ESTIMASI: Journal of Statistics and Its Application
Vol. 6, No. 1, Januari, 2025, Hal. 116-127
DOI: ….

Nigerian Population Growth Modelling and Forecasting using Univariate Time series model

Abass Ishola Taiwo[1], A A Titilola[2], Timothy Olabisi Olatayo[3], Taiwo Abideen Lasisi[4]
[123]Department of Mathematical Sciences, Faculty of Science, Olabisi Onabanjo University, Ago-Iwoye, Nigeria
[4]Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
*** Corresponding author, email: taiwo.abass@oouagoiwoye.edu.ng**

Abstract

The Nigerian population is growing rapidly, and this poses significant challenge to human existence. To have an insight, Autoregressive Integrated Moving Average (ARIMA) model was used to model and predict population growth. The results showed a yearly population mean of 99,611,692 with a standard deviation of 53,188,740. After estimation, the ARIMA(3,2,1) model was chosen for having lowest Akaike and Schwarz information criterion with the adequacy of the model attained using the Ljung-Box Statistic Autocorrelation and partial autocorrelation functions of the residuals. Coefficient and adjusted coefficient of determinations showed the model has a strong predictive accuracy, with the forecast indicating a continuous population growth increase of over 5 million annually. Conclusively, Nigerian government must plan how to curtail this explosive growth expected at over 418 million by 2050.

Keywords*: Population growth, Nigeria, Autoregressive integrated moving average model, Forecasting*

1. Introduction

Population growth is still a major concern worldwide and has elevated to the level of a public health concern, especially in nations within the African continent [1]. The population of the globe was estimated to be close to 6.1 billion at the start of the twenty-first century [2]. According to UN estimates, there will be more than 9.2 billion humans by 2050 and up to a maximum of 11 billion by 2200 where more than 90% live in developing nations across the globe [3]. Nigeria is an example with the highest population in Africa and the most populace black nation in the world [4]. Nigeria's population is growing exponentially and is estimated to hit above 350 million by 2050 [5].

Nigeria, with its roughly over 200 million inhabitants, is the eighth most populous country in the world and one of the top ten most populous countries worldwide [6]. Where the exponential growth can be attributed to the high crude birth rate at 37.12 per 1000 population with annual population growth of 2.38 per cent [7]. It is noteworthy that a multitude of variables, including lack of education, beliefs, myths, and traditions, among others, impact the fertility prevalence in Nigeria. Extended families, especially the polygamous family system are encouraged by various faiths. The practice of getting married young is very common and has greatly accelerated the country's population increase [8].

One important issue which unnecessarily strains advancement and prosperity is the quick increase in population. Since population growth has an impact on social and welfare services, this is a threat to human welfare rather than a numerical one [9,10]. The decline of facilities aridity, changing climate, extermination of creatures, lack of a place to live, delays in traffic, pollution, and loss of forests are all negatively impacted by rose population growth [11]. An analytical and critical examination of Nigeria's population growth has resulted in challenges for economic, social, and political actors. It also affects public education, health and welfare, and the standard of living environment for individuals [12,13]. Several studies have been carried out on population growth prediction in Nigerian using different approaches and these include the works of [14-23] and many more. These studies yielded good results, but the validity and predictive ability of their approaches were not determined, and the predictive evaluation metrics were not used to confirm the optimality and adequacy of the predictive values. In essence, this study will be used to model and predict Nigerian yearly population growth using the Autoregressive integrated moving average model, the model validation will be achieved using the coefficient of determination while the Adjusted coefficient of determination will be used to determine the model predictive ability. The optimality and adequacy will be attained using the following predictive evaluation metrics (Mean absolute error, Root mean square error and mean absolute percentage error). It is expected that this study will go far to provide better accurate predictive values that will assist policymakers in controlling the population explosion in Nigeria.

2. Material and Method

2.1. Data Exploration

The variation displayed by the Nigerian population series is anticipated to be identified using the time plot and the series will be described and presented using descriptive statistics.

