Modelling Nigeria Male Mortality Using Functional Data Time Series Analysis Approach.

Jean Bosco NDIKUBWIMANA (✉ ndikujeanbosco@gmail.com)
    University of Rwanda   https://orcid.org/0000-0001-5154-8218

LAWAL F.K
    University of Rwanda

James KARAMUZI
    University of Kigali (UoK)

Angelique DUKUNDE
    University of Rwanda, ACE-DS, UoK

Evariste GATABAZI
    University of Kigali (UoK)

Joel GASIGWA
    University of Kigali (UoK)

Frederic NGENDAHIMANA
    University of Rwanda, ACE-DS

MODELLING NIGERIA MALE MORTALITY USING FUNCTIONAL DATA TIME SERIES ANALYSIS APPROACH.

By:

[1]Jean Bosco NDIKUBWIMANA, [2]LAWAL F.K, [3]James KARAMUZI, [4]Angelique DUKUNDE, [5]Evariste GATABAZI, [6] Joel GASIGWA, [7] Frederic NGENDAHIMANA

[1]University of Rwanda, Rwanda; Kara University, Togo and BallsBridge University, Argentina.
[2]University of Rwanda, College of Business and Economics, Nigerian Volunteer, Rwanda.
[4, 7] University of Rwanda (UR), and African Center of Excellence in Data Sciences (ACE-DS).
[3, 4,5,6,7] University of Kigali, (UoK), Kigali - Rwanda. E-mail: ndikujeanbosco@gmail.com; ndikujeanbosco1@gmail.com, jbndikubwimana@ur.ac.rw, lawalfkray@yahoo.com, jkaramuzi@uok.ac.rw, angeliquedukunde@gmail.com, gatbzo007@gmail.com, jgasigwa@uok.ac.rw, and fngendahimana55@gmail.com
Corresponding Author e-mail: ndikujeanbosco@gmail.com or ndikujeanbosco1@gmail.com

Abstract

Incidence and mortality rates are considered as a guideline for planning public health strategies and allocating resources. Several methods have been proposed and used for modeling mortalities of various countries. Among the leading mortality, models are the Lee-Carter model which has been used in various countries and adjudged to fit the mortality of these countries well. But it came with its own limitations as the model was used in a more developed nation. In this research work, we propose functional data analysis techniques to model Nigerian Male mortality using the data obtained from the Nigeria Bureau of Statistics from 1998-2010. We compared the results obtained using some parameters such as MAPE and MSE. From the results, we discovered that the improvement of the parameters of our model shows that it is better than the Lee-Carter model in analyzing Nigerian Male Mortality.

Key words: Mortality, Lee-Carter, Functional data analysis, Forecasting, predictive Intervals, MAPE, ARIMA.

1. Introduction

Mortality forecasting date back to the French mathematician DeMoivre (1725) who proposed the survival function, thereafter, British actuary Gompertz(1825) noted that for age grouping between 20- 60 years, the force of mortality increased exponentially. His observation was based on population data from England, Sweden and France. It was found that the model works for many countries with little defect as lack of fit of the model was observed for particular age (Ol-shansky 1997). In other to overcome this anomaly of Gompertz-law many expressions have been postulated among which are Makeham (1867), Thiele (1872), Wittstein (1883), Pearson (1895), Perk (1932). Stoto (1983) reviewed most of the past mortality forecast model and discovered errors in forecasting while Keilman (1988) opined that the earlier mortality forecast have evaded some important future. He asserted old age mortality decline was underestimated and increase in life expectancy was under-projected. Many judgment decisions are involved in forecasting mortality. In fact, judgment is necessary at every stage (Alders & De Beer, 2005). Where forecast numbers of deaths are of interest, they are best derived from forecast mortality rates through population forecasting (Booth 2006).

The renewed search in mortality forecasting has been accompanied by the development of a new and more complex method. We therefore propose a new class of stochastic mortality forecasting called Functional data time series analysis for the Study. Modeling male mortality can aid in the allocation of resources to gender issues. The primary purpose of this study is to compare a proposed model called functional time series forecasting model with the standard Lee-Carter model in forecasting Nigerian male mortality.

The contents of the remainder of this article are structured as follows. In Section 2, we present data and methods. This includes Lee-Carter model and our proposed functional time series forecasting model. In section 3, we discussed application of the two models to Nigeria Male mortality. In section 4, we applied the models to our data with the results of analysis presented. In section 5, we present discussion of results while Section 6 is summary and conclusion.

