E-ISSN 2581 – 7957
P-ISSN 2277 – 3916
CVR Journal of Science and Technology, Volume 25, December 2023
DOI: 10.32377/cvrjst2516

Crude Oil Price Forecasting – ARIMA Model Approach

V. Swapna
Sr. Asst. Professor, CVR College of Engineering/H&S(Statistics)Department, Hyderabad, India
Email: swapnacvr82@gmail.com

Abstract: **The price of crude oil forecasting has become more significant in the trade of day-to-day services of energy. A good predicting model increases the efficiency of producers, and consumers, and their prices are also playing a vital role in the investment process of financial trade. To analyze and forecast time series, a well-known ARIMA model is used in this paper to model month-ahead spot prices. The proposed model has been applied a series of times which consists of the monthly prices of crude oil from the Information of the U.S. Energy Administration.**

Index Terms: **Crude oil, Box-Jenkins Methodology, MAPE, and Forecasting**.

I. Introduction

In the terminology of economics, “energy” includes all the commodities and the resources of energy by providing the ability to accomplish work. The resources of Energy commodities like crude oil, diesel fuel, gasoline, natural gas, etc., could be used to offer the services of energy as per the requirements of human actions, namely cooking, electricity, electronic activity, lighting, motive power, space, and water heating. The resources of Energy played a basic role in shaping the human lifestyle. In the present days, the need of energy has become very important for the survival of people, so the production and the consumption of energy are the utmost dwelling needs of human survival on the earth.

A. About the data

This study employed the methodology of the data series with monthly crude oil prices beginning from Jan. 1986 to Nov. 2022. The data was taken from the Department of Energy, US: Administration in Energy Information: http://www.eia.doe.gov/

II. The Methodology

In this paper, the technique of traditional time series prediction, Auto-Regressive Integrated Moving Average (ARIMA) has been discussed. It is emphasized how the best suitable ARIMA models are to be applied in the forecasting price of crude oil.

A. ARIMA Modeling

Box and Jenkins presented the model, ARIMA, for the first time in 1976, then onwards, it became one of the most widely used forecasting strategies. [11] In the model of ARIMA, it has been assumed that the value of a variable in the future will be a linear combination of values and errors of the past.

$Y_{t=}$

$$\theta_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + - - - + \emptyset_p Y_{t-p} + \epsilon_T - \theta_1 \epsilon_{t-1} - \theta_2\ \epsilon_{t-2} - \cdots - \theta_q\ \epsilon_{t-q}$$

$$\theta_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + - - - + \emptyset_p Y_{t-p} + \epsilon_T - \theta_1 \epsilon_{t-1} - \theta_2\ \epsilon_{t-2} - \cdots - \theta_q\ \epsilon_{t-q}$$

__(1)

Here Y_t is a real value, and φ, ϴ are the coefficients, p, and q are integers that are repeatedly offered as autoregressive, ϵ_t has been the casual error at time t and moving average polynomials, in that order. Basically, the procedure has three phases: (i) Model classification, (ii) Parameter estimation, and (iii) Diagnostic examination. E.g., the model, ARIMA (1, 0, 1) ought to be characterized as follows. [1]

Equation (2):

$$Y_{t=}\theta_0 + \emptyset_1 Y_{t-1} + \epsilon_T - \theta_1 \epsilon_{t-1}$$

$$\theta_0 + \emptyset_1 Y_{t-1} + \epsilon_T - \theta_1 \epsilon_{t-1} \qquad \_\_\_\_\_\_\_\_\_\_ (2)$$

The first Equation is a particular case of the family of ARIMA models. In equation (1), If q = 0, it becomes an AR model of order p, and when p = 0, it reduces to a MA model of order q. The important project of the model, [3] ARIMA construction will have to decide the appropriate order of the model (p, q). Box and Jenkins (1976) established an approach in the construction of the models of ARIMA, that has been a central effect on the prediction of applications and the analyzation of time series. [2]

The Partial Autocorrelation Function (PACF) and the Autocorrelation Function (ACF) of the collected data serve as the basic tools for recognizing the order of the model, ARIMA as stated by Box and Jenkins. Stationarity is the cornerstone of time series analysis as the essential phase to grow a new model which is to be applied for estimation. When creating an ARIMA [8] model that is used for prediction. A mean and variance of a stationary time series are constant. When the time series of experimentation exhibits trend, and heteroscedasticity. The collected data is subjected to power transformation and differencing to remove the trend and to maintain a frequent variance before fitting a model, ARIMA.

A general univariate model known as Auto Regressive Integrated Moving Averages (ARIMA) was developed on the presumption that the forecasted time series is stationary and linear.

