ISA Journal of Engineering and Technology (ISAJET)

Homepage: https://isapublisher.com/isajet/

Volume 3, Issue 6, Nov-Dec, 2026 ISSN: 3049-1843

Information-Criterion-Based Autoregressive Model Selection and Forecasting of Annual Crude Oil Production in Nigeria

Nnadede Cordelia C.[1]; Onyemachi Chris U.[2] & Nnah Bright A.[2]

[1]Abia State Bureau of Statistics, Umuahia.

[2]Department of Statistics, Abia State University, Uturu.

Received: *10.09.2026* **|** ***Accepted:*** *13.11.2026* **|** ***Published:*** *16.11.2026*

***Corresponding Author:** Nnadede Cordelia C.

DOI:*10.5281/zenodo.24279132*

Abstract **Original Research Article**

This study examined autoregressive time series modelling for forecasting annual crude oil production in Nigeria using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Annual data covering 1973–2025. The Augmented Dickey-Fuller test revealed that the original series was non-stationary ($p = 0.6964$) but became stationary after first differencing ($p = 0.01$). Among the candidate models, AR(1) recorded the minimum AIC (696.038) and BIC (699.941), leading to the selection of ARIMA(1,1,0). The Ljung-Box diagnostic test ($p = 0.5256$) indicated no significant residual autocorrelation, suggesting adequate model fit. Forecasts for 2026–2035 projected annual crude oil production at approximately 1,602.53 million barrels, with increasing prediction uncertainty over time. The study demonstrates the usefulness of information criteria in selecting parsimonious autoregressive models for crude oil production forecasting.

Keywords: crude oil production; Subset selection; autoregressive model; AIC; BIC; forecasting.

Copyright © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

1. Introduction

Time-series forecasting is important for planning because observations collected sequentially may exhibit temporal dependence. Autoregressive (AR) models describe a variable through its own past values and remain useful because they are interpretable and form part of the Box–Jenkins modelling framework [1,2]. Their performance depends on choosing an appropriate lag structure: excessive lags increase parameter uncertainty, whereas too few lags may leave serial dependence unexplained [3]. Information criteria formalize the trade-off between model fit and complexity. Akaike's information criterion (AIC) is motivated by expected information loss [4], while the Bayesian information criterion (BIC) imposes a stronger sample-size-dependent penalty and can be consistent under suitable conditions [5]. Since their penalties differ, they may select different models [6]. Subset autoregressive models allow selected non-consecutive lags (for example, 1, 4 and 8) to enter while other lags are excluded. A conventional AR(p), by contrast, includes every lag from 1 through p. Consequently, comparing ARIMA(1,1,0) through ARIMA(10,1,0) is contiguous-order selection, not an exhaustive subset-lag search [7]. This distinction is central to interpreting the present results.

Nigeria's crude oil production has experienced substantial fluctuations over several decades. A transparent statistical description and forecast can contribute to energy-sector planning, although a univariate model cannot account for operational, regulatory, security, investment or market drivers. This study therefore evaluates a reproducible information-criterion-based procedure for annual production over 1973–2025. Hence, the aim is to identify an adequate and parsimonious autoregressive model for forecasting annual crude oil production in Nigeria. The specific objectives are to: (i) examine the temporal pattern and linear trend; (ii) assess stationarity; (iii) compare candidate ARIMA(p,1,0) models using AIC and BIC; (iv) evaluate residual adequacy; and (v) generate ten-year forecasts.

2. Materials and Methods

2.1 Data and Study Design

The study used 53 annual observations of Nigerian crude oil production from 1973 to 2025, measured in million barrels. The design is retrospective and observational.

2.2 Exploratory and Trend Analysis

The annual series was inspected graphically and summarized using descriptive statistics. A linear trend model, $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$, was fitted by ordinary least squares, where t is a sequential time index. The slope coefficient and coefficient of determination were used to assess whether a simple deterministic linear trend adequately described the series.

2.3 Stationarity Assessment

Stationarity was assessed using the time plot, ACF, PACF and Augmented Dickey–Fuller (ADF) test. The ADF null hypothesis is that the series contains a unit root. At the 5% level, a p-value below 0.05 provides evidence against this null. The level series was first-differenced where necessary: $\Delta Y_t = Y_t - Y_{t-1}$. Differencing was applied to model changes in production.

