FORECASTING INTERNALLY GENERATED REVENUE OF KADUNA STATE USING ARFIMA MODEL

[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 Author: Muhammad Idris Usman, Email: baradengeri@gmail.com

Abstract

Accurate forecasting of internally generated revenue (IGR) is crucial for effective fiscal planning and sustainable economic development. This study applies the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model to forecast the IGR of Kaduna State, Nigeria. ARFIMA is particularly useful for modeling long-memory processes, which are common in financial and economic time series. The data used for the study was obtained secondarily from Kaduna State Internal Revenue Service (KADIRS). The stationarity of the data was assessed using Augmented Dickey Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. The long memory parameter d of the ARFIMA model was estimated using the Geweke and Porter-Hudak (GPH) method. The presence of a long memory structure was revealed by the sample autocorrelation function. Based on the information selection criteria, using AIC, BIC, and HQC, two optimal time series models were selected. But the prediction power of ARFIMA (3,0.423636,4) model is better and suitable for monthly periods forecasting, as such the model best fit the data. Thus, the findings can be used to provide accurate and reliable forecast of Kaduna State IGR for better revenue planning and economic policy formulation.

Keywords: *ARFIMA model, forecasting, internally generated revenue, long-memory.*

1. 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 transactions 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 among sources of revenue in Nigeria, especially placed 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 internally generated revenue can help in understanding its patterns and characteristics, which are essential for formulating effective and impactful policies to achieve good governance. Kaduna State, like many other institutions in Nigeria, generates revenue internally to complement the efforts of the Federal Government.

Time series data is a set of observations obtained by measuring a single variable over a regularly period of time. The study of Time series analysis is relevant in many fields, such as agriculture, geography, health sciences, social sciences, and economics.

The main objective of time series analysis is to decompose the data into trend, seasonal, cyclical and error components for forecasting purposes. The concept of time series is based on past observations, present and future of any series (Chatfield, 1984).

There are two main approaches to time series analysis: The time domain approach and the frequency domain approach (Hamilton, 1983).

The Autoregressive fractionally integrated moving average (ARFIMA) model is used to model time series data. ARFIMA models extend ARIMA models by allowing non-integer values for the differencing parameter, making them suitable for capturing long-memory processes.

The ARFIMA(p, d, q) model belongs to the long memory model family (Liu, Chen and Zhang, 2017). Its primary objective is to explicitly account for persistence and long-term correlations in the data.

Several studies have explored revenue forecasting using time series models. On this note, Harrison et al. (2014) employed SARIMA modelling techniques to monthly internally generated revenue of Rivers State of Nigeria. The finding revealed that adequate model for prediction was SARIMA(0, 1, 1)x$(1, 1, 1)_{12}$ model.

Gimba *et al.* (2018) studied the impact of Personal Income Tax (PIT) on Internally Generated Revenue in Kaduna State from 1988-2015 using time series data obtained from Kaduna State Internal Revenue Service (KADIRS). The study employs the Engle and Granger (1987) two-stage Co-integration estimation techniques for the long run equilibrium relationship and the associated Error Correction Mechanism (ECM) to estimate the multivariate model. The findings revealed that there is a strong positive significant impact of PIT on internally generated revenue in Kaduna State within the period of the study.

Okorie *et al.* (2018) examined the time series analysis of monthly generated revenue in Gombe Local Government of Gombe State. Ordinary least square regression and Autoregressive Average models were used for the analysis. The trend of findings shows a significant increase in the monthly internal generated revenue in the study area.

Festus, (2019) investigated the application of time series analysis on revenue generation in Adamawa State. The study used ordinary least square method of multiple regressions to establish a relation between revenue and economic growth.

Waniyos *et al.* (2020) examined the time series analysis of Internal Generated Revenue in Adamawa State. The study used least square method to determine the level of trend patterns over the study period and predicted the future generation of internal revenue in the state.

Uduma *et al.* (2021) examined the Nigerian Ports Authority (NPA) revenue generated monthly series spanning January, 2007 to December, 2019. They defined a good model and a perfect fit for the NPA revenue generated series by comparing the forecast of the transformed series with outlier and without outliers using the forecast evaluation criterion.

