Journal of the Nigerian Association of Mathematical Physics
***Volume 65, (October 2022 – August 2023 Issue*), pp87 – 94**
© J. of NAMP

MODELING AND FORECASTING LIFE EXPECTANCY AT AGE SIXTY-FIVE IN THE POPULATION OF NIGERIA USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

[1]Awariefe C and [2]Ekruyota G.

[1]Department of Statistics, Delta State University of Science and Technology, Ozoro.
[2]Department of Computer Science, Delta State University of Science and Technology, Ozoro.

Abstract

The study aims to determine an appropriate empirical model for forecasting the survival of age 65 years old male and female population of Nigeria to assess its future trend up to the year 2030. We applied several Autoregressive Integrated Moving Average (ARIMA) models to the survival to age 65 population-based secondary times series data from 1960-2020. Akaike Information Criteria (AIC) was used to select the model that has the lowest value of AIC. The proposed model and forecasted values were validated and found to be adequate in predicting the survival of the age 65 male and female populations on a year-by-year basis. From the forecasted values we discovered that the survival of the age 65 male and female populations will increase approximately by 7.8% and 8.05% respectively from 2021 to 2030 under the notion that there will be no epidemic disease that will result in an increase in mortality rate in the forecasted years.

Keywords: Life Expectancy, Survival to age 65, ARIMA Models, Stationarity, Differencing and Forecasting.

1. Introduction

Life expectancy at age 65 is the average number of years that a person at that age can be expected to live, assuming that age-specific mortality levels remain constant [1]. Life expectancy at 65 is synonymous with survival to age 65 [2]. Survival to age 65 refers to the percentage of a cohort of newborn infants that would survive to age 65 if subject to age-specific mortality rates of the specified year [3]. Life expectancy is commonly used as one of the main indicators to assess the health status of a population in developed as well as developing countries, and it is also closely associated with the degree of economic and social development of a country or a region [4]. According to [2], Nigeria has the largest population in Africa and has an elderly projected population growth rate of 3.2%, this rate according to [5] is expected to double by 2050. This trend calls for concern as it poses major economic, psychological, health, and social challenges to the Nigerian State [6]. The percentage of the total population of ages 65 and above was reported to be approximately 2.76% in 2021 [7]. [8] put Nigeria's population aged 65 and older, at 3 percent, and Life expectancy at birth for males at 55 years in 2022. Life expectancy is central to the development of Nigeria which is earnestly striving to achieve socio-economic progress through investing significantly in social sectors like health, education, manufacturing, education, environmental management and sustainability, and social safety nets. Therefore, there is a need to forecast the population of survival to age 65 in Nigeria. The study aims to determine an appropriate empirical model for forecasting the survival of the age 65 male and female population of Nigeria to assess its future trend up to the year 2030. This can be realized by employing the time series forecasting approach. Forecasting by times series is an attempt of obtaining meaningful information from past data in order to predict future events based on the time components. This study proposes the Autoregressive integrated moving average (ARIMA) model, for time series forecasting as it only needs the prior data to generalize the forecast and also increases the accuracy of the forecast while keeping the number of parameters minimum [9]. Additionally, the ARIMA model is used because of its stationarity condition and wide applicability for designing suitable models for forecasting.

Corresponding Author**:** Awariefe C., Email: awariefe@gmail.com, Tel: +2348033471856

Some studies have been undertaken by researchers to forecast life expectancy, [10] used the Autoregressive Integrated Moving Average (ARIMA) model, to project trends in Life expectancy (LE) and healthy life expectancy (HALE), and GAP across global regions from 1995 to 2025. Their paper opined that public health as indicated by LE, HALE, and GAP has improved in most countries and regions in the world since 1995, and it likely will continue to improve in the future. [11] paper empirically examined the consequences of globalization on life expectancy in Nigeria involving the periods 1986 to 2016. The Augmented Dickey-Fuller (ADF) and Johanson Co-integration test was utilized to explore the long-run relationship among the variables. The research suggests that economic globalization has a positive and significant impact on life expectancy in Nigeria. [12] used the singular value decomposition (SVD) method to model and forecast age-specific mortality in the United States from 1990 to 2065 their forecasts indicated that whereas 46% now survive to age 80, by 2065 46% will survive to age 90. However, the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and above. [13] employed a cohort-components methodology to forecast aging in America in the Twenty-first Century. The results from the article indicate that the current forecasts of the U.S. Social Security Administration and U.S. Census Bureau may underestimate the rise in life expectancy at birth for men and women combined, by 2050, from 3.1 to 7.9 years. [14] proposed a Neural network method for time series to fit and forecast the yearly population data of China, India, and Vietnam from 1960 to 2016, sourced from the World Bank. Based on the age classification, there are three age periods: the young, middle-aged, and old populations. They projected that by the year 2030, the proportion of middle-aged people (aged 15–64) will make up, respectively, 63.44%, 70.77%, and 67.26% of the populations of China, India, and Vietnam. Among the three countries, China is of the least advantage in terms of labor force population.

