International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.05, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

# Forecasting Crude Oil Price Using ARIMA Models

**J.Jaya Selvi[1], R.Kaviya Shree[2], J.Krishnan[3]**

*[1]I B.Sc. Mathematics Department of Mathematics,*

*Sri Krishna Adithya College of Arts and science, Coimbatore, Tamil Nadu(India)*

*[2]IB.Sc. Mathematics Department of Mathematics,*

*Sri Krishna Adithya College of Arts and science, Coimbatore, Tamil Nadu(India)*

*[3]Assistant Professor, Department of Mathematics,*

*Sri Krishna Adithya College of Arts and science, Coimbatore, Tamil Nadu(India)*

## ABSTRACT

*The price of crude oil means a lot more to us than just paying a few dollars more at the pump. As we can see, oil is vital to the production and shipping of many of the items that we use on a regular basis. This means that many sectors of the economy will be unfavourably affected by increasing oil prices, or helped when they go down. When we think of oil, we tend to think of fuel for our cars, trucks, and planes, and heating oil. So that estimating crude oil prices are very vital. There are many methods used to model and forecast time series data such as trend, regression, moving average etc. In this paper,ARIMA method is considered to predict inflation for upcoming years. ARIMA techniques are used to analyze time series data and have been mainly used for loading forecast due to its accuracy, mathematical soundness and flexible due to inclusion of AR and MA terms over a regression analysis. A detailed explanation of the above is presented and summarized in tables and figures using MINITAB software.*

***Keywords: Time series, Crude Oil, Forecast, AR, MA, ARIMA.***

## I. INTRODUCTION

Crude oil is one of the most economically mature commodity markets in the world. Even though most crude oil is produced by a relatively small number of companies, and often in remote locations that are very far from the point of consumption, trade in crude oil is robust and global in nature. Nearly 80% of international crude oil transactions involve delivery via waterway in super tankers. Oil traders are able to quickly redirect transactions towards markets where prices are higher. Oil and coal are global commodities that are shipped all over the world. Thus, global supply and demand determines prices for these energy sources. Events around the world can affect our prices at home for oil-based energy such as gasoline and heating oil. Oil prices are high right now because of rapidly growing demand in the developing world (primarily Asia). As demand in these places grows, more oil cargoes head towards these countries. Prices in other countries must rise as a result. Political unrest in some oil-producing nations also contributes to high prices - basically, there is a fear that political instability could shut down oil production in these countries.

**PETROLEUM**

Petroleum (from Ancient Greek word petra: “rock"+ Latin word oleum: “: oil") is a naturally occurring, yellow-to-black liquid found in geological formations beneath the Earth's surface. It is commonly refined into various types of fuels. Components of petroleum are separated using a technique called fractional distillation i.e. separation of a liquid mixture into fractions differing in boiling point by means of distillation, typically using a fractionating column. It consists of hydrocarbons of various molecular weights and other organic compounds. The name petroleum covers both naturally occurring unprocessed crude oil and petroleum products that are made up of refined crude oil. A fossil fuel, petroleum is formed when large quantities of dead organisms, usually zooplankton and algae, are buried underneath sedimentary rock and subjected to both intense heat and pressure.

## 1. Time Series Analysis

Time series is a series of data points indexed or graphed in the order of time. It is often plotted using line charts. Time series are used in statistics, signalprocessing, pattern recognition, econometrics, finance, forecasting of weather, earthquake prediction, control engineering, astronomy, and mostly in any area of applied science and engineering. Time series analysis consists of techniques for analyzing time series data in order to dig out meaningful characteristics of the data. Time series analysis can be useful to real-valued, continuous data, discrete numeric data, or discrete symbolic data. Time series forecasting is the technique used in a model to forecast future values based on previously observed values.

There are many different notations used for time-series analysis. A common notation specifying a time series $X$ that is indexed by the natural numbers is written as

$$X = \{X_1, X_2, ...\}.$$

Another common notation is

$$Y = \{Y_t: t \in T\},$$

where $T$ is the index set.

