Forecasting of bayelsa state internally generated revenue using ARIMA model and winter methods

TI Ayakeme, OE Biu, D Enegesele and N Wonu

Abstract
The study examined time series models for forecast comparison between ARIMA model and Winter methods (Additive and Multiplicative methods) of Bayelsa State Government Internally Generated Revenue (IGR) data frame of 2012 to 2018. The data used for the study were analyzed using Statistical software packages: MINITAB 18 and Micro software Excel. For a better understanding of the behavior of the data, monthly and yearly plots were done. The plot shows seasonality of order 4 with a regular first difference. Several ARMA (p, q) Models were fitted to the internally generated revenue with respective residuals as having white noise. The identified ARIMA model was ARIMA (0, 1, 1) (0, 0, 0)\(\times\) (1.1524X\(_t\) -0.1524X\(_{t-2}\) + 0.7846\(e_{t+1}\) +\(e_t\)) which has the least values of AIC and BIC amongst the fitted models. The identified ARIMA model was used to predict for 2019 to 2021. Furthermore, winters (additive and multiplicative) methods were employed in the study to model and forecast the internally generated revenue. The winters method models forecast values were compared with the identified ARIMA model forecasts using the Mean Absolute Error (MAE) and Mean square error (MSE). The ARIMA model forecast is better than winter methods: additive and multiplicative methods and was thus recommended for use.

Keywords: Model selection criteria, seasonality, ARIMA model, winters methods, additive and multiplicative model

1. Introduction
A time series is a stochastic process that is an ordered sequence of observation (a collection of observations made sequentially in time). Time series is described as a sequence of observation order by a time parameter continuously or discretely [10]. The main objectives in analyzing time series data is to interpret, understand and compute the change in the economic phenomenon in the possibility of more correctly anticipating the course of future events.

In Nigeria, there are 36 states in which Bayelsa is inclusive. At the state levels, it is saddled with the responsibilities of protecting lives and properties. This does not exclude the provision of social amenities that will make life comfortable for its citizens. Bayelsa state population is put at 1, 704, 515 (One Million, Seven Hundred and Four Thousand Five Hundred and Fifteen) as at 2006 census [7]. And with an estimated increase in population by the United Nations simply means more demand for social amenities and provision of goods and services. This underscores the importance of the need for funds to implement government policies and programmes to make life comfortable for its citizenry. Government, enterprises, and even educational institutions are therefore encouraged to augment their finances by generating revenue internally. The uncertainty in Federal Allocation Account Committee (FAAC) and the unstable oil price at the international market and with the uncontrollable increase in population, there is need for urgency in aggressive drive and improvement in IGR. The forecasting and controlling such IGR could help in the management and governance to provide basic amenities that will better the lives of its citizenry. These amenities, goods, and services require finances to provide them to have a society devoid of crime. The forecasting and control of such IGR could help in planning and providing for such amenities.

The study aims to determine suitable time series model forecasts for Bayelsa state Internally Generated Revenue (IGR) in Nigeria.
The specific objectives include
1. To obtain the series trend, yearly and monthly mean plot, and then determine the stationarity of the series.
2. To fit a suitable time series model for the Bayelsa State IGR in terms of ARIMA model and winters method.
3. To forecasts the IGR to cover from 2019 to 2021 using the identified ARIMA & winter method.
4. Interpret the forecast values for 2019-2021 years and compare the models’ forecast (ARIMA and winters method).

The rationale for this paper is that it will help policymakers to take necessary steps to increase stable state revenue for management and governance to better the lives of its citizenry. The study is restricted to Bayelsa State IGR from 2012 to 2018. This study is divided into five sections. Section 1 contains the introduction to time series forecasting, aim and objectives of this paper, and rationale for the study. Section II contains the literature on related work, Section III contains the methodology, Section IV described results and discussion and Section V concludes the research work with future direction.

2. Literature review
Revenue as defined by [1] is the source of funds expected and needed by the government to finance its activities. These funds are generated through various mediums of which one is Internally Generated Revenue. Internally Generated Revenue (IGR) is viewed as money made by federal, state, and local government within their area of operation. The IGR for state government are money generated by the state from various means like taxes (capital gain taxes, pay as you earn, direct assessment, etc.) and motor vehicle licenses, among others.