2.2. Augmented Dickey-Fuller (ADF) test

The stationarity of the series under examination is anticipated to be determined using the Augmented Dickey-Fuller (ADF) test. This will be accomplished by comparing the corresponding standard errors to the relevant values in the Dickey-Fuller table using three sets of models given as

$$\Delta x_t = (\lambda - 1)x_{t-1} + \sum_{j=1}^{j} \beta_j \Delta x_{t-j} + w_t \quad (1)$$

$$\Delta x_t = \alpha + (\lambda - 1)x_{t-1} + \sum_{j=1}^{j} \beta_j \Delta x_{t-j} + w_t \quad (2)$$

$$\Delta x_t = \alpha + \delta_t + (\lambda - 1)x_{t-1} + \sum_{j=1}^{j} \beta_j \Delta x_{t-j} + w_t \quad (3)$$

where equation (1) is a stochastic model, equation (2) has a slope variable and equation (3) combines equations (1) and (2).

2.3. Autoregressive Integrated Moving Average (ARIMA) Model

The ARIMA model is a type of univariate time series model that consists of an autoregressive component, a moving average part, and a differential order of integration (d). Given is the AR(p) and MA(q) components in equations (4) and (5).

$$x_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \cdots + \emptyset_p x_{t-p} + e_t \tag{4}$$

and

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{5}$$

In equations (4) and (5), $\emptyset$ and θ are the autoregressive and moving average coefficients, x_t is a value taken at period t and ε_t is a residual term at period t. The is independently and identically distributed with $N \sim (0,\ \sigma^2)$. An ARMA (p, q) process has AR and MA parts, where the present values of the series are termed approximately to its former values and present and, former residual terms.

The ARMA (p, q) process is defined in equation (6) as

$$x_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \cdots + \emptyset_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{6}$$

By using backward shift operator B on equation (6), this gives

$$(B)\nabla^d x_t = \theta(B) w_t \tag{7}$$

Therefore, equation (7) is an $ARMA(p, d, q)$ and $\nabla^d = (1 - B)^d$ with $\nabla^d y_t$ and d^{th} as the difference level.

2.4. Model Identification

The autocorrelation function (ACF) and partial autocorrelation functions (PACF) are used to determine the order of p and q in equation (7). The Autocorrelations function (ACF) is given as

$$\rho_k = \frac{E[(x_t - \bar{x})(x_{t-k} - \bar{x})]}{E[x_t - \bar{x}]^2} \tag{8}$$

and partial autocorrelation function (PACF) is defined as

$$x_t = \rho_0 + \sum_{k=1}^{K} \rho_{kk}\, x_{t-k} \tag{9}$$

and ρ_{kk} describe the k^{th} autoregressive parameters, $k = 1,2, \dots, K$. The smallest of the values of Akaike Information criteria (AIC) and Schwartz Bayesian Information criteria obtained after the estimation of the identified model will be chosen as the optimal model.

Akaike Information Criteria (AIC) is defined as

$$AIC = -2(\log - \text{likelihood}) + 2K \tag{10}$$

and Schwartz Bayesian Information Criteria (BIC) is expressed as

$$BIC = \ln(n) + 2\ln(L) \tag{11}$$

2.5. Parameter Estimation

The parameters of the ARIMA model are computed based on the ordinary least squares estimation method. This is achieved using

$$\hat{\emptyset} = \sum_{t=2}^{n}(x_{t-1})(x_t) \Big/ \sum_{t=2}^{n} x_{t-1}^2 \tag{12}$$

After the estimation step, the most appropriate model for the series under consideration will be selected using AIC and SBIC.

2.6. Diagnostic Checking

2.6.1 Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

The residual of the estimated model is expected to be checked for stability using the Modified Box-Pierce (Ljung-Box) Chi-Square Statistic denoted as

$$Q(m) = n(n+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j} \tag{13}$$

and n is the data points used when the difference is performed.