## II. ARIMA MODEL

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model in time series analysis. These models are fitted to time series data for understanding the data better or to forecast future points in the series. ARIMA models are applied to the data that shows evidence of non-stationarity, where an initial differencing step can be applied one or more times to remove the non-stationarity.The AR part of ARIMA shows that the variable of interest is regressed on its own lagged values. The MA part shows that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. ARIMA models can be estimated following the Box–Jenkins approach.

## 1. Autocorrelation

Autocorrelation is the correlation between a variable lagged one or more periods and itself. Autocorrelation coefficients for different time lags of a variable are used to identify time series data patterns. The formula for computing the lag *k* autocorrelation coefficient between $Y_t$ and $Y_{t-k}$, which are k periods apart, is given by

$$r(k) = \frac{\sum_{t:k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t:1}^{n} (Y_t - \bar{Y})^2}, \quad for \ \ k = 0, 1, 2, \ldots$$

where $\bar{Y}$ the mean value of the values of the series is, $Y_t$ is the observation in time period *t* and $Y_{t-k}$ is the observation *k* time periods earlier or at time period *t-k*.

## 2. Partial Autocorrelation

A partial autocorrelation at time lag *k* is the correlation between $Y_t$ and $Y_{t-k}$, after adjusting for the effects of the intervening values $Y_{t-1}, Y_{t-2,}, \ldots, Y_{t-k+1}$. In time series analysis, the partial autocorrelation function (PACF) plays an important role in data analyses aimed at identifying the extent of the lag in an autoregressive model. The use of this function was introduced as part of the Box-Jenkins approach to time series modeling, where by plotting the partial auto correlative functions one could determine the appropriate lag *p* in an *AR*(*p*) model or in an extended *ARIMA* (*p, d, q*) model.

## 3. Models

Box Jenkins Model assumes that a time series is a linear function of past actual values and error terms. That is,

$$Y_t = b_t + \varepsilon_t$$

The error terms are distributed as normally and independently distributed, having no pattern, with a mean of zero and an error variance that is lower than the variance of $Y_t$. With these assumptions, the Box-Jenkins models are classified as Auto Regressive (AR), Moving Average (MA), or a combination of the two or Auto Regressive Integrated Moving Averages (ARIMA) models. The standard notation *p* identifies the order of autoregressive, *d* identifies integration or differencing and *q* identifies the moving averages.

## 4. Auto Regressive model

It is one in which the current value of the variable is a function of previous values and an error term. The reason that this so called an auto regressive model is because $Y_t$ is being regressed on itself as

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + \varepsilon_t,$$

where $Y_t$ is the dependent variable, $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ are the independent variables based on the dependent variable lagged (*p*) specific time periods, $\Phi_0, \Phi_1, \ldots, \Phi_p$ are the computed regression coefficients and $\varepsilon_t$ is the random error term measured in time *t*.

### 5. Moving Averages model

Link the current values of the time series to random errors that have occurred in previous time periods. A moving average model is as follows

$$Y_t = \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

where $Y_t$ is the dependent variable, $\theta_0$ is the mean about which the series fluctuates, $\theta_0, \theta_1, \theta_2 \ldots, \theta_q$ are the moving average parameters to be estimated, $\varepsilon_{t-1}, \varepsilon_{t-2}, .. \varepsilon_{t-q}$ are the error terms and $\varepsilon_t$ is the random error term measured in time *t*. The highest order of the model is called *q* and refers to the number of lagged time period of the model.

### 6. ARIMA model

It is the combination of the AR and MA models. Thus, the model is given as

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} - \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

When using the ARIMA model, it is able to use a combination of past values and past errors. The order of the model is commonly written as *(p, d, q)*. To select an appropriate model for forecasting, one would depend on the autocorrelation (*AC*) and partial autocorrelation (*PAC*) statistics of time series

## III. ANALYSIS OF TIME SERIES DATA ON CRUDE OIL PRICE

This section presents the analysis of time series data on crude oil price. Appropriate time series models for making forecasts for future period are constructed using Box –Jenkins Methodology.

In order to forecast the crude oil price for future periods the time series analysis is carried out for the given data. The scatter plot of the time series data against time periods is given in fig1. From this figure, it is observed that there is upward trend in the observed time series data. Hence, the observed trend has to be removed.In order to verify whether the data is stationary time series plot has plotted and it is shown in the fig 2. It is observed that the time series data is non stationary and did not vary about a fixed level.Hence, it is necessary to transform the non-stationary data into stationary.For eliminating the trend and create a stationary series, the first differences are found and are presented in table2. The first difference series is also plotted and is displayed in fig3. The plot of the first differences against the lagged time variables indicates the stationary.

