(1,1,0) is the best model according AIC and BIC.
kth <- stl(opec.k, s.window="period")
plot(kth, main="Fig.12 STL plot of oil price")
boxplot(opec.k~cycle(opec.k),main="Fig.13 Boxplot of Price vs Month",ylab="Price",
       year.labels = TRUE,year.labels.left = TRUE,
       labelgap = 0.4, pch = 16, lwd = 1)
opec.model=arima(opec.k,order= result[minBIC, 1:3])
```

```
#Plots of the Residuals already was used but here Diagnostic test was used to know the model is adequate or no.
#Fig.14
tsdiag(opec.model, gof.lag=20)
# Fig.14 shows three of our diagnostic tools in one display—a sequence plot of
#the standardized residuals, the sample ACF of the residuals, and p-values for the
#Ljung-Box test statistic for a whole range of values of K from 5 to 20. The horizontal
#dashed line at 5% helps judge the size of the p-values. All p-values greater than alopha=0.05,
#that meaning the residuals follow white noise process. In this instance, everything looks
#very good. The estimated ARIMA(1,1,0) model seems to be capturing the dependence structure
#of the opec price time series quite well.
library(lmtest)
coeftest(opec.model)
# Diagnostic Test
#  Z-test was used  to know the model is adequate.
#the model is adequate because P-value are less than alpha=0.05, as Fig.13 showed the model is adequte.
#Now the model can be use for forecasting.
library(forecast)
opec.model=Arima(opec.k,order= result[minBIC, 1:3])
FORECAST.RESULT<-forecast(opec.model,h=24,level=50)
par(mfrow=c(1,1))
plot(FORECAST.RESULT,PI=TRUE,shaded=FALSE,pi.col=2, main = " Fig.15 Forecasts by using ARIMA(1,1,0)")
FORECAST.RESULT
```

**APPENDIX 2.** GENIŞLETILMIŞ TÜRKÇE ÖZET (EXTENDED TURKISH ABSTRACT)

## ZAMAN SERİLERİ MODELİ KULLANARAK OPEC PETROL FİYATININ 2018 YILI TAHMİNİ

OMARBL, Hawre
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Prof. Dr. Fevzi ERDOĞAN
Ekim 2017, 77 sayfa

Tez, Ocak 2005'den Şubat 2017'ye kadar OPEC Ülkesi için OPEC Ham Petrol Fiyatının aylık verilerinin doğruluğunu çalışmayı ve analiz etmeyi amaçlamıştır. Paranın istikrarında ve gelişmiş ve gelişen ülkelerin ekonomik istikrarında en önemli role sahip olduğu için bu çalışma yapılmıştır. Bu çalışmada önümüzdeki iki yılı tahmin etmek için ARIMA modelleri kullanılmıştır (Mart 2017'den Aralık 2018'e kadar olan dönem). Sonraki tahminden önce petrol fiyatıyla karşılaştırmalı olarak araştırılan bağımlı parametreler için ARIMA modellerinin seçilmesi için Akaike Bilgi kriteri (AIC) ve Bayes Bilgi Kriteri (BIC) artık varyansı kullanılmıştır. Tez, fiyat için aylık zaman serilerinin istikrarlı olmadığı ve Ocak 2005'den sonra paranın enflasyonu sebebiyle bir trendi olduğunu bulmuştur ve serileri sabit zaman serilerine değiştirmek gerekmiştir, böylece, gelecek dönem tahmini için en yetkin modelleri edinmiştir. Tez ARIMA modelini kullanarak para arzı için gelecek aylık verilerini tahmin etmiştir.

**Anahtar kelimeler:** Akaike bilgi kriteri (AIC), Bayesian Bilgi Kriteri (BIC), Genişletilmiş Dickey-Fuller Test, Zaman serisi model ARIMA, Z-testi tahmin noktası.