2.6.2. Autocorrelation function of Residual

With time series data, the value of a variable observed in the current period will likely be similar to its value in the previous period, or even the period before that, and so on. Therefore, when fitting a regression model to time series data, it is common to find autocorrelation function (ACF) in the residuals. ACF is defined as

$$r_k = \frac{\sum_{t=k+1}^{T}(y_k - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t-1}^{T}(y_t - \bar{y})^2} \tag{14}$$

2.6.3. Coefficient of Determination

The coefficient of determination is a measurement used to explain how much variability in the dependent variable is jointly explained by the independent variables. This is defined as

$$R^2 = 1 - \frac{SSE}{SST} \tag{15}$$

where SSE is the sum of the square of error and SST is the sum of squares of total.

2.6.4. Adjusted Coefficient of Determination

The ARIMA model predictive ability is going to be ascertained using the Adjusted Coefficient of Determination ($\bar{R}^2$). This is defined as:

$$\bar{R}^2 = \frac{1}{n-k}[nR^2 - k] \tag{16}$$

where R^2 is the value of the coefficient of determination, n is the of observations and k is the number of estimated coefficients.

2.7. Forecasting

There are two types of forecasting: within and out sample. The former will be utilized to increase model confidence, and the latter will be utilized to forecast future values.

2.7.1 Forecast Evaluation Metrics

The forecast evaluation metrics used for this work are mean absolute error (MAE) denoted in the forms:

$$MAE = \frac{1}{h+1}\sum_{t=s}^{h+s}(\hat{y}_t - y_t)^2 \tag{17}$$

root mean square forecast error (RMSE) is described in the form

$$RMSE = \sqrt{\frac{1}{h+1}\sum_{t=s}^{h+s}(\hat{y}_t - y_t)^2} \tag{18}$$

and the mean absolute percentage error denoted by

$$MAPE = \frac{100}{h+s}\sum_{t=s}^{h+s}\left|\frac{\hat{y}_t - y_t}{\hat{y}_t}\right| \tag{19}$$

$where\ t = s, 1+s, \dots, h+s$ [24].

3. Results and Discussion

The descriptive analysis of the Nigerian population obtained from the National Population Commission and National Bureau of Statistics websites between 1950 and 2022 signified that the average yearly population is 99,611,692 with a standard deviation of 53,188,740. The population of Nigeria ranges from 37,189,369 in 1950 to 218,541,212 in 2022. The Nigerian population dataset time plot displayed in Figure 1 demonstrated that the country's population is not stationary. This signified that the dataset exhibited a continuous rise over the years, and this is termed a secular variation.

3.1 Augmented Dickey-Fuller test Interpretation

For the Nigerian population dataset, the Augmented Dickey-Fuller test was utilized to achieve stationary. The results in Table 1 demonstrate that the series is stationary at the second level but non-stationary at the first level. I(2) is used to indicate this at levels of significance of 1, 5, and 10%, respectively. Using the ADF test results in Table 1, the Nigerian population is stationary at d = 2.

