

ISSN: 2410-1397

Master Dissertation in Actuarial Science

## Modelling and Forecasting Mortality and Longevity Risk Based on Insufficient Data: Kenyan Population

Research Report in Mathematics, Number 18, 2017

Nicholas Kibiwott Bett

August 2017





**Modelling and Forecasting Mortality and Longevity  
Risk Based on Insufficient Data: Kenyan Population  
Research Report in Mathematics, Number 18, 2017**

Nicholas Kibiwott Bett

School of Mathematics  
College of Biological and Physical sciences  
Chiromo, off Riverside Drive  
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the School of Mathematics in partial fulfillment for a degree in Master of Science in Actuarial Science

Prepared for The Director  
Board Postgraduate Studies  
University of Nairobi

Monitored by Director, School of Mathematics

## Abstract

The main risk faced with many life insurance and pensions offices is longevity and mortality risk. The aim of my research is to develop an improved demographic forecasting model of these risks in Kenya. Through my research, I intend to solve the problem of projecting mortality and longevity in presence of insufficient data and HIV/AIDS epidemic which renders the classical Lee-Carter model less coherent. The data under consideration will be obtained from the Kenya National Bureau of Statistics and United Nations. The model to be used will be an extension of Lee-Carter model developed by [Tuljapurkar et al (2004)] which is based on data inadequacy scenarios. Using the Singular Value Decomposition SVD method I fitted the data for the period 1999 to 2015 after testing for structural change and further forecasted using the random walk with drift model for the future periods of 2020 and 2030. The results were observed to be good because the model had as high as 75 percent R squared and Mean Absolute Percentage Error (MAPE) of less than 10 percent. Additionally, a close comparison between my rates with the United Nations 2017 revised ones seems to be almost alike. In application, I considered computing both the assurance and annuity factors using the projected rates to counter mortality and longevity risks.

## Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

---

Signature

Date

**NICHOLAS KIBIWOTT BETT**

Reg No. I56/76185/2014

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

---

Signature

Date

Dr Ivivi Mwaniki  
School of Mathematics,  
University of Nairobi,  
Box 30197, 00100 Nairobi, Kenya.  
E-mail: [ivivimwaniki@uonbi.ac.ke](mailto:ivivimwaniki@uonbi.ac.ke)

## Dedication

This project is dedicated to God and my family.

# Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Declaration and Approval</b> .....	<b>iii</b>
<b>Dedication</b> .....	<b>iv</b>
<b>Figures and Tables</b> .....	<b>vii</b>
<b>Acknowledgments</b> .....	<b>viii</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Background of the Study.....	1
1.1.1 Early Methods: Parametric mortality models.....	1
1.1.2 Mortality Tables.....	2
1.2 Statement of Problem.....	2
1.3 Objectives.....	3
1.4 Significance of Study.....	3
1.5 Scope of the Study .....	3
<b>2 Literature Review</b> .....	<b>5</b>
2.1 Introduction .....	5
2.1.1 Theoretical Review .....	5
2.1.2 Fitting the Mortality Index .....	5
2.1.3 Forecasting mortality Index $k_t$ .....	8
2.1.4 Estimation of Parameters .....	11
2.1.5 Parametric models .....	12
2.1.6 Cohort models.....	13
2.1.7 Life table models.....	14
2.1.8 Model Selection .....	14
<b>3 Methodology</b> .....	<b>15</b>
3.1 Basic Mortality and life table functions.....	15
3.1.1 Force of mortality.....	15
3.1.2 Crude death rate.....	15
3.1.3 Infant mortality rate .....	15
3.1.4 Age specific central death rate $m_x$ .....	16
3.1.5 Age specific death rate $q_x$ .....	16
3.1.6 Future Lifetime $T_x$ .....	16
3.1.7 Complete expected future lifetime $e_x^o$ .....	16
3.1.8 Number of persons alive at exact age $x$ .....	16
3.1.9 The probability of surviving beyond exact age $x+n$ , ${}_n p_x$ .....	16
3.1.10 Average time lived in the interval $x$ to $x+n$ , ${}_n a_x$ .....	16
3.1.11 This is the number of persons years lived, ${}_n L_x$ .....	17

---

3.2	Chow Test .....	17
3.3	Fitting Mortality Index .....	18
3.3.1	Estimation of parameters.....	18
3.3.2	<i>Forecasting Using Data of Unequal Intervals</i> .....	19
<b>4</b>	<b>Data analysis and Discussions .....</b>	<b>21</b>
4.1	Source and description of data.....	21
4.2	Description of computer programs for analysis used.....	21
4.3	General analysis of the data used .....	22
4.3.1	Mortality Rates.....	22
4.3.2	Mortality pattern in Kenya.....	22
4.4	Chow test results .....	24
4.5	Fitting the Model .....	25
4.6	Model Evaluation.....	25
4.6.1	Lee Carter model fitting for total population .....	25
4.7	Forecasting.....	27
4.7.1	Forecasted Age specific central death rates .....	27
4.7.2	Forecasted life expectancy at birth .....	27
4.8	Construction of an abridged life table for the Kenyan Population.....	29
4.9	Applying the projected rates to life assurance and annuity contracts.....	30
4.9.1	Life Assurance Contracts .....	30
4.9.2	Life annuity Contracts .....	35
<b>5</b>	<b>Conclusions and Recommendations .....</b>	<b>37</b>
5.1	Conclusions.....	37
5.2	Recommendations .....	38
	<b>Bibliography.....</b>	<b>39</b>



# Figures and Tables

## Figures

Figure 1. Mortality change 1979-2015 .....	22
Figure 2. Male, Female and Total mortality pattern .....	22
Figure 3. Female life expectancy 1979-2015 .....	23
Figure 4. Male life expectancy 1979-2015 .....	23
Figure 5. Annual crude death rates 1960-2015 .....	24
Figure 6. Annual crude death rates 1999-2015 .....	25
Figure 7. Total population fitted ax, bx and kt .....	27
Figure 8. Life Expectancy Forecasts .....	29

## Tables

Table 1. Fitted bx .....	26
Table 2. Fitted kt.....	26
Table 3. Mean Absolute Percentage Error.....	27
Table 4. R Squared.....	27
Table 5. Predicted $m_x(t)$ , t=2020 and 2030 and x are age .....	28
Table 6. Life expectancy forecasts 2020 and 2030 .....	29
Table 7. Combined life expectancy forecasts United Nations 2017 revised .....	29
Table 8. Combined life table for both male and female populations 2017-2020 .....	31
Table 9. Male life table 2017-2020 .....	32
Table 10. Female life table 2017-2020 .....	33
Table 11. Whole life and Term Assurance factors .....	35
Table 12. Whole life and Term Annuity Factors .....	36

## Acknowledgments

I would like to thank the Almighty God for giving me the knowledge and the guidance to complete this work. It is with great pleasure that I also convey the most profound gratitude to the following good people, who supported me during its preparation. Dr. Ivivi Mwaniki, lecturer and my supervisor. I am forever indebted for your assistance, helpful discussions and ideas that you gave me whenever I approached you. My class and work colleagues, your support and encouragement made the successful completion of the project possible and for this am grateful. My parents Mr. and Mrs. Bett, may the good Lord bless you abundantly for the love, care and support that you have shown me and for believing in me..... Again to you all.

Nicholas Kibiwott Bett

---

University of Nairobi, 2017.