2. Materials and Method

2.1 Data Source

This study used secondary times series data from 1960-2021 of male and female populations survival to age 65 in Nigeria. To achieve the aim of the study, secondary time series data of survival to the age 65 population was collected from [7].

2.2 Box-Jenkins ARIMA Models

To attain the aim of this study the Box-Jenkins ARIMA model is used to select the best-fitted model based on several performance benchmarks. [15] was the first to introduce the ARIMA model, the model can be used for forecasting the non-seasonal stationary time-series data. The stationary process is a requirement in developing an ARIMA model. When the time series under investigation exhibits trends and non-seasonal behavior, transformation and differencing are applied to the observed series in order to stabilize variance and to remove the trend before an ARIMA model is applied. An ARIMA model is represented by ARIMA (p, d, q) and it has three parameters: p which is the autoregressive term, denoted as AR(p); the order of differences to make the nonstationary time series stationary, denoted as d; and the number of moving average terms, denoted as MA(q). The general expression for the autoregressive process of order p, AR(p), is as: Supposed C_t is a time series random variable, then, the three basic Box-Jenkins models for C_t are:

i. Autoregressive model of order p (AR (p)):

$$C_t = \text{ძ} + \phi_1 C_{t-1} + \phi_2 C_{t-2} + \dots\dots\dots\dots + \phi_p C_{t-p} + a_t \quad (1)$$

The autoregressive part is a linear regression that relates past values of data series to future value, that is, C_t depends on its p previous values.

ii. Moving average model of order q (MA (q)):

$$C_t = \text{ძ} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots\dots\dots\dots - \theta_p a_{t-p} \quad (2)$$

The moving average part that relates past forecast errors to future values of data series. That is, C_t depends on its q previous random error terms.

Autoregressive moving average model of orders p and q ARMA (p; q)):

$$C_t = \text{ძ} + \phi_1 C_{t-1} + \phi_2 C_{t-2} + \dots\dots + \phi_p C_{t-p} + \epsilon_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots\dots - \theta_p a_{t-p} \quad (3)$$

C_t depends on its p previous values and q previous random error terms. AR and MA models are special cases of ARMA. Where a_t is typically assumed to be "white noise"; i.e., it is identically and independently distributed with a common mean zero and common variance σ^2 across all observations.

General ARIMA Representation

If C_t is integrated of order d, then, $C_t \sim I(d)$. Suppose that $C_t \sim I(d)$ and the stationary series after a d^{th} order differencing, C_t is an ARIMA (p; d; q) process, that is:

$$\phi(Bs)\nabla^d c_t = \theta(Bs) a_t \quad (4)$$

where Bs=Back shift operator, $\nabla_d = (1 - Bs)^d$ and $\phi(Bs) = 1 - \varphi_1 Bs - \varphi_2 Bs \ldots\ldots - \varphi_p Bs_p$ is an autoregressive coefficient polynomial of the smooth reversible ARMA (p, q) model, and $\Phi(B) = 1 - \vartheta_1 Bs - \vartheta_2 Bs \ldots\ldots - \vartheta_p Bs_p$ is the moving smoothing coefficient polynomial of the smooth reversible ARMA (p, q) model. The ARIMA modeling consists of four main stages: Identification, Estimation, Diagnostics Checking, and Forecasting.

Identification

The first stage in developing an ARIMA model is to establish if the series is stationary by plotting the time series and then checking for autocorrelation using the ACF and the PACF. To check if the series is stationary the Augmented Dickey-Fuller (ADF) test is employed. If the time is discovered to be non-stationery, differencing is applied to achieve stationarity. The next stage is to use Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) to determine the orders of the AR (p) and MA (q). The Akaike information criterion (AIC) or the Bayesian information criteria can used to determine the best fitted model, the model with the minimum AIC or BIC is selected as the best [16].