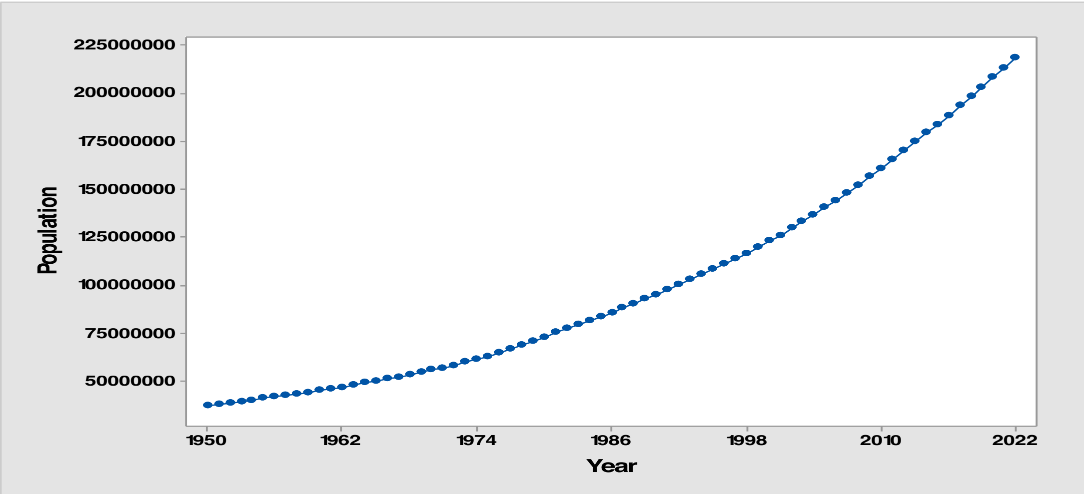

Figure 1. Time plot of the Nigerian population from 1950 to 2022

Table 1. Augmented Dickey-Fuller Unit Root Test result

		Adj. t-Stat	Prob.*
Augmented Dickey fuller test		-5.810369	0.0000
Test critical values:	1% level	-4.094550	
	5% level	-3.475305	
	10% level	-3.165046	

3.2. Identification for the ARIMA model

Using ACF and PACF, estimation models for ARIMA models are found. p = 1, 2, 3 and q = 1 or p = 3 and q = 1, 2 are chosen because the ACF plots in Figures 2 and 3 fade at lag 1 and the PACF limit after lag 3. Preliminary ARIMA models include $ARIMA(1,2,1)$, $ARIMA(3,2,1)$, $ARIMA(1,2,1)$ and $ARIMA(2,2,3)$.

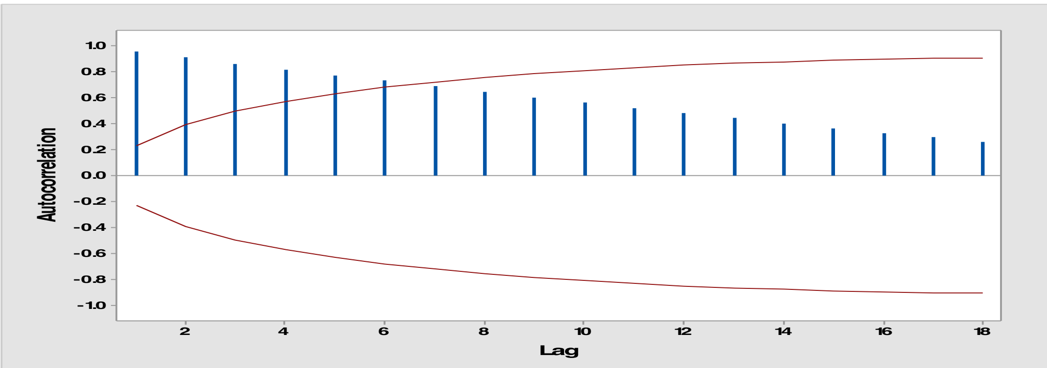

Figure 2. Autocorrelation function for the Nigerian population

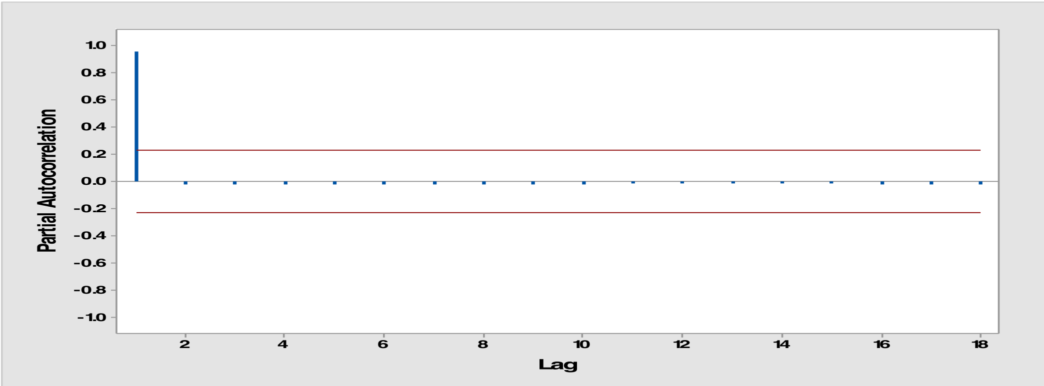

Figure 3. Partial autocorrelation function for Nigeria population

3.3. Model estimation

The four suggested ARIMA modes' parameters were estimated using the ordinary least squares approach. The least estimated values of the Akaike and Schwarz information criterion as shown in Table 2 were used to choose the ARIMA((3,2,1)) model as the better model for modelling and forecasting the Nigerian population. The estimated coefficients of the ARIMA((3,2,1)) model are given in Table 3.

