Time Series Analysis of United States of America Crude Oil and Petroleum Products Importations from Saudi Arabia

Olayan Muflih M. Albalawi
College of Science Statistic Department, Tabuk University
Kingdom of Saudi Arabia

Ayed Rheal A. Alanzi
College of Science and Human Studies at Hotat Sudair Mathematics Department, Majmaah University,
Kingdom of Saudi Arabia

Abstract

This study focuses the time series modeling of Saudi Arabian petroleum and its products importations by United States of America. This information which contains 505 observations of importations from Saudi Arabia by the United States of America on monthly basis from the first month of the year 1973 to the first month of 2015 and this information have been obtained from U.S. Energy Information Administration websites. Developing an appropriate model of time series for forecasting future value of this particular series happens to be the main objective of this study. The values up to December 2003 are taken for model building and observation from January 2004 to January 2015 is taken for model validation. Exponential smoothing techniques and Box-Henkin ARIMA techniques are used for this purpose. The accuracy of the fitted values is evaluated using model diagnostics methods and comparing the predicted values with observed values.

Keywords: Time Series Analysis, Crude Oil, Petroleum Products, Saudi Arabia Petroleum

1. Introduction

This statements focus on modeling of the Saudi Arabian petroleum and its products (Thousand Barrels per Day) importations by the United States of America on monthly basis from the first month of the year 1973 to the first month of 2015 and the sources of this information is U.S. Energy Information Administration. The modeling is mainly based on Exponential smoothing techniques and Box-Henkin ARIMA techniques. Developing an appropriate model of time series for forecasting future value of this time series happens to be the main objective of this study. The values up to December 2003 are taken for model building while the observation from January 2004 to January 2015 is taken for model validation.

Both additive and multiplicative models are considered for graphical analysis. The time series plot log transformed series reveals only seasonal variations and additive model is selected for further analysis. Time series decomposition is adopted to identify the seasonal indices. The seasonal indices suggest that the importations of Saudi Arabian petroleum and its products by United States of America are a little bit significantly seasonally affected. The import is maximum during January and minimum during March. Seasonal variations can be removed from the original data to obtain seasonally adjusted data (deseasonalized data) which is free from seasonal variations for further analysis. The seasonally adjusted series is analyzed using exponential smoothing techniques. The estimated value of alpha is 0.64198. The value of alpha telling us that more importance is given to recent observations than past forecasts. The analysis residuals suggest that the exponential model satisfy the model assumptions. Autoregressive Integrated Moving Average (ARIMA) models are tried after verifying the stationarity conditions using ADF test. Two models selected for final selection are ARIMA (1,1,1) and ARIMA(1,1,0). The principle of parsimony, AIC, BIC were employed to choose the best model. The values of AIC suggest that there is not much difference in these information criteria. Therefore, using the principle of parsimony, the ARIMA (1,1,0) is the best model. Diagnostic tests residual analysis was used to validate the model assumptions. Final model was used to forecast the values of import between January 2004 and January 2015. Both of the observed values as well as the forecasted values are found to be close.

2. Preliminary Analysis

An excellent review of the pattern of movement in the time series is recorded with the plotting of its graph. Plotting of time series graph is being employed for identifying dissimilar components of the time series, namely trend, seasonal variation and cyclical variations. The suitability of additive or multiplicative model is examined using time series plots.

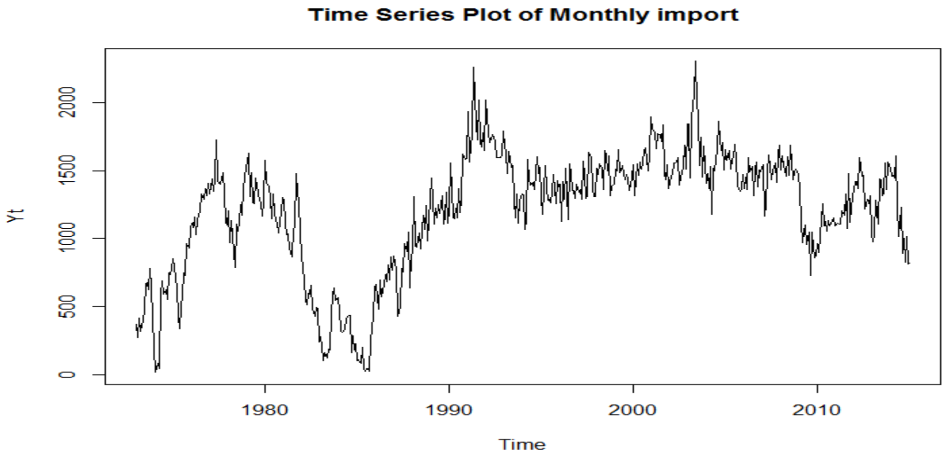

Figure 1: The time series plot of U S importation of Saudi *Arabian Petroleum and its Products.*