2. Data and Methods

Our data for this research work is a secondary data on male mortality obtained from Nigeria Bureau of Statistics from 1998 to 2010 from age zero to age 85 covering five years interval. In our analysis we shall use R- Statistical Packages to analyze our data. The statistical tools to be applied to model our data are Lee-Carter model (LC) and the proposed model Known as Functional Data Analysis (FDA).

2.1. The Lee-Carter Forecasting Model

In 1992 Lee-Carter developed a mortality model for modeling and forecasting US population. The model is a numerical algorithm used in mortality forecasting and life expectancy. The input to the model is a matrix of age specific mortality rates ordered monotonically by time, usually with ages in columns and years in row with the model given as

$$\ln(m_{x,t)} = \alpha_x + b_x k_t + \epsilon_{x,t} \tag{1}$$

or equivalently $m_{x,t} = \exp\{\alpha_x + b_x k_t + \epsilon_{x,t}\}$ (2)

Where, k_t– time parameters indexed by $t = 1, 2, \dots, T$; α_x, b_x - age - specific parameters indexed by $x = 0, 2, 7, 12 \dots, w$; $\epsilon_{x,t}$– random errors assumed to be **iid; £xt** $\sim N(0, \partial_E^2)$.

To ensure the unique representation of (2) or (3), some additional constraints are imposed (Lee, Carter 1992) in the following way: $\sum_{t=1}^{T} k_t = 0 \ \sum_{x=0}^{w} bx = 1$ so that $ax = \frac{1}{T} \sum_{t=1}^{T} \ln(m_{x,t})$.

Forecasting of mortality using LC model is based on time series prediction on the parameters, k_t. This is done using the equation 3: $\hat{\mu}_x t_{n+s} = \exp(\hat{\alpha}_x + \hat{\beta}_x^{h(0)} it_{n+s-x} + \hat{\beta}^{(1)} kt_{n+s})$, (3)

Where s >0; and it_{n+s-x} and kt_{n+s} represent the forecasted cohort and period effect respectively. To produce mortality forecast, Lee and Carter assumes that β_x remains constant over time and they use forecast of $\hat{k}_t$ from a standard univariate time series model using ARIMA with drift, that is the random walk with drift model for the parameter k_t and the model is given as

$$\hat{k}_t = \hat{k}_{t-1} + \hat{\theta} + \varepsilon_t \tag{4}$$

Where $\varepsilon_t \sim N(0, \sigma_x^2)$ and $\hat{\theta}$ is known as the drift parameter with the maximum likelihood estimate is simply $\hat{\theta} = \hat{k}_{t-1} + \hat{k}_1/t - 1$ which only depends on the first and last of the k estimates.

The performance of the Lee-Carter model has been covered in many studies (see, e.g. Lee, Miller 2001). Under this methodology, parameters, a_x, b_x, k_t are estimated with the empirical data on the considered population. Lee and Carter have used in their paper the Singular Value Decomposition (SVD) method (Lee, Carter 1992). Two other methods of estimation are also proposed, i.e. Weighted Least Squares (WLS) and the Maximum Likelihood (ML) approach (see Wilmoth 1993, Brouhns et al. 2002a,b), the estimated values of k_t form a time series with one value for each year. Because of that, the statistical methods of time series modeling can be employed. Despites the usefulness of the model, it has its own limitations. First the model was developed and used in advance countries where we have enough information. Another limitation is that it assumes errors to be constant with narrow predictive intervals (Wilmoth 1993). In order to overcome this, we proposed a new model known as functional data analysis.

2.2. Functional Data Analysis

In functional data analysis, it is assumed that there is underlying smooth function of age 0n the mortality rate in each year. Let the function $y_t(x_i)$ denote the log of the observed mortality rate of age x at year t. Assume $f_{t,F}(x)$ is the underlying smooth function obtained from the model when X represents the age continuum defined on a finite interval. We observed functional data on a set of grid points and the data are often contaminated by random noise.

$$y_t(x_i) = f_{t,i}(x) + u_{t,i} \ , \ t = 1, \ldots n, \ i = 1, \ldots, p \qquad \textbf{(5)}$$

where n denote the numbers of years and p denote the numbers of discrete data points of age observed for each functions. The errors $(u_{t,i})$ are independently and identically distributed (iid.) random variables with mean zero and variance. Smoothing techniques are thus needed to obtain each function $f_{t,i}(x)$ from a set of realizations. Hence, we used smoothen Spline to observed the smoothing function $f_{t,i}(x)$.