The values of autocorrelation and partial autocorrelation functions are given in table 4 and the plots of those values are displayed in fig4 and5 respectively. The fig4 indicates that the auto-correlations have one spike which indicates a MA (1) behavior. At the same time fig5 reveals that the partial autocorrelations have two spikes by which it is inferred that an auto-regressive of order two is appropriate for the time series data. As differencing is done one times for converting the non-stationary data into stationary data it is now possible to define an auto-regressive integrated moving average (ARIMA) model as an appropriate model in studying the given time series data. As the moving average model of order one and the autoregressive model of order two are identified on the basis of the pattern exhibited by ACs and PACs with the first order differencing, the ARIMA model is defined with reference to the parameters (*p, d, q*) = (2, 1, 1). This model can be represented as

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} - \theta_0 - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

For the given time series data, the ARIMA (2, 1, 1) model is fitted using Minitab statistical software (Version 16). The model parameters are estimated as

$\Phi_0 - \theta_0 = 0.1547,\ \Phi_1 = 0.8269\ \Phi_2 = -0.0233$ and $\theta_1 = 0.9479.$ Thus, the fitted time series model for crude oil price is given by

$$Y_t = 0.1547 + 0.8269_{t-1} - 0.0233_{t-2} - 0.9479\varepsilon_{t-1} + \varepsilon_t$$

The estimates of the model parameters are provided in table 5 along with their standard errors. It can be observed from the results that the estimate of MA1 model is highly significant that the P-values corresponding to such estimates are equal to zero. From these observations of times series analysis it is inferred that the constructed ARIMA model is adequate. The forecasts for the period from 2017 – 2021 are generated based on the model and are displayed in table 6. The fig6 displays residual plots of residuals. When the residuals are plotted against the fitted values, it is observed that the error variability is constant and the underlying relationship $Y_t$ and $t$ appears to be linear. The fig 7 shows that the time series plot of crude oil price against various time points along with the forecasted values are also plotted.

## IV. FIGURES AND TABLES

Table1:Crude Oil price data during 1946 - 2016

| Year | Crude Oil Price | Year | Crude Oil Price | Year | Crude Oil Price | Year | Crude Oil Price | Year | Crude Oil Price |
|---|---|---|---|---|---|---|---|---|---|
| 1946 | 1.63 | 1961 | 2.83 | 1976 | 13.1 | 1991 | 20.2 | 2006 | 58.3 |
| 1947 | 2.16 | 1962 | 2.85 | 1977 | 14.4 | 1992 | 19.25 | 2007 | 64.2 |
| 1948 | 2.77 | 1963 | 2.91 | 1978 | 14.95 | 1993 | 16.75 | 2008 | 91.48 |
| 1949 | 2.77 | 1964 | 3 | 1979 | 25.1 | 1994 | 15.66 | 2009 | 53.48 |
| 1950 | 2.77 | 1965 | 3.01 | 1980 | 37.42 | 1995 | 16.75 | 2010 | 71.21 |
| 1951 | 2.77 | 1966 | 3.1 | 1981 | 35.75 | 1996 | 20.46 | 2011 | 87.04 |
| 1952 | 2.77 | 1967 | 3.12 | 1982 | 31.83 | 1997 | 18.64 | 2012 | 86.46 |
| 1953 | 2.92 | 1968 | 3.18 | 1983 | 29.08 | 1998 | 11.91 | 2013 | 91.17 |
| 1954 | 2.99 | 1969 | 3.32 | 1984 | 28.75 | 1999 | 16.56 | 2014 | 85.6 |
| 1955 | 2.93 | 1970 | 3.39 | 1985 | 26.92 | 2000 | 27.39 | 2015 | 41.85 |
| 1956 | 2.94 | 1971 | 3.6 | 1986 | 14.44 | 2001 | 23 | 2016 | 36.34 |
| 1957 | 3.14 | 1972 | 3.6 | 1987 | 17.75 | 2002 | 22.81 | | |
| 1958 | 3 | 1973 | 4.75 | 1988 | 14.87 | 2003 | 27.69 | | |
| 1959 | 3 | 1974 | 9.35 | 1989 | 18.33 | 2004 | 37.66 | | |
| 1960 | 2.91 | 1975 | 12.21 | 1990 | 23.19 | 2005 | 50.04 | | |