The graphs shows clear seasonal, cyclic and trend effect in the US oil petroleum and its Products importations

The additive components model, Y =T+S+I and the multiplicative model Y =T*S*I are the two straightforward models relating the observed value Yt of a time series to be Trend (T), Seasonal (S) and Irregular (I) components are the additive components model. Y =T+S+I and the multiplicative model Y =T*S*I. When the time series being analyzed has approximately the equivalent unevenness all over the dimension of the series then the performance of the additive components model is at its best. In other words, each of the values of the series descends roughly inside a band of constant width concentrated on the trend. The best performance of the multiplicative component model is attained when the variability of the time series increases with the level.

A plot of log transformed series is useful for checking whether multiplication model is more appropriate.

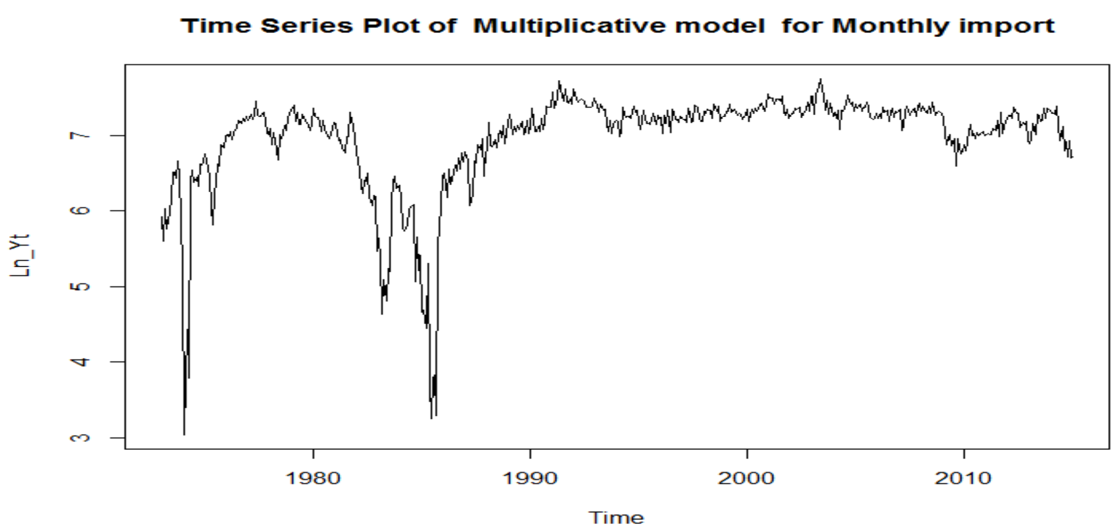

***Figure 2:** The time series plot of U S importations of Saudi Arabian Petroleum and its Products. (Multiplicative Model)*

There was no significant change in the pattern of movement of the series. The main variability in the series is the seasonal variation. The plot of log transformed series does not alter this variation. The additive model is seem to be more appropriate since there is an approximate constancy in the dimension of the periodical variation synchronously and there is no observable time series degree dependency with the arbitrary variations additionally appeared steady approximately in measurement synchronously.

2.1 Decomposition of Time Series

An effort to identify the component factors that might impact on U S Importations of Saudi Arabian Petroleum and its Products is one approach to the analyze the time series of the affected data. The process of identification is referred to as decomposition. This process involves separate identification of each component. The forecasts of the future value are obtained by combining the projections of each of the components.

Combining together inclination part, periodical part and uneven part then makes up a seasonal time series. An additive model referred one type of model which consider the worth of time series to be the addition of all parts; whereas, a multiplicative model referred one type of model treating the time to be the multiplication of all parts.
In time series decomposition, these components are estimated and removed from the original series to identify the random component. The technique of ratio to moving average is employed to obtain the periodical indicators of the specified time series using R function decompose ().The following table gives seasonal indices.

Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
37.82	-36.90	-40.54	-16.63	10.81	-10.20	11.09	9.02	21.64	-6.72	7.84	12.77

The seasonal indices suggest that there exists a little significant seasonal effect of Saudi Arabia crude oil and Petroleum Products importations on the US oil imports. The import is maximum during January and minimum during March. Seasonal variations can be removed from the original data to obtain seasonally adjusted data (de-seasonalized data) which is free from seasonal variations for further analysis. The trend and an irregular component are now contained in the seasonally adjusted series. The plot of different component's time series is given below. The graph shows no constant trend in the data but systematic seasonal variations are clear in the plot.