Received on 26-04-2023, Revised on 05-10-2023, Accepted on 30-09-2023

CVR Journal of Science and Technology, Volume 25, December 2023
DOI: 10.32377/cvrjst2516

The statistical records of data collection between July 1995 and November 2016 showed the average weekly price of crude is mapped as in Figure 1.

B. Time series plot

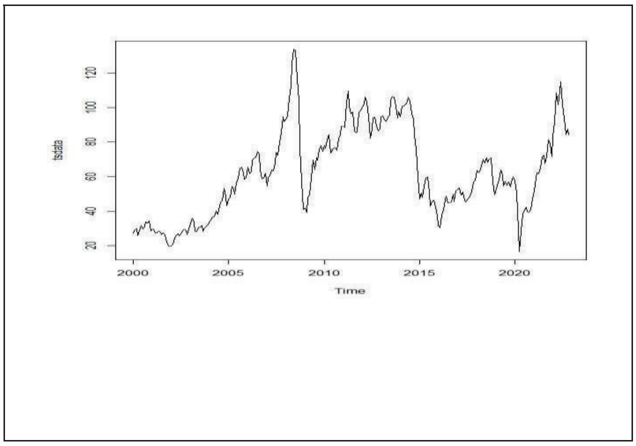

Figure 1. Time series plot of crude oil prices

It is observed as in Figure 1 that the series of time plot shows on average an increasing trend means non-stationary data. The model, ARIMA has been assessed only after converting the variable into a static series for forecasting. The natural transformation of the log is used to correct non-stationary variance and non-stationarity is corrected in an average by using suitable differentiating the data. Here, the variation of order 1 (i.e. d=1) can make data stationary in the mean. [6]

To identify the model parameters of the time series, normally we use ACF, PACF of the time series. ACF and PACF functions are used to determine the ARIMA model of a time series data. [9]

The ACF helps to identify the MA(q) hyper parameter, whereas PACF helps to identify the AR(p) hyper parameter. The difference between ACF, PACF shows the correlation between the point x(t) and a log function.

C. The time series data of ACF plot:

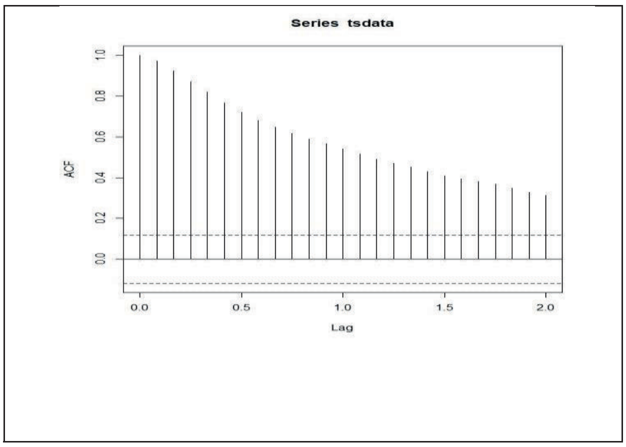

Figure 2. The prices of crude oil of the ACF plot.

D. The time series plot of PACF:

In the AR(p) process, the PACF has a significant spike at a certain lag q and PACF decay shows a sinusoidal behavior.

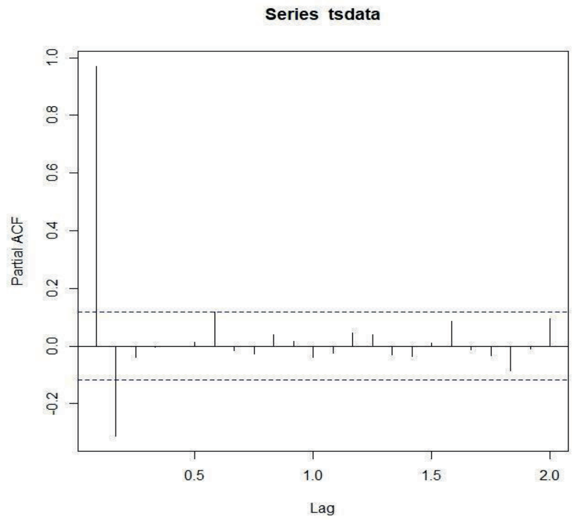

Figure 3. The prices of crude oil of PACF plot:

The above Figures 2,3 stated ACF and PACF plots show that the data is not stationary.

The given time series data is stationary at the variation of order 1 (d=1) and which is used for prediction of time series data for the year 2023.