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.05, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Table 2: First Order Differences for Checking Stationary in Crude Oil price

| Year | Fod | Year | fod | Year | fod | Year | fod | Year | Fod |
|---|---|---|---|---|---|---|---|---|---|
| 1946 | | 1961 | -0.08 | 1976 | 0.89 | 1991 | -2.99 | 2006 | 8.26 |
| 1947 | 0.53 | 1962 | 0.02 | 1977 | 1.3 | 1992 | -0.95 | 2007 | 5.9 |
| 1948 | 0.61 | 1963 | 0.06 | 1978 | 0.55 | 1993 | -2.5 | 2008 | 27.28 |
| 1949 | 0 | 1964 | 0.09 | 1979 | 10.15 | 1994 | -1.09 | 2009 | -38 |
| 1950 | 0 | 1965 | 0.01 | 1980 | 12.32 | 1995 | 1.09 | 2010 | 17.73 |
| 1951 | 0 | 1966 | 0.09 | 1981 | -1.67 | 1996 | 3.71 | 2011 | 15.83 |
| 1952 | 0 | 1967 | 0.02 | 1982 | -3.92 | 1997 | -1.82 | 2012 | -0.58 |
| 1953 | 0.15 | 1968 | 0.06 | 1983 | -2.75 | 1998 | -6.73 | 2013 | 4.71 |
| 1954 | 0.07 | 1969 | 0.14 | 1984 | -0.33 | 1999 | 4.65 | 2014 | -5.57 |
| 1955 | -0.06 | 1970 | 0.07 | 1985 | -1.83 | 2000 | 10.83 | 2015 | -43.75 |
| 1956 | 0.01 | 1971 | 0.21 | 1986 | -12.48 | 2001 | -4.39 | 2016 | -5.51 |
| 1957 | 0.2 | 1972 | 0 | 1987 | 3.31 | 2002 | -0.19 | | |
| 1958 | -0.14 | 1973 | 1.15 | 1988 | -2.88 | 2003 | 4.88 | | |
| 1959 | 0 | 1974 | 4.6 | 1989 | 3.46 | 2004 | 9.97 | | |
| 1960 | -0.09 | 1975 | 2.86 | 1990 | 4.86 | 2005 | 12.38 | | |

Table 3: Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag | 12 | 24 | 36 | 48 |
|---|---|---|---|---|
| Chi-square | 12.9 | 16.5 | 23.3 | 24.3 |
| DF | 8 | 20 | 32 | 44 |
| P- value | 0.115 | 0.683 | 0.876 | 0.993 |

Table 4:Values of ACF and PACF Coefficients along With the Values of T-statistics

| LAG | ACF | T-Statistics | LBQ | PACF | T-Statistics |
|---|---|---|---|---|---|
| 1 | -0.070539 | -0.59017 | 0.3634 | -0.070539 | -0.59017 |
| 2 | -0.035484 | -0.29541 | 0.4568 | -0.040662 | -0.34020 |
| 3 | 0.106202 | 0.88306 | 1.3052 | 0.101400 | 0.84838 |
| 4 | -0.145599 | -1.19738 | 2.9240 | -0.134532 | -1.12557 |
| 5 | -0.043833 | -0.35331 | 3.0730 | -0.056097 | -0.46938 |
| 6 | 0.259196 | 2.08550 | 8.3636 | 0.242426 | 2.02828 |
| 7 | -0.162007 | -1.22937 | 10.4633 | -0.121176 | -1.01383 |
| 8 | -0.002314 | -0.01719 | 10.4638 | -0.017659 | -0.14775 |