2.4 Candidate Models and Information Criteria

Candidate ARIMA(p,1,0) models were estimated for p = 1,…,10. Each specification models the first-differenced series as an AR(p) process: $\Delta Y_t = c + \phi_1 \Delta Y_{t-1} + \ldots + \phi_p \Delta Y_{t-p} + \varepsilon_t$. AIC and BIC were calculated as $AIC = -2\log(L) + 2k$ and $BIC = -2\log(L) + k\log(n)$, where L is the maximized likelihood, k is the number of estimated parameters and n is the effective sample size. The model with the smallest criterion was preferred within the candidate set.

2.5 Estimation, Diagnostics and Forecasting

The selected model was estimated by maximum likelihood using R's `arima` function. Residual behaviour was assessed graphically and with the Ljung–Box test at lag 10. Forecasts and 80% and 95% prediction intervals were generated for ten annual periods using the `forecast` procedure.

3. Results

3.1 Descriptive Behaviour and Trend

Production ranged from 1,241 to 2,627 million barrels. The mean (1,948 million barrels) was slightly below the median (1,993 million barrels), indicating that the distribution is not perfectly symmetric. The time plot (Figure 1) shows several distinct phases: relatively high production in the 1970s, a pronounced contraction in the early 1980s, recovery through the late 1980s and 1990s, comparatively high production during the 2000s and early 2010s, and a marked decline after 2015 followed by a modest recovery by 2025. These episodes show why a single linear trend is an incomplete description of the historical path.

Table1: Descriptive Statistics of Annual Crude Oil Production in Nigeria (1973-2025)

Statistic	Value
Number of observations	53
Period	1973–2025
Minimum	1,241
First quartile	1,716
Median	1,993
Mean	1,948
Third quartile	2,208
Maximum	2,627

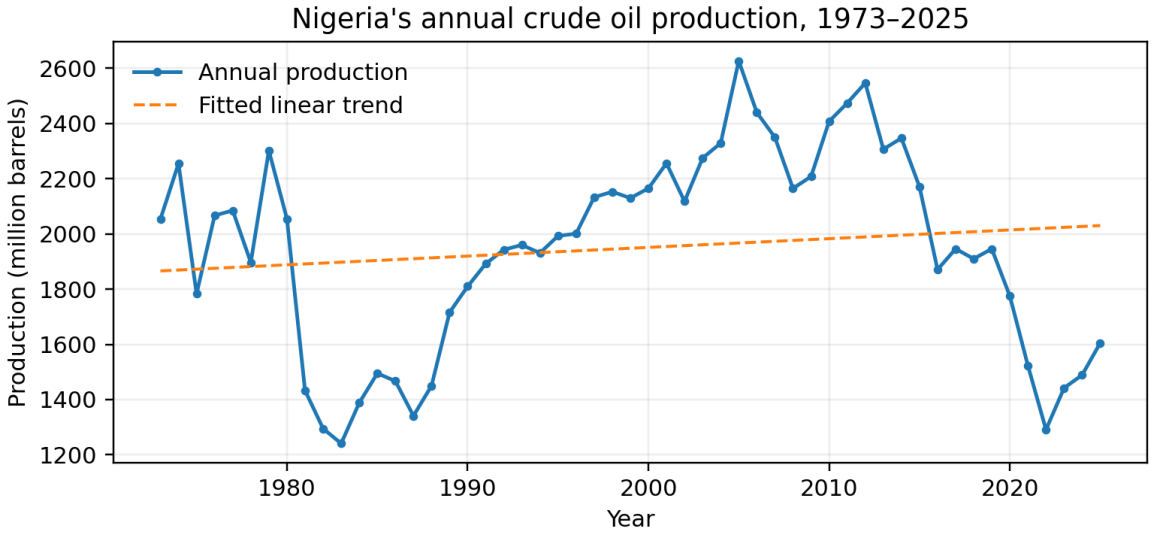

Figure 1: Annual crude oil production in Nigeria with the fitted linear trend (1973–2025).

The estimated trend was $\hat{Y}_t = 1{,}862.947 + 3.158t$. The slope was positive but not statistically significant ($t = 0.967$, $p = 0.338$), and the model explained only 1.8% of the observed variation ($R^2 = 0.01799$; adjusted $R^2 = -0.00127$). Thus, the data do not provide evidence of a statistically significant linear trend across the full period. This does not mean that production was stable: the visible falls and recoveries may reflect nonlinear movements or structural changes that a single straight-line regression cannot represent.