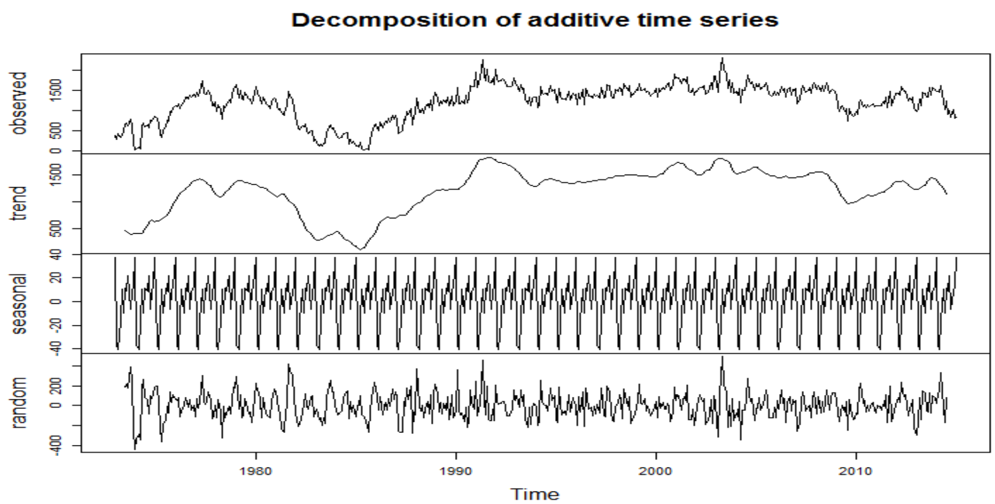

Figure 3: *plot of components showing U S importations of Saudi Arabia Petroleum and its Products.*

The plot given below gives the plot of de-seasonalized values of the series. Further analysis is carried out on these deseasonalized values of the series. The predicted values obtained from this series can be converted to the original values by adding the corresponding seasonal indices.

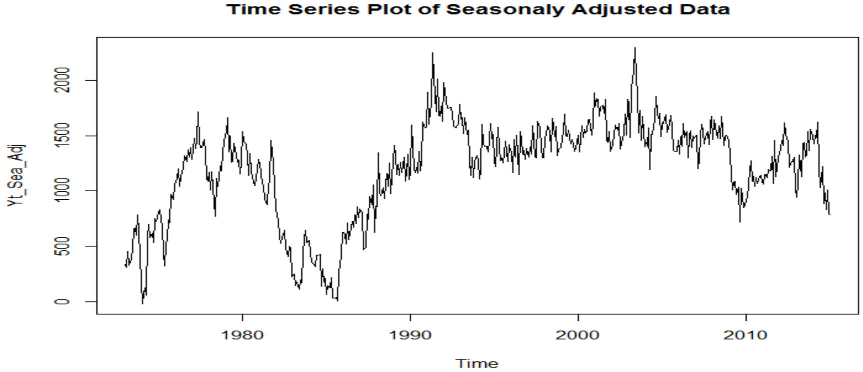

Figure 4: Plot of de-seasonalized series *showing U S importations of Saudi Arabina Petroleum and its Products.*

3. Forecasts using Exponential Smoothing

Short term forecasts for the time series can be made using exponential smoothing which is more appropriate for a trend-free and seasonality-free additive model. There is suggestion in the preceding analysis that no constant trend exists in the deseasonalized time series.

Considering the latest current understanding exponential smoothing continue to revises an estimate. An additional significance is being given to the latest current observations after consideration of many observations. A means by which the present period instant can be approximated is provided through simple exponential smoothing method. The parameter alpha is employed to control smoothing; for the current period instant approximation. 0 with 1 represent two numbers between which the value of alpha can be derived. A near 0 values of alpha denotes that little weight is positioned on the more current observations during future values forecasting. Below is exponential smoothing detail.

The exponential smoothing equation is $\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$

Holt-Winters exponential smoothing having no trend as well as seasonal component

Call:

Holt Winters(x = Yt_Sea_Adj, beta = FALSE, gamma = FALSE)
Smoothing parameters:
 alpha: 0.6419812
 beta : FALSE
 gamma: FALSE

Coefficients:
     [,1]
a 809.6127

Holt Winters () output gives expression such as projected worth of the alpha parameter is roughly 0.64198. The value of alpha telling us that more importance is given to recent observations than past forecasts. The following graph gives the joint plot of predicted values and forecasts based on single exponential smoothing. Both series are close to each other indicating that single exponential smoothing method is accurate in predicting the time values. The difference between the observed and predicted value of the series is known as error while a degree of correctness of the forecasts is used to represent the sum of squared errors (SSE).

Here this11470607 is the sum-of-squared-errors.