# 1 Introduction

## 1.1 Background of the Study

Since the commencement of the 20th-century human mortality has been on a decelerating trajectory. This phenomenon has become a key subject amongst demographers and actuaries worldwide. Research surrounding this area has gained momentum because of the immense importance of understanding human mortality trends. Government led population studies and also insurance and pension industries have increased research in this area because the findings will enable governments to plan effectively and allocate resources equitably and enable accurate pricing and risk assessments by insurances and pension plans respectively. Despite the numerous research in the area only developed countries have been well represented. On the other hand, developing countries have greatly been left out of this developments. This has led to governments and life insurance sectors to continue using less reliable forecasts of mortality and life expectancy because of acute shortage of quality data that plays a key role in most models that estimate future rates. Kenya relies heavily on international organizations such as United Nations, World Health Organization and World Bank to fill the gap in this area.

### 1.1.1 Early Methods: Parametric mortality models

Mortality law was simply a mathematical expression that described mortality as a function of age. Abraham De Moivre was the earliest recorded mathematician who made the first known contribution about mortality law [De Moivre (1725)]. From that time, numerous contributions in the field continued, notable contributors came up with Gompertz law modelled as an exponential function [Gompertz (1825)] of which [Gavrilov (1991)] defined this law of mortality as one of the most successful models. This model expressed mathematically the aging pattern of mortality a key feature as represented by an exponential pattern. However, the Gompertz model did not fit infant, child and oldest-old mortality, however, it majorly performed well in the middle ages of 30 to 90 years and so [Makeham (1860)] addressed some of these problems by including additional parameters capturing the infant mortality and younger ages. Makeham improved the fit of the Gompertz function by coming up with a Gompertz-Makeham Model the variant term is an age-independent component which captures the extrinsic mortality risk which has a more influential effect at younger ages and is often associated with early adult mortality. [Heligman and Pollard (1980)] came up with a comprehensive model that encompassed eight parameters and having three terms; the first term described the fall of mortality during childhood, the second one represented the hump feature between ages 18-24 years as a

result of accidental deaths and the last term reflected the exponential pattern of mortality at the adult ages. The Gompertz-Makeham function described in [Forfar et al (1988)] generalizes the original models proposed by Gompertz and Makeham this methodology is normally used by the Continuous Mortality Investigation Bureau CMIB to produce mortality tables. The framework used is that of linear and non-linear models. Several reviews have been written on the subject like [Benjamin and Soliman (1993)]. Despite all these tremendous progress identifiability problems emerged because the model involved parameters that were highly correlated. [Dellaportas et al (2001)] suggest a Bayesian Strategy improves model fit and numerical properties. And in recent studies stochastic mortality models have become preferable. The forecasted rates will, in turn, be used to come up with life tables that would be used by insurers and pension plans to correctly price their products in as far as the next 50 years. Abridged life tables are often used unlike the single year life tables they include methods on stable population [Coale and Demeny (1966)], the estimation of population by age, [Irwin (1988)], methods that exploit the relationship between life expectancy and demographic indices, [Mazur (1969)], methods based on the construction of abridged life tables [Chiang (1984)] and Brass-type relational methods [Brass (1971)].

### **1.1.2 Mortality Tables**

A key problem in estimating life tables in developing countries is scanty mortality data. This fact makes it hard for users to have full confidence and validity of the values to be applied and eventually projected. The Kenyan insurance industry until recently lacked mortality tables developed from its own mortality experience. And so many Kenyan insurance companies and pension schemes based their mortality basis on life tables of other countries like South Africa and the United Kingdom like the SA85-90 and A1949-52 respectively. As a result of this inadequacy, the first mortality tables specifically constructed out of the experience of assured lives in Kenya was constructed in 2009 that is the KE2001-03. In addition, the study also suggested the need to adjust regularly the mortality tables used by insurance companies so as to ensure that the mortality assumptions match mortality experience of the Kenyan population. According to The Actuarial Society of Kenya (TASK) one of the most cumbersome tasks of the mortality study and morbidity study was that data was inconsistent and incomplete in most cases and this actually brought concern of the quality of data and therefore approximations and estimations were necessitated (TASK).

## **1.2 Statement of Problem**

There is need to come up with reliable mortality forecasts despite insufficient data and catastrophes like the HIV/AIDS epidemic, this is because Lee Carter Model does not work well in areas with unreliable data and mortality shocks. This realisation frustrates the effort of coming up with accurate forecasts to be used by insurance and pension providers

---

in the country. According to a report by [OECD (2014)] failure to represent for future enhancements in mortality can expose pension and annuity providers to an expected shortfall of provisions their liabilities. Reliable methods of projecting future mortality rates is a crucial pillar of the life insurance and pensions industry.

### 1.3 Objectives

#### Overall

Model and forecast the future mortality and life expectancy rates based on insufficient data in Kenya.

#### Specific

- Estimate mortality index for the Kenyan data (fitting)
- Forecast mortality rates and life expectancy for the Kenyan population
- Construct abridged life table for the Kenyan population
- Apply projected rates to life assurance and annuity contracts for selected ages

### 1.4 Significance of Study

The gap that has been left as a result of relying on stochastic mortality models like the classical Lee Carter method is very critical in that these models are only useful for developed countries. The authors allude to that fact that these models work best where data is reliable and mortality shocks are not significant. It is, therefore, significant to perform such a study in this country referred to as developing country which faces almost these challenges of data inadequacy and mortality shocks so as to have well fitted and forecasted data that will play a role in both insurance and pensions industries to enhance viability and maintenance of a clean financial health. The potential and overall impact are immense because new and similar industries will be able to set up shop in Kenya with increased confidence of stability of profits and growth. The government also gains to benefit because it will plan more effectively in the allocation of resources to the population with the future in mind.

### 1.5 Scope of the Study

During a longevity conference in 2015 one of the lead author of the Lee- Carter model, Ronald Lee, stated that data inadequacy and AIDS epidemic in many Sub-Saharan countries make the use of the model cumbersome. Vital and census data from most of the sub-Saharan countries are insufficient Kenya being one of the countries is not left out. My study will focus on the periods of 1979-2015. HIV/AIDS on the other hand also results in coherency limitations during forecasting. This is because of the realization of the epidemic that has really affected mortality stability in form of structural distortion [Chow (1960)]. In the second chapter, I intend to give a review of previous literature that fall under this

topic and models developed so far. The third chapter will look at the methods employed for modelling demographic risk and also the construction of the life table. In chapter four I will analyze the data and give key findings and discussions and finally in chapter five I will present my conclusions and recommendations.

## 2 Literature Review

### 2.1 Introduction

This chapter looks into various past literature about stochastic mortality models used. It summarises the empirical findings of the various models under consideration.

#### 2.1.1 Theoretical Review

Mortality projections can be constructed using two approaches: Causal or explanatory models that basically explain the details of inherent mortality factors and secondly by use of extrapolative models that make use of the past data to forecast the future. According to [Oeppen and Vaupel (2002)] extrapolation methods of life expectancy has repeatedly outperformed causal based projections. These methodologies have both advantages and inadequacies. Extrapolative strategies have a tendency to depend heavily on past experience to model the future trend. However, they include less judgment in the model development process. Causal death models give more noteworthy knowledge into the potential reasons for verifiable mortality and mortality change shifts. However, the techniques utilized in the aggregation of the individual causes results to an all-cause mortality assumption which has raised doubts at times [Global mortality experience (2011)]. Insurance companies are influenced by a wide range of risks. For instance, life insurance faces two main risks: the investment risk and the demographic risk. Demographic risk can be categorized into insurance and mortality risk, mortality risk can also be extended to refer to longevity risk, living longer than expected.