Estimation

Once the orders of the model have been established the third step is the estimation of the parameters. The aim of the estimation is to determine the estimates that minimizes the mean square error (MSE) using nonlinear least squares and maximum likelihood estimates.

Diagnostics Checking

The best-fitted ARIMA model that is finally selected is the one considered best based on a set of diagnostic checking criteria. The best fitted model residual is examined to confirm that the model is adequate. The following diagnostics can be utilized: Time plot of the residuals, Plot of the residual ACF and PACF and Residual portmanteau test.

Forecasting

When a satisfactory ARIMA model is adequate, then we proceed to forecast or predict for a period or several periods ahead. However, chances of forecast errors are inevitable as the period advances. To validate whether the forecasted values are suitable the Ljung-Box test is employed to check for normality of the residuals.

3 Data Analysis and Discussion of Results

To achieve the aim of this study some time series models are fitted on population of survival to age 65 male population of Nigeria. The ACF/PACF and the 'auto-arima()' function included in the package 'forecast' (in RStudio environment) were used to select the best model and the accuracy and reliability of the best fitted model was tested based on different performance criteria.

Survival to age 65 male and female Population Forecasting Based On 1960-2021 Time Series Data

Different time series models are fitted on the survival to age 65 population using time series data 1960- 2020 and results are set out in Table 1-7 and Figure 1-6 respectively. The time series under study is plotted in order to get the feel of the data.

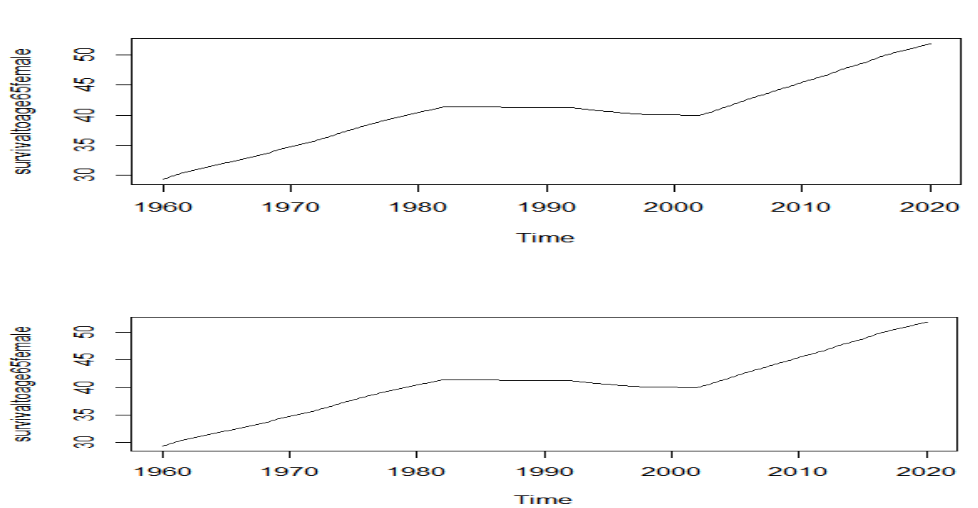

Figure 1: Time series plot of male and female survival to age 65 population of Nigeria for 1960-2020

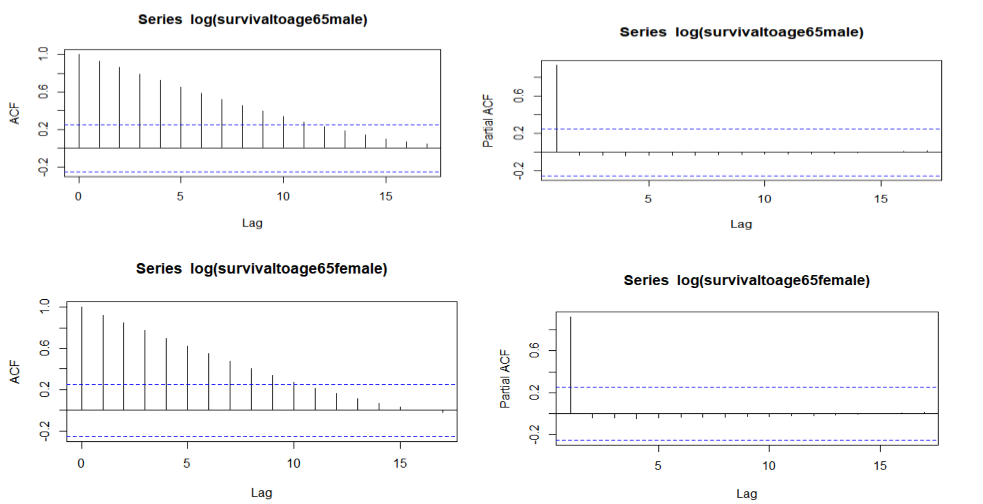

Figure 2: ACF and PACF of the survival to age 65 male and female time series data