The forecasted value of series for next 12 months can be obtained using forecast. Holt Winters () function. The forecasted values are given in the table. The forecasts are also included in a graph.
forecast. Holt Winters (Yt_Exp_F,h=12)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb. 2015	809.6127	616.0932	1003.132	513.6502	1105.575
Mar. 2015	809.6127	579.6468	1039.579	457.9101	1161.315
Apr. 2015	809.6127	548.2339	1070.992	409.8683	1209.357
May 2015	809.6127	520.2108	1099.015	367.0108	1252.215
Jun. 2015	809.6127	494.6715	1124.554	327.9516	1291.274
Jul. 2015	809.6127	471.0532	1148.172	291.8306	1327.395
Aug. 2015	809.6127	448.9784	1170.247	258.0702	1361.155
Sep. 2015	809.6127	428.1791	1191.046	226.2603	1392.965
Oct. 2015	809.6127	408.4566	1210.769	196.0975	1423.128
Nov. 2015	809.6127	389.6595	1229.566	167.3496	1451.876
Dec. 2015	809.6127	371.6683	1247.557	139.8346	1479.391
Jan. 2016	809.6127	354.3877	1264.838	113.4061	1505.819

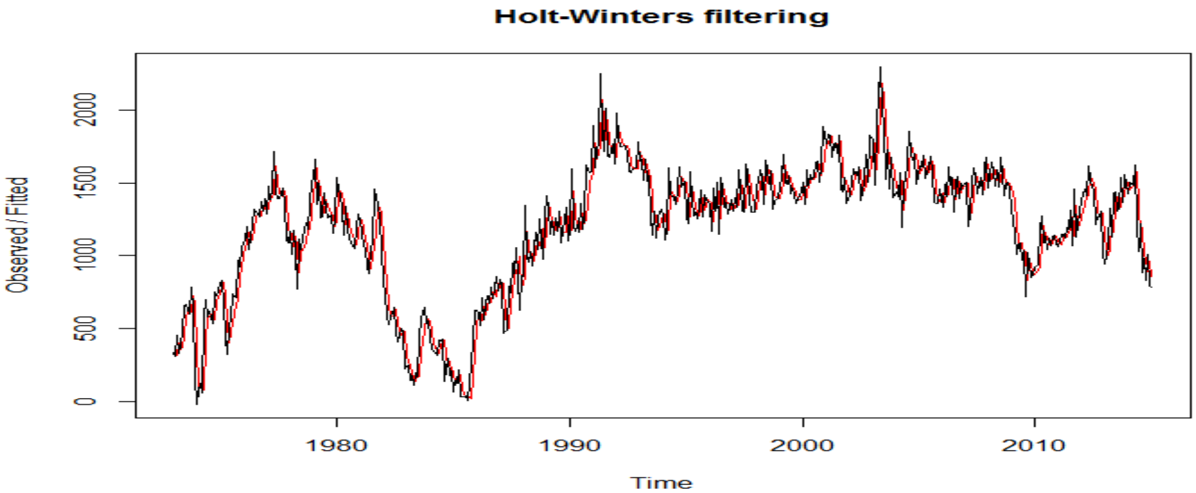

Figure 5 : Plot of observed and fitted values of the series under exponential smoothing

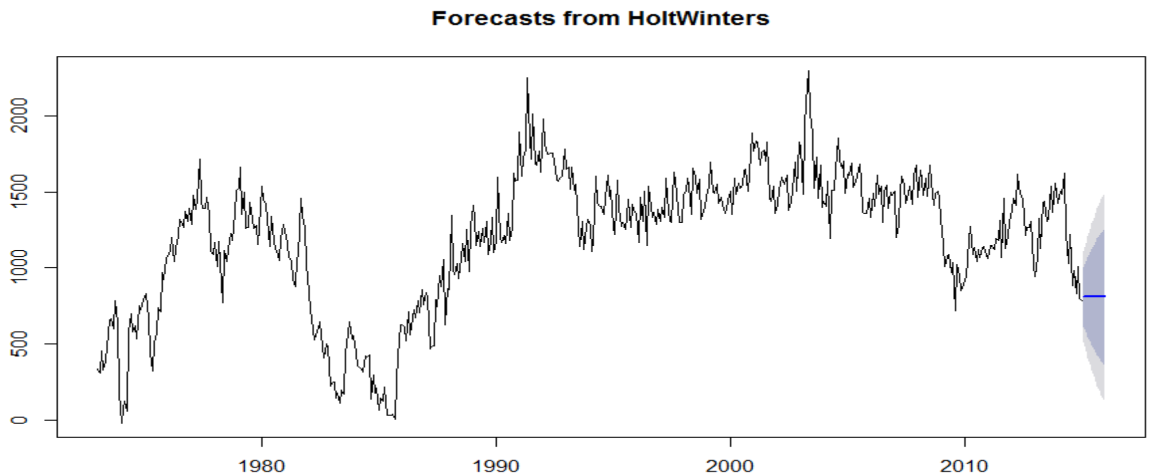

Figure 6: Plot of Forecasted series under exponential smoothing.

