



Forecasting of Male Life Expectancy in Nigeria: Box-Jenkins Approach

Oluranti G. Aje^a, Saheed B. Akanni^{b*}, Ismail A. Abdulazeez^a, Rasheedat A. Ibrahim^a, Azeezat A. Omolabake^a

^aDepartment of Statistics, Federal Polytechnic Ede, Osun State, Nigeria.

^bDepartment of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria.

ARTICLE INFO

Article history:

Received 15 December 2023

Received in revised form 24 February 2024

Accepted 05 March 2024

Keywords:

Diagnostic checking, Forecasting,
Homoscedasticity, Life expectancy, Male,
Training sets, Testing sets

MSC 2020 Subject classification:

62M08, 62M10, 62P10, 62F01

ABSTRACT

Over the past few decades, Males' life expectancy has been observed to be less than that of females. In the literature, little has been done to forecast the Life expectancy at birth for Males (LEM) in developing countries, especially in Nigeria. To understand the forecasting dynamics of LEM in Nigeria, this work therefore examines the series using the Box-Jenkins (1976) methodology. Pre-examination tests conducted on the series revealed that it is not only a difference stationary process of order one $\{I(1)\}$ but it is also best fitted by ARIMA (1, 1, 1) model. Moreover, diagnostic checking carried out on the residuals of the fitted model confirmed that the residual is a white noise error term in that it fulfills the assumptions of homoscedasticity, stationarity, and autocorrelation. To obtain a good forecast, the LEM data spanning 1960 to 2014 (55 years) were used to train the model while the remaining LEM data spanning 2015 to 2020 (6 years) were used to test the data. For the test data, the forecasted LEM has been shown to give approximate values of the actual LEM. Findings showed that the LEM is expected to increase slightly by 0.64% within the forecasted periods spanning 2021 to 2030.

1. Introduction

Globally, improving the life expectancy of people is frequently a fundamental component of the health and development goals of every country. This is because life expectancy is regarded as the key metric for assessing population health (see Roser *et al.*, 2013). Moreover, life expectancy is an important synthetic indicator for assessing the economic and social development of a country or a region (Bilas *et al.*, 2014). Besides, life expectancy is closely related to mortality and can be considered a mortality variable. Klenk *et al.* (2007) noted that life expectancy is a convenient and vital summary measure of mortality and more intuitive than mortality rates. Therefore, applying the appropriate model(s) to forecast life expectancy is crucial to improving it. Globally, and especially among developing nations, life expectancy has been at very high risk due to high rates of some macroeconomic factors such as inflation rate, unemployment, corruption index, crude oil price, exchange rate, poverty index, consumer prices, cost of living, etc. Estimates suggest that in a pre-modern, poor world, life expectancy was around 30 years in all regions of the world (Roser *et al.*, 2013). According to Geeks For geeks (2024), the pre-modern world or era lasted from the 15th century through the 18th century. In Nigeria, life expectancy at birth has increased from 37.371 years in 1960 to 52.887 years in 2020 depicting a 29.3% increment (World Bank, 2021). Over the past two centuries, life expectancy has more than doubled in many countries, for both males and females (Torri and Vaupel, 2012).

Most relevant published works on life expectancy were majorly based on time series regression or panel data techniques. For instance, Foreman *et al.* (2018) developed a three-component model of cause-specific mortality to forecast life expectancy, years of life lost, and all-cause and cause-specific mortality across countries. Their findings showed that global life expectancy is projected to increase by 4.4 years for men and 4.4 years for women by 2040, which was based on better and worse health scenarios. Also, Cao *et al.* (2020) employed Multiple Linear Regression and Autoregressive Integrated Moving Average (ARIMA) models to examine life expectancy, healthy life expectancy, and gap series for 195 countries. They projected that life expectancy and healthy life expectancy are likely to increase in most countries and regions of the world while the gap is also expected to expand. Furthermore, Olshansky (2005) examined the effect of obesity on the life

*Corresponding author. Tel.: +2348068403208

E-mail address: akannib2018@gmail.com (S. B. Akanni).

<http://doi.org/10.62054/ijdm/0101.20>

expectancy of the U.S. population by calculating the reduction in the rates of death that would occur if everyone who is currently obese were to lose enough weight to obtain an “optimal” BMI, which they defined as a BMI of 24. Their findings indicated that a steady rise in life expectancy during the past two centuries may soon come to an end. Through a review study, Seifarth (2012) highlighted biological mechanisms that may underlie the sexual dimorphism in life expectancy. They discovered that despite the noted gaps in sex equality, higher body fat percentages, and lower physical activity levels globally at all ages, a sex-based gap in life expectancy exists in nearly every country for which data exist. However, Chetty (2016) used mortality data to estimate race and ethnicity-adjusted life expectancy at 40 years of age by household income percentile, sex, and geographic area, and to evaluate factors associated with differences in life expectancy. They found that higher income was associated with greater longevity and differences in life expectancy across income groups increased over time. In his study, Mathers (2015) employed trend analysis to show that life expectancy at age 60 years has increased in recent decades in high-income countries.