We used the underlying smooth function $f_{t,F}(x)$ obtained from smooth regression as given by equation 2.6observing with error. Hence,

$$y_{t,F}(x_i) = \log\left[\mathfrak{f}_{t,F}(x_i)\right] + \sigma_{t,F}(x_i)e_{t,F,i} \tag{6}$$

Where x_i is the middle of the age-class I, (I =1, . . . , p), and $e_{t,F,i}$ is the error that is standard normal distribution, independently and identically distributed random variable and the value $\sigma_{t,F}$ allows the amount of the noise to vary with the age x, For the smooth of the data, we used p-spline regression, we there after applied product ratio method of functional data of time series analysis to the smoothed data.

We define the product ratio of functional time series analysis of the smoothed as

$$p_t(\text{x}) = \left(\mathcal{F}_{t,M}(x)\mathcal{F}_{t,F}\right)(x)^{1/2} \tag{7}$$

We model these quantities via the smoothed log rate rather than the original sex-specific mortality rates. On the log-scale, these are sums and differences which are approximately uncorrelated (Turkey 1977). We use functional time series models for $p_t(\text{x})$ (Hyndman and Ullah 2007).

$$\text{Log}[p_t(\text{x})] = \mu_p(\text{x}) + \sum_{k=1}^{K}\beta_{t,k}\,\phi_k(\text{x}) + e_t(\text{x}) \tag{8}$$

Where the function $\phi_k(\text{x})$ is the principal components obtained from decomposition of $p_t(\text{x})$ and $\beta_{t,k}$, is the corresponding principal component score. The function $\mu_p(\text{x})$ is the mean of set p_t (x). The error term given by $e_t(\text{x})$ have zero mean and are serially uncorrelated.

3. Application of the Models

We shall apply LC and FDA models to model male mortality data obtained from Nigeria Bureau of statistics from 1998 to 2010. With the application of the models to our data, we compare the forecast accuracy of our model with accuracy of the Lee-Carter forecasting model. We fit each model to the first h observations and then forecast the mortality rates for the observations in years t+1, t+2, t+3, . .., n. The value of t is allowed to vary from n_1 =50 to n-1. Subsequently, the forecast horizon varies from n= 1 to 50. Mean absolute prediction errors (MAPE) and Mean squared prediction errors (MSPE) are used as measure of point forecast errors. For each population, MAPE and MSPE can be calculate as

$$\textbf{MAPE (h)} = \frac{1}{(n-h-n_1+1)p}\sum_{n=h}^{50}\sum_{j=1}^{p}\left|\left[m_{i+h,j}(x_i)\right] - \log\left[\hat{m}_{t+h|t,j}(x_i)\right]\right| \tag{9}$$

$$\text{MSFE (h)} = \text{MSFE (h)} = \frac{1}{(n-h-n_1+1)p} \sum_{h}^{50} \sum_{i=1}^{p} \left([m_{i+h,j}(x_i)] - \log[\widehat{m}_{t+h|t,j}(x_i)]\}^2 \right) \quad (10)$$

Where, $\widehat{m}_{t+h|t,j}$ represent the h-steps ahead predictions using the first n+t-h years fitted in the model, and $m_{i+h,j}(x_i)$ denotes the true value. For our interval forecast, if $\hat{f}^u_{n+h/n}$ and $\hat{f}^i_{n+h/n}$ are the upper and the lower (1-α) 100% prediction bounds and $m_{i+h,j}$ is the realized value, the interval point x_i is given as

$$S_x(x)_j = \left[\hat{f}^u_{n+\frac{h}{n}}(x_i) - \hat{f}^i_{n+\frac{h}{n}}(x_i) \right]$$

$$+ \frac{2}{\alpha)} [\hat{f}^i_{n+h/n}(x_i) - m_{i+h,j}(x_i)] \| \left\{ m_{i+h,j}(x_i) < \hat{f}^i_{n+\frac{h}{n}}(x_i) \right\}$$

$$+ \frac{2}{\alpha)} \left[m_{i+h,j}(x_i) - \hat{f}^u_{n+\frac{h}{n}}(x_i) \right] \| \left\{ m_{i+h,j}(x_i) > \hat{f}^u_{n+\frac{h}{n}}(x_i) \right\} \qquad (11)$$