Suleiman *et al.* (2023) identified the best ARIMA time series model for monthly crude oil price in Nigeria spanning from 2006 to 2020. The finding of the study revealed that ARIMA (3,1,1) model best fits the data with minimum values of predictive measures.

Monge & Infante (2023) investigated historical data for crude oil prices using autoregressive fractionally integrated moving average (ARFIMA) model. The best specification is an ARFIMA(2,d,2) with an estimated value of d around 0.4, but its confidence interval is wide and does not allow the rejection of I(0) or I(1) hypotheses. This high level of uncertainty may be due to the presence of breaks or non-linear trends in the data.

However, the ARIMA models used in forecasting internally generated revenue are limited in handling long-range dependencies. Fractionally integrated models, such as ARFIMA, introduced by Granger & Joyeux (1980), offer a more flexible approach by allowing for fractional differencing.

Empirical applications have shown ARFIMA's superiority in modelling financial and economic data with long-memory characteristics (Baillie, 1996). Therefore, this study contributes to the literature by applying ARFIMA to Kaduna State’s IGR data.

2.0 Methodology

2.1 Autoregressive Fractionally Integrated Moving Average Process ARFIMA (p, d, q)

The ARFIMA like ARIMA models has three parameters: *p*, *d*, and *q*. The parameter corresponding to the number of lags involved in the autoregressive portion of the series is *p*. Meanwhile, the parameter for the moving average lags is *q*. If the series is fractionally integrated, with *d* takes a value in the interval of $0 < d < 1$ then, the model is referred to as an ARFIMA model.

Now, consider $\{Y_t\}$, t = 1…, is a nonstationary process with time-varying mean and variance. Then Y_t is said to be fractional integral process if

$$(1-L)^d Y_t = \varepsilon_t \tag{1}$$

where this has the interpretation as follows:

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

Here, 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) and Hosking (1981) is given by:

$$\varphi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t, 0 < d < 1. \tag{3}$$

2.2 Model Estimation

2.2.1 Unit Root Test

Kwiatkowski Phillips Schmidt and Shin Tests

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test has been proposed by Kwiatkowski et. al. (1992) with the null hypothesis that the data generating process is stationary are tested against a unit root. The test statistic is given by

$$t - statistics(t_k) = \frac{1}{T^2}\sum_{t=1}^{T}\frac{s_t^2}{\hat{\sigma}_\infty^2} \tag{4}$$

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 z_t. The null hypothesis of the test is $H_0 : \hat{\sigma}_\infty^2 = 0$ against the alternative hypothesis $H_1 : \hat{\sigma}_\infty^2 \neq 0$. Reject the null hypothesis if the test statistic is greater than the asymptotic critical values.

Augmented Dickey Fuller (ADF) Test

The ADF test statistics developed by Dickey & Fuller (1979) is given by

$$\text{ADF} = \frac{\hat{\phi}}{SE(\hat{\phi})} \tag{5}$$

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

2.2.2 Detection of Long Memory

Geweke and Porter-Hudak (GPH) test

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

$$\ln I(w_j) = \beta - d \ln\left[4\sin^2(w_j/2)\right] + n_j \tag{6}$$

Where $w_j = 2n_j/T,\ j = 1,...,n;$ 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

$$I(w_j) = \frac{1}{2\pi^T}\left|\sum_{t=1}^{T} r_1 e^{-w_j^t}\right|^2 \tag{7}$$

2.2.3 Information Selection Criteria

The popular model selection criteria are AIC due to (Akaike, 1974), HQC due to (Hannan – Quinn, 1979) and SIC due to SIC (Schwarz, 1978).

Let $L_n(k)$ be the maximum likelihood of a model with k parameters based on a sample of size n. The information criteria for selecting the most parsimonious correct model proposed by Akaike (1974) is given by:

$$Akaike: \qquad C_n(k) = -\frac{2\ln\left(L_n(k)\right)}{\frac{n+2k}{n}} \tag{8}$$

where k is the number of parameters, n is the number of observations.