An ARIMA model was developed by [4] was used in predicting the next day electricity price in the mainland span and California market. Their developed model was able to predict the 24 market clearing prices of tomorrow. Forecasting exchange rate between Ghana cedis and the US Dollar and forecast future rate using time series analysis was carried out by [5]; They use ARIMA Model was able showed that the predicted rates were consistent with the depreciating trend of the actual series. ARIMA (1,1,1) model was found as the most suitable model with the least normalized Bayesian information criterion (BIC) of 9.111, Mean Absolute Percentage Error (MAPE) of 0.915, Root Mean Square Error of 93.873, and high value of R-square of 1000. A forecast for two years from Jan 2011 to Dec. 2012 computed shows a depreciation of the Ghana cedi against the US Dollar [13, 6], use univariate time series models to forecast Thailand’s export and major trade partners. Their result showed that indirect forecasts do not outperform the direct forecast in all trade partners considered using ARIMA model [11]. Research in time series modeling of internally Generated Revenue of some local governments in Nigeria, use monthly tax data of 1999-2012. Their finding shows that ARIMA (2, 1, 2) is the suitable model for the series and The parameter coefficients indicate that the model is adequate for a good forecast. A study on time series analysis of monthly generated revenue in Gombe local government, Nigeria was carried out by [10]. They adopted secondary data collected from the Inland Revenue department from 2006-2016. They use ordinary least square and ARMA models and the result shows a steady increase in the revenue generated over the years and slight instability in the deseasonalized data [8]. Investigated forecasting of government revenue for Nepal using seasonal ARIMA and Winter Method and it was found that winter is more appropriate for forecasting government revenues [13]. Compared autoregressive moving average (ARMA), ARIMA and the autoregressive artificial neural (ARAN) models in forecasting the monthly inflow of the Dez reservoir. The inflow to the reservoir shows that the ARIMA model results match the forecast compared with the ARMA model. This paper used time series (IGR) data from 2012-2018 collected from the Bayelsa State Board of Internal Revenue after a formed request was made and its subsequent approval. A statistical software called MINITAB 18 was used to carry out the analysis.

3. Methodology
Autoregressive model [AR (P)]: The general model of an AR is given by:

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + e_t = \sum_{i=1}^{p} \phi_i X_{t-i} + e_t \]  \hspace{1cm} (1)

where : \( X_t \) is Monthly internally generated revenue, \( \phi_p \) is the autoregressive model parameters, \( X_{t-i} \) is prior observation, and \( e_t \) is the purely random process.

If \( p = 1 \):

\[ X_t = \phi_1 X_{t-1} + e_t \]  \hspace{1cm} (2)

Equation (2) is termed autoregressive model of order one [AR (1)]

If \( p = 2 \):

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t \]  \hspace{1cm} (3)

Equation (3) is called the Autoregressive model of order for [AR (2)]

An evaluation of the auto-correlation and partial autocorrelation functions would make it easier or possible in determining fitted to the data.

Moving average model [MA (q)]: The general model of a moving average with order q is given by:

\[ X_t = e_t + \sum_{i=1}^{q} \theta_i e_{t-i} \]  \hspace{1cm} (4)

\[ \theta_1, \theta_2, \ldots, \theta_q \] are the moving average model parameters.
\[ X_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} = \sum_{i=0}^{q} \theta_i e_{t-i} ; \quad \theta_0 = 1 \]  

\textbf{where} \quad X_t = \text{the series}, \quad \theta_i = \text{Moving Average Parameters} \quad \text{and} \quad e_{t-i} = \text{Prior Random shocks}

If \( q = 1 \):
\[ X_t = e_t + \theta_1 e_{t-1} \]  

Equation (3.5) is termed a Moving Average Model of order one \([\text{MA} (1)]\)

If \( q = 2 \):
\[ X_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]  

Equation (6) is called a Moving Average model of order two \([\text{MA} (2)]\)

\textbf{Autoregressive integrated moving average model (ARIMA):} The general model for ARIMA \((p, d, q)\) is given by:
\[ \phi(B)(1-B)^d X_t = \theta(B)e_t \]  

\textbf{where} \quad p = \text{Non-seasoned AR order}, \quad d = \text{Non-seasoned differencing} \quad \text{and} \quad q = \text{monthly AR order} \quad \text{and} \quad B \text{ is the shift operator} \quad \text{and} \quad d \text{ is the } d^{th} \text{ difference.}

\textbf{Model Identification}

Models are identified through the study of the behavior of auto-correlation and partial autocorrelation function.

\begin{table} 
\centering 
\begin{tabular}{|c|c|c|c|} 
\hline 
\textbf{Autocorrelation} & \textbf{AR process infinite (damped exponentials and or/damped waves). Tail off} & \textbf{MA process finite. Cuts off} & \textbf{ARMA process infinite (damped exponential and/or damped sine wave after the first q-p lags). Tails off} \\
\hline 
\textbf{Partial autocorrelation function (PACF)} & Finite cute off & Infinite demand exponential and for damped sine wave tails off & Infinite damped exponential and/or damped sine waves after the first p-q lags tails off \\
\hline 
\textbf{Reference} [3] & & & \\
\hline 
\end{tabular} 
\caption{Identification of models} 
\end{table} 