# 1. GİRİŞ

Petrol en önemli enerji kaynağıdır. 2013 BP İstatistik Raporuna göre, petrol 2012 yılında dünya ana enerji tüketiminin %33.1'ini teşkil etmiştir. Diğer enerji kaynaklarına rağmen (doğal gaz ve kömür), maliyetine göre taşıma maliyeti ihmal edilebilir seviyededir. Bu nedenle, istenen her yere taşınabilir ve uluslar arası bir piyasası vardır. Petrol İhraç Eden Ülkeler Örgütü (OPEC) ana ham petrol kaynağıdır. Dünya petrol rezervlerinin yaklaşık %72.6'sı OPEC kontrolündedir. OPEC, varsayılan %85 kullanım oranına dayalı olarak 2014'de kabaca 4 milyon varil günlük kapasitesi başta olmak üzere birçok faktör sebebiyle dünya petrol fiyatları üzeinde büyük bir etkiye sahiptir. 1 Ayrıca, Uluslar arası Enerji Kurulu (IEA)'ya göre, OPEC dışı ülkeler 2013'de günlük 35.5 milyon varil talep ettiler. Bu açık 12 OPEC üyesi tarafından sağlanan petrolle kapatılmıştır. OPEC dışı ülkelerin global petrol üretim payı toplam dünya petrol üretiminin %59.8'idir. OPEC dışı aşırı talep sebebiyle, yaygınlıkla, OPEC üreticilerinin fiyat belirleyicileri olarak hareket ettiği ve OPEC'in petrolün üretim miktarı ve fiyatlarını ayarlayarak dünya petrol piyasasında merkezi bir rol oynayabileceğine inanılmaktadır. Ham petrol ekonomi için kilit emtiadır. Artan fiyatın ve günlük dalgalanmalarının etkisi sadece ekonomileri ve finansal piyasaları etkilemez aynı zamanda kişileri de etkiler. Çünkü petrol fiyatlarındaki bir artış benzin fiyatlarında direkt etki yapacaktır; ayrıca, diğer mal ve hizmetlerin fiyatı da bundan etkilenecektir. Bu nedenle, ham petrolün tahmini fiyat dalgalanmalarının etkisini azaltmak ve yatırımcılar, risk dengeleyicileri ve enerji piyasalarında işlem yaparken bilinçli işlem yapan kişilere yardım etmek için çok önemlidir.

## 2. LİTERATÜR İNCELEMESİ

Zaman serilerini sonlu limitler toplam Cep (Sine) be (Cosine) formuna basitleştirdiğinde Fransız atletlerinin biri tarafından zaman serilerini incelenmesi ( Joseph Fourier, 1807) dikkat çekmiş ve bu çalışmalarına (Fourier Dizisi) demiştir.

$$Z_t = \sum_{i=1}^{\infty} a_i Cos \frac{2\pi_{ti}}{N} + \sum_{i=1}^{\infty} b_i Sin \frac{2\pi_{ti}}{N}$$

Burada:

ai, bi: Fourier serisi faktör.

t: zaman.

N: orijinal zaman serilerinin İzlenme Sayısı.

Her ikisi de (Stocks, 1879; Schuster, 1906 & Beveridge, 1922) seans ve gizli süreleri bulmak için zaman serisi için veri analizinde Fourier dizisini benimsemiştir. Put, Fourier dizisini basitleştirmenin (Yule, 1926) iyi bir tahmin vermediğini ve gerçek serilerin uzunluk ve büyüklüğü tutarlı olmadığından gizli süreleri tahmin için kullanılamayacağını buldu. Put (Yule, 1927), (otoregresyon) dördüncü dereceye kadar zamanlı ders otoregresyon fikrini buldu. (Walker, 1931) genişletme metodunda (Yule), mevsimsel olmayan dereceden regresyon modelini (p) biliyordu.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + \varepsilon_t \qquad (4.4.1)$$

Burada:

p: Mevsimsel olmayan otoregresyon model sıralaması.

$\phi_i (i = 1,2,..., p)$: yok Mevsimsel otoregresyon model katsayısı

$\varepsilon_t$: Rastgele hatalar.