Hannan-Quinn information criterion (HQC) is an alternative to Akaike information criterion (AIC) and Bayesian information criterion (BIC) given as;

$$Hannan - Quinn: \qquad C_n(k) = -\frac{2\ln\left(L_n(k)\right)}{\frac{n + 2kln(\ln(n))}{n}} \tag{9}$$

Schwarz information is derived using Bayesian arguments, this criterion is also known as the Bayesian Information Criterion (BIC). These criteria take the general form;

$$C_n(k) = -\frac{2\ln\left(L_n(k)\right)}{\frac{n + k\varphi(n)}{n}} \tag{10}$$

where $\varphi(n) = 2$ in Akaike case, $\varphi(n) = 2\ln(\ln(n))$ in Hannan – Quinn case $\varphi(n) = \ln(n)$ in the Schwarz case.

2.2.4 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;

Root Mean Square Error (RMSE)

The RMSE is a measure of how well the model fits the data. It is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{n}} \tag{11}$$

where the $\hat{y}_i$ are the values of the predicted variable when all samples are including in the model formation, and n is the number of observations. RMSE

Mean Absolute Error (MAE)

The MAE is a quantity used to measure how close predictions are to the eventual outcomes.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{12}$$

It is an average of the absolute errors. i.e. $|e_i| = |f_i - y_i|$, where f_i is the prediction and y_i is the true value.

Mean Absolute Percentage Error (MAPE)

The MAPE is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$\text{MAPE}=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t-F_t}{A_t}\right|*100 \quad (13)$$

Where, A_t is the actual value and F_t is the forecast value.

The difference between A_t and F_t is divided by the Actual value A_t again.

3.0 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.

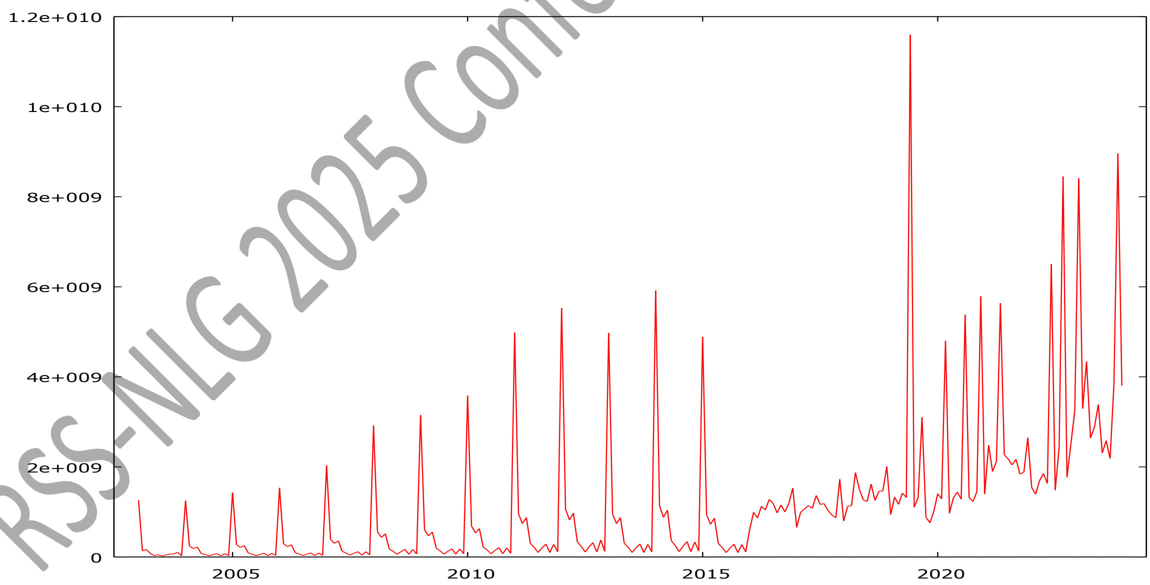

Fig. 1. Time plot of internally generated revenue

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). 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 confirmed whether the data is stationary or not.