Table 2. The corroboration results of the model for setting orders for the population

	$ARIMA$ $(1,2,1)$	$ARIMA$ $(3,2,1)$	$ARIMA$ $(2,2,1)$	$ARIMA$ $(1,2,3)$
R-squared	0.912217	0.935237	0.911367	0.918723
Adjusted R-squared	0.873679	0.905982	0.879281	0.880358
Akaike Info. Criterion	25.23594	25.19010	25.23594	29.38254
Schwarz Info. Criterion	25.39528	25.38131	25.39528	29.57376

Table 3. Coefficient of ARIMA(3,2,1) model for the Nigerian population

ARIMA(3,2,1)	Coefficient	Std. error	t-statistics	p-value
AR(1)	0.9234	0.2260	4.0	0.000
AR(2)	-0.4193	0.1846	-2.27	0.026
AR(3)	0.4146	0.1197	3.47	0.001
MA(1)	0.3506	0.2439	1.44	0.155

3.4. Model diagnostic check

The computed Modified Box-Pierce (Ljung-Box) Chi-Square Statistic values shown in Table 4 were used to determine the ARIMA(3,2,1) model's suitability for population forecasting in Nigeria. This was further supported by the error term's ACF and PACF in Figures 4 and 5 where the residuals lacked structure and were significant at =0.05. In essence, ARIMA(3,2,1) is adequate when forecasting the population of Nigeria. The ARIMA(3,2,1) model is therefore, the most suitable and adequate model for modelling and forecasting the Nigerian population. The was further ascertained and validated using the estimated coefficient of determination and adjusted coefficient of determination for all the estimated models. The results are given in Table 2 where the coefficient of determination for the ARIMA(3,2,1) model signified the variation in the dependent variable is explained above 83% by the independent variables and the adjusted coefficient of determination signified that the ARIMA(3,2,1) model has the predictive ability of more than 91%. Based on these, ARIMA(3,2,1) is optimal for the forecast population of Nigeria than all other models estimated.

Table 4. Modified Box-Pierce (Ljung-Box) Chi-Square Statistic result

Lag	12	24	36	48
Chi-square	9.0	12.7	26.6	28.6
Degree of freedom	8	20	32	44
P-value	0.342	0.888	0.738	0.965

