FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 6, June, 2025, pp 193 – 201
DOI: https://doi.org/10.33003/fjs-2025-0906-3666

ASSESSING THE PERFORMANCE OF ARIMA AND ARFIMA MODELS IN FORECASTING INTERNALLY GENERATED REVENUE OF KADUNA STATE

*[1]Muhammad Idris Usman, [2]Tasi'u Musa and [1]Auwalu Ibrahim

[1]Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, Nigeria.
[2]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.

*Corresponding authors' email: baradengeri@gmail.com

ABSTRACT

Internally generated revenue (IGR) is an important source of revenue that can be used to fund public services and infrastructure projects. Accurate forecasting of IGR is essential for effective budgeting and financial planning. This study assessed the performance of ARIMA and ARFIMA models in forecasting internally generated revenue of Kaduna State. The study uses monthly IGR data from January 2003 to December 2023. The stationarity of the data was assessed using Augmented Dickey Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. The findings showed that both ARIMA and ARFIMA models perform well in forecasting IGR, but ARFIMA model outperforms ARIMA model in terms of mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The generated forecast values for 24 months using the model revealed that out-sample IGR forecasts fluctuated (decreasing and increasing). Thus, the study recommends the use of ARFIMA model for forecasting IGR in Kaduna State for better revenue planning and economic policy formulation.

Keywords: ARIMA Model, ARFIMA model, Forecasting, Internally generated revenue

INTRODUCTION

Internally Generated Revenue (IGR) refers to the income that a government or organization generates from its own activities within its jurisdiction, excluding external sources such as Federal allocations or grants. This revenue is typically derived from taxes, fees, license, and other charges imposed on businesses, individuals, and transaction conducted within the entity's geographical boundaries (Uduma *et al.,* 2021). The internally generated revenue (IGR) is an important source of revenue that can be used to fund public services and infrastructure projects. It has taken the second position in sources of revenue when Nigeria put heavy reliance on oil (Okorie *et al.,* 2018). Every institution is encouraged to augment its finances by generating revenue internally. The forecasting and control of such internally generated revenue could help in knowing the patterns and characteristics of the revenue for formulating a very good and impactful policies to achieve good governance. Kaduna State of Nigeria just as any other institution generates revenue internally to complement the efforts of the Federal Government.

Time series modeling and forecasting are essential methods used in different fields, especially in economic and financial trends, due to their ability to manage risk and increase investment in financial and industrial markets (Suleiman *et al.,* 2023). Therefore, a nontraditional and accurate statistical techniques called autoregressive integrated moving average (ARIMA) and fractional autoregressive integrated moving average (ARFIMA) models are considered to describe changes in internally generated revenue series.

Autoregressive integrated moving average (ARIMA) models are popular and widely used class of univariate time series models for forecasting and analyzing economic and financial data (Zhang, 2013). The ARIMA model is a combination of three key components: Autoregressive (AR), which uses past values of the time series to forecast future values. The AR component is based on the idea that the current value of a time series is a function of past values. Integrated (I), which differences the time series to make it stationary. A time series is said to be stationary if its mean and variance remain constant over time. Moving average (MA), this component uses the errors (residuals) from past forecasts to improve future forecasts. The MA component is based on the idea that the errors in a time series are correlated with each other.

Autoregressive fractionally integrated moving average (ARFIMA) models are extension of the ARIMA model that allows for fractional differencing. This means that the degree of differencing (d) can be a fraction, rather than an integer. Fractional differencing is useful for modeling time series data that exhibits long-range dependence, which means that the data is correlated over long periods of time (Liu, Chen and Zhang, 2017). ARFIMA models are particularly useful for modeling financial time series data, such as stock prices, internally generated revenue, and exchange rates.

ARIMA models have been widely used for forecasting time series data, including revenue (Box & Jenkins, 1976). However, ARFIMA models have been shown to outperform ARIMA models in certain contexts, particularly when the data exhibits long-range dependence (Hosking, 1981). Several studies have used ARIMA and ARFIMA models for forecasting economic and financial data. For example, Alireza and Ahmad (2009) used ARIMA and ARFIMA Model to explore the long memory of the Stock Price Index. The findings showed that the ARFIMA is a significantly better model in this regard after comparing the forecasting performance of the two models. Shittu and Yaya (2009) assessed the forecasting performance of ARIMA and ARFIMA models for stationary type series that exhibited long memory properties. The ARFIMA model showed more realistic forecast values that reflected current economic realities in the countries studied. Omekara *et al*. (2016) examined the accuracy of ARIMA and ARFIMA in forecasting the liquidity ratio of Nigerian commercial banks. For each of the ARFIMA and ARIMA models, the optimal model was determined based on least AIC values.

Hamzaoui & Regaieg (2017) examines the structure of the daily Euro to US dollar forward premium types of exchange rate using ARFIMA model. The results of the analysis confirmed the evidence of LM and fractional dynamics of the forward premium data. The ARFIMA model adequately fitted the data. Liu, Chen & Zhang (2018) compared and evaluated four components of the ARFIMA model which are simulation, fractional order difference filter, estimation and forecast. The result of the study showed that the ARFIMA model gives a better fitting result, especially for the data with long rage dependence (LRD) or long memory property. Elmezouar *et al.* (2021) examined the accuracy of ARIMA, ARFIMA and NNAR in modeling and forecasting the total fisheries production data in India. Augmented Dickey Fuller (ADF) test was used in testing the fundamental assumption of stationarity. The results indicated that ARIMA and NNAR models were outperformed by ARFIMA model in forecasting the total fisheries prediction.

