Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 7
Open Access

FORECASTING WORLD CRUDE OIL PRICE: USING ARIMA MODEL

Seema Mahendra Attarde and Dr. Swati Desai
[1]Research Scholar, Shri JJTU University, Rajasthan
[2]Department of Statistics, Shri JJTU University, Rajasthan
[1]attardeseema13@gmail.com

Cite this paper as: Seema Mahendra Attarde and Dr. Swati Desai (2024) Forecasting World Crude Oil Price: Using Arima Model. *Frontiers in Health Informatics, 13* (7), 782-789

ABSTRACT
A serious concern has been risen much interest into the investigation of its price swing as well forecasting. The study used the dynamics of monthly Brent oil price from Nov 1994 to Dec 2011. The data grouped into two parts. The first twelve years used for the model construction and the next twelve months (Jan 2012 –Dec 2012) used for validating forecasting accuracy. They were subject to log transformation as well differencing to make stationary, besides testing of autocorrelation and residual analysis to determine among family of ARIMA models. ARIMA (2,0,1) was fitting well for forecasting the volatilities price. The study suggested choosing alternative models such as ARCH and GARCH models to have a best accuracy of forecasting oil prices due to prevailing outliers.

Keywords*: Brent, ARIMA, Stationarity, Residual analysis, forecasting*

INTRODUCTION

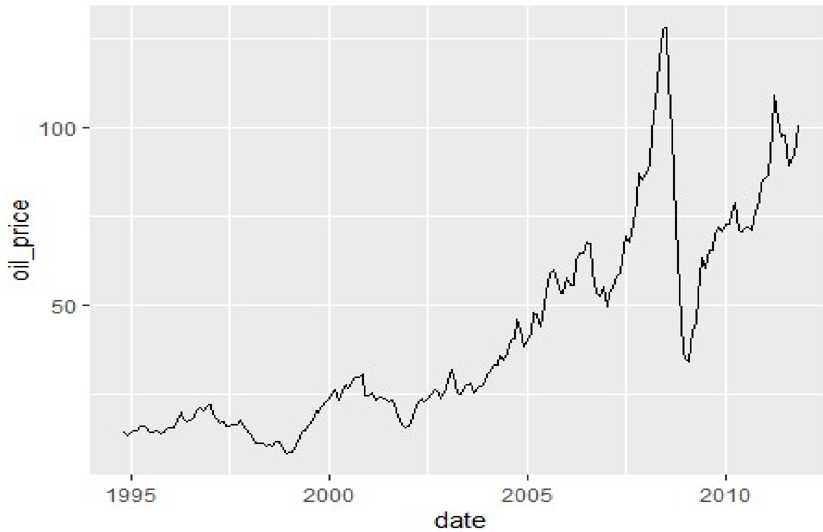
Crude oil is an important commodity in the world trade (**Selvi, Kaviya Shree, and Krishnan, 2018**). A larger part of day-to-day life depends on oil for industrial and households' usage. Apart from that, it influences greatly on the manufacturing and other industries come from the oil producing sector. It is estimated that products derived from crude oil provides about 33% of the energy needs globally **(Energy Information Administration, 2013**) and have implication on the economy to a larger extent (**Fondo et al., 2021**).

At the user lever, the price swings will have direct effect on food supply, detergents, prescription drugs, and household appliances at rural and urban segments. Therefore, we may see both negatively and positively affect on economic variables (**Mensah, 2015**). Added to that, price volatility will have significant impact on import and export front, besides in the financial market. For instance, an increase in price induces higher cost of production and changes capacity utilization of firms. Such issues are usually passed on to consumers through soaring prices of consumer goods.

The global demand for crude oil (including befouls) in 2022 expected to have 99.57 million and expected to increase to 101.89 million in 2023. There was a slowdown in 2020, due to impact of the corona virus pandemic, and shutdowns of economy around the world. However, the forecast for oil demand by OPEC will be 109.8 million by 2045, including gasoline and diesel (**www.statista.com**).

Globally, over the past two to three decades, oil price has been experiencing volatile. The recent decline has left many industry players much concern on the future price. Though research studies on price fluctuations unabated, we explore the ARIMA models forecasting. In particular, with prone to volatilities in the financial market, we try parsimonious ARIMA model that has the best forecasting ability amidst the volatilities price. The spot price (North Sea-Europe) is used in the study. It is also classified as light crude oil together with benchmark for world oil pricing and trading.