Woolf and Schoomaker (2019) reviewed the relationships between life expectancy and mortality rates in the United States between 1959 and 2017. They discovered that the United States life expectancy increased for most of the past 60 years, but the rate of increase slowed over time and life expectancy decreased after 2014. Torri and Vaupel (2012) examined forecasting life expectancy in an international context for some countries using the ARIMA model, discrete geometric Brownian motion, and discrete model of geometric mean-reverting processes. Their findings established that the life expectancies forecast for different countries are positively correlated because of their tie to the forecast best-practice line. Levantesi *et al.* (2022) applied simultaneous forecasting and functional clustering techniques to forecast the multivariate time series of life expectancy at birth of the female populations in some developed and developing countries of the world. They found out that the evolution of developed countries follows a homogeneous pattern and supports the persisting homogeneity within the high longevity cluster over time.

Pascariu *et al.* (2018) applied the Double-Gap model to forecast the life expectancy at age 0 and the remaining life expectancy at age 65 for some developed countries. He established that the model should be considered as a promising available forecasting tool. Using the Li-Lee model, Van Baal *et al.* (2016) showed that life expectancy (LE) is likely to increase for all educational groups whereas LE between educational groups will widen. Bennett *et al.* (2015) examined the future of life expectancy and life expectancy inequalities in England and Wales using Bayesian Spatiotemporal Forecasting techniques. Their findings showed that life expectancy will reach or surpass 81.4 years for men and reach or surpass 84.5 years for women in every district by 2030. Additionally, Nigri *et al.* (2021) employed a long short-term memory approach to forecast life expectancy and disparity in Australia, Italy, Japan, Sweden, and the USA. Their predictions were found to be coherent with historical trends and biologically reasonable providing a more accurate portrait of the future life expectancy and lifespan disparity. Levantesi *et al.* (2022) *investigated clustering-based simultaneous forecasting of life expectancy time series through long-term short-term memory neural networks in forty-one (41) countries*. Their results showed that the evolution of developed countries follows a homogeneous pattern and supports the persisting homogeneity within the high longevity cluster over time. Rabbi *et al.* (2018) studied mortality and life expectancy forecast for nine comparatively high mortality Central and Eastern European (CEE) countries using seven different variants of the Lee–Carter method and the Bayesian Hierarchical Model. Their results revealed that the use of the probabilistic forecasting technique from the Bayesian framework resulted in a better forecast than some of the extrapolative methods but also produced a wider prediction interval for several countries.

Kontis *et al.* (2017) utilized the Bayesian ensemble model to study the future of life expectancy in 35 industrialized countries. Their findings established that life expectancy is projected to increase in all 35 countries with a probability of at least 65% for women and 85% for men. Levantesi *et al.* (2023) examined the multi-country clustering-based forecasting of healthy life expectancy using multivariate forecasting techniques. Their findings established that the predictive analysis in a multi-population perspective to obtain more accurate information by exploiting the similarities between countries that have shown similar trends. Bergeron-Boucher *et al.* (2019) investigated the impact of the choice of life table statistics using extrapolative methods. The results show that forecasting based on death rates and probabilities of death leads to more pessimistic forecasts than using survival probabilities, life table deaths, and life expectancy when applying existing models based on linear extrapolation of (transformed) indicators. Based on the abovementioned relevant literature, it was observed that little has been done on the forecasting dynamics of Life expectancy at birth for Males (LEM). Consequently, this study aimed to forecast the future values of LEM using the Box-Jenkins (1976) methodology.

2. Materials and Methods

In this study, a low-frequency time series data spanning 1960 to 2020 (61 years) was collected on life expectancy at birth for Males (LEM). The LEM data was extracted from the repository of the World Governance Index (WGI) via their website <http://data.worldbank.org>. To ensure a good forecast of the LEM series, these data sets were further divided into two sets namely the training sets and tests sets. The training sets represent 90% of the total datasets which spanned 1960 to 2014 (55 years) whereas the test sets represent 10% of the total datasets which spanned 2015 to 2020.

The Box-Jenkins (1976) methodology otherwise known as ARIMA (p, d, q) model is desirable for modeling and forecasting the univariate LEM since it is measured at a very low frequency. According to Pankratz (1983), the model describes the relationship between its current time series and its past values. The mathematical specification of the ARIMA (p, d, q) model is stated as equation (1):

$$LEM_t = \mu + \phi_1 LEM_{t-1} + \phi_2 LEM_{t-2} + \dots + \phi_p LEM_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + \varepsilon_t \quad (1)$$

Where:

LEM_t = LEM at current time t , LEM_{t-1} = LEM at immediate past period $t-1$, LEM_{t-2} = LEM at two years past value $t-2$, LEM_{t-p} = LEM at period $t-p$, ε_t = Random shock at period t , e_{t-1} = Random shock at immediate past period $t-1$, e_{t-q} = Random shock at period $t-q$. μ , ϕ_p , and θ_q are parameters to be estimated. Based on (Hipel and McLeod, 1994; Flaherty and Lombardo, 2000), equation (1) can be re-written using lag polynomial as follows:

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

where $\phi(L)$ and $\theta(L)$ give the set of autoregressive and moving average parameters and ε_t remains the white noise.