Azza et al. (2021) assessed the performance of ARIMA and ARFURIMA models Kijang Emas monthly average prices in Malaysia. The findings revealed that ARFIMA model performed better in forecasting the Kijang Emas prices in Malaysia compared to the ARIMA model. Jibrin et al. (2021) compare the performance of ARIMA and ARFURIMA models using monthly Nigeria stock index. The findings indicated that ARFURIMA was the best fitted and accurate model.

Nwakuya and Biu (2022) applied ARFIMA models on COVID-19 daily deaths in Nigeria. The estimates of d parameter for the ARFIMA models was obtained using Geweke Porter-Hudak estimator (GPH). Suleiman *et al.* (2023) identified the best ARIMA time series model for monthly crude oil price in Nigeria spanning from 2006 to 2020. The best model was selected using the criteria of mean square error, root mean square error, and mean absolute error. Monge and Infante (2023) investigate historical data for crude oil prices using autoregressive fractionally integrated moving average (ARFIMA) models to determine whether shocks in the series have transitory or permanent effects.

Kelkar *et al.* (2021) used Seasonal Autoregressive Integrated Moving Average (SARIMA) models to forecast American Southwest Airlines' revenue for 2020. The study identifies SARIMA (0,1,0) (0,1,1)4 as the best-fitted model with the lowest AICs. Diagnostic tests confirm the model's accuracy and a solvency risk analysis is conducted to assess Southwest's financial performance during the COVID-19 pandemic.

Atoyebi *et al.* (2023) investigated forecasting currencies in circulation (CIC) in Nigeria using the Holt- Winters exponential smoothing methods, both additive and multiplicative. The analysis uses data from January 1960 to December 2022 to determine the optimal forecasting approach and the most effective smoothing parameters. Their results revealed that the multiplicative Holt-Winters method outperformed the additive method in accuracy, with significantly lower MAPE, MAD, and MSD values. Ajisola (2023) analyzed monthly data from the Federal Inland Revenue Service (FIRS) spanning 2010 to 2021, exploring three models for tax revenue forecasting: Multivariate Linear Regression (MLR), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Multivariate Long Short Term Memory Networks (LSTM). The study finds that LSTM and MLR perform well due to their ability to predict using multiple independent variables. LSTM achieves a high R^2 score of 98.9% and an adjusted R^2 score of 98.8%, suggesting its efficacy in forecasting tax revenue.

Tasi'u *et al*. (2024) assessed the performance of SARIMA and Holt-Winters models in forecasting the tax revenue of Nigeria. Following Box and Jenkins model identification, estimation, and forecasting procedures, SARIMA $(3,2,1)_4(0,1,1)_4$ model was selected based on the minimum AIC value, outperforming other models. The Multiplicative Holt-Winters model was also chosen for similar reasons. An in-sample forecast with 80% training and 20% validation set revealed that the SARIMA model outperformed Holt-Winters based on RMSE and MAE.

David *et al*. (2024) studied the symmetric and asymmetric characteristics as well as the persistence of shocks in the Nigerian crude oil return, utilizing monthly and daily crude oil prices spanning from January, 2006 to September, 2022 and November 3, 2009 to November 4, 2022 respectively. Descriptive statistics, normality measures, time plots and Dickey Fuller generalized least squares unit root tests were employed to analyze the series properties. Symmetric ARMA (1,1) - GARCH (2,1) and Asymmetric (1,1) – TARCH (2,1) models for monthly and daily returns. Models selection criteria including AIC, SIC and HQC and log likelihood guided the order and error distribution selection. Result revealed non-normal distributions for both monthly and daily prices and returns, non-stationarity in prices, and weak stationarity in log returns with ARCH effects detected in both returns.

In Nigeria, several studies have used regression and univariate time series models for forecasting internally generated revenue. For example, Patrick and John (2013) applied ARIMA Modeling to internally generated revenue of Akwa Ibom State, Nigeria. The best model was selected using minimum values of mean square error, root mean square error, and mean absolute error. Harrison et al. (2014) employed SARIMA modelling techniques to monthly internally generated revenue of Rivers State. Examination of the series reveals a seasonal nature of annual periodicity. Okorie *et al.* (2018) used ordinary least square regression and autoregressive average models to model and forecast the monthly generated revenue of Gombe State. Festus (2019) employed ARIMA methodology on internally generated revenue of Adamawa State. The best model was selected using the criteria of mean square error, root mean square error, and mean absolute error.

Studies including, (Patrick *et al.* (2013); Harrison *et al.*, 2014; Okorie *et al.,* 2018; Festus, 2019), employed regression and ARIMA models to describe internally generated revenue records. However, the accuracy of these methods are weak when extreme fluctuations occur and are unable to provide accurate results when the data exhibits long-range dependence. In this paper, a method that is capable of modelling and forecasting extreme fluctuations and long memory time series will be used. The method will be compared with the existing ARIMA model and studies the monthly Kaduna State internally generated revenue.

MATERIALS AND METHODS

Autoregressive Integrated Moving Average ARIMA (p, d, q)

Autoregressive integrated moving average (ARIMA) model is a statistical model proposed by Box and Jenkins (1976) to forecast and analyze time series data by integration process. Consider nonstationary time series Y_t and suppose the stationary d^{th} order difference of Y_t is denoted by $\Delta^d Y_t$, then an ARIMA model of order *p*, *d* and *q*, denoted by ARIMA (p,d,q) is given as

$$\phi(L)\Delta^d Y_t = \theta(L)\varepsilon_t \qquad (1)$$

where, L is the backward shift operator, ε_t is a white noise process and *d* is the integration parameter.