| | | | | | |
|---|---|---|---|---|---|
| 9 | -0.144916 | -1.07667 | 12.1989 | -0.224416 | -1.8776 |
| 10 | -0.124903 | -0.91298 | 13.5094 | -0.056692 | -0.47432 |
| 11 | -0.021423 | -0.15476 | 13.5486 | -0.055230 | -0.46209 |
| 12 | -0.046778 | -0.33781 | 13.7387 | -0.118433 | -0.99088 |
| 13 | -0.052082 | -0.37550 | 13.9786 | -0.030906 | -0.25857 |
| 14 | -0.015088 | -0.10857 | 13.9990 | -0.089507 | -0.74887 |
| 15 | -0.058651 | -0.42194 | 14.3143 | 0.016561 | 0.13856 |
| 16 | -0.006629 | -0.04757 | 14.3184 | -0.048495 | -0.40573 |
| 17 | 0.023217 | 0.16660 | 14.3696 | -0.005526 | -0.04623 |
| 18 | 0.005891 | 0.04225 | 14.3730 | -0.020580 | -0.17219 |

Table 5: Model Parameters

| Type | Coef | SE Coef | T | P |
|---|---|---|---|---|
| AR 1 | 0.8269 | 0.163 | 5.07 | 0.000 |
| AR 2 | -0.0233 | 0.1451 | -0.16 | 0.873 |
| MA 1 | 0.9479 | 0.1466 | 6.47 | 0.000 |
| Constant | 0.1547 | 0.1317 | 1.17 | 0.224 |

Table 6: Forecast Values for crude oil price

| Period | Forecasts |
|---|---|
| 2017 | 39.9565 |
| 2018 | 43.2298 |
| 2019 | 46.0071 |
| 2020 | 48.3822 |
| 2021 | 50.4363 |