To investigate the stationarity of the ARIMA (p, d, q) process, all the roots of $\phi(L)$, the characteristic equation must lie outside the unit circle.

Equation 2 can also be expressed as equation 3 below:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1-L)^d Y_t = (1 + \sum_{i=1}^q \theta_i L^i)\varepsilon_t \quad (3)$$

Where p , d , and q are integers greater than or equal to zero and refer to the order of the Autoregressive (AR), Integrated (I), and Moving Average (MA) parts of the model respectively. The integer "d" indicates the level of differencing. Moreover, the basic four-step procedures of the Box-Jenkins (1976) methodology are given below.

Step 1: Model Identification

Here, we first employ the correlogram and Akaike Information Criterion (AIC) to jointly identify the orders p and q components of the AR and MA parts of the ARIMA (p, d, q) model. The correlogram is used to determine the range of values of p and q using the Sample Autocorrelation Function (SACF) and Sample Partial Autocorrelation Function (SPACF). However, the AIC was further used to determine the best values of p and q from the range of values earlier suggested by the correlogram. According to Burnham and Anderson (2004), the AIC can be computed using the following mathematical relations:

$$AIC(p) = n \ln \left(\frac{\hat{\sigma}_{\varepsilon_t}^2}{n} \right) + 2p \quad (4)$$

where: p is the number of parameters, and n is the number of observations.

Furthermore, the best model would be selected from the competing or trained models using the Mean Square Error (MSE) and Root Mean Square Error (RMSE) evaluation metrics. The MSE and RMSE mathematical formulae are stated as equations (5) and (6)

$$MSE = \frac{1}{n} \sum_{i=1}^n (LEM_i - \hat{LEM}_i)^2 \quad (5)$$

where n is the number of observations, LEM_i is the actual value of the i th observation, and $\hat{LEM}_i$ is the predicted value of the i th observation.

$$RMSE = \sqrt{MSE} \quad (6)$$

Step 2: Model Estimation

Having determined the appropriate lag orders of p and q to be included in the model, the next thing is to estimate the model using these lags. Here, the trained data will be used to estimate the best model.

Step 3: Diagnostic Checking

For this study, the time series diagnostics (tsdiag) in the R package will be employed to conduct the residual diagnostic of the best model. The tsdiag plots usually check for the presence of homoscedasticity, stationarity, and autocorrelation. Once the fitted model passes these tests, the next thing is forecasting.

Step 4: Forecasting

Here, two types of forecasts will be considered. The first forecast is the in-sample and this is applied to the trained sets which spans 2015 to 2020. However, the second forecast will be based on the out-sample forecast which spans 2021 to 2030.

4. Results and Discussion

In this section, we present the results of the analyses carried out on the LEM series using the latest version R 4.3.1 package.

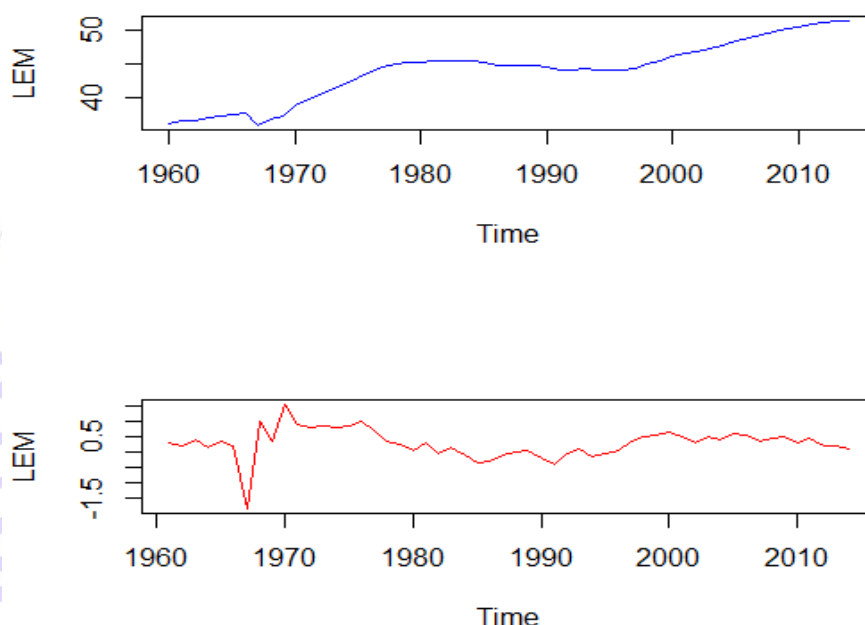


Figure 1: Time series plots of the trained LEM series at level and first difference, (1960-2014) yearly

The plots in Figure 1 represent the time series plots of the LEM series at the level and first difference correspondingly. For the first-time series plot above, it is seen that the LEM series are trending up deterministically at a level with very little or zero fluctuations. This suggested that the LEM is not a level stationary time series and needs to be differenced at least once to become stationary. On the other hand, the time series plot of the differenced LEM series revealed that it is now stationary after the first difference {i.e. $I(1)$ }. That is, the order of integration of LEM is suggested to be $I(1)$.