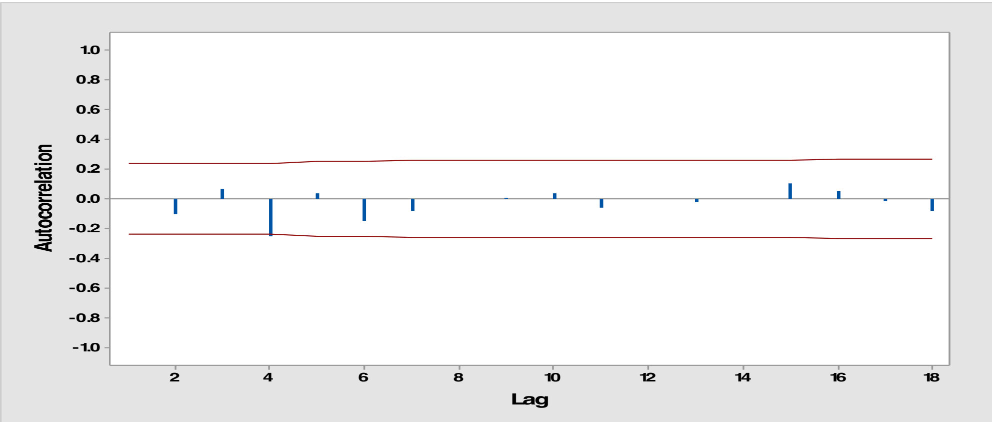

Figure 4. ACF of Residuals for population

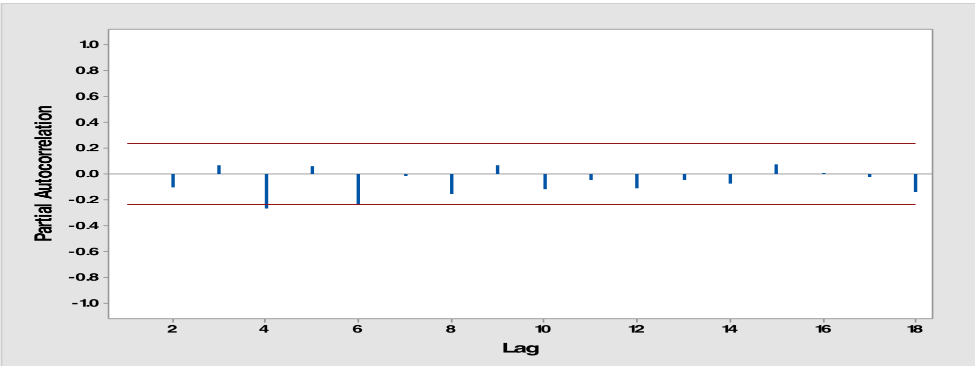

Figure 5. PACF of Residuals for population

3.5. Forecasting with ARIMA(3,2,1) model

In Figure 6, the forecast outcome from ARIMA(3,2,1) model is shown. The forecast, upper and lower limit values for the Nigerian population are also discussed using Table 5. Figure 6 and Table 5 showed that the population forecast for Nigeria using ARIMA(3,2,1) signified a consistent increase on an annual basis. This suggests that the population of Nigeria will continue to increase, and based on the upper and lower bounds, the population of Nigeria may range from 339,2377,74 to 418,347,670 in the next 28 years. Based on the forecast evaluation metrics that is, RMSE, MAE, and MAPE values given in Table 6 signified relatively low values. Therefore, the ARIMA(3,2,1) model has a reasonable degree of predictive accuracy for forecasting the Nigerian population growth.

