

Forecasting Nigeria's Crude Oil Production Using Box and Jenkins Methodology

Modupe Stella Omotayo-Tomo & Nwokeocha Divine Chinelo

Department Of Mathematical Sciences,
Olusegun Agagu University of Science and Technology,
Okitipupa Ondo State.

Email: modupe.omotayo@oaustech.edu.ng

<https://doi.org/10.56201/ijasmt.vol.12.no9.2026pg77.81>

Abstract

This research analyzed and predicted the crude oil output of Nigeria utilizing the Autoregressive Integrated Moving Average (ARIMA) methodology, rooted in the Box–Jenkins framework. The aim was to pinpoint a suitable time-series model for capturing production trends and producing dependable short-term predictions. To test for stationarity, the Augmented Dickey–Fuller (ADF) unit root test was applied, while residual diagnostic assessments were conducted to check the suitability of the chosen model. Findings revealed that crude oil output was non-stationary at the levels, with an ADF statistic of -1.743 and a p -value of 0.6838 . Nevertheless, upon first differencing, the series turned stationary, achieving an ADF statistic of -7.8812 , suggesting integration of order one, $I(1)$. The multiplicative $ARIMA(4,1,4)_{12}$ model emerged as the most fitting model through the model selection criteria. Residual diagnostics demonstrated no considerable deviation from normal distribution at the 1% significance threshold (Shapiro–Wilk statistic = 0.98927 , $p = 0.1184$) and no notable residual autocorrelation (Ljung–Box $p = 0.8562$). The model produced a root mean square error (RMSE) of 0.1058 . Predictions spanning from July 2023 to June 2027 estimated monthly crude oil production to fall within the vicinity of 1.1 to 1.2 million barrels, based on the findings. The analysis advocates for the utilization of ARIMA modeling as an effective approach for forecasting and planning crude oil output in Nigeria.

Keywords: Forecasting, Crude, Oil, Production, Box, Jenkins & Methodology

INTRODUCTION

In Africa, Nigeria was the leading oil producer in 2019. The oil production amounted to 101.4 million metric tons (mmt) in the country. Angola and Algeria followed with 69.1 and 64.3 mmt respectively. Nigeria was not only the largest oil producing country in the continent; it was also the largest crude oil exporter in the continent. The oil producing states in Nigeria include Akwa Ibom, Rivers, Delta, Ondo, Lagos, Bayelsa, Edo, Imo and Abia with each of the states producing 504 Thousand barrels per day (BPD), 346 Thousand BPD, 344 Thousand BPD, 290 Thousand BPD, 60 Thousand BPD, 40 Thousand BPD, 33 Thousand BPD, 17 Thousand BPD and 11 Thousand BPD respectively, with Akwa Ibom State being the largest oil producer and Abia State being the least oil producer (National Bureau of Statistics, 2017).

In the past twenty years, the oil industry made major contributions to the economy of Nigeria (Alemho, Deebom, & Lekara-Bayo, 2026). The oil industry created more employment opportunities, contributed to the government revenues at every level, increased the gross domestic

product, contributed to the local expenditures, foreign exchange reserves and the supply of energy. The production and exportation of petroleum play a dominant role in Nigeria's economy (Deebom, Lekara & DaWariboko, 2022). The quantity to either be refined and sold or be exported depends on the quantity of crude oil produced. The pre knowledge of the quantity of crude oil that will be produced can help the country have a viable economy. A drastic fall in the quantity of crude oil produced is fatal to the Nigerian economy since Nigeria is a mono-product economy country and depends heavily on the oil sector. Many researchers (Omekara et al., 2015, Fatoki et al., 2017, Eriga, 2013) have fitted different classical statistical models to the crude oil production in Nigeria but the application of machine learning models on crude oil production has been grossly understudied. The population is becoming increasingly reliant on crude oil and its products. Crude oil has become one of the most traded commodities on a global scale because of this rising demand.

Literature Review

This literature review's goal is to provide a thorough overview of the body of knowledge already in existence while navigating the complex world of crude oil production and forecasting. The literature review section provides an overview of the existing research and studies related to the analysis of crude oil production using Box-Jenkins methodology. It highlights the key findings and methodologies, setting the foundation for the current project (Omotayo, Ajiboye, & Adeoti, 2024).