#### 2.1.2 Fitting the Mortality Index

##### Lee Carter Model (LC)

[Lee-Carter (1992)] considers available information and builds up a stochastic model of mortality, which accounts for a mortality level that varies over time as per a single index. The new technique proposed is extrapolative and does not consider medicinal, behavioral or social factors on mortality change. The qualities are that it combines a rich yet parsimonious model of analyzing mortality. It is anchored firmly on long term past trends from 1900 to 1988 for the case of United States (US) information, and it gives probabilistic certainty regions to its forecasts. This strategy varies from others in that it permits age-specific death rates to decay exponentially unbounded. Further, [Lee (2000)] a follow-up

to the paper [Lee-Carter (1992)] lays out the extensions, constraints and applications of the Lee-Carter model. The technique can also be featured with disaggregation by sex, geographical locations, cause, and lower limits for death rates. The model structure is given by:

$$\ln(m_x(t)) = \alpha_x + \beta_x k_t + \varepsilon_{x,t} \quad (1)$$

$$\varepsilon_{x,t} \sim N(0, \sigma^2)$$

$m_x(t)$ : observed central death rate

$\alpha_x$ : mean age-specific level of mortality

$k_t$ : A time-trend index of general mortality level

$\beta_x$ : relationship factor from the age of profile as the  $k_t$  varies

$\varepsilon_{x,t}$ : The error term at age  $x$  and time  $t$ .

The time index  $k_t$  captures the key time trend on the logarithmic central death rates at every age. The age element  $\beta_x$  alters the main time pattern in accordance to the rate of change at a specific age against the main trend. The model assumes that  $\beta_x$  does not change over time. However, the LC model in equation (1) is overly parameterized in the extent that it leads to identifiability problems and so to obtain a unique solution the following modification must be made to the equation:

Two constraints are imposed,

$$\sum_x b_x = 1 \text{ and } \sum_t k_t = 0 \quad (2)$$

$\alpha_x$ , is obtained to be the average over time of the natural log of  $m_x(t)$  given as:

$$\alpha_x = \frac{1}{T} \sum_{t=t_1}^T \ln(m_x(t)) \quad (3)$$



[Lee-Carter (1992)] used a random walk with drift model which is an ARIMA (0,1,0) model which is expressed as:

$$k_t = k_{t-1} + \mu + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, \sigma^2) \quad (4)$$

Where  $\mu$  is the drift parameter and quantifies the mean annual change in the series, and  $\varepsilon_t$  is an uncorrelated error. The RWD model provides good outcomes in many scenarios [Tuljapurkar et al (2000), Lee-Miller (2001)]. From the forecast of the principal component scores, the projected age-specific log mortality rates are arrived at using the computed age parameters,  $\alpha_x$  and  $\beta_x$ .

### **Lee-Miller LM Model**

It is an extension of the LC method, and differentiates itself from LC model in the following ways:

1. Fitting period begins in the second half of the twentieth century (1950).
2. Adjustment of  $k_t$  requires fitting to the life expectancy  $e_0$  in year  $t$ .
3. The jump-off rates are the real rates in the jump-off year in place of the fitted values.

In their assessment of the LC strategy, [Lee-Miller (2001)] found a mismatch between fitted rates for the most recent year of the fitting time frame and real rates in that year; this jump-off error added up to 0.6 years in future for males and females combined.

Actual rates were used to remove the jump-off error in the year of jump-off. Furthermore, the level of change in death rates was not steady over time, which is a key assumption in the LC technique. As a result, the alteration of historical principal component scores added up a larger error. To by-pass this, [Lee-Miller (2001)] embraced 1950 as the beginning year of the fitting time frame because of different age patterns of change for 1900-1949 and 1950-1995. This arrangement has been used by [Tuljapurkar et al (2000)]. Moreover, the change of  $k_t$  was accomplished by fitting observed life expectancy in year  $t$ .

## Lee-Carter Model fitted to limited data

[Tuljapurkar et al (2004)] developed here an extended LC way to deal with circumstances in which mortality information are available at just a few points in time, and at unevenly divided intervals, circumstances frequently experienced in developing nations. [Tuljapurkar et al (2004)] noted China and South Korea as countries with scanty data and attempted to compare to developed countries at that time. The sex-combined age-particular death rates of South Korea were accessible for the years 1972, 1978, and afterward every year for 1983 through to 2000 making a total of twenty data points while for China was for 1974, 1981 and 1990, three data points. The explained ratio of the fitted SVD model was 0.84 and 0.96 respectively, inferring that the adjustments in the age pattern of mortality were more coherent for both countries, however, the relative error value was 0.139 for South Korea which was less than the case for China by around 40 percent despite China having a higher explained ratio. In the presence of fewer years of data (less than 20), these variations are assumed to be random fluctuations. For South Korea, with a period of 28 years, the model is more robust unlike for China.

### 2.1.3 Forecasting mortality Index $k_t$

Forecasting is the process of projecting the future in reference to the historical and present data ["Forecasting"]. It refers to a formal statistical method employing time series data it goes on to state that uncertainty is central to forecasting. In any case, the data must be up to date in order for the forecast to be as accurate as possible.

#### Random Walk with Drift (RWD)

[Lee-Carter (1992)] projected the mortality index using a standard time series model which is an ARIMA model, (0,1,0), also referred to as a random walk with drift RWD. It is given by the following equation

$$k_t = k_{t-1} + \mu + \varepsilon_t \quad (5)$$

Such that  $\varepsilon_t \sim N(0, \sigma^2)$

$k_t$  represents the mortality index at time t,

$\mu$  is the drift term and

$\varepsilon_t$  is the error term

## Properties of Random Walk with Drift

In principle  $\mu$  may be a positive or a negative value. We will use back substitution to derive the key properties of this model and start from the main equation;

$$\begin{aligned}
 k_t &= k_{t-1} + \mu + \varepsilon_t \text{ For } t > 0 \\
 \text{But } k_{t-1} &= k_{t-2} + \mu + \varepsilon_{t-1} \text{ For } t > 1 \\
 \text{Replacing, we get } k_t &= k_{t-2} + 2\mu + \varepsilon_t + \varepsilon_{t-1} \\
 &\vdots \\
 \text{And so } k_t &= k_0 + t\mu + \sum_{i=0}^{t-1} \varepsilon_{t-i} \text{ for } t > 0
 \end{aligned}$$

We set the initial condition  $k_0=0$ , then we get

$$E[k_t] = \mu t \text{ and } Var[k_t] = t\sigma^2 \quad (6)$$

since  $\sum_{i=0}^{t-1} \varepsilon_{t-i}$  are iid's. We note that the mean and the variance are both non-stationary time series because they depend on time  $t$ .