Autoregressive Fractionally Integrated Moving Average (ARFIMA) Model

The ARFIMA models are extension of ARIMA models which allows modeling time series data that exhibits long-range dependence. The model has three parameters: *p*, *d*, and *q*, where the parameter p represent the number of lags involved in the autoregressive portion in the series, q is the parameter representing the moving average portion, and d is the fractionally integrated parameter which usually takes a value in the interval of $0 < d < 1$..

Consider Y_t to be a nonstationary process with mean and variance changing over time. Then Y_t is said to be fractional integral process if

$$(1-L)^d Y_t = \varepsilon_t \quad (2)$$

where this has the interpretation as follows:

$$Y_t - dY_{t-1} + \frac{d(d-1)}{2!}Y_{t-2} - \frac{d(d-1)(d-2)}{3!}Y_{t-3} + \ldots = \varepsilon_t \quad (3)$$

where, L is the backward shift operator, ε_t is a white noise process and d is the long memory parameter such that $0 < d < 1$.

The general form of an ARFIMA model of Granger and Joyeux (1980) is given as

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t, 0 < d < 1. \quad (4)$$

ARFIMA process is said to be nonstationary when $d \geq 0.5$, while the process is said to exhibit long memory if $0 < d < 0.5$. The process shows short memory when $d = 0$ and anti-persistence when $d < 0$.

Detection of Long Memory

Geweke and Porter-Hudak (1983) proposed a semi-parametric approach to test for long memory, using the following regression;

$$\ln I(w_j) = \beta - d\ln\left[4\sin^2(w_j/2)\right] + n_j \quad (5)$$

where $w_j = 2n_j/T, j = 1, \ldots, n$; n_j is the residual term and denotes Fourier frequencies. $I(w_j)$ represent the periodogram of a time series r_1 and it is defined as

$$I(w_j) = \frac{1}{2\pi^T}\left|\sum_{t=1}^{T} r_1 e^{-w_j^t}\right|^2 \quad (6)$$

Model Building Strategy

Unit Root Test

The data is checked for stationarity using tests such as Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Augmented Dickey Fuller (ADF) tests.

The KPSS test statistic proposed by Kwiatkowski et al. (1992) with the null hypothesis that the data generating process is stationary is given as

$$t - statistics(t_k) = \frac{1}{T^2}\sum_{t=1}^{T}\frac{s_t^2}{\hat{\sigma}_\infty^2} \quad (7)$$

where $s_t = \sum_{j=1}^{t} w_j$ with $w_j = y_t - y$ and $\hat{\sigma}_\infty^2$ is an estimator of the long-run variance of the process.

The ADF test statistic developed by Dickey and Fuller (1979) is given as

$$\text{ADF} = \frac{\hat{\varphi}}{SE(\hat{\varphi})} \quad (8)$$

where $SE(\hat{\varphi})$ is the standard error for $\hat{\varphi}$, and $\hat{\varphi}$ denotes estimate. The null hypothesis of unit root is accepted if the test statistic is greater than the critical values.

Model Selection

The popular model selection criteria are AIC due to (Akaike, 1974), HQC due to (Hannan – Quinn, 1979) and SIC due to SIC (Schwarz, 1978). The expressions for the three information selection criteria are, respectively, given as

$$Akaike: \quad C_n(k) = -\frac{2\ln(L_n(k))}{\frac{n+2k}{n}} \quad (9)$$

$$Hannan - Quinn: C_n(k) = -\frac{2\ln(L_n(k))}{\frac{n+2k\ln(\ln(n))}{n}} \quad (10)$$

$$Schwarz\ information: C_n(k) = -\frac{2\ln(L_n(k))}{\frac{n+k\varphi(n)}{n}} \quad (11)$$

where k is the number of parameters, n is the number of observations. $\varphi(n) = 2$ in Akaike case, $\varphi(n) = 2\ln(\ln(n))$ in Hannan – Quinn case $\varphi(n) = \ln(n)$ in the Schwarz case.

Forecasting Evaluation

Performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) were considered to assess the forecast accuracy;

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{n}} \quad (12)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (13)$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100 \quad (14)$$

where y_i is the true value, $\hat{y}_i$ is the predicted values when all samples are include in the model formation, n is the number of observations.

RESULTS AND DISCUSSION

The study uses monthly internally generated revenue data of Kaduna State, obtained from Kaduna State Internal Revenue Service (KADIRS). The dataset covers a sufficiently long period spanning from January 2003 to December 2023 to analyze long-memory effects. R and Gretl statistical softwares were used in conducting the analysis. The pattern and behavior of the data was studied by Time plot, ACF and PACF as shown below.

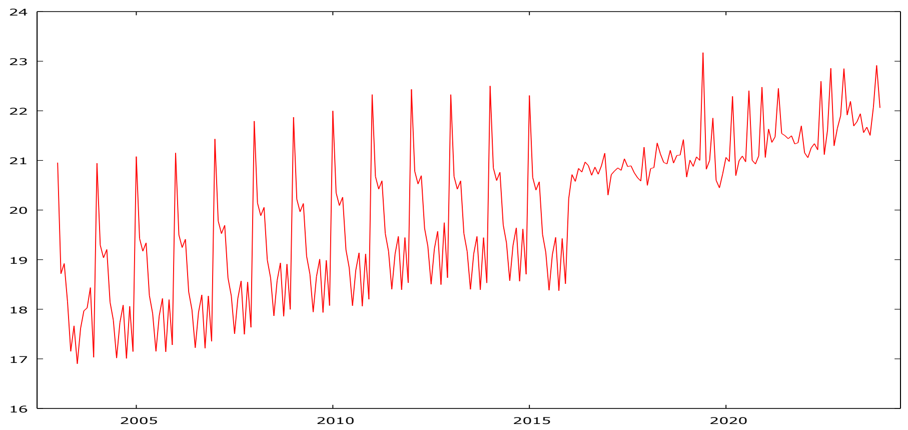

Figure 1: Time plot of internally generated revenue

Figure 1 revealed that the monthly average internally generated revenue is increasing at the beginning of each year (between January and April) and subsequently decreasing (between May and December). This suggested a form of seasonality in the series. The IGR series increase from 2010 to 2015, then decrease from 2016 to 2019, with a further increase from 2020 to 2023; which is the peak. This indicates that the series consists of trend, meaning it's not yet stationary. Further unit root tests such as KPSS and ADF tests were used to confirm the stationarity of the data.