To check autocorrelation and confirm whether the series is stationary, we used ADF test.

Table 1: ADF Test of the survival to age 65 male time series data

Variable	ADF Statistics	Probability	Alternative Hypothesis	Level of significance (α)
Survivaltime(male)	-3.0006	0.1701	Stationary	5% (0.05)
Survivaltime(female)	-3.0916	0.1333	Stationary	5% (0.05)

The ADF test in Table 1 shows that the time series data is not stationary, since p-values of 0.1701 and 0.1333 respectively are greater than $\alpha = 0.05$, the null hypothesis is accepted; this implies there is presence of unit root indicating non-stationarity of both male and female series.

To remove non-stationarity, differencing is employed and ACF and PACF plots are used to check autocorrelation, then, ADF test is used to check stationarity.

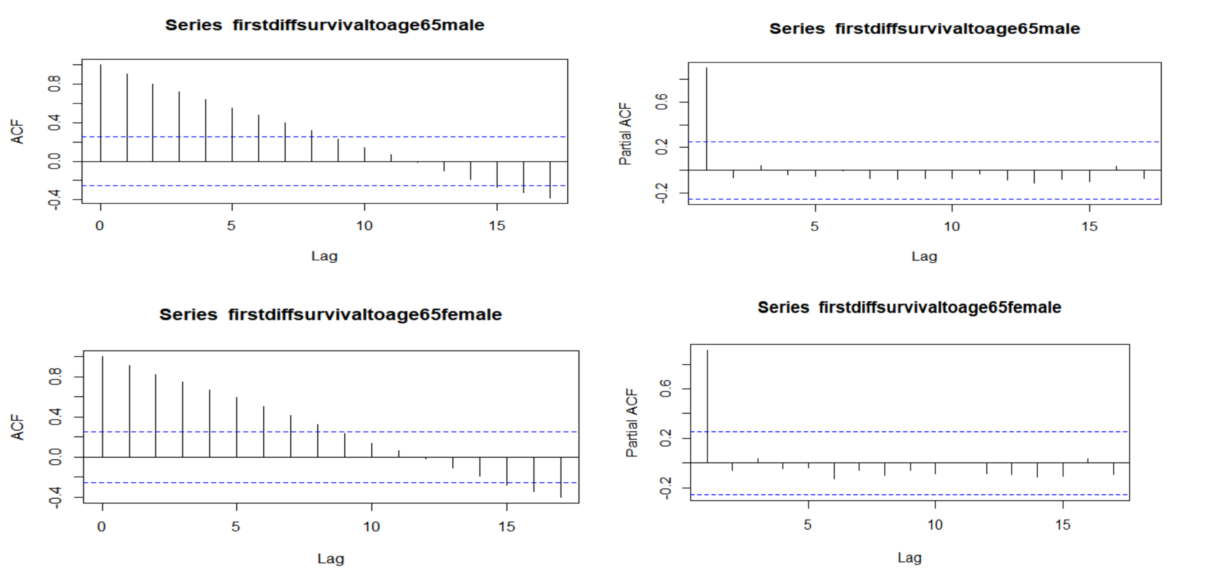

Figure 3: ACF and PACF of the 1st diff survival to age 65 male and female time series data

Table 2: ADF Test of the 1st difference survival to age 65 male and female time series data

Variable	ADF Statistics	Probability	Alternative Hypothesis	Level of significance (α)
Survivaltime (male)	-1.632	0.7235	Stationary	5% (0.05)
Survivaltime (female)	-1.518	0.7692	Stationary	5% (0.05)