## Cairns-Blake-Dowd (CBD) Model

The Cairns-Blake-Dowd, the CBD model, was designed for modelling mortality at higher ages. Especially post retirement ages that are around age 60 years and above in the United Kingdom. The two-factor model is a key variant of the Lee-Carter model is given by:

$$\log it(q_{x,t}) = k_t^1 + k_t^2(x - \bar{x}) \quad (7)$$

The logit function,  $\log it(x) = \log\left(\frac{x}{1-x}\right)$ ,  $x \in (0, 1)$ , assumes the linearity of the logit of one-year death probabilities at advanced ages. For a given year the logit of the one-year death probability takes the intercept and slope parameters across years as a stochastic process. The  $k_t^1$  represents the level of the logit-transformed mortality curve and  $k_t^2$  represents the steepness of the logit transformed mortality curve. This article also examined the advancement of post-age 60 mortality curve in the UK and its effect on the pricing of longevity risk. The principal factor influences death rate progression at every age similarly, while the second factor influences death rate elements at higher ages. The suggested alteration includes an allowance for the underlying stochastic mortality and compensation for the parameter risk. According to [Cairns et al (2006)], mortality risk ought to be taken to mean all forms of uncertainty in future mortality rates. Longevity risk ought to be deciphered as uncertainty in the long haul trend in death rates and survival function of an individual and short-term risk ought to be translated as the risk that, over a brief timeframe, death rates are significantly higher or lower than would normally be expected. Examples of such include the HIV/AIDS epidemic of 1984 in Kenya which can be taken into account. Once the epidemic has passed, we expect mortality rates to resume the previous levels. This model, however, is suitable for countries in the developed countries where data for the post-60 year population is available and complete unlike our situation in Kenya.

### **P splines Smoothing Model**

[Currie et al (2004)] developed another stochastic mortality model applying P-splines model which uses penalized splines to impose smoothness across years and ages and then account for that when fitting. A critical component of this technique is that forecasting is a characteristic result of the smoothing procedure. In this model, the force of mortality is assumed to be modeled as a linear association of smooth functions across age and time the model is given below:

$$\ln(m_{x,t}) = \sum_{ij} \theta^{ij} \beta_{x,t}^{ij}(\mathbf{8})$$

$\theta$  refers to the rate of mortality for  $i$ th and  $j$ th entry in the matrix

$\beta$  Refers to B-spline's consisting of cubic polynomial components for  $i$ th and  $j$ th entry in the matrix.

### Booth Maindonald Smith (BMS) method

Being a variation of the LC technique, the BMS strategy varies as follows:

1. Fitting time frame is resolved on the premise of a factual goodness of fit criteria, under the presumption that the essential part where  $k_t$ 's are linear.
2. The alteration of  $k_t$  includes fitting to the age dispersion of deaths instead of the aggregate deaths.
3. The jump-off rates are the fitted rates.

[Booth et al (2002)] observed the linear time series to be traded off by structural distortion. By first expecting the linearity of vital part scores. The Booth et al strategy looks to accomplish the ideal goodness of fit by choosing the ideal fitting time frame from all conceivable fitting periods.

#### 2.1.4 Estimation of Parameters

Estimated parameters once obtained are interpretable and a basic RWD has by and large been proper for the single extrapolated parameter. The technique proposed includes less subjective in enhancements to the Lee-Carter estimation premise. [Wilmoth (1993)] displayed a weighted least squares for fitting the Lee-Carter:

#### Weighted Least Squares

The model is given by:

$$\log q_x(t) = a_x + b_x \kappa_t + \varepsilon_x(t), \text{ and he needed to estimate } a_x, b_x \text{ and } \kappa_t$$

.

To achieve a unique solution the following restrictions were used;

$$\sum_{x=\min}^{x=m} b_x^2 = 1 \text{ (a) } \sum_{t=\min}^{t=n} \kappa_t = 0 \text{ (b) ,}$$

these restrictions make no modifications to the model.

The estimation is calculated from the observed one-year death probabilities. To get a starting value, we assume that is independent of  $x$ , where  $m$  is the number of ages or

age groups. In the classic paper by Lee and Carter, they make a re-estimation to get the observed number of deaths equivalent to the fitted number of deaths, that is

$$D_t = \sum_x \exp(a_x + b_x \kappa_t) N_{x,t}$$

Where  $D_t$  is the total number of deaths in the year  $t$  and  $N_{x,t}$  is the population of age  $x$  in the year  $t$ . No analytic method is accessible so it must be done by seeking over a range of  $k$ . Be that as it may, this second phase of estimation does not have any effect on whether the Lee-Carter model could be fitted to the information or not.

### Maximum Likelihood Estimation MLE

An MLE solution that is tied down on the assumption of a Poisson distribution of deaths was developed; this Poisson log-bilinear model has in this way been connected by; [Brouhns et al (2002)]. [Koissi and Shapiro (2006)] executed not a very clear way of Lee-Carter to address infringement of the Lee-Carter assumption of constant error variance crosswise over age. [Wolf (2004)] built up a Lee-Carter variation connected to first differences in log-death rates that combine estimation of the Lee-Carter and time series models. [Li and Chan (2005)] proposed an outlier-adjusted Lee-Carter method that reinforces the estimation.

#### 2.1.5 Parametric models

They incorporate parameterized functions, for example, the laws of mortality, and models inside GLM's. Non-parametric strategies are additionally connected, for example, the principal components approach which addresses the dimensionality issue. Time series methods are oftentimes applied in extrapolative forecasting. They possess the benefit of being stochastic, hence assist in the computation of probabilistic expectation intervals for the predicted values. On account of a zero-factor underlying model, ARIMA representation has usually been applied. [Alho and Spencer (2005)] suggest that the data series ought to be longer than the predicted timeframe, and ideally twice or thrice as long. Among the one-factor models used are parameterized functions and the Brass relational model. A two-factor model utilized as a part of mortality estimating is the Lee-Carter method. The Lee-Carter model can be reached out to incorporate higher order terms and so doing greater flexibility is achieved in forecasting and modelling univariate ARIMA representations given by [Renshaw and Haberman (2003)] as:

$$\ln(m_x(t)) = \alpha_x + \beta_x^1 k_t^1 + \beta_x^2 k_t^2 + \dots + \beta_x^p k_t^p + \varepsilon_{x,t} \text{(9)}$$

### 2.1.6 Cohort models

The one and two-factor models might be connected to cohort data to come up with models of mortality over the life time. In the instance of one and-two factor models, the cohort approach has the benefit of being free of "tempo" misrepresentations caused by changes in timing: when mean age at death is getting larger [Bongaarts and Feeney (2002)]. In principle, cohort models give a superior premise in predicting than period models. The main constraint of all cohort models is that they require substantial data demands and relevance. These issues are lessened when just part of the age range is complete. However, vague data lead to statistical problems in modelling [Renshaw and Haberman (2006)]. One main example model is the age period cohort model expressed by:

$$\ln(\mu_x(t)) = \alpha_x + \beta_x^1 k_t + \beta_x^2 k_{t-x} + z_{x,t} \text{(10)}$$

$x$ 's are ages available in the data such that:  $x \in [x_{\min}, \dots, x_{\max}]$   $t$ 's are the observed periods such that:  $t \in [t_{\min}, \dots, t_{\max}]$  Under the assumption:  $\mu_{x+k_1, t+k_2} = \mu_{x,t} = m_x(t) = \frac{d_x(t)}{E_x(t)}$  for  $0 \leq k_1, k_2 < 1$  is the unsmoothed death rates.  $d_x(t)$  is the realized number of deaths aged  $x$  at time  $t$  while

$E_x(t)$  is the comparable observed exposure to risk.