# 3. UYGULAMA BÖLÜMÜ

## 3.1. Veriler

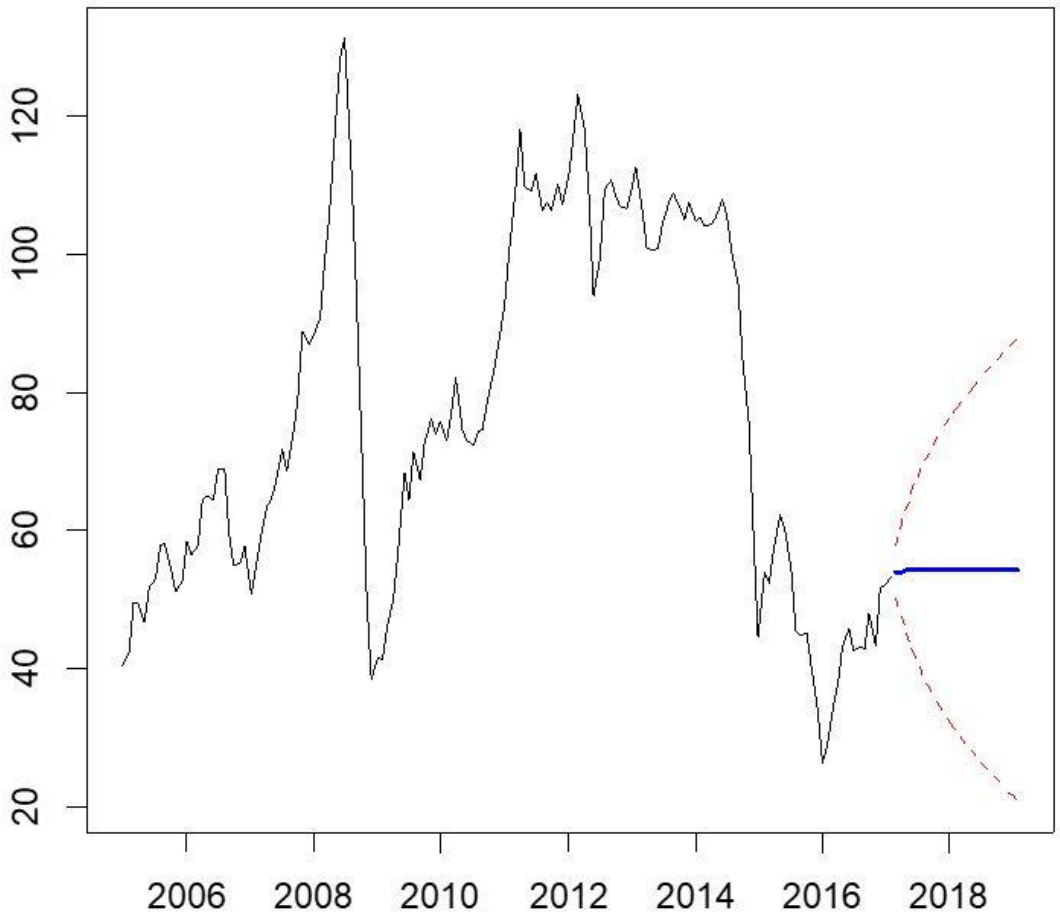
Ham petrol fiyatlarını Ocak 2005 yılından Şubat 2017'ye kadar analiz ettik ve petrol fiyatının 12 aylık mevsimsel döngüsü olduğuna karar verdik.
Kasım-Şubat arasında en düşük, Haziran-Eylül arasında en yüksek
2007-2008 önemli yukarı trend ve 2014'den beri aşağı trend
Veri kaynağı: opec.org
Zaman çerçevesi: Ocak 2005 - Şubat 2017

OPEC verileri özeti
Maksimum değer: 131.22
Minimum değer: 26.5
1'inci Çeyrek: 52.78
Medyan: 72.59
Ortalama: 76.18
3'üncü çeyrek: 104.91
Standart Bölüm: 26.42

Özet tablosunda gösterildiği şekilde, 2005 ila 2017 arası fiyatlar OPEC ham petrol fiyatlarına dayalıdır. Ham petrol fiyatı varil başına $26.5 ve $131.22 arasında oynamıştır. Bu ham petrol fiyatının dönemden döneme anormal şekilde değişeceğini göstermiştir. Tablo ayrıca, ham petrol verilerinin ortalama ($76.18) ve Standart Bölüm ($26.42) 'ye sahip olduğunu göstermektedir

### 3.2. Tahmin

Türettiğimiz modeli kullanarak 2017 Mart ayından 2018 Aralık ayına kadar tahmin veriyoruz. Tahmin yıl içi fiyatlarını veriyor.