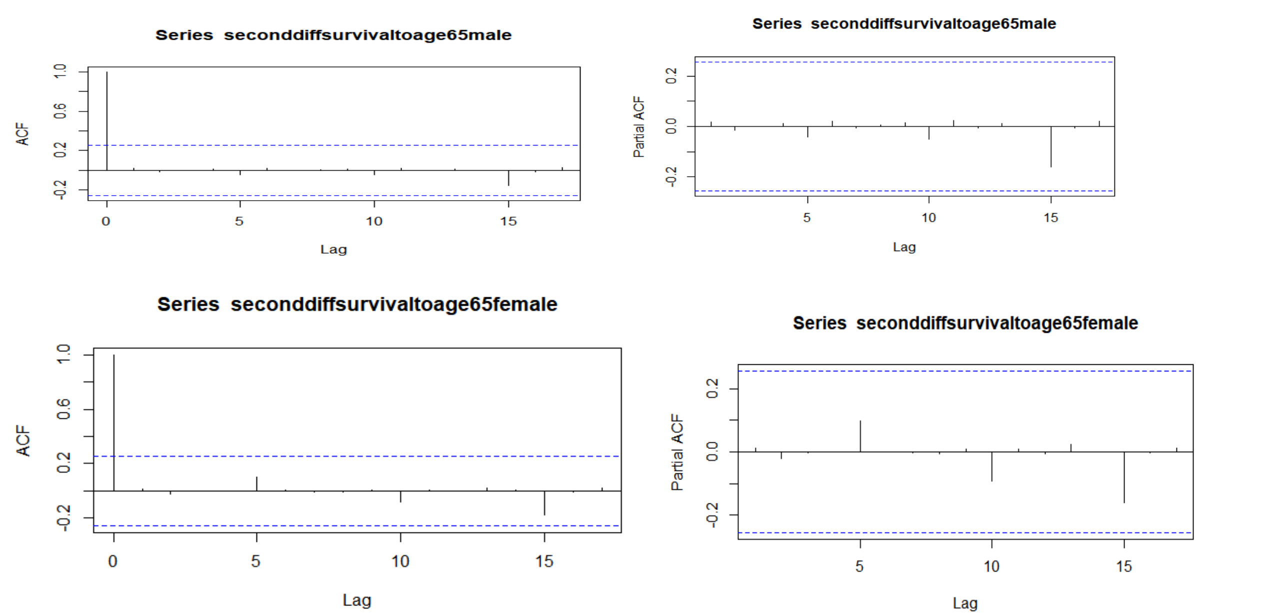

Figure 4: ACF and PACF of the 2nd difference survival to age 65 male and female time series data

Table 3:ADF Test of the 2nd difference survival to age 65 male and female time series data

Variable	ADF Statistics	Probability	Alternative Hypothesis	Level of significance (
Survival time (male)	-3.515	0.04806	Stationary	5% (0.05)
Survival time (female)	-3.5826	0.04237	Stationary	5% (0.05)

The ADF test in Table 3 shows that both male and female time series data are non-stationary, since their p-values of 0.7235 and 0.7692 are greater than $\alpha = 0.05$, the null hypothesis is accepted; this implies that there is presence of unit root indicating non-stationarity of the series. From the inspection of the ACF, PACF and ADF test in Figure 3 and Table 2 above the time series data are non-stationary. The ACF, PACF and ADF test of the second difference series in Figure 4 and Table 3 showed that the series are stationary.

Selection of the Best ARIMA model

Based on the ACF and PACF plots in Figure 4 and the application of 'auto.arima()' function in R environment above the possible ARIMA models for the male and female time series are listed in Table 4

Table 4: Evaluation of ARIMA models

Male		Female	
Model	AIC		
ARIMA (0,2,0) ARIMA (1,2,0) ARIMA (0, 2,1) ARIMA (1,2,1)	-83.43427 -81.45381 -81.45423 -79.43548	ARIMA (0,2,0) ARIMA (1,2,0) ARIMA (0,2,1)	-71.45349 -69.46599 -69.46646

From Table 4 above ARIMA (0,2,0) is the best model for both the male and female time series because it has the lowest AIC value. This was confirmed by the AUTO-ARIMA Function in R. Best model: ARIMA (0,2,0)

Table 5: Test for Normality of the Residuals

Test (Box Pierce)	**Test statistic**	**p-value**
Male (forecast residuals) Female (forecast residuals)	0.017729 0.011162	0.9159 0.9159