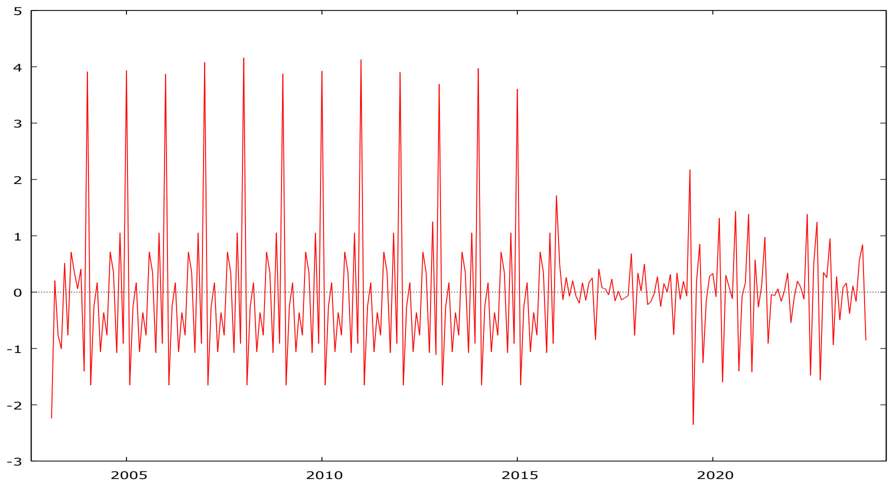

Figure 2: Time plot of fractional differencing

Figure 2 presented the time plot of the fractional differencing for the IGR data. The plot after fractional differencing shows the mean and variance were stabilized. This means that the IGR series is stationary at fractional differencing.

ACF and PACF Plot of the IGR Series

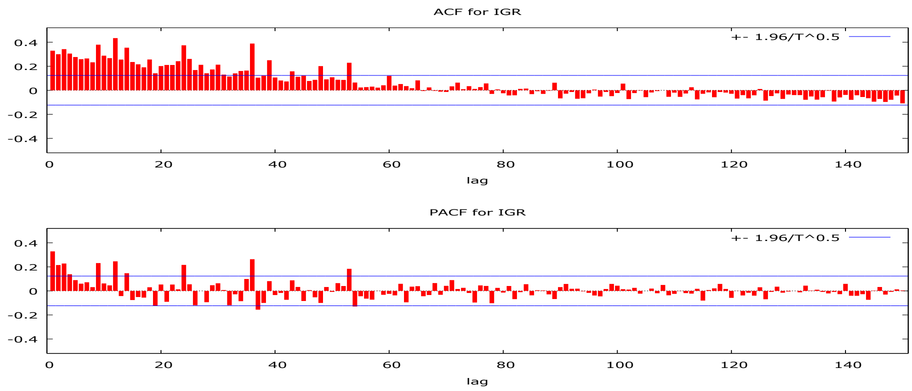

Figure 3: ACF and PACF Plot of the IGR Series

Figure 3 displayed the sample ACF and PACF of the IGR series. The autocorrelation function of IGR decreases slowly at a hyperbolic rate, an indication of long memory (or long-range dependence), which is also conformed to a fractionally integrated series. The PACF is significant at lag 54 but decays very slowly to zero.

Unit Root Test and Long Memory Test

Table 1: ADF and KPSS Tests of the Data

Test	Lag Order	Unit Root of the Original Data		Unit Root at First Difference	
		T-Statistic	P-Values	T-Statistic	P-Values
ADF Test	5	-1.32359	0.0638	-9.76023	1.94e-018
	12	0.38859	0.9825	-9.8177	1.27e-018
	20	1.95997	0.9999	-5.87809	2.34e-007
KPSS Test	5	2.88736	0.0013	0.024352	0.8472
	12	1.61586	0.0025	0.056119	0.7488
	20	1.08174	0.0044	0.089088	0.6022

ADF (Unit Root Test)
H_0: Presence of unit root in the IGR series
H_a: No presence of unit root in the IGR series

KPSS (Test of Stationary)
H_0: IGR series is stationary
H_a: IGR series is non-stationary

From Table 1, the KPSS test result showed that the data was not stationary before first differencing, but was stationary after first differencing at 5% level of significance. Table 1 also displayed the ADF test result which revealed that there was presence of unit root in the data before first differencing but there was no presence of unit root in the data after first differencing at 5% level of significance.

Table 2: Long Memory Estimate of IGR Series

Test	Estimate (d)	Z statistic	p-value
Geweke and Porter-Hundlak (GPH)	0.42363579	4.31967	0.0000

The GPH test result is presented in Table 2. The estimate of the fractional parameter d is obtained to be 0.4236. GPH provides fractional difference parameter values which lies within the conventional long memory parameter.