Figure 1: Monthly Crude Oil Price ($) (1994-2011)



The observation shows (Figure 1) a record highest price of $128.08 recorded in August 2008 mainly economic crises in that period and the lowest price of $8.03 in December 1998. The main cause of lowest price was due to successive decisions by OPEC and a few non-OPEC exporters announcing production cuts were met with disbelief. Furthermore, there was a lack of trust between protagonists, Saudi Arabia and Venezuela, Iran and Gulf countries. The twelve years monthly data (November 1994-December 2011) used for the model construction and 12 months (January 2012– December 2012) used for validating forecasting accuracy of the model. Data on monthly crude oil price (figure-1) shows a time series plot of a non-stationary series. Clearly there is seemingly increasing trend coupled with fluctuations (1994 -2011).

2. REVIEW OF LITERATURE

According to Laban Gasper and Haika Mbwambobe (2023), based on the performances of various price models, ARIMA (0, 1, 1) was the best model of capturing the underlying volatility, despite the corona virus and the Ukraine war having a considerable impact. A study by Rodhan and Jaaz, (2022), utilized for predictions of WTI prices (January 1990-March 2021) with ARIMA approach. The ARIMA (1,1,4) model was shown to be the most accurate forecasting model.

Shah and Kiruthiga (2020) employed ARIMA to find out the nonlinear feature of prices and found ARIMA (0,1,4) was the most appropriate to make predictions. Another study by Selvi et al., (2018) also used the ARIMA. Their projections for the years 2017 through 2021, suggested that prices should be stabilized and monitoring as a steady increase in oil costs could be a significant problem in the future.

A study by Bichanga, (2018) attempted different models, including ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,1,2). He reported that ARIMA (1,1,0) had the lowest AIC values which was found to be more accurate at predicting petroleum prices in Kenya. Mensah (2015) reported ARIMA (1,1,1) model found to be the most effective forecasting model using the MSE and MAE technique.

GARCH and ARCH models were used by Suleiman, et.al (2015) on Nigerian oil prices during 1999-2013 showed that the best models for predicting data series were ARIMA (3, 1, 1) and GARCH (2, 1) and their projection for another 6 months period indicated a sharp increase as compared to historical averages. An empirical study by Fondo et al., (2021) based on ARIMA VAR models on Kenyan prices confirmed that VAR model performed better than other models. Ahmed and Shabri, (2014) made forecasting using Support Vector Machines (SVM) in comparison to the performance to ARIMA and GARCH models. The findings showed that SVM method outperforms other two methods. Shah and Kiruthiga, (2020) also had similar conclusions on SVM method.

3. DATA AND METHODOLOGY

The study used secondary data on monthly price (1994- 2012). The data are obtained from the U.S Energy Information Department and R-Studio software used for analysis.

3.1 Stationarity

The price may exhibit non-stationarity at their levels. For the estimation of its model, it becomes imperative to detrend the data before go for further analysis. A stationary series can be said to be having a constant variance and, absence of autocorrelation over time and no periodic fluctuations (**Brockwell and Davis, 2002**). The plot shows a non-constant mean and variance. A technique for making constant mean and variance of series stationary is the differencing.

$y_t = \triangle \log(x_t) = \log(x_t) - \log(x_{t-1}) = \log(p_t/p_{t-1})$

where $x_t = p_t$ represent the price and y_t is it's differenced series

Figure 2: Differenced log monthly price (1994-2011)

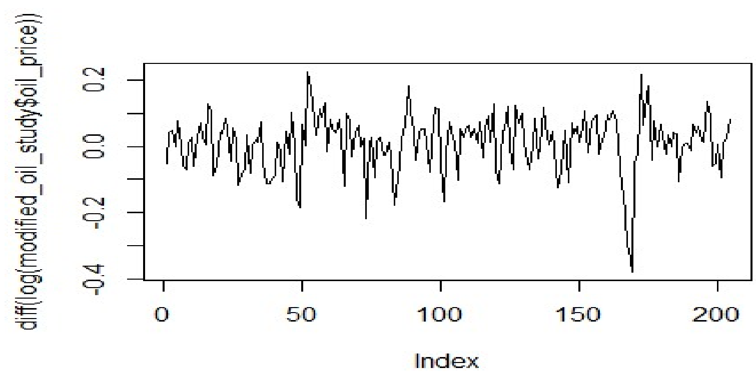


Figure.2 exhibits that the series with log differencing is a realization of stationary process. After making differencing of log price from 1994 to 2011, we could get constant mean and constant variance, which is desirable for forecasting except few lags.