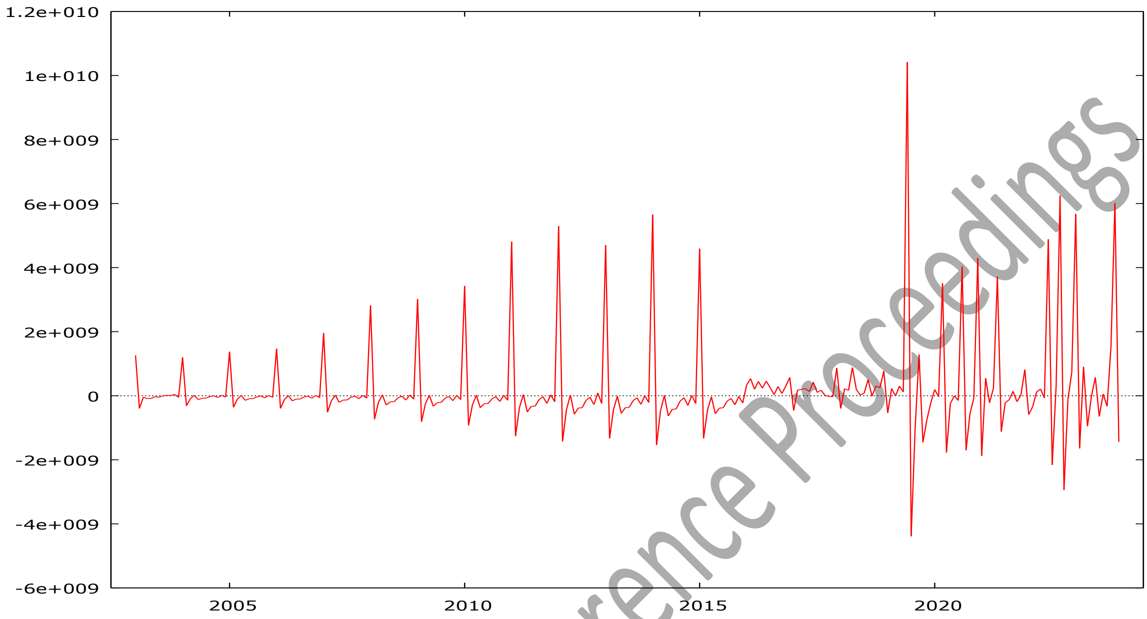

Fig. 2. Time plot of fractional differencing

Figure 2 shows the time plot of the ffractional differencing of the original series. The diagram that plots the series after fractional differencing shows that the variability of the series appears to be stable. The time plot of the series appears to be stationary for both mean and variance suggesting that the time series is stationary.

3.1 Stationarity Test of fractional differencing

Table 1: ADF and KPSS tests of the data

		Unit Root of the Original Data		Unit Root of the Fractional Differencing	
Test	Lag Order	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

From Table 1, the KPSS test result showed that the data was not stationary before fractional differencing since the p-values corresponding to the KPSS test are less the level of significance 5%, implying that the null hypothesis (H_0) was rejected. However, the data was stationary at fractional differencing since the p-values are greater than the level of significance 5%, implying that we fail to reject H_0. Similarly, from table 1, the ADF test result revealed that there was presence of unit root in the data before fractional differencing since the p-values are greater than the 5% level of significance. But there was no presence of unit root in the data after fractional differencing since the p-values are less than 5% level of significance.