3.2 Stationarity and Autocorrelation

Table 2: Stationarity and Autocorrelation

Series	ADF statistic	Lag order	p-value	Decision at 5%
Level production	−1.6971	3	0.6964	Do not reject unit-root null
First difference	−4.5075	3	0.0100*	Reject unit-root null

The R `tseries:adf.test` output gives p = 0.01 as a lower-bound; reported as p ≤ 0.01 treated as probability. The level series failed to reject the unit-root null (ADF = −1.6971, p = 0.6964). The first-differenced series rejected the null (ADF = −4.5075, reported p = 0.01), supporting one difference before autoregressive modelling. Figure 2 displays the annual changes. Large positive and negative changes remain visible, particularly around the early 1980s, so stationarity of the mean should not be confused with an absence of volatility or unusual observations.

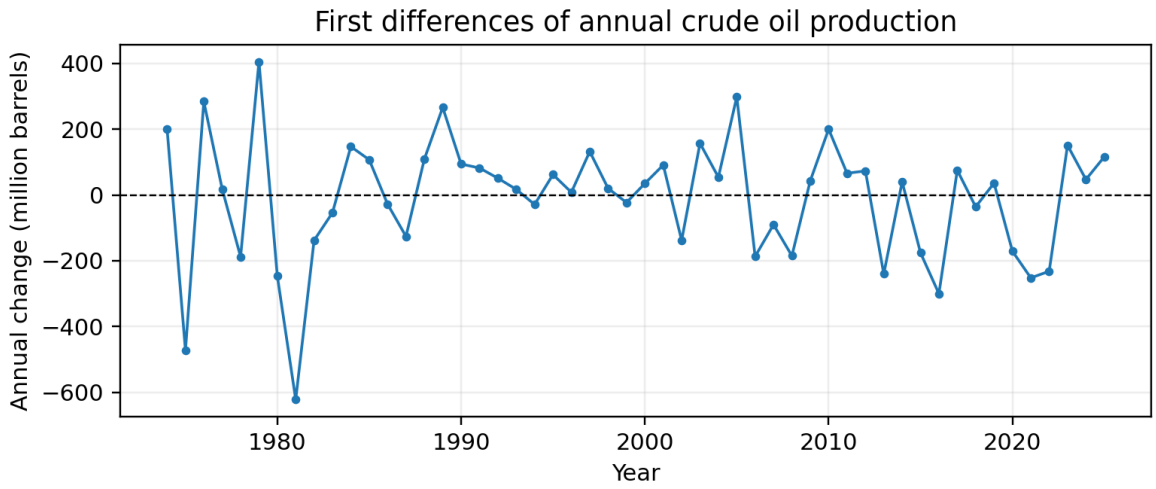

Figure 2: First differences of annual crude oil production. Positive values indicate year-on-year increases; negative values indicate decreases.