## 4. SONUÇ

Bu çalışmanın ana hedefi 2018 yılı için beklenen petrol fiyatını bulmaktır.

R-programı modelini kullanarak orijinal serilerden bir örnek ACF be PACF hesaplanmıştır be grafikleri çizilmiştir. Bu modeller, uygun modeli tespit etmek için kullanılmıştır. İstikrarı sağlamak için seriler bir farkla dönüştürülmüştür. Fark serilerinin çizimi serinin ortalama etrafında eşit olarak dağıldığını göstermiştir. Fark sersisinin (ACF ve PACF) dağıtılmasından sonra, ARIMA modeli AREVIA (1,1,0) tespit edilmiştir. Verilen modelin parametreleri tahmin edilmiştir. Daha sonra model, ADF test istatistiklerini kullanarak bir istatistik teste tabi tutulmuştur. Analiz, modelin kayda değer şekilde farklı, yeterli ve yetkin olduğunu kanıtlamıştır. Verilen model, OPEC petrol fiyatı değerlerini tahmin için kullanılmıştır. Beklentiler önemli şekilde artmayan yad a azalmayan orijinal verileri iyi temsil etmektedir.

Asgari BIC be AIC sebebiyle ARIMA modeli AREVIA (1,1,0) uygulanmıştır. Son olarak, fiyat tahmini sonraki 22 ay için yapılmıştır.

Ancak fiyat AREVIA (1, 1, 0) modele göre önceki modele kıyaslı bir sabitle daha yüksek olarak tahmin edilmiştir. Böylece 2018 için ortalama fiyat olarak $54 tahmin edilmiştir. Fiyatların yıldan yıla artıp azalmadığı tahmin edilmiştir.

**CURRICULUM VITAE**

Hawre Asaad OMARBL was born in 1993 (Erbil/Iraq), finished his secondary and high school education from Azadi Boys' in 2010. The same year got accepted in Statistics Department of Administration and Economics College at Salahaddin University – Erbil. In 2014 got graduated. In February 2014, started her postgraduate study in the department of Statistics, Institute of Natural and Applied Sciences at Van Yüzüncü Yıl University – VAN.

Date: ...../...../.....

Thesis Title: **THE 2018 FORECASTING OF OPEC OIL PRICE BY USING TIME SERIES MODEL**

The title of the mentioned thesis, above having total 25 pages with cover page, introduction, main parts and conclusion, has been checked for originality by **Turnitin** computer program on the date of ..... and its detected similar rate was 3% according to the following specified filtering

Originality report rules:

- Excluding the Cover page,
- Excluding the Thanks,
- Excluding the Contents,
- Excluding the Symbols and Abbreviations,
- Excluding the Materials and Methods
- Excluding the Bibliography,
- Excluding the Citations,
- Excluding the publications obtained from the thesis,
- Excluding the text parts less than 7 words (Limit match size to 7 words)

I read the Thesis Originality Report Guidelines of Yuzuncu Yil University for Obtaining and Using Similarity Rate for the thesis, and I declare the accuracy of the information I have given above and my thesis does not contain any plagiarism; otherwise I accept legal responsibility for any dispute arising in situations which are likely to be detected.

Sincerely yours,

Date and signature

Name and Surname: Hawre OMARBL

Student ID#: 149702329

Science: Statistics Department

Program: .....

Statute: M. Sc. ■ Ph.D. □

**APPROVAL OF SUPERVISOR**
SUITABLE

Prof. Dr. Fevzi ERDOĞAN

**APPROVAL OF THE INSTITUTE**
SUITABLE

Prof.Dr.Süat ŞENSOY
Enstitü Müdürü