3.2 ARIMA Model

ARIMA (p,d,q)-Box Jenkins Model **(1976)** is a common method for forecasting. Once data is stationary, we begin to explore the different ways we can have a fitting model. Autoregressive (AR) is written as:

$$y_t=\delta+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}+\epsilon_t \text{ ---(1)}$$

In Moving Average (MA), the variable of interest (price of crude) is modeled via its own imperfectly predicted values of current and previous times written in terms of error terms:

$$y_t=\mu+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\cdots+\theta_q\epsilon_{t-q}+\epsilon_t \text{ ---- (2)}$$

The differenced ARIMA (p,d,q) model became ARMA(p,q) process. The ARMA (p,q) process has the following mathematical form:

$$y_t=\delta+\{\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}\}+\{\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\cdots+\theta_q\epsilon_{t-q}\}+\epsilon_t \text{ ….. (3)}$$

Where p and q refer to the order of the autoregressive terms y_t and moving average terms ϵ_t respectively and ϕ, and θ are their respective coefficients

3.2.1 Model Specification

A prerequisite for the Box-Jenkins approach is data should have I(0) after first difference. ARIMA model has the following steps. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the two commonly used techniques for choosing the correct model. We start the model identification by plotting the ACF and PACF against different lags to determine the model.

3.2.2 Model Selection

The study employed AIC and BIC criterion. AIC is a penalized-likelihood criterion to measure the distance between the fitted likelihood function and the real likelihood of data. If both AIC and BIC is lower, it will be more accurate (**Suleiman et al., 2015**).

3.2.3 Parameter Estimation

For small to moderate sample sizes, it may be advantageous to ML estimation using all data. Also, many large-sample results are known about the sampling distribution of ML estimators.

3.2.4 Diagnostic Check

This technique is the Box Jenkins-Methodology to determine residuals white noise. To do this, the serial correlation is observed if it exists in the residuals. Additionally, the residuals are assumed to follow a normal distribution with a mean of zero and a constant variance (**Nyongesa and Wagala, 2016**).

4. RESULTS AND DISCUSSION

4.1 Model Selection ACF and PACF

By inspection approach for detecting stationarity, figures 1 and 2 also provides extremely helpful information suggesting not stationary. To find out the proper order (p, q) for our model, we begin by visualizing various lags (Figure 4.1)

Figure 4.1: ACF and PACF log differenced crude oil price

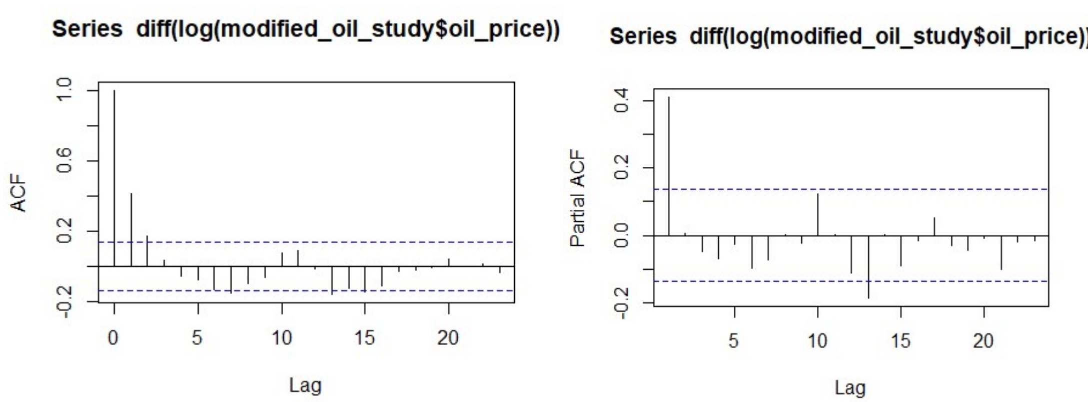


Table 1: BIC and AIC information criteria

Order	(0,0,1)	(0,0,2	(1,0,0)	(1,0,1)	(1,0,2)	(2,0,0)	(2,0,1)	(2,0,2)
AIC	-447.28	-450.51	-453.04	-451.05	-449.34	-451.05	**-457.97**	-456.44

	17	13	94	04	08	06	**32**	19
BIC	-437.3274	-437.2388	-443.0950	-437.7779	-432.7502	-437.7781	**-443.1826**	-436.5332

We estimated eight ARIMA models (Table 1) of which model ARIMA (2,0,1) was found a better model based on minimum values of AIC (-457.9732) and BIC (-443.1826).