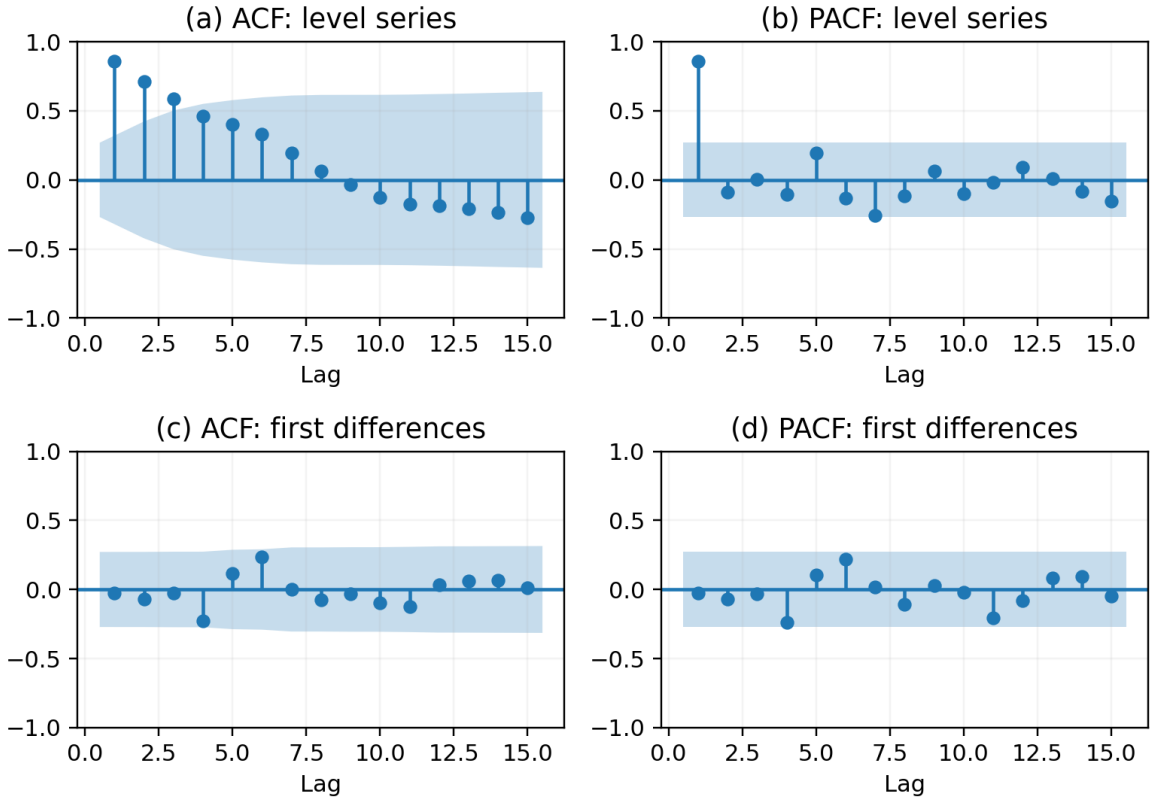

Figure 3: ACF and PACF of the level and first-differenced series. The level ACF decays slowly, while differencing substantially weakens the serial dependence.

In the original series, the ACF was reported to have a high positive lag-one autocorrelation (approximately 0.80) and a gradual decay, consistent with strong persistence. After differencing, most ACF and PACF coefficients were reported to be small and within or close to the 95% confidence bounds. The reconstructed plots provide a visual check using the supplied annual observations. ACF/PACF patterns are diagnostic aids.

3.3 Information-Criterion Model Selection

Table 3: Information-Criterion Model Selection

Candidate model	AIC	BIC
ARIMA(1,1,0)	696.0384	699.9409
ARIMA(2,1,0)	697.7561	703.6099
ARIMA(3,1,0)	699.7374	707.5424
ARIMA(4,1,0)	698.1424	707.8986
ARIMA(5,1,0)	698.7762	710.4836

ARIMA(6,1,0)	697.6724	711.3311
ARIMA(7,1,0)	699.4904	715.1003
ARIMA(8,1,0)	700.7181	718.2793
ARIMA(9,1,0)	702.4835	721.9959
ARIMA(10,1,0)	703.7767	725.2404