\textbf{The backward shift operator:} The backward shift operator, \( B \) is defined by;
\[ B X_t = X_{t-1} \]  

Hence
\[ B^m X_t = X_{t-m} \]  

\textbf{Forward Shift Operator, } \( F \): The operator is carry out by forward shift operation \( F = B^{-1} \) and is given by;
\[ F X_t = X_{t+1} \]  

Hence
\[ F^m X_t = X_{t+m} \]  

\textbf{Backward Difference operator:} Another important operator that is important to time series work is the backward difference operator, which can be written in terms of \( B \);
\[ \nabla X_t = (1-B)X_t = X_t - X_{t-1} \]  

Hence
\[ \nabla^d X_t = (1-B)^d X_t \]
Non-stationary Series and Models

Consider the ARIMA model given below;

\[
\phi(B)X_t = \theta(B)e_t
\]

where \( \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \) and \( \theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \) \hspace{1cm} (14)

To ensure stationarity the roots \( \phi(B) = 0 \) must lie outside the unit circle and exhibit explosive non-stationarity if the roots lie inside the unit circle.

Criteria for model selection

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) model selection was considered in the paper [5]. Both formula are defined as;

Akaike Information Criterion (AIC)

\[
\text{AIC} = n \times \left( \frac{RSS}{n} \right) + 2k
\]

Bayesian Information Criterion (BIC)

\[
\text{BIC} = n \ln \left( \frac{RSS}{n} \right) + k \ln (n)
\]

where \( n \) = the number of data points in \( X_t \) (or the number of observations). \( k \) = the number of free parameters to be obtained.

RSS is the residual sum of squares or \( \sigma^2 \) is the error variance of the model which is a biased estimator for the true variance. In term of the residual sum of squares (RSS); \( ln(n) \) is the natural logarithm.

Winters’ method

Winters’ method is of two types: additive and multiplicative model, where the multiplicative model is used when a seasonal component, and possibly trend is apparent in the data. However, winters’ method smoothes the data using the Holt-Winters’ exponential smoothing and provides short to medium-range forecasting. Furthermore, winters’ method calculates dynamic estimates for three components: level, trend, and seasonal. The two competing models are multiplicative and additive.

Multiplicative

The multiplicative model is chosen when the seasonal pattern in the data depends on the size of the data. In other words, the magnitude of the seasonal pattern increases as the series goes up, and decreases as the series goes down. The multiplicative model equation is:

\[
\hat{Y}_t = (L_{t-1} + T_{t-1}) S_{t-p}
\]

where

\[
T_t = \gamma [L_{t-1} - L_{t-1}] + (1 - \gamma) T_{t-1}
\]

\[
S_t = \alpha \left( \frac{Y_t - L_{t-1}}{S_{t-p}} \right) + (1 - \alpha) S_{t-p}
\]

\[
L_t = \beta \left( \frac{Y_t - S_{t-p}}{L_{t-1}} \right) + (1 - \beta) [L_{t-1} + T_{t-1}]
\]

\( L_t \) is the level at time \( t \), \( \alpha \) is the weight for the level. \( T_t \) is the trend at time \( t \), \( \gamma \) is the weight for the trend. \( S_t \) is the seasonal component at time \( t \), \( \beta \) is the weight for the seasonal component. \( p \) the seasonal period. \( Y_t \) is the data value at time \( t \). \( \hat{Y}_t \) is the fitted value, or one-period ahead forecast at time \( t \).

Additive

The additive model is chosen when the seasonal pattern in the data does not depend on the size of the data. In other words, the magnitude of the seasonal pattern does not change as the series goes up or down. Also, the Additive model is given by:

\[
\hat{Y}_t = L_{t-1} + T_{t-1} + S_{t-p}
\]

where

\[
T_t = [L_{t-1} - L_{t-1}] + (1 - \gamma) T_{t-1}
\]

\[
S_t = \alpha \left( \frac{Y_t - L_{t-1}}{S_{t-p}} \right) + (1 - \alpha) S_{t-p}
\]

\[
L_t = \beta \left( \frac{Y_t - S_{t-p}}{L_{t-1}} \right) + (1 - \beta) [L_{t-1} + T_{t-1}]
\]
Forecasting
Winters’ methods use the trend and seasonal component to generate a forecast.