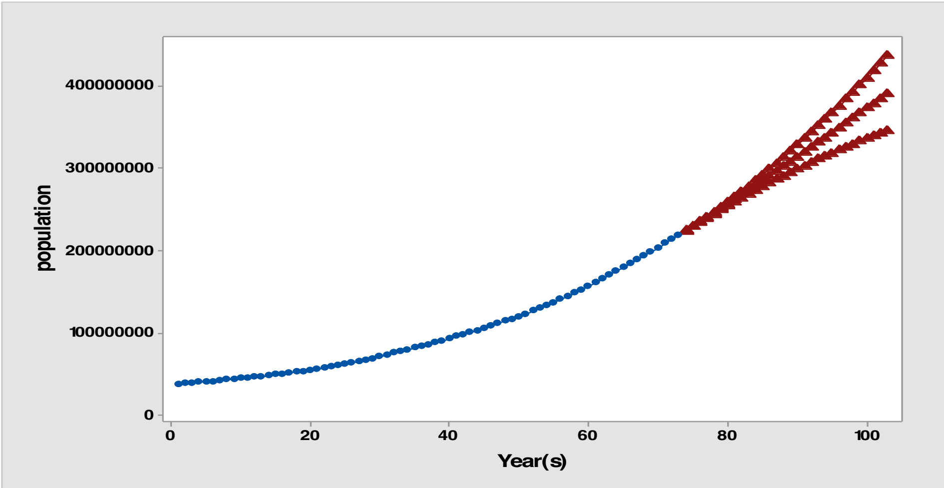

Figure 6. Forecast of Nigerian population from 1950 to 2050

Table 5. Forecast of Nigerian population from 2023 to 2050

Year(s)	Forecast values	Lower limit	Upper limit
2023	223751713	223633578	223869848
2024	229020900	228694809	229346991
2025	234342022	233742824	234941219
2026	239715767	238768240	240663295
2027	245140661	243749914	246531408
2028	250612251	248679318	252545184
2029	256127331	253556170	258698492
2030	261684196	258377073	2649913192
2031	267280769	263136163	271425375
2032	272914519	267829470	277999568
2033	278583272	272454756	284711787
2034	284285221	277009693	291560748
2035	290018559	281491799	298545320
2036	295781476	285899112	305663839
2037	301572306	290230171	312914441
2038	307389523	294483686	320295359
2039	313231664	298658509	327804819
2040	319097329	302753724	335440935
2041	324985205	306768631	343201780
2042	330894061	310702678	351085443
2043	336822728	314555442	359090014
2044	342770101	318326632	367213571
2045	348735139	322016081	375454197

2046	354716858	325623720	38389995
2047	360714329	329149569	392279088
2048	366726674	332593732	400859616
2049	372753066	335956387	409549745
2050	378792722	339237774	418347670

Table 6. Forecast Evaluation Metrics for ARIMA Model

Forecast Evaluation Metrics	$ARIMA(3,2,1)$
MAE	4.480962
RMSE	5.882820
MAPE	32.69064

4. Conclusion

The Autoregressive Integrated Moving Average model is utilized in this study to model and forecast the historical annual time series dataset for the Nigerian population from 1950 to 2050. A time series model for the historical Nigerian population was built using the study. According to the descriptive analysis result, the annual population of Nigeria had a mean of 99, 611, 692 with a standard deviation of 53, 188, 740. The population of Nigeria ranges from 37, 189, 369 in 1950 to 218, 541, 212 in 2022. The Nigerian population's time plot demonstrated the presence of secular variation, and this signified the series is not stationary. The Augmented Dickey-Fuller test was used to attain stationarity at the second difference (d = 2). The tentative ARIMA models were identified using autocorrelation and partial autocorrelation function plots of the Nigerian population. The ARIMA(3,2,1) model was selected for having the lowest values of Akaike and Schwarz information criteria after model estimation with the ordinary least squares method.

The Modified Box-Pierce (Ljung-Box) Chi-Square Statistic, autocorrelation, and partial autocorrelation of the residual were used to ensure the suitability of the ARIMA(3,2,1) model for forecasting the Nigerian yearly population series. Also, ARIMA(3,2,1) was validated using the coefficient of determination and adjusted coefficient of determination where the former signified above 83% of the variation in the dependent variable is explained by independent variables and the latter indicated the model has above 91% predictive ability. The ARIMA(3,2,1) model forecast, upper and lower bound values signified that the Nigerian population will have a continuous increase every year. The 28-year forecast values indicated a population range between 339, 237, 774, and 418, 347, 670. The forecast evaluation metrics RMSE, MAE, and MAPE) which are relatively low signifying that the ARIMA(3,2,1) model has a strong predictive accuracy. In conclusion, the Nigerian population growth forecast indicated an increase of more than 5 million every year and the Nigerian government must develop agreeable plans and recommendations to address the difficulties that will result from this explosive growth which is expected to hit above 418 million by 2050.