Omekara et al. (2015) proposed a multiplicative SARIMA (1, 1, 1)(0, 1, 1)₁₂ model as the high performance and best model to be fitted to the crude oil production data in Nigeria from 2005-2012. The forecasted values from the fitted model were close to the actual values. Fatoki et al. (2017) fitted the ARIMA model to the crude oil production in Nigeria. The order (1, 2, 2) of the ARIMA model adequately fitted the data. The forecast from the model suggested that there will be continuous increase in crude oil production for 2014–2023 but the actual crude oil production experienced a drastic fall after 2015, and Nigeria recorded her lowest crude oil million barrels per day (1.5 mmbd) production between 2006 to 2020 in August 2016. This was a 46.7% drop in the crude oil production when compared to the quantity produced two years from then (2.2 mmbd). The crude oil production increased after August 2016 though was barely above 2 mmbd. It started another decline in the second half of 2020. This is contrary to the forecast for the ARIMA (1, 2, 2) model.

Materials and Methods

Classical time series model in form of Autoregressive Integrated Moving Average (ARIMA) was developed by Box and Jenkins (1976) often refers to as Box-Jenkins methodology; it is considered as one of the best forecasting models. This model has recorded significantly in financial and other academic research during the last three decades. The method is suitable for short-term forecasting a non-stationary series and seasonal time series data with large sample observations. The general form of ARIMA model (p, d, q) is specified as:

$$y_t = \mu + \sum_{i=1}^p \phi_i (y_{t-i} - \mu) + E_t + \sum_{i=1}^q \theta_i e_{t-i}$$

Where p is the order of autoregressive represents the number of lagged values for y (non-stationary) to turn to stationary series.

q is the order of the moving average, represents the number of lagged values of the error term and e_t stands as the random error, and θ the number of times μ is differenced to produce a stationary series.

RESULTS

Table 4.1 Unit Root Test of Stationarity; H_0 : The Series has a Unit Root

Variable	ADF @ Levels	P-value	ADF @ First Difference	P-value	Remarks
Crude Oil Production	-1.743 [5]	0.6838	-7.8812[5]**	0.01	I(1)

ADF Critical Value at 5% = -2.95;

[8] Indicates that a maximum lag length of 5 was included in the tests.

** indicates significant at 1%

These results provide evidence that the crude oil production data includes a unit root. As a result, we can draw the conclusion that the series contains a unit root (Omotayo, Ajiboye, & Adeoti, 2024). Differentiation is the sole method that can be used to calibrate the two series for the purpose of model building development. This is because ARIMA modelling cannot achieve white noise with non-stationary data. Stationarity was reached for the first order differencing of this series, and it was determined that order I(1) was appropriate. **Diagnostic Check on the fitted ARIMA (4.1.4)₁₂ model**

Table 4.2: Shapiro-Wilk Test of Residual Normality of Nigeria Crude Oil Production

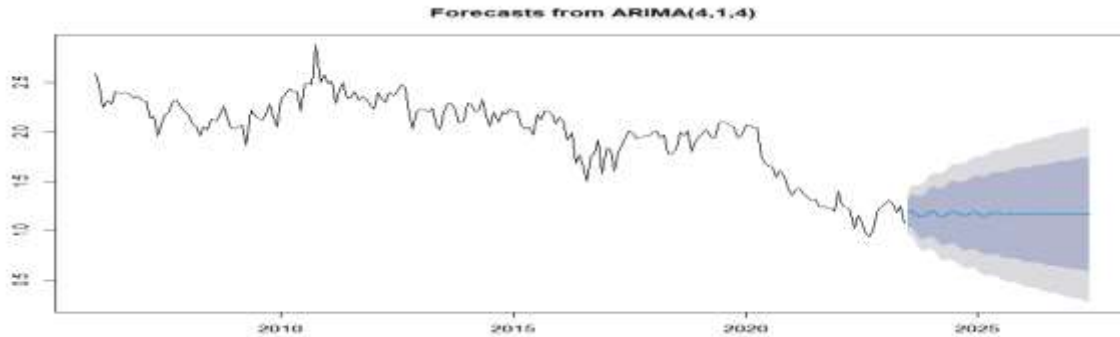
Shapiro-Wilk	p-value
0.98927	0.1184

The Shapiro-Wilk test of normality in Table 4.2 yields a test statistic of 0.9892, with a corresponding p-value of $0.1184 > 0.01$, indicating that the normality of residuals of the best-fitting ARIMA(4, 1, 4) is not rejected at 1% significance levels. The residuals are therefore normally, independently, and identically distributed (NIID).