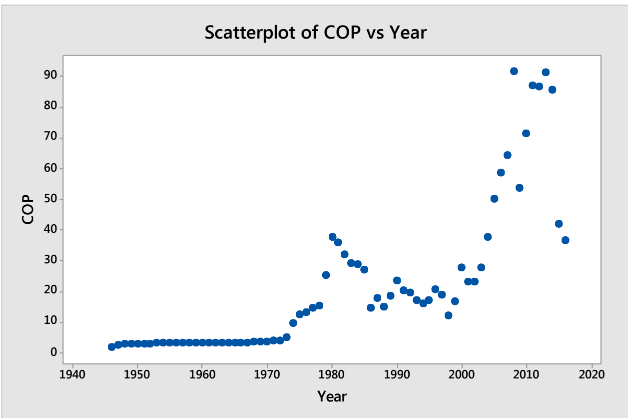


Fig1: scatter plot for crude oil price against time

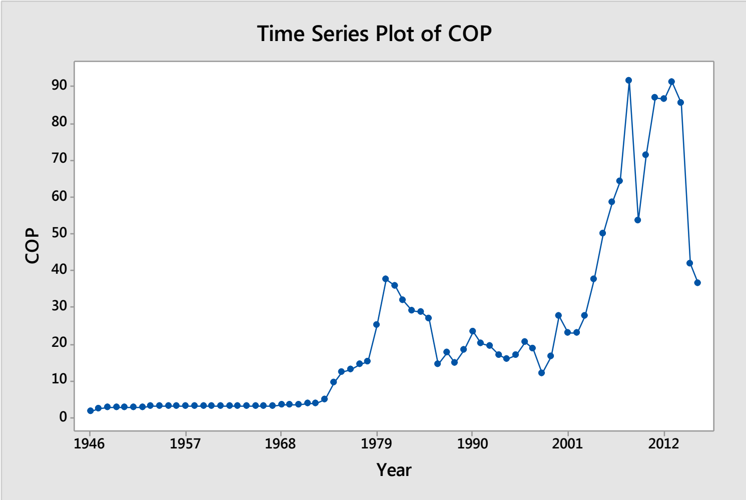


Fig 2: Time series plot of crude oil price against time

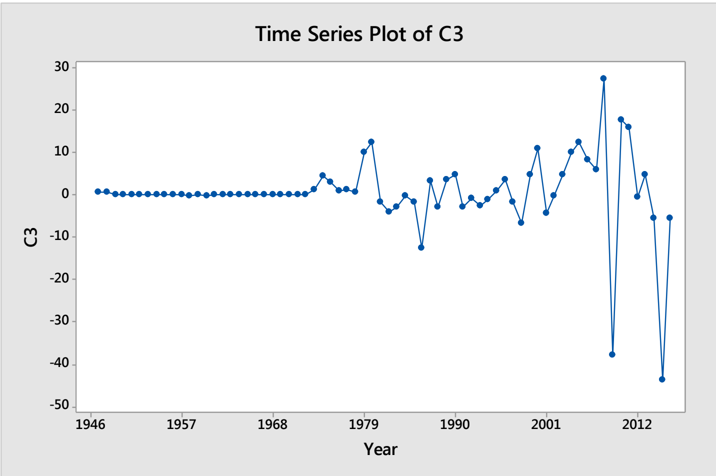


Fig 3: plot of first order differences against time lag

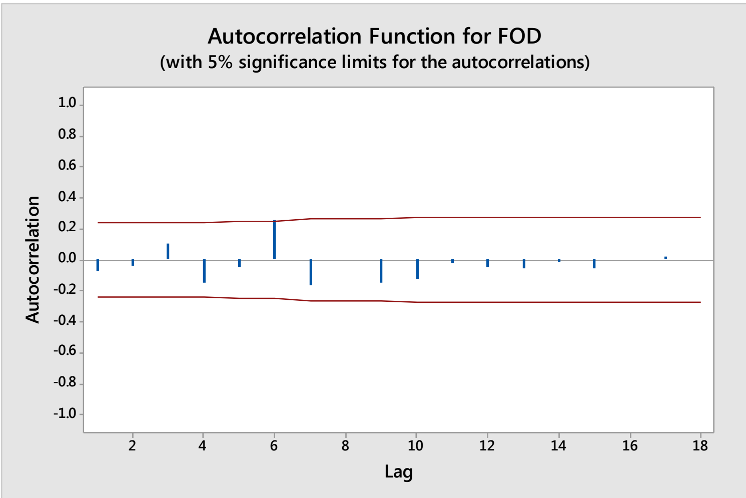


Fig 4: plot of Auto correlation against time lag

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.05, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