Table 1: Summary unit root test results for the LEM series at a level

ADF tests for LEM at the level				
	t-test	5pct	p-value	Order of Integration
LEM	2.4201	-2.6	0.0191 *	$I(0)$
KPSS tests for LEM at the level				
t-test	10pct	5pct	2.5pct	1pct
0.1712	0.119	0.146	0.176	0.216
KPSS tests for LEM at first difference				
t-test	10pct	5pct	2.5pct	1pct

0.1171 0.119 0.146 0.176 0.216

Note: 1pct, 2.5pct, 5pct and 10pct represent the 1%, 2.5%, 5% and 10% critical values

Observably, the ADF tests in Table 1 reported a different result from what the level time series plot in Figure 1 earlier reported. That is, the ADF tests reported LEM as a level stationary series {i.e. $I(0)$ } whereas the level time series plot in Figure 1 suggested that it is not. In other words, we reject the null hypothesis of no stationarity for LEM at level under ADF ($p\text{-value} < 0.05$).

However, for appropriate determination of the true order of integration of the series, the KPSS was also applied to test the presence of unit roots in the series. Based on the results of the KPSS tests reported in the same Table 1, it is clear that truly LEM is not $I(0)$ but $I(1)$. In other words, the KPSS t-test ($= 0.1712$) is greater than the 10% and 5% critical values of 0.119 and 0.146 but less than the 2.5% and 1% critical values of 0.176 and 0.216. This indicates our failure to accept the null hypothesis of stationarity for LEM at the level. Furthermore, the KPSS t-test ($= 0.1712$) is less than all the critical values after the first difference. As a result, we fail to reject the null hypothesis of stationarity for LEM after the first difference. The ADF test developed by Dickey and Fuller (1979) usually tests the hypothesis that the series is stationary whereas the KPSS test developed by Kwiatkowski *et al.*, (1992) usually tests the null hypothesis that the series is stationary. That is, these tests can be regarded as two-mirror opposite hypotheses. The essence of using these two tests in this study is to adequately determine the true order of integration of the LEM series.

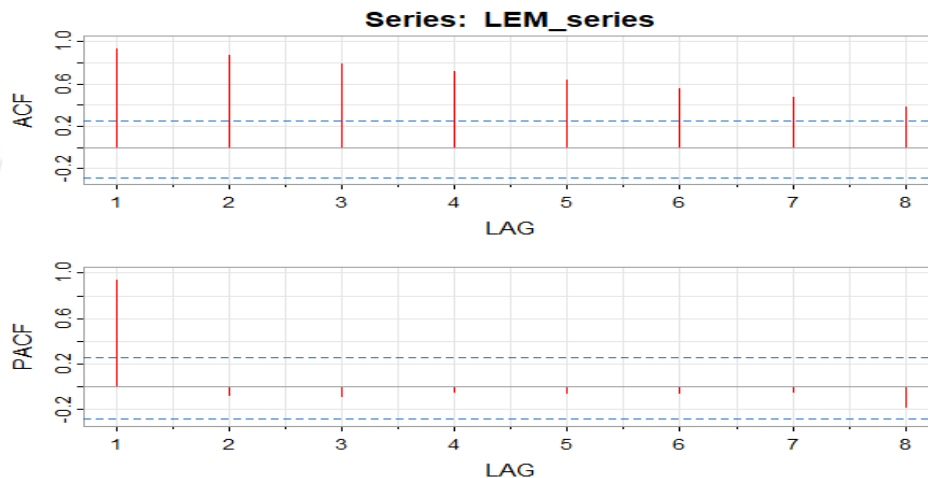


Figure 2a: Correlogram of the LEM series at level form

The correlogram in Figure 2a revealed that the spikes of ACF are decaying geometrically across the lags while a spike of the PACF cuts off abruptly at lag 1 for the PACF (i.e. lag 1 is statistically significant since it is not within the 95% confidence bounds). This indicates that either ARIMA (1, 1, 0) or ARIMA (1, 1, 1) model is suggested to fit the LEM series. In other words, ARIMA (1, 1, 0) and ARIMA (1, 1, 1) are the two competing models for analysing the series.

Moreover, the competing models were subjected to further evaluation using the Akaike Information Criteria (AIC), Mean Square Error (MSE), and Root MSE (RMSE). The essence of this evaluation is to determine the best model for forecasting the LEM series.