### 2.1.7 Life table models

International organisations like the United Nations UN, World Health Organisation WHO and the World Bank comes up with their forecasts for different regions in the world. The methods used produces only model life tables which are just estimations of [Coale and Demeny (1966)] regional model life tables where the North Model is common for developing countries like Kenya. Other notable ones include the [United Nations (1982)]. [Olivieri (2001)] in their paper dealt with the application of forecasted mortality rates in life insurance, specifically to life annuities and term assurances the Heligman-Pollard principle is adopted in these projections.

In Kenya notable life tables developed is the KE2001-03 which was adopted by insurance and pensions industries after usage of England and Wales's life tables A1949-52 and South Africa's SA 1985-90. This KE2001-03 was constructed basing on data available in the year 2001 and 2003. The Principle of correspondence was the key assumption. They used census reports to come up with the crude rates and later performed graduation to smoothen the crude death rates. HIV/AIDS was acknowledged to have fundamental effect on the mortality rates and so two options were presented one with no HIV/AIDS and the other with HIV/AIDS. For the former SA 85-90 was incorporated as the base level prior to smoothening. For ages between 55-100 years (older ages), female mortality was assumed to represent 80 percent of the male rates, this assumption was because of the lower death rates noticed among females.

### 2.1.8 Model Selection

From our literature review the Singular Value Decomposition, the Weighted Least Square method and the Maximum Likelihood Estimation methodology are the three widely used methods for obtaining the model's parameters during fitting. [Koissi et al (2005)] reviewed these methods and concluded that the Singular Value Decomposition was the best alternative for obtaining the mortality index  $k_t$ .



## 3 Methodology

### 3.1 Basic Mortality and life table functions

#### 3.1.1 Force of mortality

The force of mortality  $\mu_x(t)$  at age  $x$  in year  $t$  is defined as:

$$\mu_x(t) = \lim_{\Delta x \rightarrow 0} \frac{[x < T_{0,(t-x)} \leq x + \Delta x | T_{0,(t-x)} > x]}{\Delta x} \quad (11)$$

Where  $T_{0,(t-x)}$  is the outstanding life time of an individual born at time  $t-x$ . This individual will die at age  $x + T_{0,(t-x)}$  in year  $t + T_{0,(t-x)}$

Based on the assumption:

$$\mu_{x+k_1,t+k_2} = \mu_{x,t}, \quad 0 \leq k_1, k_2 < 1 \quad (12)$$

Uniform distribution of death UDD.

#### 3.1.2 Crude death rate

Ratio of the number of deaths in a year to the total population at mid-year multiplied by 1,000

#### 3.1.3 Infant mortality rate

Number of infant deaths under one year old per 1,000 live births.

### 3.1.4 Age specific central death rate $m_x$

Ratio of the number of deaths in the year to the average or mid-year total number of person years lived over the year

$$m_x = d_x / L_x \quad (13)$$

### 3.1.5 Age specific death rate $q_x$

The initial rate of mortality  $q_x$  measures the probability of death over the next year of person aged  $x$ , it applies to the age at the start of the interval.

$$q_x = d_x / l_x \quad (14)$$

### 3.1.6 Future Lifetime $T_x$

It is a random variable accounting for the future life time of an individual aged  $x$ .

### 3.1.7 Complete expected future lifetime $e_x^o$

$e_x^o$  in continuous time refers to the complete expectation of future lifetime after age  $x$ , so that, for a life of exactly aged  $x$ , the probable age at death is  $(x + e_x^o)$

$e_x^o = [T_x]$  where  $T_x$  has a pdf given by  $f_x(t) = {}_t p_x \mu_{x+t}$

$e_x$  is the discrete case and is approximated in terms of the continuous case as:  $e_x = e_x^o - 1/2$

### 3.1.8 Number of persons alive at exact age $x$

It is denoted as  $l_x$ .  $l_0$  is the radix and by convention, it starts as 1, 1000 or 100,000.

### 3.1.9 The probability of surviving beyond exact age $x+n$ , ${}_n p_x$

The probability of surviving between exact ages  $x$  and  $x+n$  it is the complement of  ${}_n q_x$

$${}_n p_x = l_{x+n} / l_x = 1 - {}_n q_x \quad (15)$$

### 3.1.10 Average time lived in the interval $x$ to $x+n$ , ${}_n a_x$

It is the average share of the time lived in the interval  $x$  to  $x+n$  by the deceased during that interval. According to [Newell (1988)] it is common to use  $a_o = 0.1$  in low mortality countries and use  $a_o = 0.3$  for high mortality countries. We will use 0.3 for the Kenyan case.

### 3.1.11 This is the number of persons years lived, ${}_nL_x$

Number of person years lived between exact age  $x$  and  $x+n$

$${}_nL_x = n(l_{x+n} + {}_na_x * {}_nd_x) \quad (16)$$

## 3.2 Chow Test

Suppose we model our data as:

$$y_t = a + bx + e_t \quad (17)$$

splitting our data into two parts we get:  $y_{t_1} = a_1 + b_1x + e_{t_1} \dots \dots 1$  and  $y_{t_2} = a_2 + b_2x + e_{t_2} \dots \dots 2$ .

Where  $e_t$  's is i.i.d with  $N(0, \sigma^2)$

Hypothesis:

$$H_0: a_1 = a_2, \text{ and } b_1 = b_2$$

$$H_1: a_1 \neq a_2, b_1 \neq b_2$$

Let  $ST$  be sum of squared residuals from the pooled groups

Let  $S_1$  be sum of squared residuals from the first group

Let  $S_2$  be sum of squared residuals from the second group

Further, let  $N_1$  and  $N_2$  be the number of observed realizations in each group and  $k$  be the total number of parameters,  $k=2$ .

And so the Chow test, at 95 percent level, is an approximation of an F test distribution and is given in statistic form as:

$$[(ST - (S_1 + S_2))/K]/(S_1 + S_2)/(N_1 + N_2 - 2K) \sim F(K, N_1 + N_2 - 2K)$$

### 3.3 Fitting Mortality Index

#### 3.3.1 Estimation of parameters

Parameter vector estimate is obtained as the mean over time of the logarithm of the central rate of death.

$$\alpha_x = \frac{1}{T} \sum_{t=t_1}^T \ln(m_x(t)) \quad (18)$$

Then apply SVD to the matrix  $Z_{x,t}$  such that

$$Z_{x,t} = \ln(m_x(t)) - \hat{\alpha}_x \quad (19)$$

In order to obtain the matrix in the following form:

$$ULV' = SVD(Z_{x,t}) = L_1U_{x1}V_{t1} + L_2U_{x2}V_{t2} + \dots + L_XU_{xX}V_{tX} \quad (20)$$

Approximating the first term gives the following estimates:

$\hat{\beta}_x = U_{x1}$  and  $\hat{k}_t = L_1V_{t1}$ . The following are the Singular Value Decomposition steps:

1.  $\hat{\alpha}_x = \frac{1}{T} \sum_{t=t_1}^T \ln(m_x(t))$

2. Formulate a matrix  $Z_{x,t}$  for obtaining  $\beta_x$  and  $k_t$  where

$$Z_{x,t} = \ln(m_x(t)) - \hat{\alpha}_x = \beta_x k_t$$

3. Implement SVD to the matrix  $Z_{x,t}$  which disintegrates the matrix of  $Z_{x,t}$  to the product of three matrices:

$$ULV' = SVD(Z_{x,t}) = L_1U_{x1}V_{t1} + L_2U_{x2}V_{t2} + \dots + L_XU_{xX}V_{tX}$$