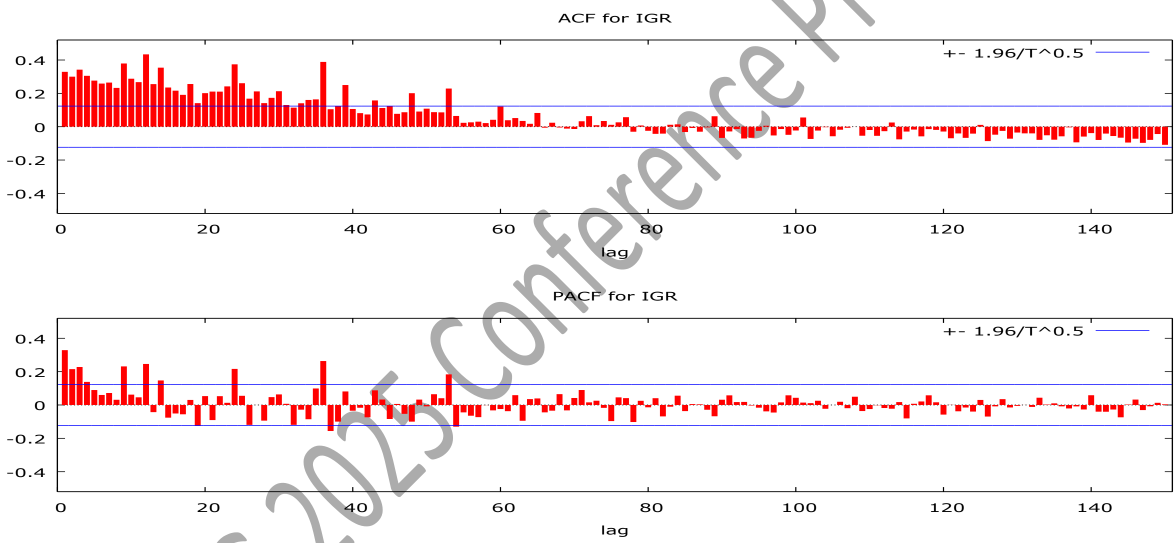

Fig. 3. ACF and PACF Plot of the IGR Series

Fig 4.3 shows that the sample ACF and PACF of 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.

3.2 Long Memory Parameter Estimation

Table 2: Long memory estimate of IGR series

Test	Estimate (d)	Z statistic	p-value
Geweke and Porter-Hundlak (GPH)	0.42363579	4.06361	0.0004

The GPH estimated d parameter to be 0.4236. GPH provides fractional difference parameter values which lies within the conventional long memory parameter.

3.3 ARFIMA Model Identification

In this section, the best model among the candidate of ARFIMA models for the IGR data will be selected. The fraction difference parameter estimated (d=0.4236) is chosen because the estimate lies within the conventional long memory parameter ($-0.5 < d < 0.5$).

Table 3: Result of ARFIMA model identification and selection

MODEL	AIC	BIC	HQC
ARFIMA(0,0.4236,0)	11356.81	11363.87	11359.65
ARFIMA(0, 0.4236,1)	11328.87	11359.46	11333.13
ARFIMA(0, 0.4236,2)	11328.94	11353.06	11334.62
ARFIMA(0, 0.4236,3)	11330.97	11358.62	11338.07
ARFIMA(0, 0.4236,4)	11332.06	11353.24	11340.58
ARFIMA(1, 0.4236,0)	11336.63	11357.22	11340.89
ARFIMA(1, 0.4236,1)	11328.83	11352.94	11334.51
ARFIMA(1, 0.4236,2)	11330.91	11358.56	11338.01
ARFIMA(1, 0.4236,3)	11332.77	11353.94	11341.29
ARFIMA(1, 0.4236,4)	11333.39	11358.09	11343.33
ARFIMA(2, 0.4236,0)	11330.07	11354.19	11335.75
ARFIMA(2, 0.4236,1)	11330.78	11358.42	11337.88
ARFIMA(2, 0.4236,2)	11332.62	11353.80	11341.14
ARFIMA(2, 0.4236,3)	11334.06	11358.77	11344.01
ARFIMA(2, 0.4236,4)+	**11317.04**	**11345.27**	**11328.40**
ARFIMA(3, 0.4236,0)	11331.61	11359.26	11338.71
ARFIMA(3, 0.4236,1)	11331.91	11353.08	11340.43
ARFIMA(3, 0.4236,2)	11333.88	11358.59	11343.82
ARFIMA(3, 0.4236,3)	11335.96	11364.19	11347.32
ARFIMA(3, 0.4236,4)+	**11318.35**	**11350.11**	**11331.13**

Table 3 displayed the results of ARFIMA model selection. Twenty (20) models were tested based on the information selection criteria, and thus two models were selected for further examination namely ARIMA(2,0.4236,4) and ARFIMA (3,0.4236,4) models since they have the minimum values of AIC, BIC and HQC.