Identification of ARIMA and ARFIMA Models

In this subsection, the identification, estimation and diagnostic test of the ARIMA and ARFIMA models will be discussed. Applying the principle of parsimony with d=1, p,q = (0, 1, 2, 3), sixteen (16) models were generated. The models will be selected based on the information selection criteria and the chosen models will be estimated and examine to ascertain their adequacy.

Table 3: AIC, BIC and HQC for Candidate Models

ARIMA				ARFIMA			
(p, d, q)	**AIC**	**BIC**	**HQC**	**(p, d, q)**	**AIC**	**BIC**	**HQC**
(0,1,0)	795.4907	802.5416	798.3282	(0,0.4236,0)	919.1966	926.2555	922.0370
(0,1,1)	706.0353	716.6117	710.2915	(0, 0.4236,1)	841.6057	852.1940	845.8662
(0,1,2)	696.9869	711.0887	702.6618	(0, 0.4236,2)	769.7757	783.8934	775.4563
(0,1,3)	679.5707	699.1980	689.6643	(0, 0.4236,3)	762.5294	780.1766	769.6303
(1,1,0)	723.7519	734.3283	728.0081	(1, 0.4236,0)	759.0637	769.6520	763.3242
(1,1,1)	723.8648	737.9666	729.5397	(1, 0.4236,1)	722.4480	736.5657	728.1286
(1,1,2)	688.8381	706.4653	695.9317	(1, 0.4236,2)	717.1288	734.7760	724.2297
(1,1,3)	677.5571	698.7098	686.0694	(1, 0.4236,3)	716.1032	733.7503	723.2040
(2,1,0)	724.0943	738.1961	729.7692	(2, 0.4236,0)	714.1996	728.3174	719.8803
(2,1,1)	681.1881	698.8154	688.2817	(2, 0.4236,1)	695.1032	722.7503	707.2040
(2,1,2)	681.7638	702.9165	690.2762	(2, 0.4236,2)	717.8984	739.0750	726.4194
(2,1,3)	679.3045	703.9827	689.2357	(2, 0.4236,3)+	696.7058	721.4118	706.6469
(3,1,0)	726.0750	743.7022	733.1686	(3, 0.4236,0)	716.0827	733.7299	723.1836
(3,1,1)	678.7497	699.9024	687.2621	(3, 0.4236,1)	715.1475	736.3241	723.6686
(3,1,2)	684.5500	709.2282	694.4812	(3, 0.4236,2)	700.2729	724.9789	710.2141
(3,1,3)+	624.9349	653.1385	636.2847	(3, 0.4236,3)	698.4644	726.6998	709.8257

Table 3 presented the results of information selection criteria of the ARIMA and ARFIMA models. Tweny (20) models were tested based on the Alkaike information criteria (AIC); Bayesian information criteria (BIC); Hannan-Quinn information criteria (HQC), where ARIMA(3,1,3) and ARFIMA(2, 0.4236,3) models possessed the minimum values of AIC, BIC and HQC and therefore, were selected for further examination.

Estimation of ARIMA and ARFIMA Models

Table 4: Estimation of ARIMA (1,1,2) and ARFIMA(2, 0.4236,4) Models

Parameters	ARIMA (1,1,2)		ARFIMA (2,0.4236,4)	
	Coefficient	**Z-Statistic**	**Coefficient**	**Z-Statistic**
Const	0.0154399	11.79(4.61e-032)	4.47183	1.425 (0.1542)
phi_1	−1.21184	−20.15(2.49e-09)	−0.0151841	-0.69(0.4872)
phi_2	−0.131530	−1.335(0.1819)	0.978885	44.98(0.0000)
phi_3	0.491651	8.644(5.44e-018)	-	-
theta_1	0.607407	15.84(1.68e-056)	0.389457	5.17(2.33e-07)
theta_2	−0.684905	−16.14(1.28e-058)	−0.321850	3.59(0.0003)
theta_3	−0.922502	−36.19(9.04e-287)	0.253802	3.33(0.0009)

Note: p-values are in parenthesis

The estimated parameters of ARIMA (3,1,3) and ARFIMA (2, 0.4236,3) models are presented in Table 4. The parameters (phi-1, phi-3, theta_1, theta_2 and theta_3) in the ARIMA (3,1,3) model are statistically significant to the model at 5% level of significance. The parameters (phi-2, theta_1, theta_2, and theta_3) in the ARFIMA (2, 0.4236,4) model are statistically significant.

Diagnostic Checking of ARIMA and ARFIMA Models

Test for autocorrelation and partial autocorrelation, Jarque–Bera and Ljung-Box tests were employed to ascertain the adequacy of the model.