U accounts for the age element, L accounts for the singular values and V accounts for the time element

4. By running the SVD using the `lca()` function in R program we obtain the following estimates:

$\hat{k}_t = L_1 V_{t1}$  is the first vector of time element matrix,  $\hat{\beta}_x = U_{x1}$  is the first vector of the age element matrix

5. Lca () function approximates a further matrix  $\hat{Z}_{x,t}$  by the product of the estimated parameters  $\hat{\beta}_x$  and  $\hat{k}_t$  to obtain:

$$\hat{Z}_{x,t} = \hat{\beta}_x \hat{k}_t$$

6. Approximate the logarithm of central death rate  $\hat{\beta}_x, \hat{k}(u_0), \hat{k}(u_1), \hat{k}(u_2), \hat{k}(u_3), \hat{k}(u_4), \hat{k}(u_5), \hat{k}(u_6)$  and  $\hat{k}(u_7)$  by the following equation:

$$\ln(\hat{m}_{x,t}) = \alpha_x + \hat{\beta}_x \hat{k}_t \quad (21)$$

### 3.3.2 Forecasting Using Data of Unequal Intervals

One of the key assumption is that a RWD model holds because of the nature of a stable decrease in mortality rates over time which is basically an ARIMA (0,1,0) model given by:

$$k_t = k_{t-1} + \mu + \varepsilon_t \text{ Such that } \varepsilon_t \sim N(0, \sigma^2),$$

$k_t$  accounts for the mortality index at time t,  $\mu$  is the drift term and  $\varepsilon_t$  is the error term. we are working on unequal intervals  $k_t$  will be represented as:

$$k(u_t - u_{t-1}) = \mu [u_t - u_{t-1}] + \sigma [\varepsilon_{u_{t-1}+1} + \varepsilon_{u_{t-1}+2} + \dots + \varepsilon_{u_t}] \quad (22)$$

where  $\varepsilon_t/s \sim N(0, \sigma^2)$

The drift is estimated as:

$$\hat{\mu} = \left\{ \frac{\sum_{t=1}^T \{k(u_t) - k(u_{t-1})\}}{\sum_{t=1}^T (u_t - u_{t-1})} \right\} = \frac{k(u_T) - k(u_0)}{u_T - u_0}$$

Due to lack of independence the variance is derived as:

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \{[k(u_t) - k(u_{t-1})] - \hat{\mu} [u_t - u_{t-1}]\}^2}{\{u_T - u_0\} - \frac{\sum_{t=1}^T [u_t - u_{t-1}]^2}{u_T - u_0}}$$

Mean Forecasts can be carried out for  $t > T$  using:

$$k_t = k_T + \mu(t - T) + \hat{\sigma} \sum_{s=T+1}^t e(s) \quad (25)$$

And subsequently obtain the logarithm of the central death rates estimates by:

$$\ln(m_{x,t}) = \ln(m_{x,T}) + \beta_x [k(t) - k(T)] \quad (26)$$

---

## 4 Data analysis and Discussions

### 4.1 Source and description of data

My research contains data of age specific mortality rates derived from Kenya National Bureau of Statistics KNBS, World Health Organisation, WHO and [United Nations (2015)] Mortality data in Kenya is obtained from census data which has been collected every 10 years: 1979,1989,1999,2009 by KNBS for the periods: 1979, 1989, 1999 and 2009. These rates are adjusted for HIV/AIDS and analysed for completeness after which they are then presented as analytical reports published immediately after analyzing and compiling the census reports. Additionally World Health Organisation obtains their own data separately for the periods 2010-2015, their data is regarded as of good quality. Additionally, my research will also include established estimates of crude mortality rates from combined population from the United Nations since 1960 to 2015 for purposes of checking for structural distortions. The age specific central death rates for male, female and total population will be in form of abridged life table and limited to age 80+ years that is 0, 1-4 and 5-79 (quinquennial) and 80+. To maintain uniformity in our data, we have maintained in annual periods having eight points.

### 4.2 Description of computer programs for analysis used

We have equally used Microsoft Excel and R program to perform most of the data analysis such as computations and summary of data. Specific functions of excel and installed R packages such as demography were key in estimating parameters used.

#### **Assumptions**

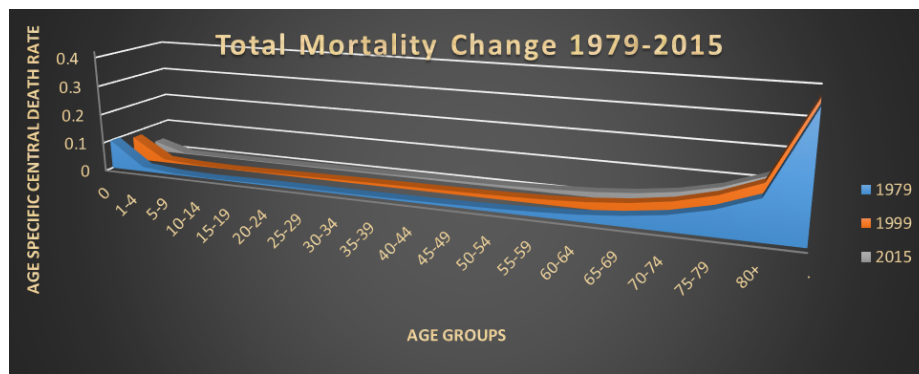
General key assumptions for this model are:-

- Future mortality trends will be a representation of historical trend shown by the data
- High epidemic scenario HIV/AIDS presence

-Random walk with drift model was assumed to follow the mortality pattern

## 4.3 General analysis of the data used

### 4.3.1 Mortality Rates

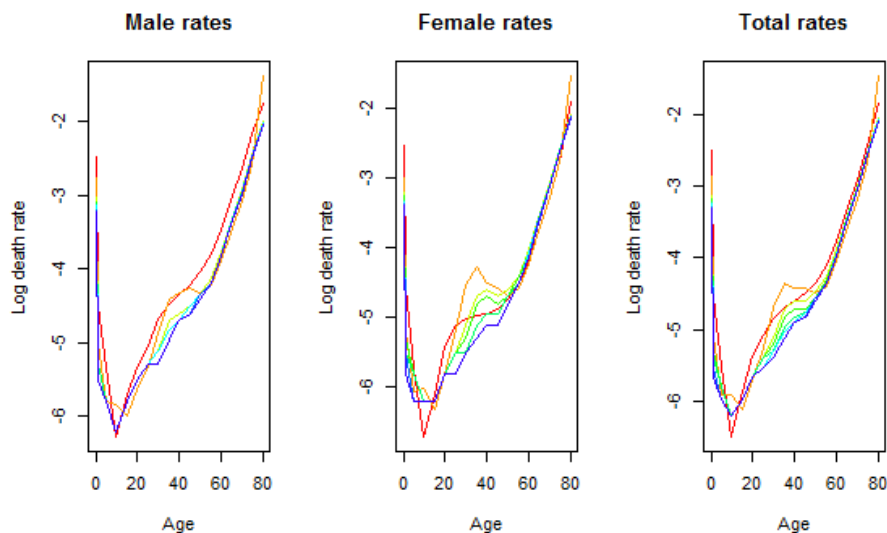


1.png 1.png

Figure 1. Mortality change 1979-2015

Over time the mortality rates seem to be on a downward trend. The red area represents the period between 1990-1999 when mortality rates increased due to deaths related to HIV/AIDS.