3.4 Model Estimation

Table 4: Result of ARFIMA (2,0.4236,4) Model Estimation for the IGR Series

Parameter	Coefficient	Std. Error	Z-Statistic	P-Value
Constant	2.03416e+08	4.11725e+07	4.941	7.79e-07 ***
phi_1	1.64370	0.0262905	62.52	0.0000 ***
phi_2	−0.953044	0.0242775	−39.26	0.0000 ***
theta_1	−2.07771	0.0699511	−29.70	7.17e-194 ***
theta_2	1.58904	0.156601	10.15	3.41e-024 ***
theta_3	−0.217912	0.152161	−1.432	0.1521
theta_4	−0.141387	0.0645650	−2.190	0.0285 **

Table 4 shows that the parameters (phi-1, phi-2, theta_1, theta_2, and theta_4) are statistically significant to the model at 5% level of significance. However, the parameter theta_3 is statistically insignificant to the model. Thus, the ARFIMA (2,0.4236,4) model is fitted as;

$$(1-L)^{0.4236}\left(Y_t + 1.6437Y_{t-1} - 0.9530Y_{t-2}\right) = \varepsilon_t + 2.0777\varepsilon_{t-1} + 1.5890\varepsilon_{t-2} - 0.1414\varepsilon_{t-4} \quad (4.3)$$

Table 5: Result of ARIMA (3,0.4236,4) Model Estimation for the IGR Series

Parameter	Coefficient	Std. Error	Z-Statistic	P-Value
Constant	2.05908e+08	4.10980e+07	5.010	5.44e-07 ***
phi_1	1.11887	0.302002	3.705	0.0002 ***
phi_2	0.111575	0.525496	0.2123	0.8319
phi_3	−0.614768	0.295654	−2.079	0.0376 **
theta_1	−1.55943	0.294357	−5.298	1.17e-07 ***
theta_2	0.264395	0.624407	0.423634	0.6720
theta_3	0.801111	0.476940	1.680	0.0930 *
theta_4	−0.316371	0.116046	−2.726	0.0064 ***

Table 5 revealed that the parameters (phi-1, phi-3, theta_1, theta_3, and theta_4) are statistically significant to the model at 5% level of significance. However, the parameters (phi-2, theta_2, and theta_3) are statistically insignificant to the model at 5% level of significance. Thus, the ARFIMA (3,0.4236,4) model is fitted as;

$$(1-L)^{0.4236}\left(Y_t + 1.1189Y_{t-1} - 0.6148Y_{t-3}\right) = \varepsilon_t - 1.5594\varepsilon_{t-1} - 0.316\varepsilon_{t-4} \quad (4.4)$$

3.4 Model Checking

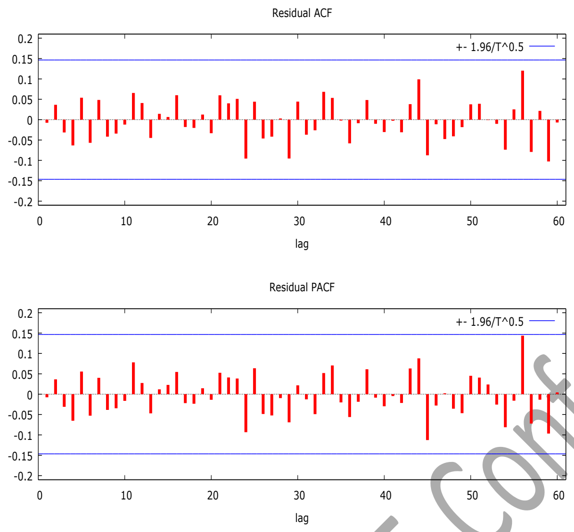

Fig. 4. ACF and PACF of the residual of ARFIMA (2,0.4236,4)

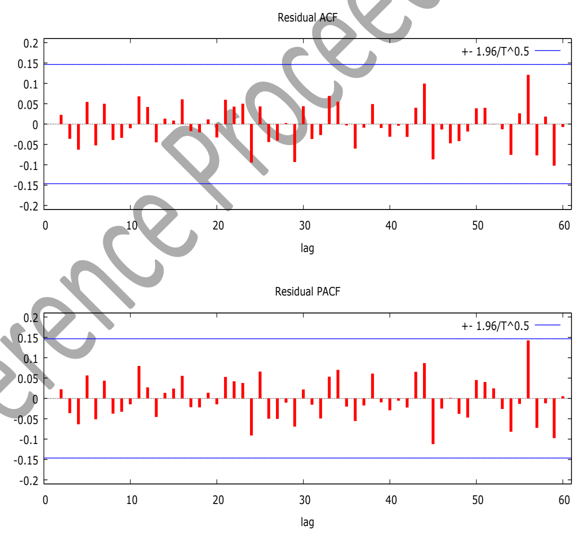

Fig. 5. ACF and PACF of the residual of ARIMA (3, 0.4236,4)

Fig 4 and Fig 5 confirmed that there is no form of correlation amongst the residuals, this means that the ARFIMA (2,0.4236,4) and ARFIMA (3,0.4236,4) models have passed the standard test criteria of being white noise, since the residuals are uncorrelated and stationary.