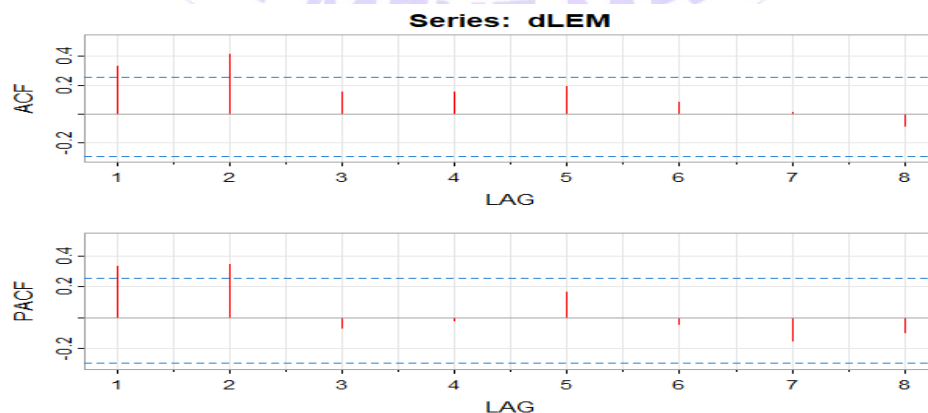


Figure 2b: Correlogram of the LEM series after the first difference

The correlogram in Figure 2b is also instrumental in determining the true order of integration of the LEM series in that it supported the results earlier reported by the KPSS tests which claimed that the series is $I(1)$. Careful inspection of the ACF and PACF in this correlogram revealed that the spikes of both the ACF and PACF are not statistically significant which means that the series is truly an $I(1)$ process.

Having determined the level of differencing d to be 1 (i.e. $d = 1$), the next thing is to subject the competing models to further selection criteria.

Table 2: Selection criteria for competing models

S/N	(p,d,q)	AIC	MSE	RMSE
1	(1,1,0)	76.72	0.1388899	0.3726793
2	(1,1,1)	70.838*	0.07410197*	0.2722168*

Based on the results of the selection criteria presented in Table 2, it is clear that only the ARIMA (1, 1, 1) model reported the least values of the AIC, MSE, and RMSE. Consequently, it is chosen as the best model for analyzing and forecasting the LEM series.

Table 3: Estimates of ARIMA (1, 1, 1) model

Dependent variable:	
LEM	
ar1	0.885 (0.083)***
ma1	-0.554 (0.147)***
Observations	54
Log Likelihood	-32.419
sigma2	0.193
Akaike Inf. Crit.	70.838

Note: The *** denotes the extremely significant variables in the model

The estimates of the ARIMA (1, 1, 1) model is stated as equation (7)

$$LEM_t = 0.885LEM_{t-1} - 0.554e_{t-1} + \varepsilon_t \quad (7)$$

Based on the results presented in Table 3, both the immediate past period $t-1$ (2013) of the AR1 term of LEM and the immediate past period $t-1$ (2013) of the MA term has significant impacts on LEM at the current period t (2014).

After model estimation, the next thing is to conduct the diagnostic checking on the residuals of the fitted ARIMA (1, 1, 1) model. Here, basic assumptions like homoscedasticity, stationarity, and autocorrelations were checked using the time series diagnostic (tsdiag) plots. If the time model passed these assumptions, then its residual is regarded as white noise. Otherwise, it is not. The tsdiag plot is given in Figure 3b below.

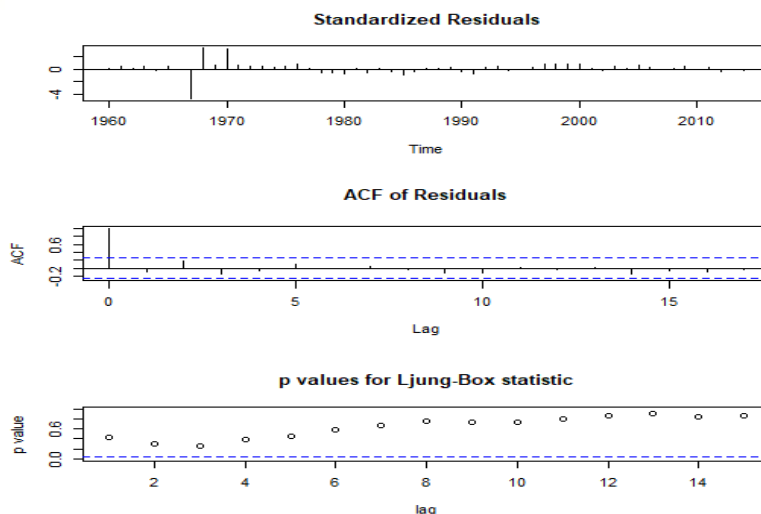


Figure 3b: Plots of time series diagnostics (tsdiag) for the ARIMA (1, 1, 1) model

Figure 3b represents the plots of time series diagnostics (tsdiag) which comprise standardized residuals, ACF of residuals, and p-values for Ljung-Box statistic.