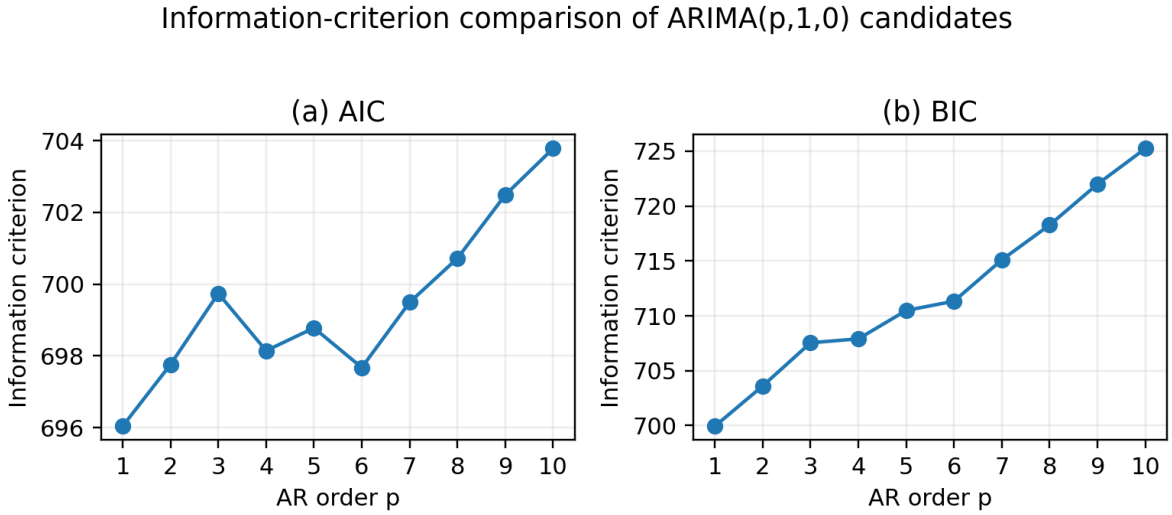

Figure 4: AIC and BIC across the ten contiguous ARIMA (p, 1, 0) candidate orders. Lower values indicate a preferred trade-off between fit and complexity.

ARIMA(1,1,0) achieved the minimum AIC (696.0384) and BIC (699.9409). Both criteria therefore selected the same model within the specified candidate set. The BIC increased more sharply as additional lags were introduced, reflecting its stronger penalty for model complexity. Importantly, the agreement between AIC and BIC strengthens the choice only relative to the ten models evaluated.

3.4 Parameter estimation and residual diagnostics

Table 4: Parameter Estimation

Parameter	Estimate	Standard error
AR(1)	−0.0217	0.1394
Innovation variance	35,261	—

The selected model can be written as $\Delta\hat{Y}_t = -0.0217\Delta Y_{t-1} + \varepsilon$. The estimated AR(1) coefficient of −0.0217 indicates a very weak negative relationship between the current annual change in crude oil production and the change recorded in the preceding year. This suggests that an increase in production change in the previous year is associated with a negligible tendency towards a decrease in the current year's production change, and vice versa. However, the magnitude of the coefficient is very small, while its standard error (0.1394) is considerably larger, suggesting that the estimated first-order autoregressive effect is not statistically strong. The innovation variance of 35,261 represents the estimated variability of the random shocks or unexplained fluctuations in the differenced production series. This indicates that changes in crude oil production are influenced by substantial variations not explained by the selected autoregressive structure.

Table 5: Residual Diagnostics

Diagnostic	Statistic	p-value	Interpretation
Ljung–Box, lag 10	Q = 9.0688	0.5256	No significant residual autocorrelation detected