Forecast accuracy measures (FAM) or forecasting errors
Since there are always errors in forecasting, to measure that error, and to ensure the utmost accuracy in forecasting, two techniques were applied, Mean Absolute Error (MAE), and Mean square error (MSE).

4. Results and Discussion
This paper uses the ARIMA model and winter methods (Additive and Multiplicative methods) to analyze the internally generated revenue (2012-2018) data of Bayelsa State, Nigeria. The data was obtained from the Board of Internal revenue of the Bayelsa State Government of Nigeria (Appendix). The series plot, monthly, and yearly mean plots were done to determine the behavior of the series.

4.1 Plots and ARIMA model Identification

Figure 1 shows an upward trend, which indicates that the average internally generated revenue is increasing as the years’ increase. The yearly mean plot increase from 2012 to 2014, then decrease from 2014 to 2016, with a further increase from 2016 to 2018; which is the peak. Figure 2 shows that the monthly mean plot has a constant trend with quarterly decrease and increases after four months period; which suggests a seasonality of order 4.

Figure 3 showed three trend curves (linear, quadratic, and cubic), which indicated that the yearly mean of the internally generated revenue of Bayelsa state has a cubic trend as the years went since it has the highest $R^2$ value of 82.5%.
The ACF plot in Figure 5 shows two spikes cut off, suggesting MA(q) process, and Figure 4.7 also shows two spikes cut off for PACF plot, suggesting AR(p) process.

![Time Series Plot of 1st Diff IGR](image)

**Fig 4:** First Difference Series Plot of Bayelsa state IGR from 2012-2018

Figure 4 shows that the series (IGR) is stationary after the first difference. The ACF plot in Figure 5 shows two spikes cut off, suggesting MA(q) process, and Figure 4.7 also shows two spikes cut off for PACF plot, suggesting AR(p) process.

![Partial Autocorrelation Function for IGR](image)

**Fig 5:** ACF plot of Bayelsa state IGR

![Partial Autocorrelation Function for IGR](image)

**Fig 6:** PACF plot of Bayelsa state IGR

However, several ARMA (p, q) model as shown in Table 2 was fitted to the internally generated revenue of Bayelsa State and the AIC and BIC were used to select the optimum model that fit the series.

**Table 2:** ARMA Models and its Parameter Estimates and Model selection Criteria

<table>
<thead>
<tr>
<th>MODELS</th>
<th>AR</th>
<th>MA</th>
<th>SAR</th>
<th>SMA</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,1,0)(0,0,0)4</td>
<td>-0.5426</td>
<td></td>
<td></td>
<td></td>
<td>2926.48</td>
<td>3301.0</td>
</tr>
<tr>
<td>ARIMA(2,1,0)(0,0,0)4</td>
<td>-0.3040</td>
<td></td>
<td></td>
<td></td>
<td>2921.545</td>
<td>3298.596</td>
</tr>
<tr>
<td>ARIMA(0,1,0)(1,0,0)4</td>
<td></td>
<td>0.1006 (0.369)</td>
<td></td>
<td></td>
<td>2960.296</td>
<td>3342.208</td>
</tr>
<tr>
<td>ARIMA(0,1,0)(0,0,1)4</td>
<td></td>
<td>-0.0842 (0.454)</td>
<td></td>
<td></td>
<td>2960.431</td>
<td>3342.343</td>
</tr>
<tr>
<td>ARIMA(0,1,1)(0,0,0)4</td>
<td>0.7578 (0.000)</td>
<td></td>
<td></td>
<td></td>
<td>2909.02</td>
<td>3283.639</td>
</tr>
<tr>
<td>ARIMA(0,1,2)(0,0,0)4</td>
<td></td>
<td>-0.1588 (0.157)</td>
<td></td>
<td></td>
<td>2910.202</td>
<td>3287.252</td>
</tr>
<tr>
<td>ARIMA(0,1,1)(1,0,0)4</td>
<td>0.7846 (0.000)</td>
<td>0.1524 (0.195)</td>
<td></td>
<td></td>
<td>2907.02</td>
<td>3279.639</td>
</tr>
</tbody>
</table>

**Footnote:** The bold model has the least AIC and BIC.

The suitable model that predicts the internally generated revenue of Bayelsa is ARIMA (0, 1, 1) (1, 0, 0)4. Hence, the mathematical expression is given by:

\[
(1 - \phi_{1,4} B) \nabla X_t = \theta_{1} e_{t-1} + e_t
\]

(25)
\[
(1 - \phi_4 B)(1 - B)X_t = \theta e_{t-1} + e_t
\]
\[
(1 - B - \phi_4 B + \phi_4 B^2)X_t = \theta e_{t-1} + e_t
\]
\[
X_t = (1 + \phi_4)X_{t-1} + \phi_4X_{t-2} = \theta e_{t-1} + e_t
\]

By substitution
\[
X_t - 1.1524X_{t-1} + 0.1524X_{t-2} = 0.7846e_{t-1} + e_t
\]
or
\[
X_t = 1.1524X_{t-1} - 0.1524X_{t-2} + 0.7846e_{t-1} + e_t
\]

4.2 Winter’s Methods (Multiplicative and Additive)

Figure 7 and 8 shows the multiplicative and additive models of the Winters Method along with their respective forecast.