4.2 Model Estimation

The model representation is given below:

$y_t = 0.0105 + 1.3785y_{t-1} - 0.4402y_{t-2} - 1.001u_{t-1} + \epsilon_t$

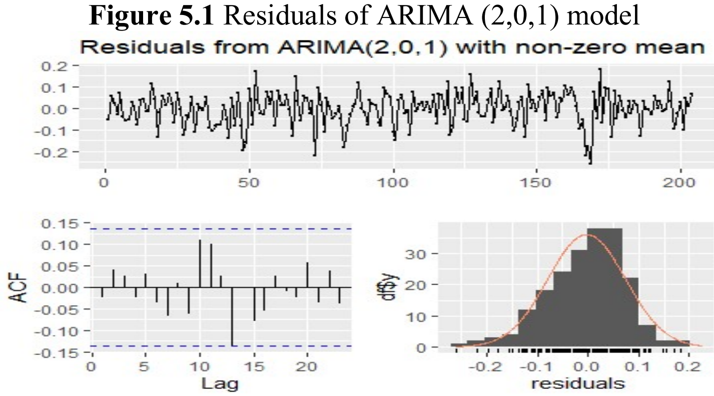
Table 2 MLE for the ARIMA (2,0,1) of monthly log return

Model	**Coefficient**	**Standard error**	**z-statistics**	**p-value**
AR1	1.3785	0.0626	22.0207**	0.000
AR2	-0.4402	0.0629	6.9984**	0.000
MA1	-1.001	0.015	66.6667**	0.000

4.3 Residual Analysis

If the Box-Jenkins model selected is good enough data, the residual realization would be white noise. That is residual must be independent following its normal distribution. It is analyzed graphically the residual plot, correlogram and the normal plot. We also performed the Ljung-Box test autocorrelation.

Figure 5.1 Residuals of ARIMA (2,0,1) model



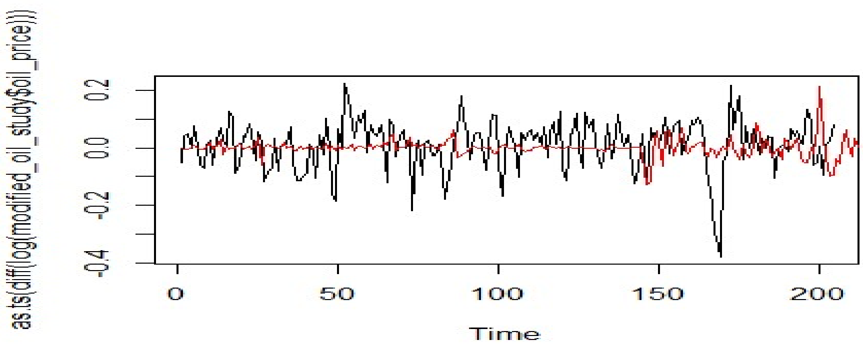
Ljung-Box Test

Residuals from ARIMA (2,0,1) Q = 5.6378, df = 7, p-value = 0.5826

From the figure 5.1, the residuals are stationary. The plot is closely normally distributed. Also, the Ljung-Box test suggests p-value which is 0.5026, validating the normality of the residuals.