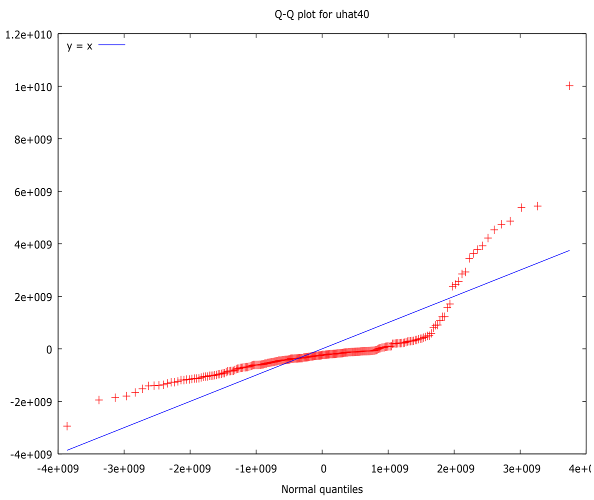

Fig. 6. Normal probability plot from ARFIMA (2,0.4236,4) model

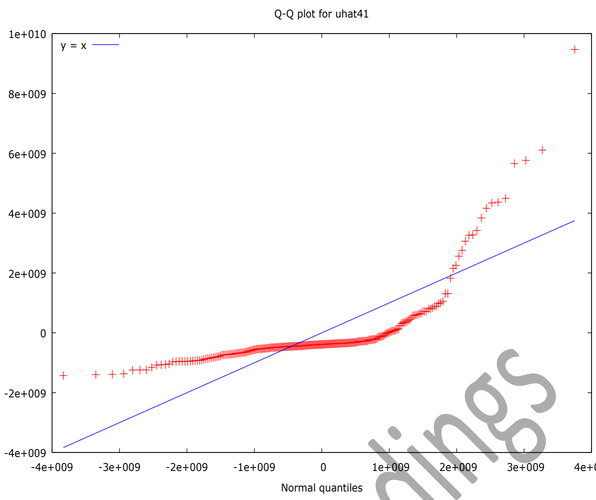

Fig. 7. Normal probability plot from ARFIMA (3, 0.4236,4) model

From Figures 6 and 7 above, it is observed that the relationship between the theoretical percentiles and the sample percentiles is approximately linear. Therefore, the Normal Probability Plot of the residuals of the data using the two models suggests that the error terms are indeed normally distributed.

3.5 ARFIMA Model Forecasting

Table 4.14: Forecasting Results using selected ARFIMA Models

Year	**ARFIMA(2,0.4236,4) model**			**ARFIMA(3,0.4236,4) model**		
	Prediction	Lower Bound	Upper Bound	Prediction	Lower Bound	Upper Bound
2024:1	1.81203e+008	-2.74627e+009	2.38386e+009	9.87875e+008	-3.54771e+009	1.57196e+009
2024:2	6.72894e+008	-2.12333e+009	3.46912e+009	2.16772e+008	-3.01402e+009	2.58047e+009
2024:3	4.04755e+008	-2.39851e+009	3.20802e+009	4.69706e+008	-3.28293e+009	2.34352e+009
2024:4	6.37891e+006	-2.80239e+009	2.81515e+009	1.43344e+008	-2.95661e+009	2.66992e+009
2024:5	3.12339e+008	-3.12348e+009	2.49880e+009	1.02125e+008	-2.81675e+009	2.81596e+009
2024:6	4.56545e+008	-3.26777e+009	2.35468e+009	3.51464e+008	-2.46515e+009	3.16808e+009
2024:7	3.89825e+008	-3.20201e+009	2.42236e+009	5.60457e+008	-2.25622e+009	3.37713e+009
2024:8	1.42723e+008	-2.95849e+009	2.67305e+009	7.45670e+008	-2.07299e+009	3.56433e+009
2024:9	1.99852e+008	-2.62065e+009	3.02036e+009	7.59908e+008	-2.06041e+009	3.58023e+009
2024:10	5.27443e+008	-2.29620e+009	3.35109e+009	6.68021e+008	-2.15440e+009	3.49044e+009
2024:11	7.39416e+008	-2.08494e+009	3.56377e+009	4.52938e+008	-2.37029e+009	3.27617e+009
2024:12	7.75628e+008	-2.04882e+009	3.60007e+009	1.93281e+008	-2.63009e+009	3.01665e+009