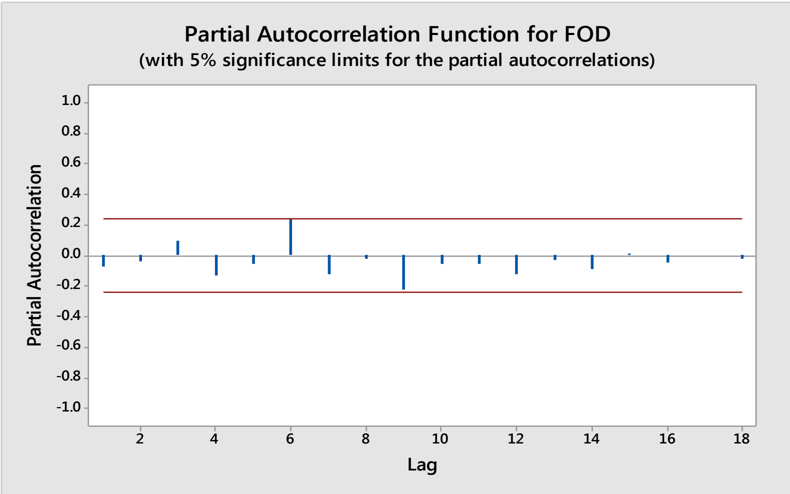


Fig 5: plot of Partial Auto correlation against time lag

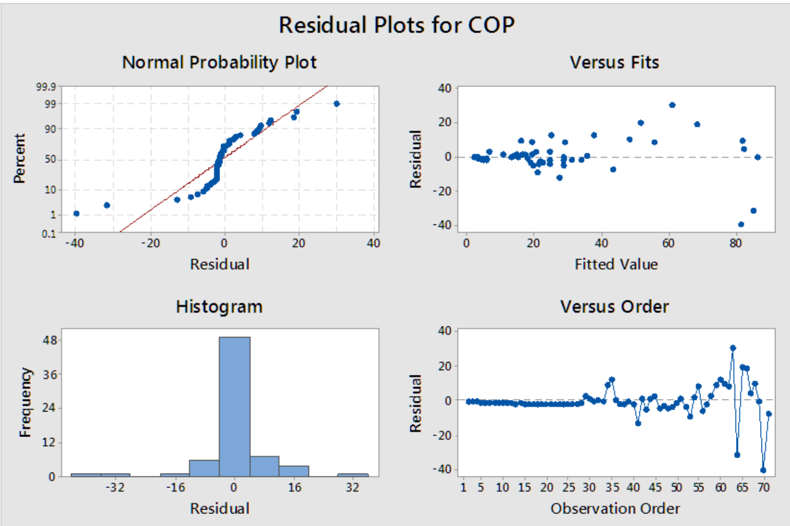


Fig 6: Residuals plots for crude oil price

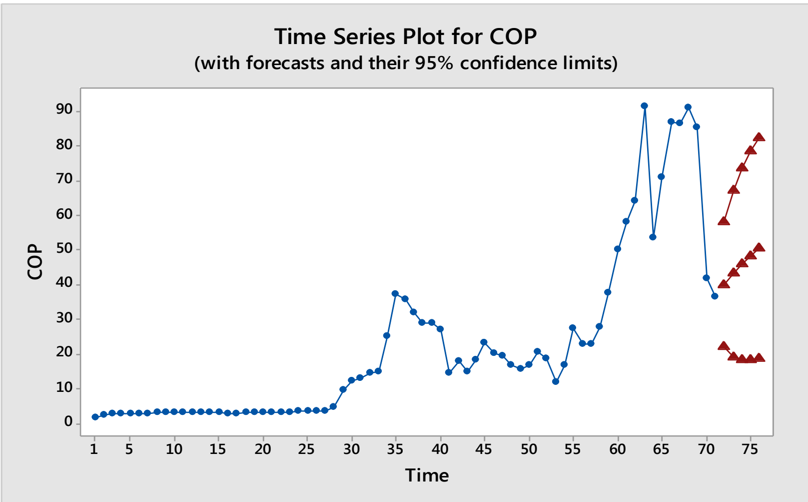


Fig 7: Time series plots for crude oil price with forecasting values

## V. CONCLUSION

ARIMA is one of the useful techniques in forecasting the data. In this paper, the data is analyzed using ARIMA model. The results are summarized in time series plot and ARIMA model and the future values are forecasted using the model. The prices of crude oil are normal during 1946 – 1972, which was identified as "The early postwar era". Then afterwards the prices of crude oil were increasing due to major oil crisis, this occurred in 1973 when Arab members of OPEC decided to quadruple (i.e.,) increased to four times the price of oil to almost $12 a barrel. It is substantially less important and remains economically in the nineteenth century than it is today. The prices of crude oil in the upcoming year will in an increasing manner. Even though the use of oil is different, the prices of oil should be maintained or at some instance, it will lead to a huge problem to the economy of a country. Hence, the prices should be normalized and special care should be given in monitoring the prices of oil.

## REFERENCES

[1] O. D. ANDERSON, *Identifying Box-Jenkins Processes*, *Computer Applications* vol. 2, 1974, p. 275-283.

[2] Paul Newbold, *ARIMA model building and the time series approach to forecasting* Journal of forecasting, vol.2, 23-35, 1983

[3] T. AMEMIYA and R. Y. WU, *The Effect of Aggregation on Prediction in the Autoregressive Model* J.A.S.A., vol. 67, 1972, p. 628-632.

[4] O. D. ANDERSON, *an inequality with a time series application* J. Econom.,vol. 2, 1974, p. 189-193

[5] M.A. Adelman, *the World Petroleum Market,* Baltimore: Johns Hopkins
University Press, 1972.

[6] Barsky, R. B. and Kilian, L. *Do We Really Know that Oil Caused the*
*Great Stagflation? A Monetary Alternative,* in B.S. Bernanke and K. Rogoff (Eds.) *NBER Macroeconomics Annual 2001*. Cambridge, MA: MIT Press, 2001.

[7] Caplinger, M. W., Allegheny *Oil Heritage Project: A Contextual*
*Overview of Crude Oil Production in Pennsylvania'*, 1997, HAER No. PA-436.

[8] James D. Hamilton, *Historical Oil Shocks*, Department of Economics, University of California, San Diego.