The selection of the best fitted model is based on AIC criteria. The best fitted model for both male and female time series is ARIMA (0, 2, 0) as revealed in Table 4. The model has the minimum AIC compared to other models. Therefore, ARIMA (0, 2, 0) is the suitable model for forecasting the survival to age 65 male and female populations of Nigeria. A model is regarded good or adequate if its residuals satisfy the assumption of normality. The results of the Box Pierce test in Table 5 shows that the residual of the model belongs to the normal distribution. After checking the suitability of the model forecasting, the next step is to forecast the survival to age 65 population. We now forecasted the population based on the selected ARIMA (0, 2, 0) from 2021 to 2030. The result of the forecasts is presented in Table 6. The validity of the forecasted values was tested for normality using the Box test and the results presented in Table 7. The test results from Table 7 shows that the forecasted values belong to the normal distribution ($p > 0.05$)

Table 6: Yearly survival to age 65 male and female population forecasts of Nigeria

	Male			Female		
Year	Forecast	95% Lo	95% Ho	Forecast	95% Lo	95% Ho
2021	47.60	47.36993	47.83007	52.36	52.10530	52.61470
2022	48.05	47.53555	48.56445	52.87	52.30048	53.43952
2023	48.50	47.63915	49.36085	53.38	52.42702	54.33298
2024	48.95	47.68985	50.21015	53.89	52.49497	55.28503
2025	49.40	47.69375	51.10625	54.40	52.51113	56.28887
2026	49.85	47.65527	52.04473	54.91	52.48036	57.33964
2027	50.30	47.57777	53.02223	55.42	52.40640	58.43360
2028	50.75	47.46393	54.03607	55.93	52.29222	59.56778
2029	51.20	47.31596	55.08404	56.44	52.14024	60.73976
2030	51.65	47.13569	56.16431	56.95	51.95251	61.94749

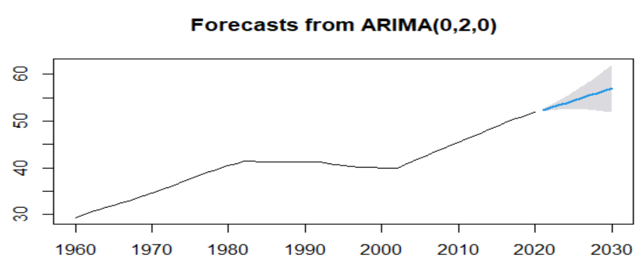

Figure 5: Forecasting plot for predicting survival to age 65 male population of Nigeria

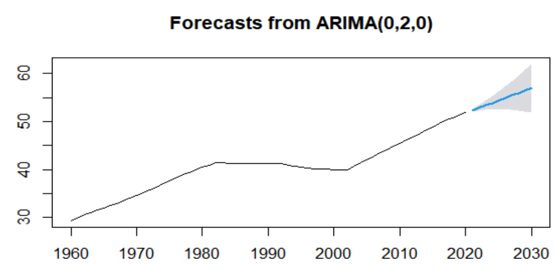

Figure 6: Forecasting plot for predicting survival to age 65 female population of Nigeria

Table 7: Test for normality of the forecast residuals

Ljung Box Test	Test statistic	p-value
Male	0.28139	0.998
Female	0.6623	0.985

It is shown in Table 7 that the survival to age 65 of male and female population of Nigeria is projected to be 47.60 and 52.36 in 2021 respectively. The forecasted survival to age 65 populations for male and female are expected to increased steadily to about 51.65 and 56.95 respectively in 2030 with the assumption that there is no epidemic disease which results in a decrease in survival to the age of 65 population in Nigeria as a result of mortality of the diseases.

4. Conclusion

To achieve the aim of the study, a secondary time series data of survival to age 65 male and female population of Nigeria collected from the World bank publication from 1960-2020 was applied to several ARIMA models. AIC was used to select the best model, ARIMA (0, 2, 0) was discovered to be the best fitted model for the both populations, as it has the minimum value of AIC criteria. The proposed model and forecasted values were validated and found to be adequate in predicting the survival to age 65 male and female populations on a year-by-year basis. From the forecasted values we discovered that the survival to age 65 male and female populations will increase approximately by 7.8 % and 8.05% respectively from 2021 to 2030 barring any unforeseen epidemic disease that will result in increase of mortality rate. Future research work may be undertaken to study the survival to age 70 or 80. The outcomes of this study could be valuable for government health agencies of Nigeria, especially when it comes to budgetary appropriation for medium-term and long-term planning.