For the standardized residuals in the tsdiag plot, the residuals obtained from the ARIMA (1, 1, 1) model are homoscedastic, because there is a constant spread of the fitted values or points around zero. Moreover, for the ACF in the tsdiag plot, the residuals obtained from the fitted ARIMA (1, 1, 1) model are stationary in that all the spikes of the ACF are statistically insignificant (i.e. all spikes are found within the 95% confidence bounds). Lastly, for the p-values of the Ljung-Box statistic in the tsdiag plot, the p-values fall above the horizontal line which indicates that we fail to reject the null hypothesis of no autocorrelation in the residuals of the fitted ARIMA (1, 1, 1).

SHaving fulfilled the assumptions of homoscedasticity, stationarity, and autocorrelation, the residual of the model is therefore regarded as white noise which means it is good for forecasting. Our forecasting comprises two parts namely the in-sample and out-sample forecasts. The in-sample forecast was used to forecast the remaining 5% of the datasets which spanned 2015 to 2020. The essence of this forecast was to check if the trained ARIMA (1, 1, 1) model was able to forecast the remaining 6 years that were left out correctly. The trained model is said to forecast the series correctly if and only if the forecasted series is very close to the actual LEM series. Otherwise, the trained model does not forecast the series correctly.

Table 4: In-sample forecast from 2015 to 2020

		ARIMA (1,1,1)		
Year	LEM	Forecast	Lo 95	Hi 95
2015	51.477	51.51849	50.6584	52.3786
2016	51.848	51.63835	50.20663	53.0701
2017	52.017	51.74439	49.74455	53.7442
2018	52.331	51.83819	49.26522	54.4112
2019	52.651	51.92117	48.77197	55.0704
2020	52.456	51.99458	48.26928	55.7199

The results of the in-sample forecast presented in Table 4 revealed two things. Firstly, the forecasted LEM is very close to the actual LEM in that its forecast gives the approximate values of the actual LEM. Secondly, just like the actual LEM data shows a slight increment from 2015 through 2020, the forecasted LEM also shows a slight increment from 2015 through 2020. Besides, the 6-year in-sample forecasts are within the 95% confidence bounds (i.e. Lo 95 and Hi 95).

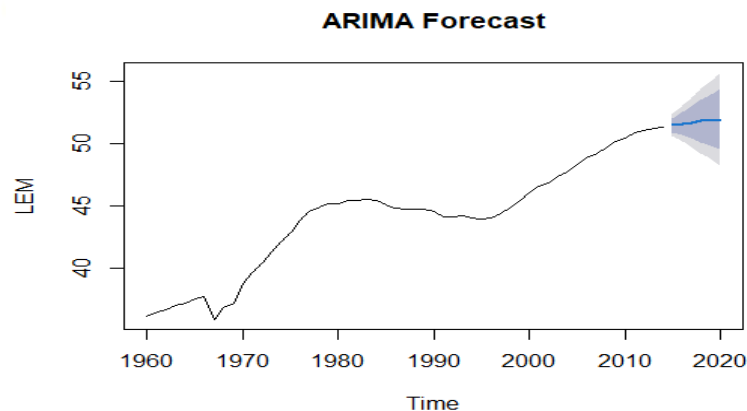


Figure 4: Forecast plot for the trained data from 2015 to 2020

The forecast plot in Figure 4 also supported the forecast results reported in Table 4 in that the forecasted LEM was seen within the forecasted period between 2015 and 2020.

The forecast plot in Figure 4 also supported the forecast results reported in Table 4 in that the forecasted LEM was seen within the forecasted period between 2015 and 2020.

Table 5: Out-sample forecast from 2021 to 2030

ARIMA (1,1,1)			
Year	Forecast	Lo 95	Hi 95
2021	52.05952	47.7612	56.35783
2022	52.11696	47.25112	56.9828
2023	52.16778	46.74173	57.59384
2024	52.21274	46.23515	58.19033
2025	52.25251	45.73305	58.77198
2026	52.28769	45.23667	59.33871
2027	52.31882	44.74699	59.89065
2028	52.34635	44.26469	60.42801
2029	52.37071	43.79027	60.95115
2030	52.39226	43.32407	61.46044