Where α is the level of significance, the forecast mean interval across the horizon is given as

$$\overline{S_\alpha}(\text{h}) = \frac{1}{(n-h1)p} \sum_{n=h}^{50} \sum_{i=1}^{p} \overline{S_\alpha} \left| \left[\hat{f}^u_{n+\frac{h}{n}}(x_i) \right], \hat{f}^i_{n+\frac{h}{n}}(x_i); m_{i+h,j}(x_i) \right| (12)$$

Where p denotes the number of age groups and h denotes the forecast horizons

4. Analysis and Results

In this section we shall presents the results of the analysis with the models. The results of the analysis are as presented here below.
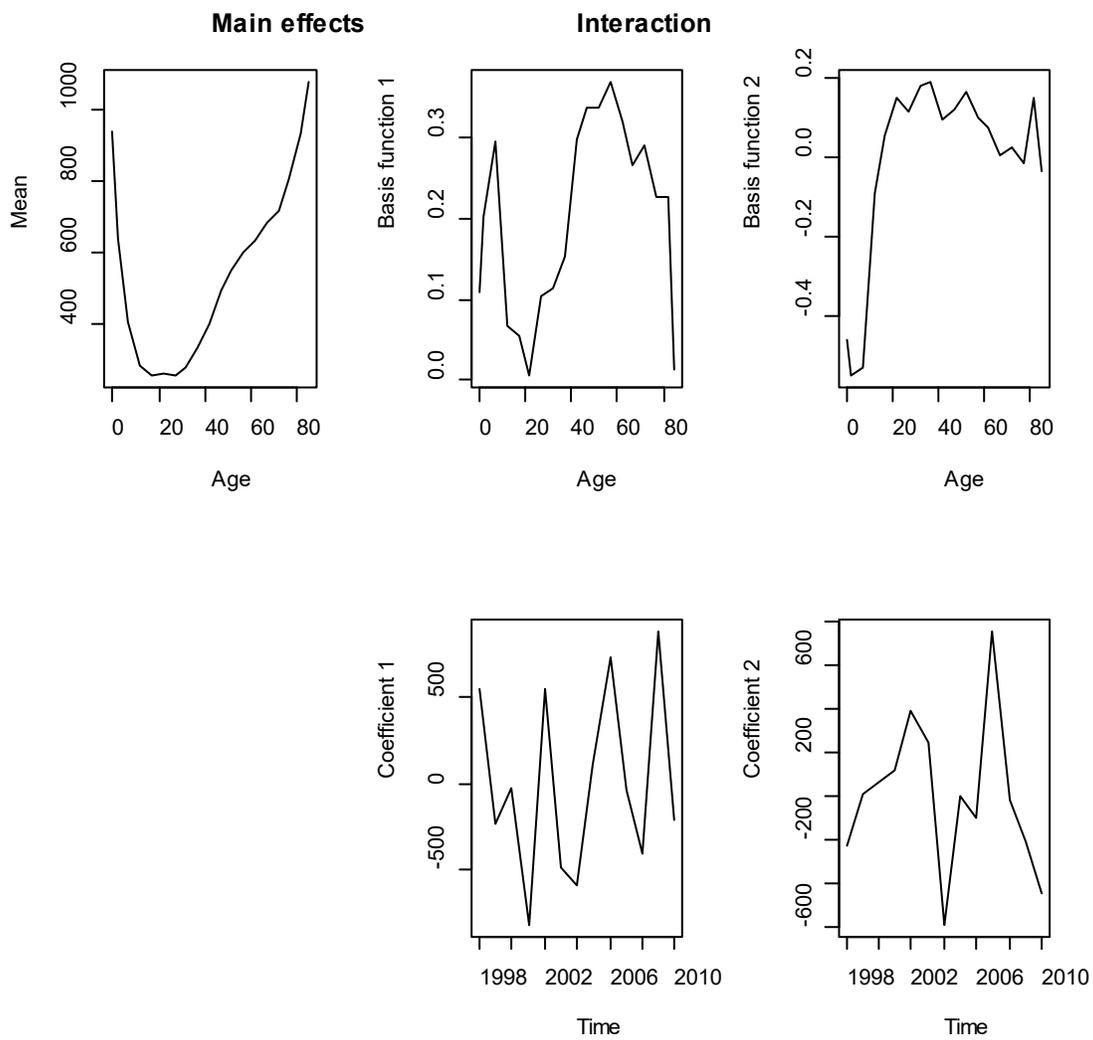
Figure1: Basis and coefficients for male mortality of LC model
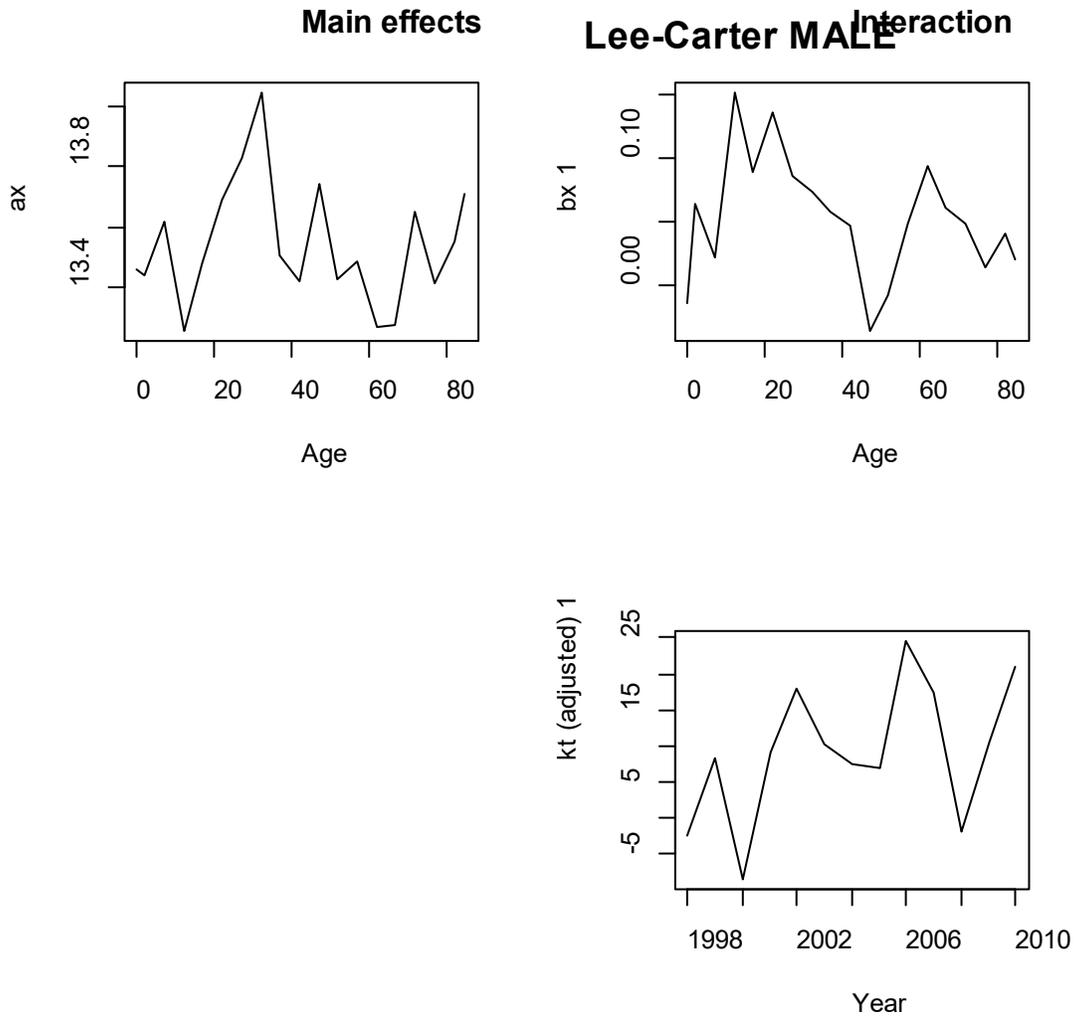
Sources: Data from Nigeria National Bureau of Statistics (1998-2010)

Researchers' computation (2020).