Authorship contributions: AC proposed the materials and methods of the study. AC and EG wrote the manuscript. AC sourced all the data. EG installed the R studio and assisted in the statistical analysis. AC and EG finalized, reviewed and edited the paper.

Competing interests: The authors declare that they have no competing interests

Funding: No funding was received for this study.

Appendix

RStudio Codes Utilized in the Work

```
library(readsurvivaldata)
> survivaldata <- read_excel("survivaldata.xlsx")
> View(survivaldata)
> class(survivaldata)
[1] "tbl_df"     "tbl"        "data.frame"
survivaltime=ts(survivaldata$male,start=min(survivaldata$Year),end = max(survivaldata$Year),frequency = 1)
> survivaltime
> class(survivaltime)
[1] "ts"
> plot(survivaltime)
> acf(survivaltime)
> pacf(survivaltime)
> library(tseries)
> library(forecast)
> adf.test(survivaltime)
logsurvival=log(survivaltime)
> plot(logsurvival)
> acf(logsurvival)
> pacf(logsurvival)
> adf.test(logsurvival)
```

```
firstdiff=diff(ts(logsurvival))
> acf(firstdiff)
> pacf(firstdiff)
> adf.test(firstdiff)
seconddiff=diff(ts(firstdiff))
> acf(seconddiff)
> pacf(seconddiff)
> adf.test(seconddiff)
survivalmodel=auto.arima(survivaltime,ic='aic',trace=TRUE)
> survivalmodel
> mysurvivaltimeforecast
 Box.test(mysurvivaltimeforecast$residuals,lag=5,type = 'Ljung-Box')
```

References

[1] Organization for Economic Co-operation and Development (OECD). (2022). Life expectancy at 65 (indicator). doi: 10.1787/0e9a3f00-en (Accessed on 29 August 2022)

[2] United Nations Population Division. (2020). World Population Prospects: Revision.

[3] United Nations Population Division. (2019). World Population Prospects: Revision.

[4] Bilas, V., Franc, S and Bošnjak, M. (2014). Determinant factors of life expectancy at birth in the European Union countries. *Coll Antropol*. 38(1): 1-9.

[5] Mbah, P. O. (2016). The Neoliberal state and administrative reforms in Nigeria. *Afro Asian Journal of Social Science*, VII(III), 1–30

[6] Tanyi, P, L., André, P and Mbah, P. (2018). Care of the elderly in Nigeria: Implications for policy, *Cogent Social Sciences*, 4:1, 1555201.

[7] World Bank Development Indicators. (2021). World Bank Development Indicators Reports.

[8] United Nations Population Fund. (2022). United Nations Population Dashboard Nigeria.

[9] Sharma, D., & Phulli, K. (2020). Forecasting and Analyzing the Military Expenditure of India Using Box-Jenkins ARIMA Model. Papers 2011.06060, *arXiv.org*. Handle: Repec: arx:2011.06060. https://ideas.repec.org/p/arx/

[10] Cao, X., Hou, Y., Zhang, X., Xu, C., Jia, P., Sun, X., Sun, L., Gao, Y., Yang, H., Cui, Z., Wang, Y and Wang, Y. (2020). A comparative, correlate analysis and projection of global and regional life expectancy, healthy life expectancy, and their GAP: 1995-2025. *Journal of global health*. Vol. 10 No. 2.

[11] Popoola, T. (2018). Impact of Economic Globalization on Life Expectancy in Nigeria. *Health Econ Outcome Res Open Access 4*.

[12] Lee, R. D and Carter, R. L. (1992). Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association*, 87:419, 659-671.

[13] Olshansky, S.J., Goldman, D. P., Zheng, Y and Rowe, J.W. (2009). Aging in America in the Twenty-first Century: Demographic Forecasts from the MacArthur Foundation Research Network on an Aging Society. The Milbank Quarterly, Vol. 87, No. 4 (pp. 842–862) The Milbank Memorial Fund. *Published by Wiley Periodicals Inc.*

[14] Wei Y, Wang Z, Wang H, Li Y and Jiang Z. (2019). Predicting population age structures of China, India, and Vietnam by 2030 based on compositional data. *PLOS ONE* 14(4).

[15] Box, G. E., and Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. San Francisco. Calif: Holden-Day.

[16] Tsay, R.S. (2005). *Analysis of financial time series*. John Wiley & Sons.