The ACF function can be used as a diagnostic tool which will compute the correlogram (plot of autocorrelation values) of the forecast errors. The ACF of the residuals decline quickly to zero indicating that the residuals form a stationary series.

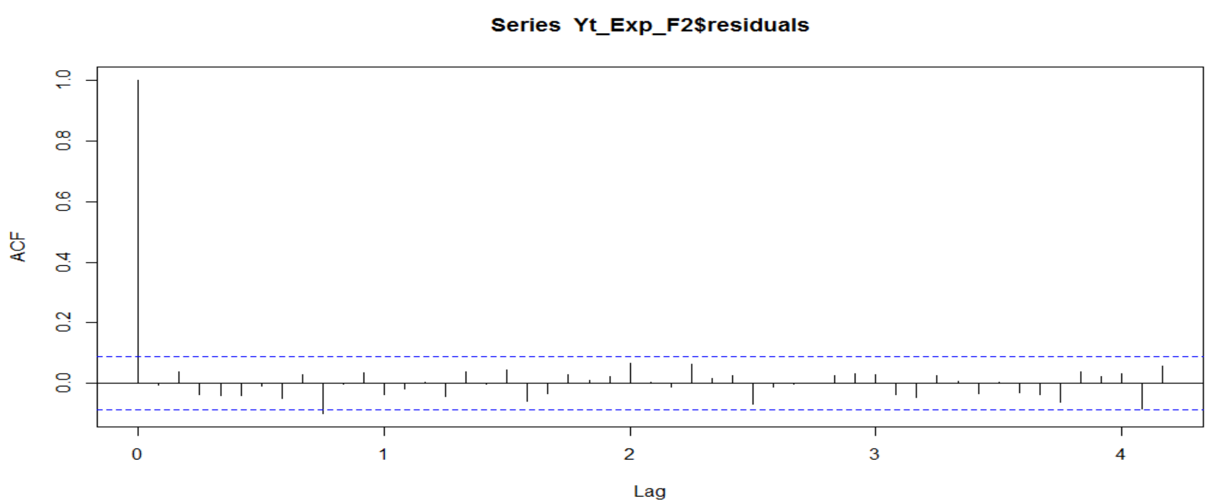

Figure 7 : Plot of Auto correlation function

The plots of ACF suggest the residuals autocorrelation and assumption are not violated for a valid time series model. Box –Ljung test can be used a formal test for auto correlation among residuals.

In a formal test of Box-Ljung test

Data: Yt_Exp_F2$residuals

X-squared = 16.8533, df = 20, p-value = 0.6625

16.8533is the Ljung-Box test figure, whereas, 0.6535 is the p-value, therefore, the in-sample forecast inaccuracy on lags 1-20 contains little proof of non-zero autocorrelations.

4. ARIMA Modeling and Forecasting

A specific time series models aimed at the unequal part of the time series with non-zero auto correlations in the unequal part are included in the autoregressive integrated Moving Average (ARIMA) models. Stationary time series are being defined using models of ARIMA. Thus, it is necessary to initially 'differentiate 'the time series pending getting a motionless time series, when starting off with a non-motion provided that there is the variation in the time series d times for achieving a series that is motionless.

Here seasonally adjusted time series is divided into two groups. The values up to December 2003 is taken as Model building Period and observation from January 2004 to January 2015 is taken for Model validation period.
A formal test for stationary is achieved using the augmented Dickey-Fuller Test (ADF) test. ADF test for the original series suggest that the series is not motionless. But there is significant stationary test for the first difference gives indication as first differenced series is motionless.

The Original Series

Augmented Dickey-Fuller Test
Data: Yt_L
Dickey-Fuller = -2.1797, Lag order = 7, p-value = 0.5012
Alternative hypothesis: stationary

Differenced series

Augmented Dickey-Fuller Test
Data: Yt_L_d1
Dickey-Fuller = -7.6617, Lag order = 7, p-value = 0.01
Alternative hypothesis: stationary

4.1 Model Identification

The previous analysis indicated that the first difference of the series is stationary. Now we are applying graphical techniques to identify the other parameter (p,q) of the ARIMA model. ACF and PACF plots are used for this purpose. There is evidence of decay on the ACF plot with ensuing oscillation which is typical of an AR(1) model. In the meantime, after lag 1 the value of auto correlogram becomes zero while the partial auto correlogram decreases to zero and therefore a moving average model of order q=1.