5. FORECASTING

Figure 5.2: Fitting the actual versus forecasted diff log oil price

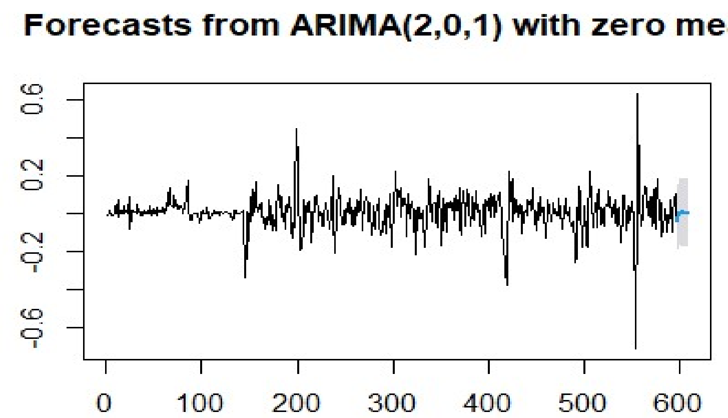


From Figure 5.2, the red line indicates forecasted line while the black line shows actual values. It appears that ARIMA (2, 0, 1) model is closely fitting data. The twelve months forecasted values (Jan 2012 to Dec 2012) had little variation than actual values for the said period. A point forecast with 95 % confidence level for lower and upper intervals was used.

Table 3: Twelve months ahead crude oil forecast (Jan 2012-Dec 2012)

Month	Point forecast	95% Low confidence level	95% High confidence level
January 2012	103.83459	96.80969	110.8595
February	105.87949	93.52428	118.2347
March	106.26909	88.99706	123.5411
April	105.57791	84.23713	126.9187
May	104.21255	79.71525	128.7099
June	102.54418	75.74370	129.3447
July	100.85750	72.46814	129.2469
August	100.85750	72.46814	129.2469
September	99.35545	69.92558	128.7853
October	98.16312	68.08150	128.2447
November	97.33933	66.86137	127.8173
December	96.89073	66.17232	127.6091

Figure 5.3: Forecasts from ARIMA (2,0,1) with zero mean



CONCLUSION

A popular tool is the ARIMA model which can predict well in some series, its forecasting performance can be woefully affected due to the presence of outliers, measurement errors and volatilities. The present study is used the Box- Jenkins methodology to examine a best model and its forecast ability. Considering AIC and BIC values, ARIMA (2, 0, 1) model has the better forecasting model among a group of ARIMA models considered. However, a review of volatility studies supports that the ARIMA model may not so accurate well due to high volatilities. The outliers were higher number in the monthly oil price data. Hence, the ARCH model by Engle (1982) and its variants (GARCH, EGARCH etc.) may provide best accuracy for such volatile models. The study recommends that getting better forecasting results, daily oil price data can be used in place of monthly data.

REFERENCES

Box, G.E.P. and G. Jenkins, (1976). Time Series Analysis: Forecasting and Control. 1st Edn., Holden-Day, San Francisco, ISBN-10: 0816211043, pp: 575.

BP Statistical Review of World Energy (2022), U.S. Energy Information Administration, International Energy Statistics, USA

Brockwell, P. J. & Davis, R. A. (2006), Introduction to time series and forecasting, Springer Science & Business Media

Demand for crude oil worldwide: https://www.statista.com/statistics/271823/global-crude-oil-demand/

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation, Econometrica, vol.50, pp.987-1007

FACTS Global Energy, Asia Pacific Databook 1 Demand, Spring 2022

Fondo, K. S., Onago, A. A., Kiti, L. A., & Otulo, C. W. (2021). Modeling of Petroleum Prices in Kenya Using Autoregressive Integrated Moving Average and Vector Autoregressive Models. IOSR Journal of Mathematics, 17(6), 18–27. https://doi.org/10.9790/5728-1706011827

Majumdar, Rumki (2022), "India Economic Outlook" Deloitte Insights, Deloitte, 12 July

Mensah, E. K. (2015). Box-Jenkins Modeling and forecasting of Brent crude oil price. MPRA, Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Italy.

Nyongesa, D. N., & Wagala, A. (2016). Non Linear Time Series Modelling Of the Diesel Prices in Kenya. International Journal of Academic Research in Economics and Management Sciences, 5(4). https://doi.org/10.6007/ijarems/v5-i4/2368

Selvi, J. J., Kaviya Shree, R., and Krishnan, J. (2018). Forecasting Crude Oil Price Using ARIMA Models. International Journal of Advanced Research in Science and Engineering, 7(5), 334–343.

Suleiman, S., Usman, U., & Adamu, U. (2015). Modelling and Forecasting the Crude Oil Price in Nigeria. International Journal of Novel Research in Marketing Management and Economics, 2(1), 1–13.

U.S. Energy Information Administration, Short-Term Energy Outlook, August 2022.