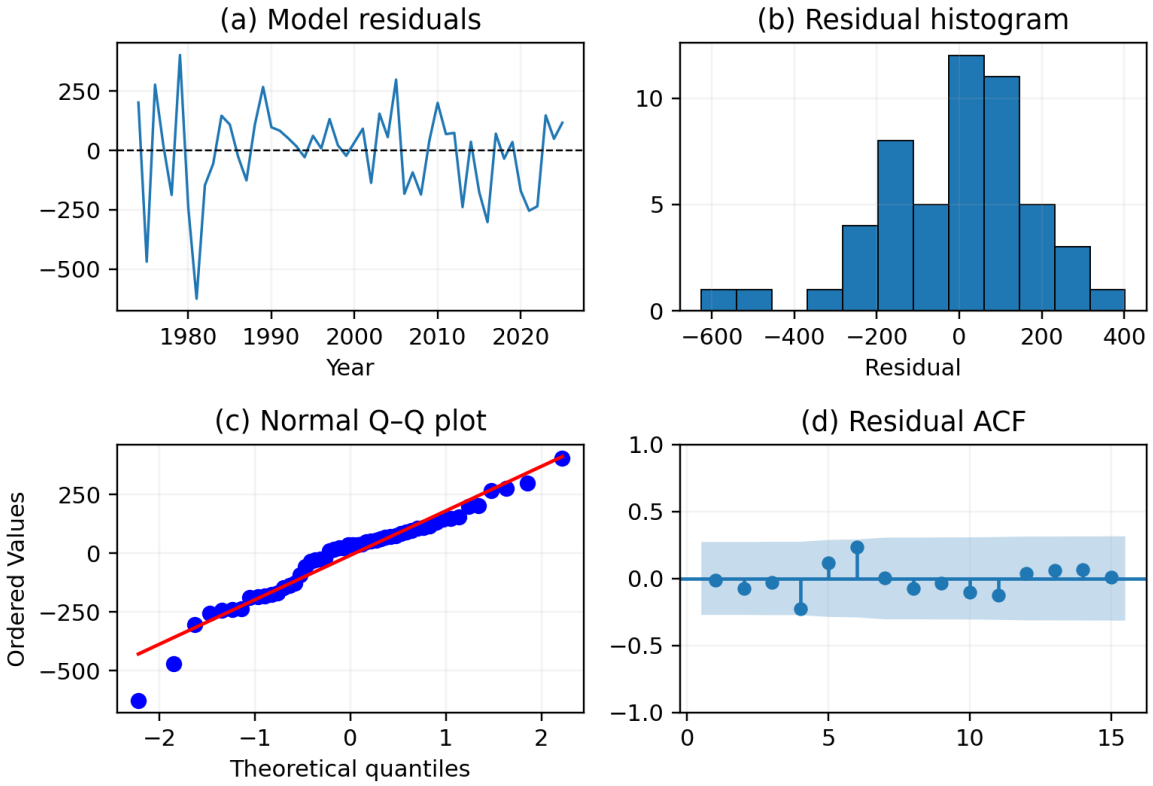

Figure 5: Residual diagnostics for an ARIMA (1, 1, 0) fit to the supplied series: residual time plot, histogram, normal Q–Q plot and residual ACF.

The diagnostic adequacy of the selected ARIMA(1,1,0) model was assessed using the Ljung–Box test at the 5% level of significance. The test produced a Q-statistic of 9.0688 and a p-value of 0.5256. Since the p-value is greater than the significance level of 0.05, the null hypothesis of no significant residual autocorrelation up to lag 10 is not rejected. This indicates that there is insufficient statistical evidence of remaining serial correlation in the model residuals. The result suggests that the selected model has adequately captured the major linear temporal dependence in the differenced crude oil production series. Furthermore, the residual autocorrelation function (ACF) presented in Figure 5 shows no prominent systematic pattern, providing additional descriptive support for the Ljung–Box test result. The residual histogram and Q–Q plot were also examined to assess the distributional behaviour of the residuals. The diagnostic findings provide no significant evidence of residual autocorrelation at the 5% level, supporting the adequacy of the ARIMA(1,1,0) model in terms of residual independence.

3.5 Forecasts for 2026–2035

Table 6: Forecasts for 2026–2035

Year	Forecast	95% lower	95% upper
2026	1602.48	1234.437	1970.522
2027	1602.534	1087.667	2117.402
2028	1602.533	974.165	2230.902
2029	1602.533	878.239	2326.828
2030	1602.533	793.609	2411.457
2031	1602.533	717.031	2488.035
2032	1602.533	646.568	2558.498
2033	1602.533	580.954	2624.113
2034	1602.533	519.306	2685.76
2035	1602.533	460.983	2744.083

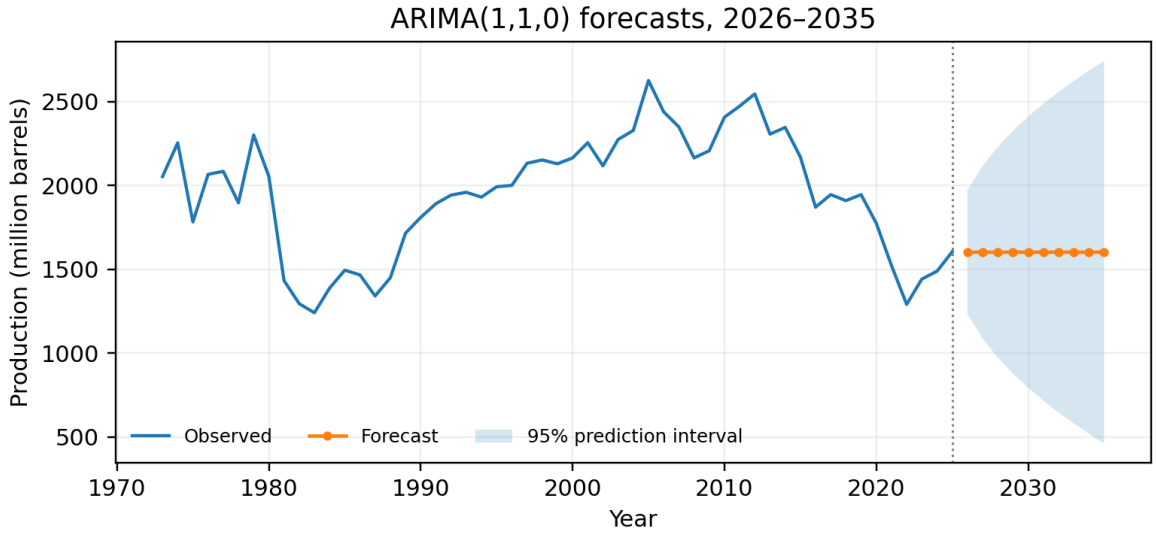

Figure 6: Ten-year forecasts from ARIMA (1, 1, 0), with 95% prediction intervals.