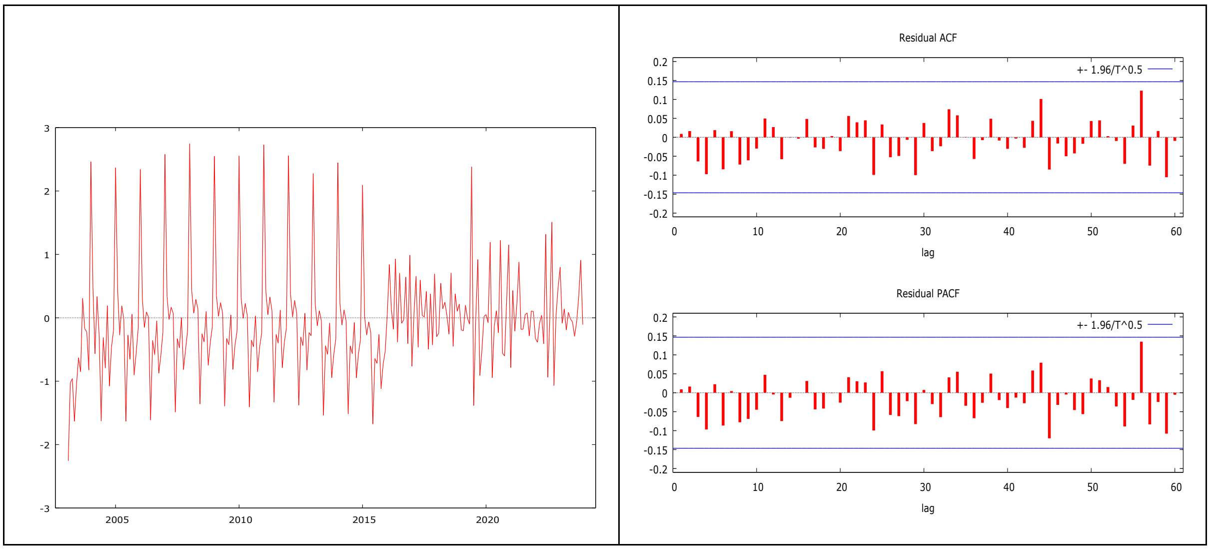

Figure 4: Diagnostic Plots of the ARIMA(3,1,3) Model fitted to IGR series

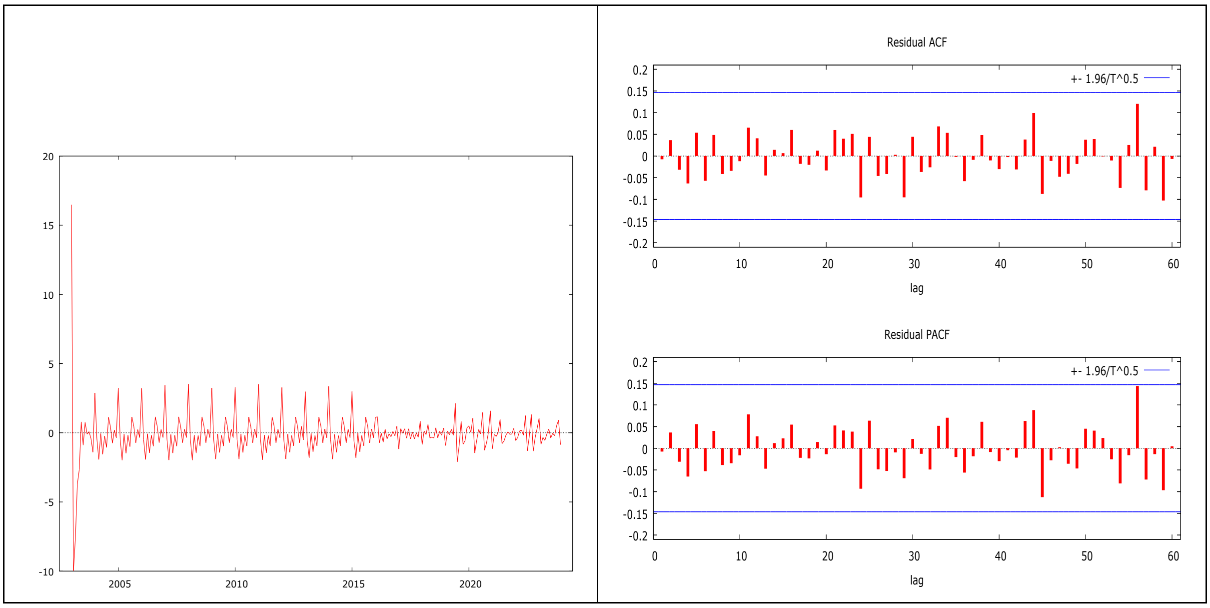

Figure 5: Diagnostic Plots of the ARFIMA (2, 0.4236,3) Model fitted to IGR series

The diagnostic plots ARIMA(3,1,3) and ARFIMA (2, 0.4236,3) models are displayed in Figure 4 and Figure 5, respectively. Both the figures revealed that there is no form of correlation amongst the residuals from the ACF and PACF plot. It was also observed that the residuals of the IGR series are stationary from the time plot. Therefore, ARIMA (3,1,3) and ARFIMA (2, 0.4236,3) models passed the standard criteria of being white noise, since the residuals are uncorrelated and stationary.

Table 5: Diagnostic Tests of the ARIMA(1,1,2) and ARFIMA (2, 0.4236,4) Models

Models	Jarque-Bera Test	ARCH-LM Test
ARIMA (1,1,2)	1.1215(0.1308)	1.5037(0.1122)
ARFIMA (2, 0.4236,4)	1.5532(0.1326)	0.6422(0.6523)

Note: p-values are in parenthesis

The diagnostic test result using Jarque-Bera test and ARCH-LM test of the ARIMA (3,1,3) and ARFIMA (2, 0.4236,3) models are presented in Table 5. The result indicated evidence of normality and homoscedasticity in the error terms since the p-values are greater than 5% level of significance.

Table 6: Forecast Accuracy Measures of ARIMA and ARFIMA Models

Models	MSE	RMSE	MAE	MAPE
ARIMA (3,1,3)	0.6763	0.8224	0.5533	2.7706
ARFIMA (2,0.4236,3)	0.6282	0.7926	0.5318	2.5600

The forecast accuracy results of the ARIMA and ARFIMA models are presented in Table 6. The smaller the value of the error, the better the forecasting performance of the model. The ARFIMA (2,0.4236,3) model possessed the minimum values of mean square error (0.6282), root mean square error (0.7926), mean absolute error (0.5318), and mean absolute percentage error (2.5600) compared to ARIMA (3,1,3) with mean square error (0.6763), root mean square error (0.8224), mean absolute error (0.5533), and mean absolute percentage error (2.7706). This may be attributed to the presence of long-range dependence in the IGR series. Therefore, the prediction power of ARFIMA (2,0.4236,3) model is better and suitable for monthly periods forecasting, as such the model best fit the data.