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>ARIMA (0,1,1) (1,0,0) Model Forecasts</th>
<th>Lower</th>
<th>Upper</th>
<th>Additive Method Forecasts</th>
<th>Lower</th>
<th>Upper</th>
<th>Multiplicative Method Forecasts</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>Jan</td>
<td>1,123,418,868.00</td>
<td>537,708,006.00</td>
<td>709,129,730.00</td>
<td>1,352,653,947.00</td>
<td>1,340,280,822.00</td>
<td>1,946,941,522.00</td>
<td>1,944,546,622.00</td>
<td>1,709,129,730.00</td>
<td>1,709,129,730.00</td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td>1,134,295,344.00</td>
<td>555,155,797.00</td>
<td>733,434,891.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td>1,182,983,630.00</td>
<td>570,679,852.00</td>
<td>795,227,409.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
<td>1,199,497,733.00</td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>1,108,477,750.00</td>
<td>456,771,174.00</td>
<td>1,733,609,865.00</td>
<td>1,104,287,137.00</td>
<td>665,645,116.00</td>
<td>1,733,609,865.00</td>
<td>1,733,609,865.00</td>
<td>1,733,609,865.00</td>
<td>1,733,609,865.00</td>
</tr>
<tr>
<td></td>
<td>May</td>
<td>1,132,096,602.00</td>
<td>670,903,172.00</td>
<td>1,833,434,352.00</td>
<td>1,350,233,794.00</td>
<td>1,350,233,794.00</td>
<td>1,350,233,794.00</td>
<td>1,350,233,794.00</td>
<td>1,350,233,794.00</td>
<td>1,350,233,794.00</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>1,133,753,701.00</td>
<td>856,771,174.00</td>
<td>1,810,736,228.00</td>
<td>1,116,592,038.00</td>
<td>648,206,099.00</td>
<td>1,764,977,976.00</td>
<td>1,764,977,976.00</td>
<td>1,764,977,976.00</td>
<td>1,764,977,976.00</td>
</tr>
</tbody>
</table>

Figure 7: Winter Method Plot for Multiplicative of Bayelsa state IGR

Fig 8: Winter Method Plot for Additive of Bayelsa state IGR
Appendix A

Below is the data used for this study work

<table>
<thead>
<tr>
<th>Year</th>
<th>Data</th>
<th>ARIMA(0,1,1)</th>
<th>Additive Method</th>
<th>Multiplicative Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>2013</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

Fig 9: Forecasts Comparisons Plot for both ARIMA and Winters Models

<table>
<thead>
<tr>
<th>Year</th>
<th>Data</th>
<th>ARIMA(0,1,1)</th>
<th>Additive Method</th>
<th>Multiplicative Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>2013</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

*Figures and data are placeholders for demonstration purposes.*
4.3 Models Forecasts comparisons

Figure 9 shows the model forecast comparison for the three models considered in the study. Using forecast accuracy measure, the ARIMA model was the most stable and it has the least value of the Mean Square Error (MSE) accuracy measure. ARIMA model shows a steady trend than the winters additive and multiplicative methods.

5. Conclusion

In this study, suitable IGR data Improvement forecasts and model has been identified. This result of monthly and yearly means plots shows an upward trend, indicating that the average internally generated revenue is increasing as the years went. However, the plot of the monthly average has a constant trend with quarterly decrease and increases after four months period, which suggests a seasonality of order 4. Therefore, a suitable ARIMA model was identified and was used to forecasts. The winters additive and multiplicative methods were employed in the study; the additive and multiplicative models forecasts were obtained with their respective confidence limit (upper and lower). The obtained winters additive and multiplicative forecast values were compared with the identified ARIMA model forecasts using forecast accuracy measures. The ARIMA model forecasts were better than winter methods: additive and multiplicative methods forecast for the years 2019-2021 and was thus adopted in the study.
6. References


7. Nigeria population in 2018 Source; [CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0] Retrieve on the 10/10/2019.