### 4.3.2 Mortality pattern in Kenya

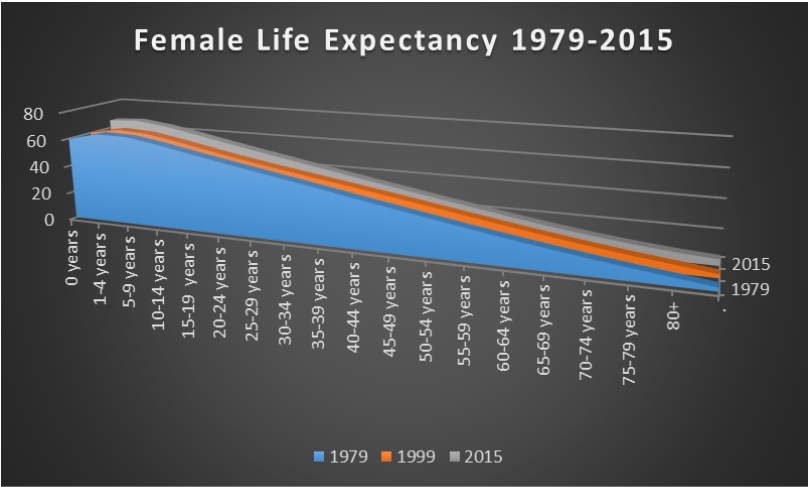


2.png 2.png

Figure 2. Male, Female and Total mortality pattern

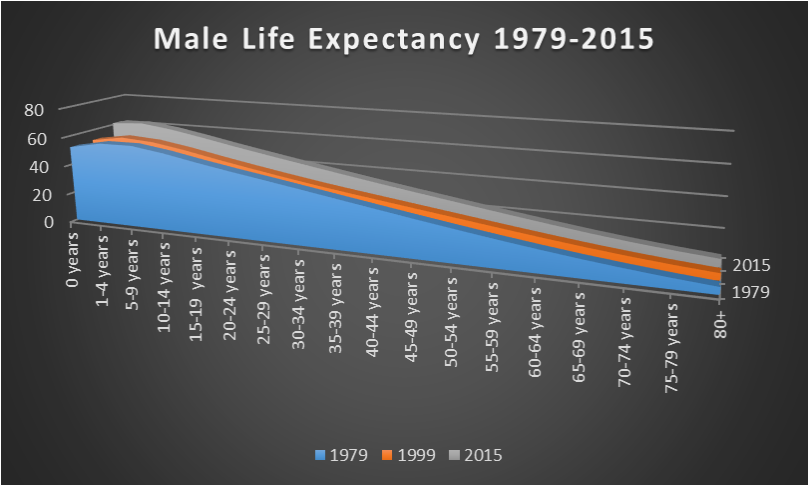


To demonstrate the improvement in Kenyan mortality in further we can check, we have plotted age group specific logs of central death rates for Kenyan male, female and total populations. Overall there has been a decline in mortality rates in Kenya during the period 1979-2015 represented by the different colour schemes. Blue being the latest and red being the oldest. Orange lines represent the highest mortality for ages 24-45 years.



3.png 3.png

Figure 3. Female life expectancy 1979-2015



4.png 4.png

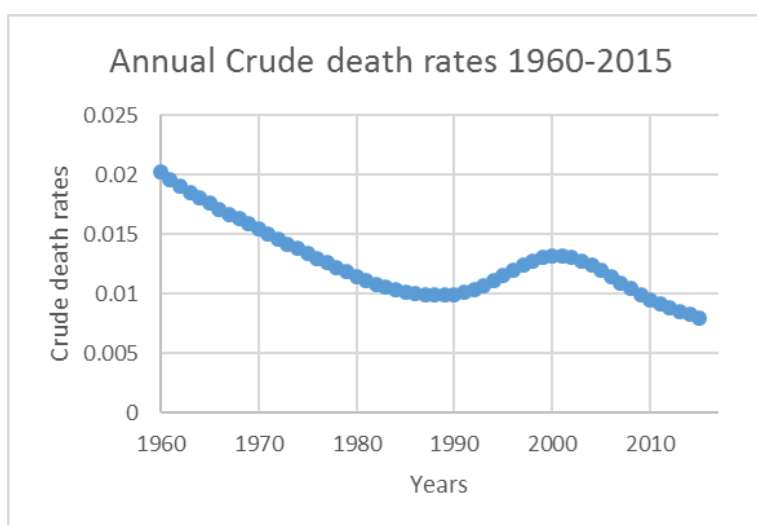
Figure 4. Male life expectancy 1979-2015

Life expectancy is also shown to be improving throughout this period for both male and female populations. The 1990's is the only period with a decrease because of the HIV/AIDS epidemic, however, the upward trend is again beginning to manifest this is evidenced by

[Cairns et al (2006)] mortality shock description of short term risk because of epidemics.

#### 4.4 Chow test results

From our data, crude death rates given annually for the years 1960-2015 were obtained from United Nations Development and Population. I decided to pick 1999 as a separation point because of the highest death rates experienced, approximately 180,000 died from causes related to HIV/AIDS this is according to [Global Aids Epidemic Report (2000)] and also around this period HIV/AIDS was regarded as a national disaster in Kenya. Therefore our separated structural periods will be S1960-1999, S2000-2015 and S1960-2015.

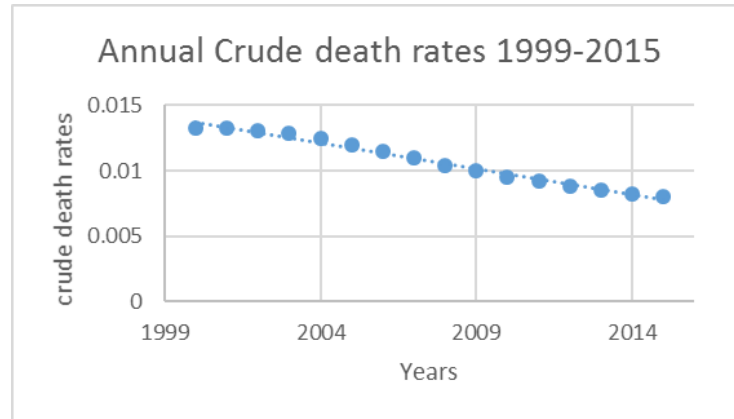


5.png 5.png

**Figure 5. Annual crude death rates 1960-2015**

Running the test we obtain the computed F Statistic value to be 17.964 and our tabulated value of F (2, 56) to be 3.15 at 95 percent level of significance and so we reject the null hypothesis and assume there is indeed a break point and result in a structural change in terms of mortality trend between 1960 and 2015. From this test, we can note that HIV/AIDS has brought a significant alteration of the mortality pattern. I, therefore, decide to forego the periods [1960, 1999) and concentrate for the period 1999-2015 since this point is also a break point. Graphical result is as fig 6 and we note the linear trend evident:

As shown from fig 5 and fig 6, the break point test, we are left with only eight points of data: - 1999, 2009, 2010-2015. Clearly, these points are unequally uniform and by standards of the Lee-Carter model, it is insufficient to fit this kind of data. [Tuljapurkar et al (2004)]



6.png 6.png

**Figure 6. Annual crude death rates 1999-2015**

solves this problem by extending the lee carter model to such situations.