The stationary time series data projection as follows at d=1

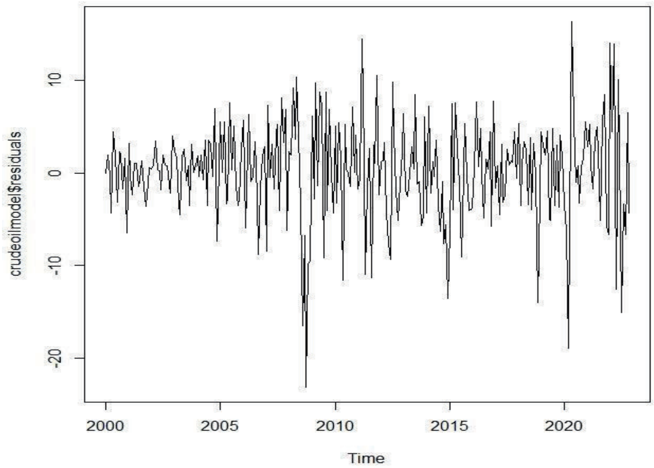

Figure 4. Residual plot of the crude oil prices

The Figure 4 shows that that the data has become stationary. The p value as 0.01 which proves this data has become stationary as given by Dickey-Fuller Test. Plotting from the ACF to the above differencing is given in Figure 5. [3].

E-ISSN 2581 – 7957
P-ISSN 2277 – 3916

DOI: 10.32377/cvrjst2516

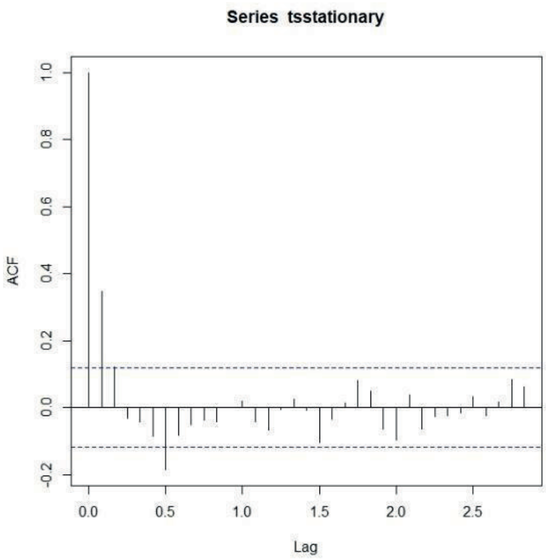

Figure 5. ACF of first difference data

For checking correlation between residuals, it gets the ACF. ACF has a significant spike at a certain lag p and ACF decay shows a sinusoidal behavior.

TABLE I.
ERROR MEASURES

Ljung Box Q test	Error measures						
Statistics	DF	P-value	MAPE	MAE	RMSE	BIC	AIC
17.058	20	0.6492	7.186	4.111	5.443	1718.39	1711.16

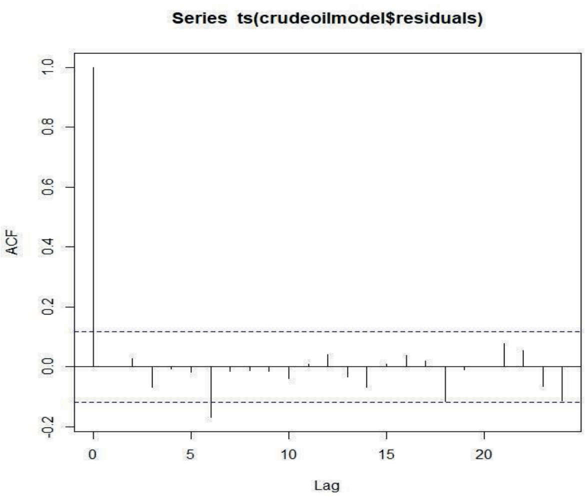

Figure 6. ACF of Residuals

The Figure 6 shows clearly that there is a little correlation between the residuals, hence, this forecasting is a good model.

Crude oil forecast:

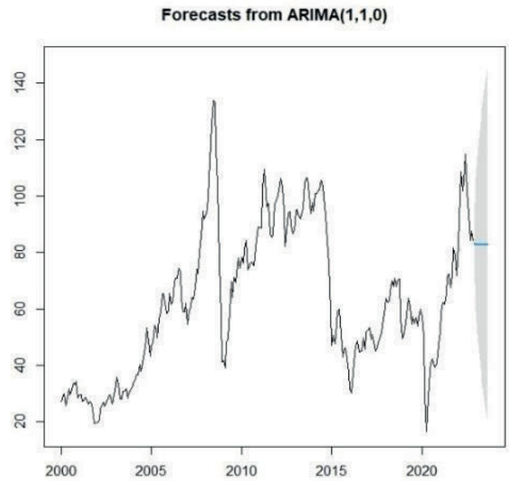

Figure 7. The prices of crude oil forecast:

The forecast values are given in Figure 7.