2025:1	6.33129e+008	-2.19302e+009	3.45928e+009	6.47497e+007	-2.88826e+009	2.75876e+009
2025:2	3.64392e+008	-2.46519e+009	3.19397e+009	2.50198e+008	-3.07456e+009	2.57416e+009
2025:3	5.84769e+007	-2.77434e+009	2.89129e+009	3.26852e+008	-3.15292e+009	2.49921e+009
2025:4	1.88237e+008	-3.02248e+009	2.64600e+009	2.74681e+008	-3.10253e+009	2.55317e+009
2025:5	3.02212e+008	-3.13651e+009	2.53209e+009	1.10853e+008	-2.93986e+009	2.71815e+009
2025:6	2.54421e+008	-3.08926e+009	2.58041e+009	1.25396e+008	-2.70391e+009	2.95470e+009
2025:7	6.72463e+007	-2.90414e+009	2.76965e+009	3.75933e+008	-2.45338e+009	3.20525e+009
2025:8	1.94867e+008	-2.64480e+009	3.03454e+009	5.81896e+008	-2.24791e+009	3.41170e+009
2025:9	4.47318e+008	-2.39422e+009	3.28885e+009	6.95057e+008	-2.13602e+009	3.52613e+009
2025:10	6.12464e+008	-2.22951e+009	3.45444e+009	6.90627e+008	-2.14210e+009	3.52335e+009
2025:11	6.43320e+008	-2.19870e+009	3.48534e+009	5.71678e+008	-2.26233e+009	3.40569e+009
2025:12	5.36645e+008	2.30634e+009	3.37963e+009	3.68528e+008	-2.46600e+009	3.20305e+009

Table 4.14 displayed the forecast values of Kaduna State IGR series using ARFIMA(2,0.4236,4) and ARFIMA(3,0.4236,4) models. The forecasts values along with 95% upper and lower bound of the IGR series was generated for the next two years or 24 months, starting January 2024 until December 2025.

Table 6: Forecast Evaluation Statistics

	MSE	RMSE	MAE	MAPE
ARFIMA (2,0.423636,4)	1.734226e+18	1.3169e+009	7.745e+008	1129.1
ARFIMA (3,0.423636,4)	1.727647e+18	1.3144e+009	7.6082e+008	1077.2

The smaller the value of the error, the better the forecasting performance of the model. From Table 6, it could be seen that both the two selected models have shown good result (minimum predictive measures). But forecast result of ARFIMA (3,0.423636,4) model is closer to the actual series. Therefore, the prediction power of ARFIMA (3,0.423636,4) model is better and suitable for monthly periods forecasting, as such the model best fit the data.

4.0 Conclusion and Policy Implications

The findings demonstrate ARFIMA's effectiveness in forecasting IGR, highlighting its potential for improving fiscal management in Kaduna State. 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

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

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1), 5-59.

Chatfield, C. (1984). The analysis of time series. Chapman and Hall Publisher.

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

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.

Gimba, V. K., Ahmed, M. & Musa, A. O. (2018). Analysis of impact of personal income tax on internally generated revenue in Kaduna State, Nigeria. *KASU Journal of Management Sciences*, 9(1), 102-115.

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.

Hamilton, J.D. (1983). Oil and Macro Economy since World War II. *The Journal of Political Economy*, 91: 228-248.

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.

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

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.

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

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.

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.

Waniyos, U. H., Bakari, M. A., Buba, M. S. & Adams, J. Y. (2020). Empirical Analysis of Internal Generated Revenue in Adamawa State Nigeria. *Confluence Journal of Economics and Allied Sciences*, 3(2), 46-57.