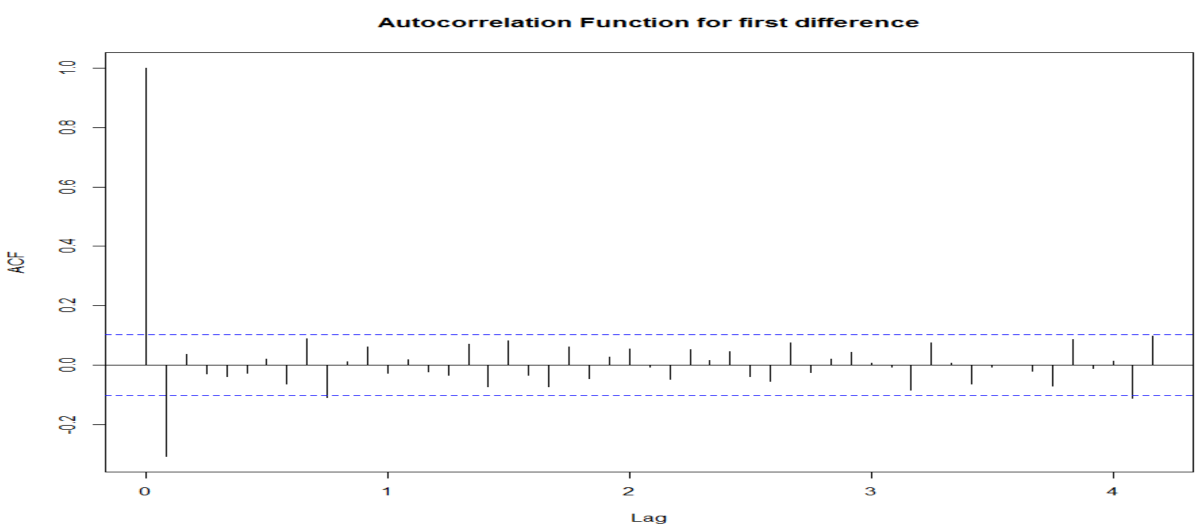

Figure 8 : Auto correlogram for the first difference

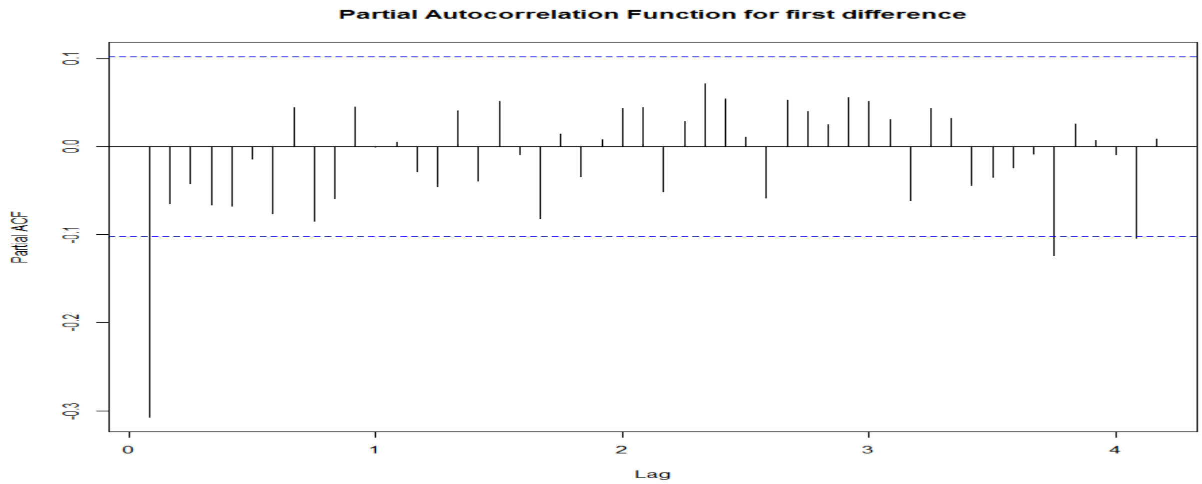

***Figure 9* :** Partial auto correlogram for the first difference

We consider two models for the final selection ARIMA (1, 1, 1) and ARIMA (1,1,0). Principle of parsimony, AIC, BIC is used to choose the best model. The principle of parsimony is very important in the modeling time series. Representing the systematic structure with the minimum possible parameters is what the principle signifies. Basically, this indicate that ,if both are adequate, minimal representations of a time series process are more desirable than more complex ones.

THE ARIMA (1,1,1)

Call:
arima(x = Yt_L, order = c(1, 1, 1))
Coefficients:
ar1 ma1
0.0017 -0.3353
s.e. 0.2381 0.2324

The value 24053 is the approximation of sigma^2: log likelihood = -2397.81, AIC = 4801.62

ARIMA (1,1,0)

Call:
arima(x = Yt_L, order = c(1, 1, 0))
Coefficients:
ar1
-0.3082
s.e. 0.0495

The value 24205 is the approximation of sigma^2: log likelihood = -2398.98, AIC = 4801.95

The values of AIC suggest that there is not much difference in these information criteria. Therefore, using the principle of parsimony, the ARIMA(1,1,0) is the best model. The values of the number of barrels for the year 2004 is given below.