TABLE II.
ARIMA MODEL PARAMETERS

Model	**AIC**
ARIMA(1,1,0)	1711.16
ARIMA(0,1,1)	1715.99
ARIMA(0,1,0)	1744.41
ARIMA(1,1,1)	1713.16

From the above Table II of the ARIMA model parameters, it is observed that ARIMA (1,1,0) is fitting well for the data. [8] The parameters of the selected model are given in Table III.

TABLE III.
ARIMA MATHEMATICAL MODEL PARAMETER

Model	Parameter	P-value	Significance
AR(1)	0.3470	0.01	significant
MA(1)	0.0565	0.01	significant

The Mathematical model of ARIMA time series data for forecasting time series is

$$Yt = \mathbf{0.3470Y_{t-1} + 0.0565}$$

[8] For forecasting the prices of crude oil, the suitable model is.

TABLE IV.
FORECASTS OF THE PRICES OF CRUDE OIL

Year & Month	Prices (Dollars per Barrel)	Lower limit	Upper limit
Dec 2022	83.26	69.19	97.33
Jan 2023	82.88	59.27	106.48
Feb 2023	82.75	51.38	114.11
Mar 2023	82.70	44.82	120.57
Apr 2023	82.68	39.16	126.21
May 2023	82.68	34.13	131.23
Jun 2023	82.68	29.57	135.79
Jul 2023	82.68	25.36	139.99
Aug 2023	82.68	21.45	143.90
Sep 2023	82.68	17.76	147.59

From Table IV, it is observed from the forecasts that crude oil prices would reach 83 dollars per Barrel by September 2023 with the maximum scope of 147.59 Dollars per Barrel.

III. CONCLUSIONS

Among all the models used ARIMA (1,1,0) gave the best execution. There are theoretical reasons on why the model, ARIMA is a good prediction of Time Series data: [9]

The model, ARIMA, includes an amalgamation of both components of Auto Regressive (AR) and Moving Average (MA). The parameters of the ARIMA model have been estimated by minimizing the fitting errors. The ARIMA also determines the level of differencing used in making the data stationary to some extent.

Time series data forecasts may be accurate using a variety of sophisticated methods, but there are times when they significantly over fit the data. Additionally, most of them are deficient in flexibility. [11] The ARIMA has the advantage of being adaptable enough in capturing observed trends and associations as per the collected data while remaining simple enough but not to over fit the collected data.

The best prediction is provided by the neural network model when the analysis is extended to R.

In the proposed ARIMA model it is observed from the value of RMSE (5.443).

REFERENCES

[1] Ashok K. Nag and Amit Mitra (2002): "Forecasting Daily Foreign Exchange Rates Using Genetically Optimized Neural Networks", Journal of Forecasting, 21, 501-511.

[2] Box, G.E.P. and G.M. Jenkins (1970): "Time series analysis: Forecasting and control", San Francisco: Holden-Day.

[3] C.M. Kuan and T.Liu, "Forecasting exchange rates using feedforward and recurrent neural networks," J. Appl. Econometrics, vol. 10, 1995, pp. 347–364.

[4] Ciobanu Dumitru, Vasilescu Maria, Advantages and Disadvantages of Using Neural Networks for Predictions (https://ideas.repec.org/a/ovi/oviste/vxiiy2012i1p444449.html)

[5] Honlun Yip, Hongqin Fan, Yathung Chiang (2014): "Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models", Automation in Construction, Vol.38, Pages. 30–38.

[6] Kanthala Sampath Kumar, V. Swapna, S.A. JyothiRani , V.V. Hara Gopal, (2018). Volatility of daily and Weekly Crude Oil Prices – A Statistical Perspective. International Journal of Scientific Research in Mathematical and Statistical Sciences, 5(5), 108-113.

[7] Makridakis, S., S.C. Wheelwright, and R.J. Hyndman (1998): "Forecasting: Methods and Applications", New York: John Wiley & Sons.

[8] Raghavender, M., & Guguloth, R. (2015). Forecasting of maize production in Telangana.Rani, S. J., & Haragopal, V. V. Forecasting Exchange Rates using Neural Neworks.

[9] Ruslana Dalinina, Introduction to Forecasting with ARIMA in R https://blogs.oracle.com/datascience/introduction-to-forecasting-with-arima-in-r)

[10] V. Swapna, Kanthala Sampath Kumar, V.V. Hara Gopal, (2020). Statistical An alysis of the Costs of OPEC Crude Oil. International Journal of Scientific Research in Mathematical and Statistical Sciences, 7(4), 1-9.