The point forecasts rise marginally from 1,602.480 million barrels in 2026 to approximately 1,602.534 million barrels in 2027, after which they remain effectively constant. The widening 95% prediction intervals are a more important feature than the near-flat point path: the 2026 interval is approximately 1,234–1,971 million barrels, whereas the 2035 interval is approximately 461–2,744 million barrels. The growing interval width reflects accumulated uncertainty as the forecast horizon increases. These are model-based conditional projections, not guarantees or policy-adjusted production targets.

4. Discussion

This study examined the application of autoregressive time series models in forecasting annual crude oil production in Nigeria using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as model selection tools. The dataset consisted of annual crude oil production observations covering the period 1973 to 2025. The study commenced with an exploratory analysis of the series through graphical examination and trend analysis. The time series plot revealed substantial fluctuations in crude oil production over the study period, while the fitted trend model indicated the absence of a statistically significant linear trend. The estimated trend coefficient was positive but insignificant ($p = 0.338$), suggesting that time alone could not adequately explain variations in crude oil production. Stationarity of the series was investigated using the Augmented Dickey-Fuller (ADF) test. The original series was found to be non-stationary, with an ADF test p-value of 0.6964. To achieve stationarity, first differencing was applied. The ADF test performed on the differenced series yielded a p-value less than 0.01, confirming that the transformed series was stationary and suitable for autoregressive modelling. The autocorrelation structure of the stationary series was examined using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Several candidate autoregressive models were subsequently estimated and compared using AIC and BIC.

The results showed that the AR (1) model produced the smallest AIC (696.038) and BIC (699.941) values among all competing models. Consequently, the ARIMA (1,1,0) model was selected as the most appropriate model for the crude oil production series. The estimated AR (1) coefficient was −0.0217, indicating a weak autoregressive

relationship in the differenced series. Diagnostic checking was performed using residual analysis and the Ljung–Box test. The residuals exhibited characteristics of white noise, while the Ljung–Box test produced a p-value of 0.5256, indicating the absence of significant residual autocorrelation. These results confirmed the adequacy of the fitted model.

5. Conclusion and Implications

The study demonstrated the usefulness of autoregressive modelling techniques in analysing and forecasting annual crude oil production in Nigeria. The findings revealed that the crude oil production series is non-stationary in its original form but becomes stationary after first differencing. Using both Akaike Information Criterion and Bayesian Information Criterion as model selection criteria, the ARIMA (1,1,0) model was identified as the optimal forecasting model. The selected model provided an adequate representation of the series as confirmed by diagnostic tests, which showed that the residuals behaved as white noise. The forecasting results suggest that crude oil production in Nigeria is likely to remain relatively stable in the coming years, with no substantial upward or downward movement expected under current conditions. Therefore, it can be concluded that the ARIMA(1,1,0) model is an appropriate and reliable model for forecasting annual crude oil production in Nigeria and that AIC and BIC are effective criteria for selecting parsimonious autoregressive models.

References

[1] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). John Wiley & Sons.

[2] Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed.). Springer.

[3] Lütkepohl, H. (2005). New introduction to multiple time series analysis. Springer.

[4] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

[5] Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136

[6] Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. Biometrika, 92(4), 937–950. https://doi.org/10.1093/biomet/92.4.937

[7] McLeod, A. I., & Zhang, Y. (2006). Partial autocorrelation parameterization for subset autoregression. Journal of Time Series Analysis, 27(4), 599–612. https://doi.org/10.1111/j.1467-9892.2006.00485.x

[8] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.

[9] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts.

[10] Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. Journal of Econometrics, 2(2), 111–120. https://doi.org/10.1016/0304-4076(74)90034-7

[11] Hamilton, J. D. (1994). Time series analysis. Princeton University Press.

[12] Chatfield, C. (2016). The analysis of time series: An introduction (7th ed.). Chapman & Hall/CRC.

[13] Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). Springer.