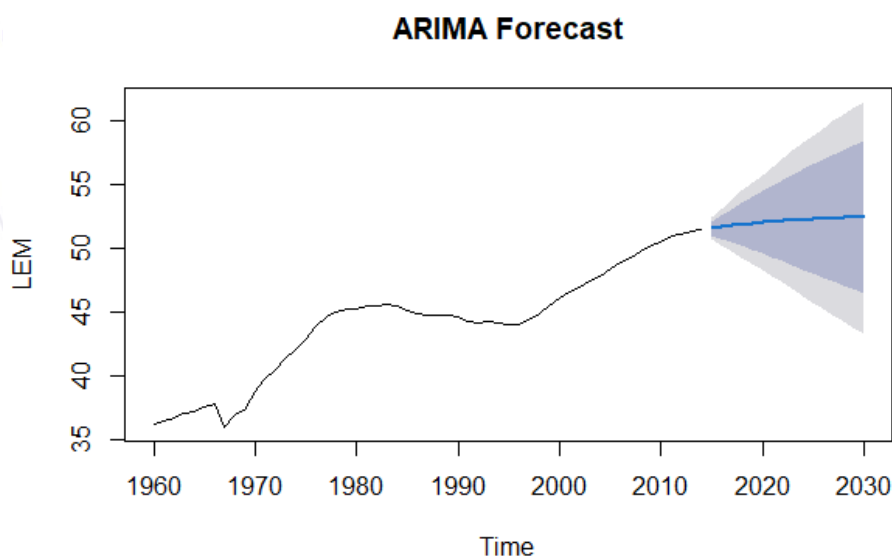


Figure 5: Forecast plot for the LEM data from 2021 to 2030

Forecast results reported in Table 5 and Figure 5 showed that the LEM series is projected to increase slightly from 52.05952 years in 2021 to 52.39226 years in 2030 depicting approximately a 0.64% increment. Equally, the forecasted LEM series are seen within the 95% confidence interval; which indicates that the predicted values are highly reliable.

4.1 Summary of Findings

This work has applied the univariate time series techniques to model and forecast the Life expectancy at birth for Males (LEM) in Nigeria as described by Box and Jenkins (1976). As usual, the modeling of the LEM series started by subjecting the series to pre-estimation evaluations such as the time series plots and unit root analysis. Results of

the time series plots in Figure 1 suggested that the LEM series are trending upward deterministically at a level with very little or zero fluctuations which is an indication of non-stationary process. However, for appropriate determination of the stationarity status or order of integration of the series, we employed the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests for the unit root analyses. The results of Augmented Dickey-Fuller (ADF) tests presented in Table 1 reported LEM as a level stationary process {i.e. $I(0)$ } which contradicts the results earlier reported by the time series plots in Figure 1. Moreover, the results of KPSS tests presented in the same Table 1 later confirmed that truly LEM series is not an $I(0)$ process but an $\{I(1)\}$ process. This is evident from the rejection of the hypothesis of stationarity at $I(0)$ and acceptance of the hypothesis of stationarity at $I(1)$ for the series under KPSS tests. Furthermore, the correlogram in Figure 2a revealed that the spikes of ACF are decaying geometrically across the lags while a spike of the PACF cuts off abruptly at lag 1 for the PACF which suggested that either ARIMA (1, 1, 0) or ARIMA (1, 1, 1) model are two candidates that can be fitted to the series. Based on the results of the selection criteria presented in Table 2, the ARIMA (1, 1, 1) model was selected as the best model for fitting the series. Besides, the results of diagnostic checking from the time series diagnostics (tsdiag) plots in Figure 3b indicated that residuals of the fitted ARIMA (1, 1, 1) model do not violate the assumptions of homoscedasticity, stationarity, and autocorrelation. Consequently, the residual of the model is therefore a white noise which means it is good for forecasting the series. Furthermore, in-sample and out-sample forecasts were made on the series. Firstly, the forecasted LEM is very close to the actual LEM in that its forecast gives the approximate values of the actual LEM. Secondly, just like the actual LEM data shows a slight increment from 2015 through 2020, the forecasted LEM also shows a slight increment from 2015 through 2020. Besides, the 6 years' in-sample forecasts are within the 95% confidence bounds (i.e. Lo 95 and Hi 95) which is an indication of a good forecast. Lastly, out-sample forecast results reported in Table 5 and Figure 5 jointly showed that the LEM series is projected to increase slightly from 52.05952 years in 2021 to 52.39226 years in 2030 depicting approximately 0.64% increment. Similarly, the forecasted LEM series are seen within the 95% confidence interval; which indicates that the predicted values are highly reliable.

5. Conclusion and Recommendations

This work investigates the forecasting dynamics of Life expectancy at birth for Males (LEM) in Nigeria using the low-frequency univariate time series techniques otherwise known as Autoregressive Integrated Moving Average (ARIMA) modeling techniques. However, the study contributes significantly to the literature on life expectancy in two ways. First, little or no work has been done on LEM. Secondly, most of the previous works failed to train their data and ARIMA model before forecasting the series under study. Conclusively, the results of the analysis and summary of findings established that the LEM series is projected to increase slightly from 52.05952 years in 2021 to 52.39226 years in 2030 depicting approximately 0.64% increment. In essence, LEM remains 52 years within the forecasted period which spans 2021 to 2030. This study recommends that Governments at all levels in collaboration with policymakers should formulate policies that will lead to a significant reduction in macroeconomic factors like unemployment, inflation, consumer price index, poverty index, exchange rate, mortality rate, etc.