Point	**Forecast**	**Lo 80**	**Hi 80**	**Lo 95**	**Hi 95**
Jan 2004	1482.264	1282.878649	1681.649	1177.330476	1787.197
Feb 2004	1456.056	1213.607841	1698.505	1085.263319	1826.850
Mar 2004	1464.133	1175.358051	1752.908	1022.489746	1905.777
Apr 2004	1461.644	1135.754810	1787.533	963.239479	1960.049
May 2004	1462.411	1102.462449	1822.360	911.917123	2012.905
Jun 2004	1462.175	1071.338181	1853.011	864.441811	2059.908
Jul 2004	1462.248	1042.729516	1881.766	820.650065	2103.845
Aug 2004	1462.225	1015.882701	1908.568	779.603290	2144.847
Sep 2004	1462.232	990.583171	1933.881	740.907336	2183.557
Oct 2004	1462.230	966.566482	1957.893	704.178108	2220.282
Nov 2004	1462.231	943.663041	1980.798	669.149970	2255.311
Dec 2004	1462.230	921.728579	2002.732	635.604212	2288.857

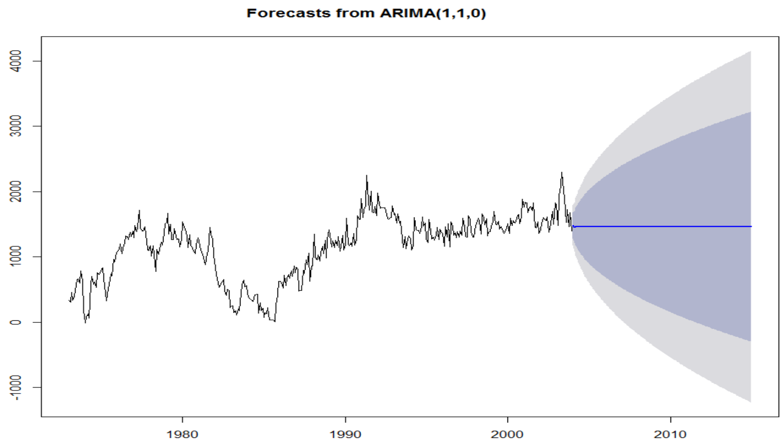

Figure 10 : Time Series plot of fitted values from ARIMA (1,1,0)

4.2 Diagnostics

The verification of the ARIMA (1, 1, 0) model goodness-of-fit as compared to the original series are done using the residual diagnostic tests and the over fitting process.

The ACF function and Box-Ljung test being used as a diagnostic test is employed in the computation of the correlogram of the inaccuracy in the forecast derived from ARIMA (1, 1, 0) model.

Box-Pierce test
data: arima2forecast$residual
X-squared = 0.1645, df = 1, p-value = 0.6851
The correlogram and Box-Ljung test suggest that model assumptions are valid for the ARIMA model.

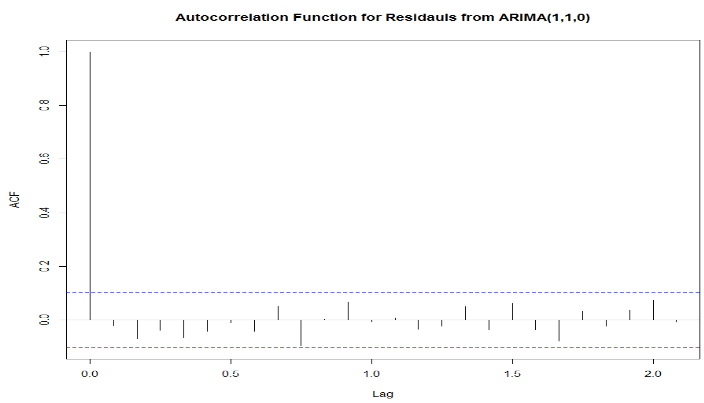

***Figure 11:** ARIMA (1, 1, and 0) is the source of this ACF of residuals*

An indication that the residuals are independent and identically distributed around the value zero are given by normal Probability Plot, Histogram and index plot of residuals from ARIMA(1,1,0).