References

[1] Harvard International Review Public Health and Overpopulation: The United Nations Takes Action. https://hir.harvard.edu/public-health-and-overpopulation
[2] Chamie, J. Future Challenges. In Population Trends: Humanity in Transition. *International Peace Institute*, pp. 2–10, 2007. http://www.jstor.org/stable/resrep09597.5
[3] Stuart, B., Wolfgang, L., & Sergei, S. Very Long Range Global Population Scenarios to 2300 and the Implications of Sustained Low Fertility. *Demographic Research*. 28, 39, 1145-1166, 2013.
[4] West Africa Gateway. *Nigeria*, 2012. https://www.oecd.org/swac/publications/Nigeria_e-version_en_light.pdf
[5] The Economist. *Africa's population will double by 2050*, 2020. https://www.economist.com/special-report/2020/03/26/africas-population-will-double-by-2050.
[6] Nigeria Multidimensional Poverty Index. 2022, https://www.nigerianstat.gov.ng/pdfuploads/NIGERIA%20MULTIDIMENSIONAL%20POVERTY%20INDEX%20SURVEY%20RESULTS%202022.pdf
[7] The Global Economy. *Nigeria Economy Indicator*, 2023, https://www.theglobaleconomy.com/nigeria/
[8] Nyoni, K. Population Explosion in Nigeria by 2050: Overwhelming Empirical Evidence from the Box-Jenkins Arima Approach. *Dimorian Review*, 6(2), 15-27, 2019.
[9] Sinding, S. W. Population, poverty and economic development. *Philosophical Transaction of Royal Society B*, 364(1532), 3023-3030, 2009.
[10[United Nations. *Why population growth matters for sustainable development, Online Edition*, 2022. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd_2022_policy_brief_population_growth.pdf
[11] Dominic A., Oluwatoyin, M. A., & Fagbeminiyi, F. F. The determinants of population growth in Nigeria: a co-integration approach, *The International Journal of Humanities and Social Studies*, 4(11), 38-44, 2016.
[12] Aliyu A. A., & Amadu, L. Urbanization, cities, and health: The challenges to Nigeria - A review. *Annal of African Medicine*, 16(4), 149-158, 2017.
[13] Adejumo, O., O. Growth Limits: A Conceptual Analysis for Sustainable Development in Nigeria. *SAGE Open*, *10*(2), 1-12, 2020.

[14] Folorunso, O., Akinwale, A., T. Asiribo, O. E., & Adeyemo, T. A. Population prediction using artificial neural network. *African Journal of Mathematics and Computer Science Research,* 3(8), 155-162, 2010.

[15] Olatayo, T. O., & Adeboye, N. O. Predicting Population Growth through Births and Deaths Rate in Nigeria. *Mathematical Theory and Modeling,* 3(1), 13-29, 2013.

[16] Olayiwola, O. M., Lawal. G. O., Amalare, A. A., Agboluaje, S. A., & Fantola, J. O. Modelling Nigeria Population Growth Rate. *Journal of Advances in Mathematics* 10(6), 72-81, 2015.

[17] Aidi, H. O., Emecheta, C., & Ngwudiobu, I. M. Population Dynamics and Economic Growth in Nigeria. *Journal of Economics and Sustainable Development,* 7(15), 16-24, 2016.

[18] Agog, N. S., Bako, S. S., Bamanga, M. A., Peter, M., & Byeli, U. On Logistic Growth Model for Forecasting Nigeria's Population. *Science World Journal*, 15(4), 112-115, 2020.

[19[Odimientimi, A., Nnamdi, E., Omokri, A., & Adigwe, A. A Comparative Study of Some Time Series Models on the Population of Nigeria. *Global Scientific Journals,* 9, 2454-2465, 2021.

[20] Lasisi, T. A., Oladimeji, O. A., & Ogunbanwo, T. S. Detection of Appropriate Model for Nigeria Population Growth Using Root Mean Square Error (Rmse). *Researcher,* 13(11), 35-42, 2021.

[21] Ogunjobi, E. O., Odusina, M. T., & Adeoti, O. A. Population Growth Model to Project Human Population in Nigeria. *African Scholars Journal of Science Innovation and Technology Research*, 24(9), 89-98, 2022.

[22] Oladimeji, O. A., Akomolafe, A. A., Lasisi, T. A., Ojo, T. O., Adesina, O. A., & Egbedokun, G. O. Detection of Appropriate Model for Nigeria Population Growth Using Root Mean Square Error (RMSE). *International Journal of Systems Science and Applied Mathematics*, 7(3), 46-51, 2022.

[23] Orukpe, A., Imianvan, A., & Akazue, M. A. A Comparative Model for Predicting Population Census in Nigeria. *International Journal of Computer Applications*, 185(26), 27-30, 2023.

[24] Olatayo, T. O., Taiwo, A. I., & Lasisi, T. A. Population prediction modelling with autoregressive integrated moving average (ARIMA) model. *Journal of the Nigerian Association of Mathematical Physics,* 27(7), 189-200, 2014.