References

- Aburto, J. M., Villavicencio, F., Basellini, U., Kjærgaard, S., and Vaupel, J. W. (2020). Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences*, 117(10), 5250-5259.
- Bennett, J. E., Li, G., Foreman, K., Best, N., Kontis, V., Pearson, C., ... and Ezzati, M. (2015). The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *The Lancet*, 386(9989), 163-170.
- Bergeron-Boucher, M. P., Kjærgaard, S., Oeppen, J., and Vaupel, J. W. (2019). The impact of the choice of life table statistics when forecasting mortality. *Demographic Research*, 41, 1235-1268.
- Bilas, V., Franc, S., and Bošnjak, M. (2014). Determinant factors of life expectancy at birth in the European Union countries. *Collegium antropologicum*, 38(1), 1-9.
- Box, G. E., & Jenkins, G. M. (1976). Time series analysis, control, and forecasting. *San Francisco, CA: Holden Day*, 3226(3228), 10.
- Burnham, K. P., and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods and research*, 33(2), 261-304.

- Cao, X., Hou, Y., Zhang, X., Xu, C., Jia, P., Sun, X., ... and Wang, Y. (2020). A comparative, correlated analysis and projection of global and regional life expectancy, healthy life expectancy, and their GAP: 1995-2025. *Journal of Global Health*, 10(2).
- Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., ... and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *Jama*, 315(16), 1750-1766.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431.
- Geeks Forgeeks (2024). Definition of pre-modern world. Retrieved from <https://www.geeksforgeeks.org/the-pre-modern-world/>, 30th January 2024.
- Foreman, K. J., Marquez, N., Dolgert, A., Fukutaki, K., Fullman, N., McGaughey, M., ... and Murray, C. J. (2018). Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*, 392(10159), 2052-2090.
- Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier.
- Klenk, J., Rapp, K., Büchele, G., Keil, U., and Weiland, S. K. (2007). Mortality and life expectancy: Increasing life expectancy in Germany: quantitative contributions from changes in age- and disease-specific mortality. *European Journal of Public Health*, 17(6), 587–592. <https://doi.org/10.1093/eurpub/ckm024>.
- Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K., and Ezzati, M. (2017). Future life expectancy in 35 industrialized countries: projections with a Bayesian model ensemble. *The Lancet*, 389(10076), 1323-1335.
- Kwiatkowski, D. P., Phillips, C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationary against the alternative of a unit root,” *Journal of Econometrics*, 54, 159-178.
- Levantesi, S., Nigri, A., and Piscopo, G. (2022). Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. *International Journal of Approximate Reasoning*, 140, 282-297.
- Levantesi, S., Nigri, A., Piscopo, G., and Spelta, A. (2023). Multi-country clustering-based forecasting of healthy life expectancy. *Quality and Quantity*, 1-27.
- Flaherty, J., and Lombardo, R. (2000, January). Modeling private new housing starts in Australia. In *A paper presented at the Pacific-Rim Real Estate Society conference* (pp. 24-27).
- Mathers, C. D., Stevens, G. A., Boerma, T., White, R. A., and Tobias, M. I. (2015). Causes of international increases in older age life expectancy. *The Lancet*, 385(9967), 540-548.
- Nigri, A., Levantesi, S., and Marino, M. (2021). Life expectancy and lifespan disparity forecasting: a long short-term memory approach. *Scandinavian Actuarial Journal*, 2021(2), 110-133.
- Mathers, C. D., Stevens, G. A., Boerma, T., White, R. A., and Tobias, M. I. (2015). Causes of international increases in older age life expectancy. *The Lancet*, 385(9967), 540-548.
- Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., ... and Ludwig, D. S. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine*, 352(11), 1138-1145.
- Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins: Concepts and Cases*, John Wiley and Sons, Inc. New York.

- Pascariu, M. D., Canudas-Romo, V., and Vaupel, J. W. (2018). The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*, 78, 339-350.
- Rabbi, A. M. F., and Mazzuco, S. (2018). Mortality and life expectancy forecast for (comparatively) high mortality countries. *Genus*, 74, 1-28.
- Roser, M., Ortiz-Ospina, E., and Ritchie, H. (2013). Life expectancy. *Our world is in data*.
- Roser, M., Ritchie, H., Ortiz-Ospina, E., & Rod  s-Guirao, L. (2013). World population growth. *Our world in data*. Retrieved from <https://ourworldindata.org/life-expectancy>, 19th February, 2024.
- Seifarth, J. E., McGowan, C. L., and Milne, K. J. (2012). Sex and life expectancy. *Gender medicine*, 9(6), 390-401.
- Torri, T., and Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28(2), 519-531.
- Van Baal, P., Peters, F., Mackenbach, J., and Nusselder, W. (2016). Forecasting differences in life expectancy by education. *Population studies*, 70(2), 201-216.
- Woolf, S. H., and Schoomaker, H. (2019). Life expectancy and mortality rates in the United States, 1959-2017. *Jama*, 322(20), 1996-2016.
- World Bank (2021). Statistics on Life expectancy. Retrieved from <http://data.worldbank.org>, 12th December 2022.