Figure 2: Basis and coefficients for male mortality of FDA model
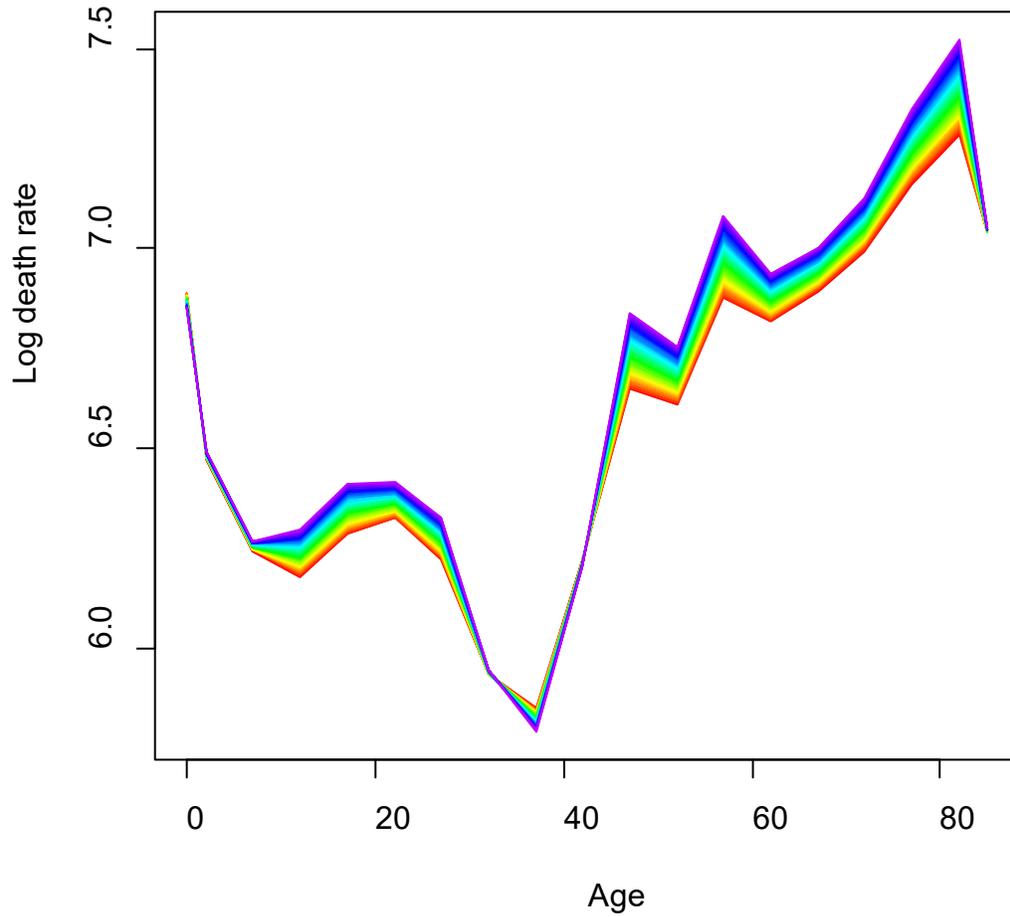

Sources: Data from Nigeria National Bureau of Statistics (1998-2010)

Researchers' computation (2020).