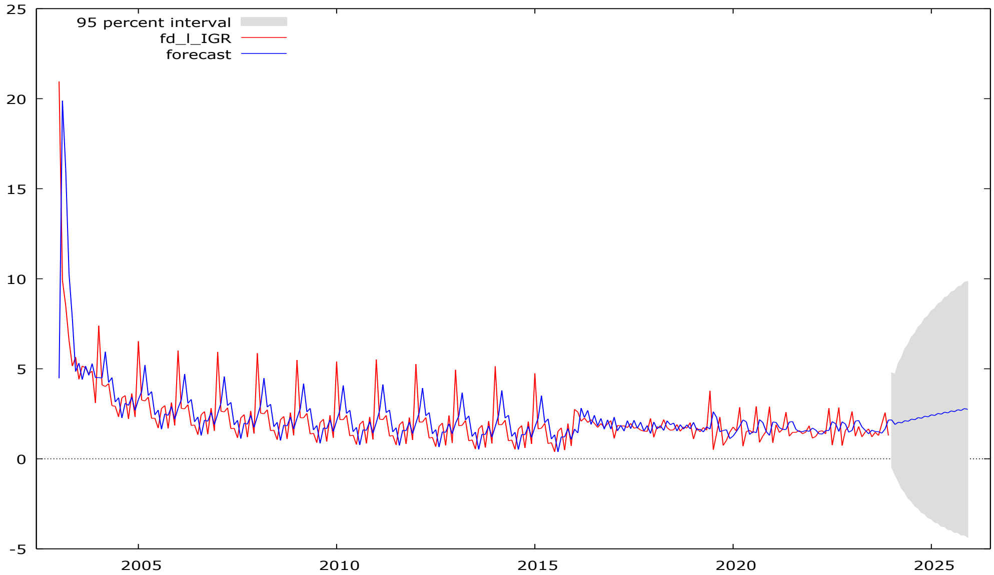

Figure 6: Forecast of IGR Series using ARFIMA (2,0.4236,3)

Table 7: Forecast and 95% Confidence Interval of the IGR SERIES Using ARFIMA (2,0.4236,3)

Year	Prediction	Lower Bound	Upper Bound
2024:1	2385274383	744563703.2	25193075202
2024:2	185849915.5	493361420.6	18828961021
2024:3	653119061.1	492748699.6	22124490511
2024:4	463102977.1	586572597.9	19958441035
2024:5	1593119690	531227540.2	18231432471
2024:6	1116706986	733733361.1	24753184892
2024:7	3762915903	477099005.2	16710698836
2024:8	2608914314	623057718.8	25537454117
2024:9	8616286256	566080987.5	19392980351
2024:10	5912058999	611646354.4	21844434242
2024:11	7044197040	706363121.8	24868455752
2024:12	4785804609	533299106.4	18789330143
2025:1	5595228400	801304033.2	28352356924
2025:2	3765852123	553310563.8	19715618985
2025:3	4322500970	730719336.9	26202456431
2025:4	7837925252	685267615.8	24642637222
2025:5	8837059301	612210661.9	22019207992
2025:6	5845278042	835698951.2	30184953501
2025:7	6476914004	556648879.1	20825618271
2025:8	4249937016	839236405.4	30436850371
2025:9	4630357889	668801595.9	24370178281
2025:10	8196556795	717575693.2	26214606441
2025:11	8784810610	838377472.2	30246776612
2025:12	5679846246	633511687.3	23358139652

The forecasts values along with 95% upper and lower confidence interval for 24 months (January 2024 to December 2025) out-sample for the of the IGR series analyzed using the ARFIMA model are presented in Figure 4 and Table 8, respectively. The plots and forecast values revealed that out-sample forecasts fluctuated (decreasing and increasing). This signifies that the ARFIMA (2,0.4236,3) is the appropriate model for modelling and forecasting the monthly internally generated revenue of Kaduna State, Nigeria.

CONCLUSION

The study assessed the performance of ARIMA and ARFIMA models in forecasting internally generated revenue of Kaduna State. The findings demonstrate ARIMA and ARFIMA effectiveness in forecasting IGR series. The results showed ARFIMA model outperforms ARIMA model with minimum values of mean square error (0.6282), root mean square error (0.7926), mean absolute error (0.5318), and mean absolute percentage error (2.5600). Policymakers can leverage these forecasts for informed decision-making and resource allocation. Future research may explore hybrid models integrating ARFIMA with machine learning techniques for enhanced accuracy.

REFERENCES

Ajisola, A. S. (2023). *An Efficient Time Series Model for Tax Revenue Forecasting: A Case of Nigeria* (Doctoral dissertation, AUST).

Atoyebi, S. B., Olayiwola, M. F., Oladapo, J. O., & Oladapo, D. I. (2023). Forecasting Currency in Circulation in Nigeria Using Holt-Winters Exponential Smoothing Method. *South Asian Journal of Social Studies and Economics*, 20(1), 25-41.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *I.E.E.E. Transaction on Automatic Control,* 19, 716-723.

Alireza, E. & Ahmad, Y.S. (2009). Long Memory Forecasting of Stock Prices Index using a Fractionally Differenced ARMA Model. *Journal of Applied Sciences Research,* 10, 1721-1731.