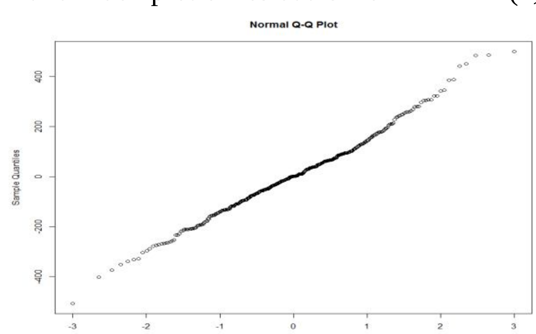

Figure 12: Normal Probability Plot of residuals

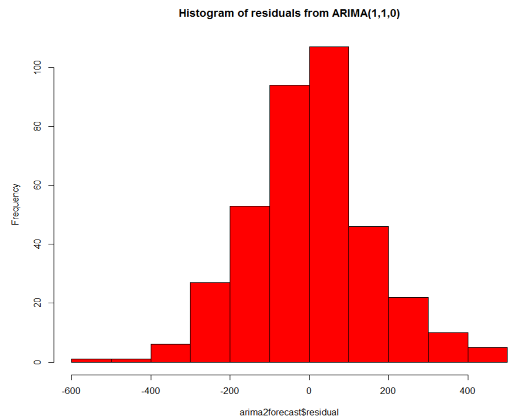

Figure 13: *Histogram of residuals from ARIMA (1,1,0)*

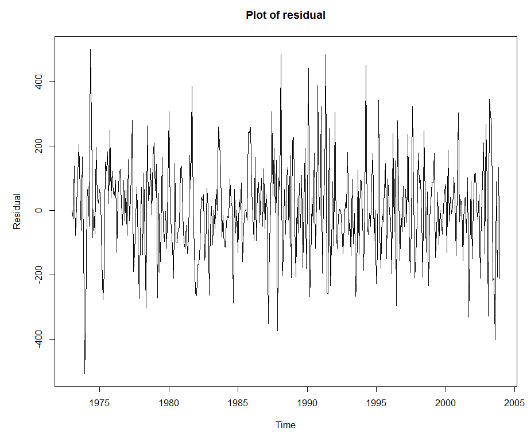

Figure 14: *Index Plot of residuals from ARIMA(1,1,0)*

4.3 The Model Validation

Values predicted are compared with the true values of the series in the validation period. The following table gives a comparison of actual and forecasted values during the initial months of the data validation period.

	Actual	Forecast	L. CI 95	U. CI.95
Jan ,2004	1477	1482.264	1177.33	1787.197
Feb, 2004	1369	1456.056	1085.263	1826.85
Mar, 2004	1531	1464.133	1022.49	1905.777
April, 2004	1177	1461.644	963.2395	1960.049
May,2004	1519	1462.411	911.9171	2012.905

The above table suggests that the forecasted values of number of barrels are very close to the true value.

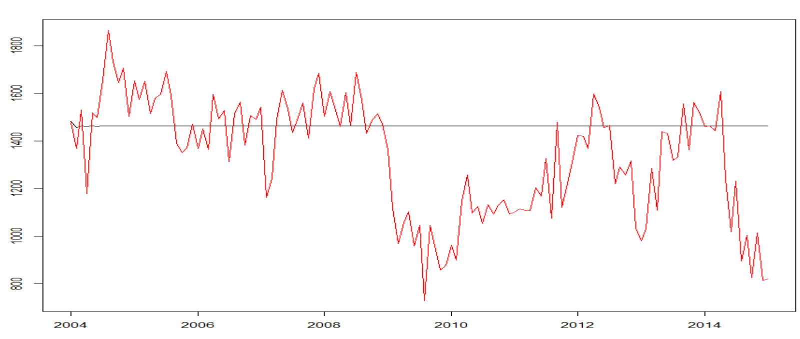

Figure 15: Plot of Time series values in the data validation period.

5. Conclusions

This study focuses the time series modeling of U S importations of Saudi petroleum and its products and sources of this information is the United States Energy Information Administration (EIA).Exponential smoothing techniques and Box-Henkin ARIMA techniques are used for this purpose. The accuracy of the fitted values is evaluated using model diagnostics methods and comparing the predicted values with observed number of barrels.

The best fitted model for the series as suggested by the analysis suggest ARIMA (1, 1, 0). The analysis of residuals from the model suggests that the residuals satisfy the model assumptions perfectly. The predicted from the model is compared with the values of the series in the validation period. It was found that the forecasted values are very close to the original values of the series in the validation period.

Reference

Christopher Chatfield (2010). The Analysis of Time Series: An Introduction, Sixth Edition, Taylor & Francis.

Cowpertwait P.S.P., Metcalfe A.V. (2009), Introductory Time Series with R, Springer.

Robert H. Shumway, David S. Stoffer (2010). Time Series Analysis and Its Applications: With R Examples. Springer Texts in Statistics, Springer.

Shumway R. H., Stoffer (2011), Time Series Analysis and its Applications, Springer.

U.S. Energy Information Administration (EIA). URL http://www.eia.gov/.

Woodward, W. A., Gray, H. L. & Elliott, A. C. (2012), Applied Time Series Analysis, CRC Press.