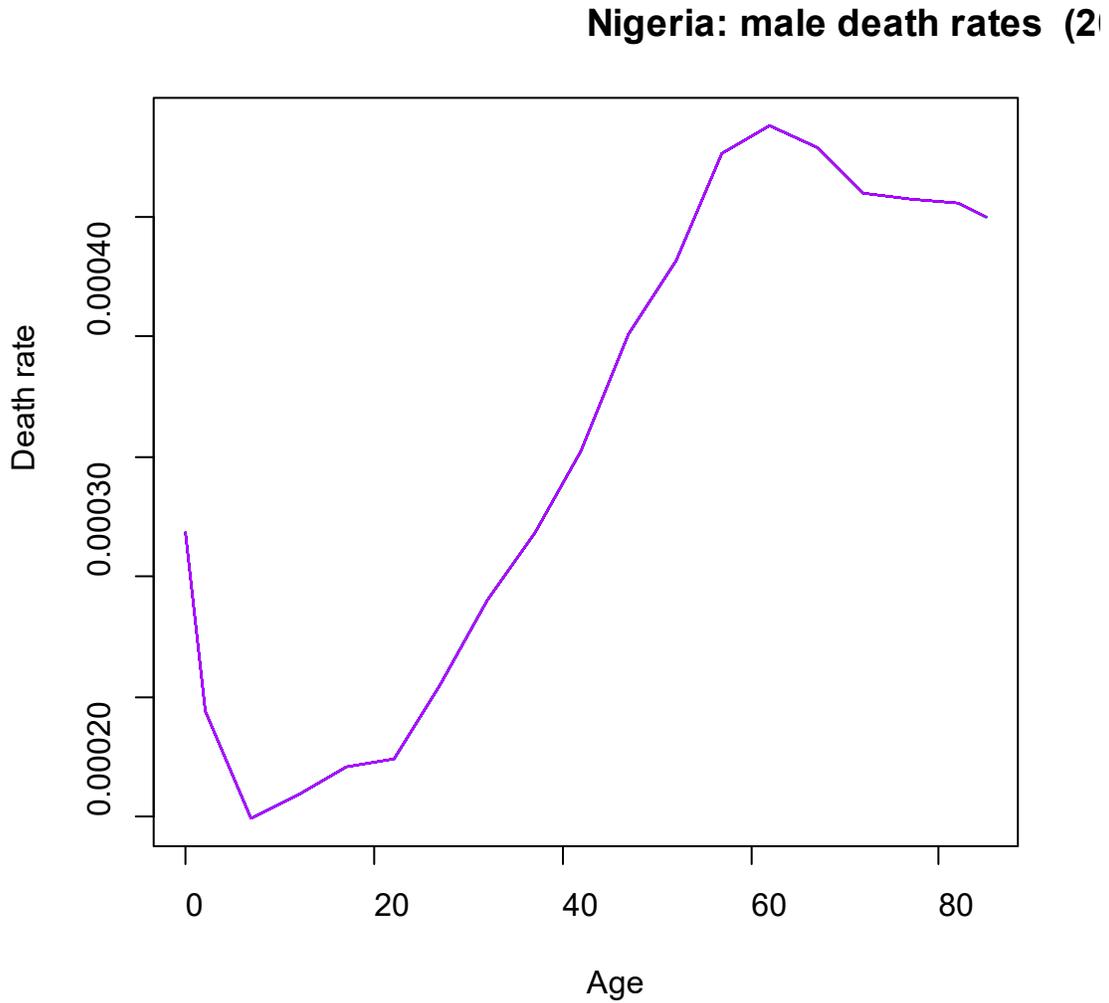**Figure 3:** Forecasted graph for LC mortality model

Nigeria: male death rates (2

Sources: Data from Nigeria National Bureau of Statistics (1998-2010)

Researchers' computation (2020).

Figure 4: Forecasted graph for FDA mortality model

Nigeria: male death rates (2

Table 1: PARAMETERS OF THE MODELS

1.1. Errors Measure Based on Mortality Rates

PARAMETERS	LC	FDA
ME	-6.322	-97.63083
MSE	36.07	0.8567
MPE	0.13388	0.03527
MAPE	0.35167	0.47585

1.2. Errors Measure Based on Log Mortality

PARAMETERS	LC	FDA
ME	0.0302	-0.17778
MSE	0.17277	0.04345
MPE	0.01006	-0.02245
MAPE	0.05202	0.08073

Researchers' computation (2020).

5. Discussion of Results

The graphs in figure one shows that the value of the Lee-Carter parameters, ax, t is the slope of the model at age x in time t. The value of k_t captured the mortality trend at time t, and bx describes the amount of mortality change at a given age for a unit of yearly total mortality change. We can see that the movement of the graphs portrait a zig-sag movements, showing that there have been variations on the volume of mortality experience at different age and years. The graphs in figure two are the parameters of FDA model which also produce the same pattern of movements with LC model. While, figure three and four show that the forecasted graph of LC and FDA model respectively. From the graphs, we can see that our model produced a smooth forecasted graph compare to that of LC graph. The graph also shows that there is a high death rate between age zero and also high at the older ages.

Basing on table one which shows the parameters of the two models. We can see that the errors measures base on death rate of FDA are lower than that of LC model. So, also are the errors obtained, and these errors measures base on log rate except the Maximum Average percentage Error (MAPE) of FDA that is higher than that of LC for both errors of measurement.