Azza, A. M. A., Auni, F. M. Z. & Wee, P. M. (2021). The Comparison between ARIMA and ARFIMA Model to Forecast Kijang Emas (Gold) Prices in Malaysia using MAE, RMSE and MAPE. *Journal of Computing Research and Innovation*, 6(3), 22-33.

Box, G. E. P. & Jenkins, G. M. (1976). Time Series Analysis: *Forecasting and Control*. San Franciso: Holden- Day.

David, A. K., Enobong, F. U., & Damian, O. (2024). Volatility Analysis of Crude Oil Prices in Nigeria. *FUDMA Journal of Sciences (FJS),* 8(1), 125 – 134. https://fjs.fudutsinma.edu.ng

Dickey, D. A. & Fuller, W. A. (1979). Distributions of the Estimators for Autoregressive Timse Series with a Unit Root. *Journal of the American Statistical Association,* 74: 427-431.

Elmezouar, C., Ibrahim. M. A., Ahmad, I. & Laksaci, A. (2021). Comparison of ARFIMA, ARIMA and Artificial Neural models to Forecasting the Total Fisheries Production in India. *Journal of Animal & Plant Sciences*, 31(5), 1477-1484.

Festus, A. (2019). Application of Time Series Analysis on Revenue Generation in Adamawa State, A Case Study of Adamawa State Board of Internal Revenue Yola, Nigeria. Adapoly Printing Venture, Yola.

Geweke, J. & Porter-Hudak, S. (1983). The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis*. 4, 221-237.

Granger, C. W. J. & Joyeux, R. (1980). An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1(1), 15-29.

Hannan, E. J. & Quinn, B. G. (1979).The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society*, 41, 190-195.

Harrison, E. E., Uchenna, I. A., Simon, I. A., Yellow, M. D. & Chinedu, R. I. (2014). Application of Seasonal Box-Jenkins Techniques for Modelling Monthly Internally Generated Revenue of Rivers State of Nigeria. *International Journal of Innovative Science, Engineering & Technology (IJISET)*, 1(7), 122-126.

Hamzaoui, N. & Regaieg, B. (2017). The Long Memory Behaviour of the EUR/USD Forward Premium. *International Journal of Economics and Financial Issues*, 7(3), 437-443.

Hosking, J.R.M. (1981). Fractional Differencing. *Biometrika*, 68(1), 165-176.

Jibrin, S.A., Musa, Y., Samail, M. & Abdul, A. (2021). Comparing the performance of ARIMA and ARFURIMA Models in Forecasting Nigeria Stock Price Index. *Sule Lamido University Journal of Science and Technology (SLUJST)*, 2(2), 1-15

Kelkar, M., Borsa, C., & Kim, L. (2021). Time-Series Statistical Model for Forecasting Revenue and Risk Management. *Journal of Student Research*, 10(3), 1-14.

Kwiatkowski, P .C. B. Phillips, P. Schmidt, & Yongoheol, S. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometric*, 54: 159-178.

Liu, K., Chen, Y. & Zhang, X. (2017). An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average). *Programs Axioms*, 6(16), 1-16.

Monge, M., & Infante, J. (2023). A Fractional ARIMA (ARFIMA) Model in the Analysis of Historical Crude Oil Prices. *Energy Research Letters*, 4(1), 112-121.

Nwakuya, M. T. & Biu, E. O. (2022). ARFIMA Modelling and Long Memory: The Case of Covid-19 Daily Deaths in Nigeria. *International Journal of Innovative Mathematics, Statistics & Energy Policies*, 10(4), 47-59.

Schwarz, K. J. (1978).The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society*, 41, 190-195.

Shittu, O.I. & Yaya, O.S. (2009). Measuring Forecast Performance of ARMA and ARFIMA Models: An Application to US Dollar/UK Pound Foreign Exchange Rate. *European Journal of Scientific Research*, 32(2):167-176.

Suleiman, M., Muhammad, I., Zakar, A. A., Zakari, Y. Iliyasu, R., Muhammad, A., Adamu, I. & Abdu, M. (2023). Modelling Nigeria Crude Oil Prices using ARIMA Time Series Models. *NIPES Journal of Science and Technology Research*, 5(1), 230-241.

Tasi'u, M., Abdulrazak, U. and Hussaini, G. D. (2024). Modelling and Forecasting Nigeria's Tax Revenue: A

Comparative Analysis of SARIMA and Holt-Winters Models. *UMYU Scientifica*, 3(3), 118 – 129.

Okorie, C.E., Ossai, F.C. & Ben, J. (2018). Time Series Analysis of Monthly Generated Revenue in Gombe Local Government. *International Journal of Scientific and Innovative Mathematical Research*, 6(2), 17-24.

Omekara, C.O., Okereke, O.E. & Ukaegeu, L.U. (2016). Forecasting Liquidity Ratio of Commercial Banks in Nigeria. *Scientific Series*, 3, 144-159.

Patrick, U. U. and John, E. E. (2013). Modeling Internally Generated Revenue (IGR) of Local Governments in Nigeria. *Mathematical Theory and Modeling*, 3(14), 23-29.

Uduma, E. A., Iwueze, S. I., Arimie, O. C. & Biu, O. E. (2021). Modelling the Nigerian Ports Authority Revenue Generated Series with and without Outliers. *Academic Journal of Statistics and Mathematics (AJSM)*, 7(7), 16-39.

Wiri, L. & Lebari, G. T. (2022). Modelling of Nigeria Exchange Rate using Autoregressive Fractional Integrated Moving Average. *Asian Journal of Pure and Applied Mathematics* 4(1), 28-35.

Zhang, G. P. (2003). Time Series Forecasting using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 159–175.