Table 4.3 Ljung-Box Portmanteau Test of No serial Correlations

Chi-Squared	Degree of Freedom	P-value
0.032818	1	0.8562

The Chi-square test statistic value and large p-values in the Box test of the residuals of the fitted ARIMA model suggested that the null hypothesis that all autocorrelation functions are zero could not be rejected. In other words, we can conclude that there is no (or almost no) evidence for non-zero autocorrelations in the fitted model's residuals.



Accuracy: $RMSE = 0.1058$

Based on the objectives of this study, the ARIMA(4,1,4)12 MODEL provided the most accurate forecasts. Consequently, the shaded zones of forecast represent the 80% and 95% lower and upper side projections of prediction intervals, as the forecast lines indicated that crude oil production between July 2023 and June 2027 will be between 1,100,000 and 1,200,000 barrels.

DISCUSSION

The correlation research and the partial correlation study both suggested that there were very few significant spikes at the appropriate lags. Considering this, it is possible to demonstrate that the series is stationary when compared to the initial difference. The results of the Dickey-Fuller test and the related p-values at a level of significance greater than 0.05. These results provide evidence that the crude oil production data includes a unit root. As a result, we can draw the conclusion that the series contains a unit root. Differentiation is the sole method that can be used to calibrate the two series for the purpose of model building development. This is because ARIMA modelling cannot achieve white noise with non-stationary data. Stationarity was reached for the first order differencing of this series, and it was determined that order I(1) was appropriate. Analysis suggests that ARIMA order (4,1,4) is best adapted for the crude oil production in Nigeria. The ARIMA orders with the parameters examined for significance in table 4.4 below were selected using the rule of parsimony, which states that the simplest model should be chosen provided all necessary conditions are met.

CONCLUSION

The study reveals that the most suitable model is the multiplicative ARIMA (4, 1, 4)12. This study showed a minor variation in comparison to the research conducted by Ette et al. (2012) and Omekara et al. (2015). The observed variance may be ascribed to the inclusion of an extended time in the dataset. The predictions for the 48-month period from July 2023 to June 2027, as determined by the fitted model, exhibited a statistically significant agreement with the observed values, with a confidence level of 95%. Hence, it is advisable to utilize the model. The Box-jenkins methodology is hereby recommended as the best model for estimating the production of the Nigerian crude oil.

REFERENCES

- Alemho, J. E., Deebom, Z. D., & Lekara-Bayo, I. B. (2026). Markov switching VAR-artificial neural network hybrid model for studying monetary aggregates in Nigeria. *Benin Journal of Statistics*, 9, 44–54.
- Badejo, O. M., Omotayo-Tomo, M. S., Francis, S. D., & Awolere, I. T. (2026). Copula modeling of inflation–exchange rate dependency amid democratic consolidation and economic struggles in Nigeria (2003–2024). *Coast Journal of the School of Science*, 8(1), 1612–1622.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Rev. ed.). Holden-Day.
- Deebom, Zorle Dum, LekaraBayo Ifeoma Better & DaWariboko, Asiki Yvonne (2022). The Nexus between Real Sector Diversification and Sustainable Economic Growth in Nigeria ARDLECM approach. *Studies in Economics & International Finance*, Vol. 2, No. 2, pp. 205–223. <https://DOI: 10.47509/SEIF.2022.v02i02.05>
- Etuk, E. H., & Amadi, E. H. (2013). Multiplicative SARIMA modelling of Nigerian monthly crude oil domestic production. *Journal of Applied Mathematics & Bioinformatics*, 3(3), 103–112.
- Fatoki, O., Ilo, H. O., & Ugwoke, P. O. (2017). Time series analysis of crude oil production in Nigeria. *Advances in Multidisciplinary & Scientific Research Journal*, 3(3), 1–10.
- National Bureau of Statistics. (2017). *Selected petroleum statistics: Oil and gas production, drilling and development* (2012–2016).
- Omekara, C. O., Okereke, O. E., Ire, K. I., & Okamgba, C. O. (2015). ARIMA modeling of Nigeria crude oil production. *Journal of Energy Technologies and Policy*, 5(10).
- Omotayo, M. S., Ajiboye, A. S., & Adeoti, O. A. (2024). Dependence structure and portfolio value-at-risk for some selected Nigerian stocks using a copula-based volatility model. *International Journal of Computer Science and Mathematical Theory*, 10(1), 102–?
- Omotayo, M. S., Ajiboye, A. S., & Adeoti, O. A. (2024). Determinants of the optimal and symmetry of the volatility model for some selected Nigerian stocks. *International Journal of Applied Science and Mathematical Theory*, 10(1), 29–44.