6. Summary and Conclusion

We have been able to use the Lee-carter model and our proposed model to analyze Nigeria male mortality data. Though both models produce a good results on male mortality but we have seen that our model has an improvement over the LC model by producing a smooth forecast curve and have least errors for both two errors of measurements except one which is the MAPE. We recommend that the model could be extended to modelling both male and female mortalities and compare the results. The model could also be used to project life table for Nigerian insurance companies.

7. Conflict of interest

There is no conflict of interest to publish this paper, no funding, data are secondary from Nigeria National Bureau of Statistics (1998-2010)

8. References

ALDERS, M., & DE BEER, J. (2005). An expert knowledge approach to stochastic mortality
        Forecasting in the Netherlands. In N. Keilman (Ed.), *Perspectives on Mortality*
        Forecasting (Vol. II. Probabilistic models, pp. 39-64. Swedish Social Insurance.
Booth, H., Hyndman, R. J., Tickle, L. & de Jong, P. (2006), Lee-Carter mortality forecasting:
        Amulti-country comparison of variants & extensions, Demographic Research15 (9),289-310.
Brouhns, N.; Denuit, M.; Vermunt, J.K. A Poisson log-Bilinear regression approach to
        the construction of projected lifetables. Insur. Math. Econ.2002 3, 373–393. [CrossRef]
Carter,L.R.; Lee, R.D, Modeling and Forecasting US sex differentials in Modeling. Int. J.
        Forecast. 1992, 8, 393 – 411.
DeMoivre, A. 1725. Annuities on lives: Or, the valuation of annuities upon any number of lives

as also of reversions. William Person, London

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical transactions of the Royal Society of London 115, 513–583.

Hyndman, R.J.; Booth, H.(2008). Stochastic Population forecasts using functional data models for mortality fertility and migration, Int. J. forecast 2008, 24, 323-342.

Hyndman, R. J. & Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: a functional data approach', Comput. Stat. Data Anal, 2007 51, 4942–4956.

Hyndman, R.J.; Shang, H.L Rainbow plots, bagplot, and boxplots for Functional data. J. Comput. Graph. Stat. 2010, 19, 29-45.

Hyndman, R.J.; Booth, H; Yasmeen, F. Coherent (2013). Mortality forecasting: The product –ratio method with functional time series models. Demography 2013,50, 261-283.

Hyndman, R.J,; Khandakar, Y. Automatic Time Series: The forecast Package for R. J Stat. Softw. 2008, 27, doi: 10.18637/jss.v027,103.

KEILMAN, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics,* 13(3), 245-277.

Lee, R. D. (2000), 'The Lee-Carter method for forecasting mortality, with various extensions and applications', North American Actuarial Journal 4 (1), 80–92.

Lee, R. D. & Carter, L. R. (1992), 'Modeling and forecasting US mortality', Journal of the American Statistical Association 87 (419), 659–671.

Lee, R. D. & Miller, T. (2001a), 'Evaluating the performance of the Lee-Carter method for forecasting mortality', Demography 38 (4), 537–549.

Lee, R. D. & Miller, T. (2001b), 'Evaluating the performance of the Lee-Carter method for forecasting mortality', Demography 38 (4), 537–549.

OLSHANSKY, S. J. (1987). Simultaneous/Multiple Cause-Delay (SIMCAD):an epidemiological approach to projecting mortality. *Journal of Gerontology* 42(4), 358-365

Pearson, K. 1897.The chances of death, and other studies in evolution. Folklore Enterprises, Ltd. 8 (4), 370–372.

Stoto, M., 1983. Accuracy of Population Projections. Journal Statistical Association, 78, 13-20.

Thiele, T. 1872. On a Mathematical Formula to Express the Rate of Mortality Throughout Life Journal of the Institute of Actuaries, 16: 313-329

Tukey, J. W. (1977), Exploratory Data Analysis, Addison-Wesley, London.

URL: http://www.amazon.com/Exploratory-Data-Analysis-Wilder-Tukey/dp/B0007347RW

URL: http://www.sciencemag.org/cgi/reprint/296/5570/1029.pdf

URL:http://hdl.handle.net/10256/742

URL: robjhyndman.com/software/demography

Wilmoth, J. R. (1995), 'Are mortality projections always more pessimistic when disaggregated by cause of death?', Mathematical Population Studies 5 (4), 293–319.

Wittstein, T. and D. Bumsted. 1883. The mathematical law of mortality. Journal of the Institute of Actuaries and Assurance Magazine 24 (3), 153–173.