## 4.5 Fitting the Model

Fig 7 shows the total population fitted  $ax$ ,  $bx$  and  $kt$ :

We will apply the LC method based on data with unequal intervals Tuljapurkar et al (2004) and fit our model to the following data set.  $u(0)=1999$ ,  $u(1)=2009$ ,  $u(2)=2010$ ,  $u(3)=2011$ ,  $u(4)=2012$ ,  $u(5)=2013$ ,  $u(6)=2014$  and  $u(7)=u(T)=2015$ . Applying SVD approach on the lee carter model we estimate the following parameters of  $bx$  and  $kt$

We obtain the fitted mortality index,  $kt$  for the period Table 2:

## 4.6 Model Evaluation

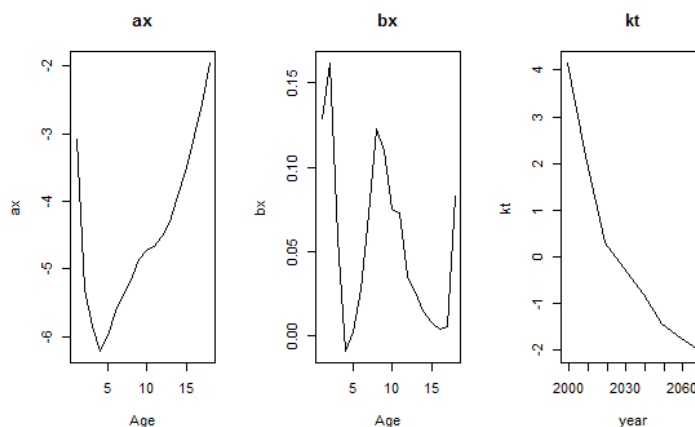
### 4.6.1 Lee Carter model fitting for total population

Table 1. Fitted bx

Age groups(years)	Female	Male	Total
0	0.12598	0.11057	0.12867
1-4	0.15439	0.14873	0.16152
5-9	0.06129	0.06245	0.05824
10-14	-0.02577	0.01014	-0.00893
15-19	-0.00328	0.00363	0.00269
20-24	0.03482	0.01801	0.02802
25-29	0.13415	0.03165	0.07230
30-34	0.16983	0.08791	0.12258
35-39	0.15902	0.00803	0.11001
40-44	0.09674	0.06732	0.07535
45-49	0.08909	0.06558	0.07278
50-54	0.02386	0.04775	0.03568
55-59	-0.00598	0.05006	0.02590
60-64	-0.01631	0.04026	0.01500
65-69	-0.02707	0.03737	0.00759
70-74	-0.03142	0.03620	0.00435
75-79	-0.02676	0.03504	0.00534
80+	0.08742	0.06699	0.08290

Table 2. Fitted kt

Years	female	male	total
u(0)=1999	3.285095	5.124415	4.415705
u(1)=2009	2.619431	1.734368	2.154127
u(2)=2010	2.461997	1.529948	1.968678
u(3)=2011	2.304562	1.325528	1.783229
u(4)=2012	2.147128	1.121108	1.597781
u(5)=2013	1.989694	0.916688	1.412332
u(6)=2014	1.832259	0.712268	1.226883
u(7)=2015	1.674825	0.507848	1.041434



7.png 7.png

Figure 7. Total population fitted ax, bx and kt

Table 3. Mean Absolute Percentage Error

	Total	Male	Female
<b>Mortality rates</b>	0.055	0.051	0.087
<b>Log mortality rates</b>	0.014	0.014	0.02

Lee-Carter analysis

The Mean Absolute Percentage Error are all within the 10 percent bound both for mortality and log mortality rates, Table 3, and according to [Lewis (1982)] it can be interpreted to be a highly accurate forecast. More so, the R squared values are in the region of 70 percent, Table 4, and hence our model seems to explain the level of model variation quite well.

## 4.7 Forecasting

### 4.7.1 Forecasted Age specific central death rates

### 4.7.2 Forecasted life expectancy at birth

Table 4. R Squared

	Total	Male	Female
<b>R squared</b>	0.74	0.76	0.66

Table 5. Predicted  $m_x(t)$ , t=2020 and 2030 and x are age

Predicted $m_x(t)$ Age (years)	2020			2030		
	total	male	female	total	male	female
0	0.0319	0.0341	0.0310	0.0243	0.0248	0.0273
1-4	0.0030	0.0032	0.0028	0.0021	0.0021	0.0024
5-9	0.0024	0.0027	0.0019	0.0021	0.0023	0.0018
10-14	0.0020	0.0020	0.0020	0.0021	0.0019	0.0021
15-19	0.0025	0.0030	0.0020	0.0025	0.0030	0.0020
20-24	0.0034	0.0039	0.0029	0.0032	0.0037	0.0028
25-29	0.0037	0.0048	0.0028	0.0032	0.0044	0.0025
30-34	0.0040	0.0044	0.0037	0.0031	0.0034	0.0031
35-39	0.0053	0.0069	0.0046	0.0042	0.0068	0.0039
40-44	0.0069	0.0082	0.0057	0.0059	0.0067	0.0052
45-49	0.0074	0.0091	0.0057	0.0064	0.0075	0.0052
50-54	0.0096	0.0112	0.0079	0.0089	0.0098	0.0077
55-59	0.0126	0.0140	0.0110	0.0120	0.0121	0.0111
60-64	0.0187	0.0208	0.0161	0.0181	0.0185	0.0164
65-69	0.0293	0.0313	0.0264	0.0288	0.0281	0.0271
70-74	0.0473	0.0484	0.0447	0.0469	0.0436	0.0461
75-79	0.0766	0.0780	0.0730	0.0757	0.0705	0.0750
80+	0.1141	0.1189	0.1129	0.0958	0.0980	0.1034

Table 6. Life expectancy forecasts 2020 and 2030

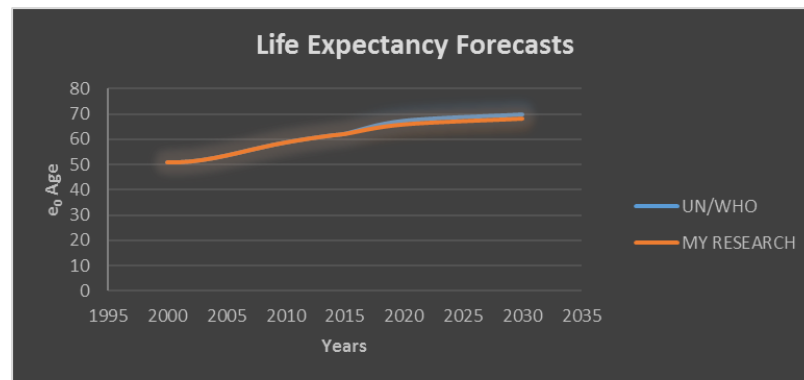
Predicted $m_x(t)$	2020			2030		
	total	male	female	total	male	female
$e_0$	65.9	63.9	67.84	68.21	66.9	68.83

Table 7. Combined life expectancy forecasts United Nations 2017 revised

Period year	2020	2030
$e_0$	67.3	69.8

Using the [Chiang (1984)] method on the age specific central death rates we obtain the life expectancy values,  $e_0$ , at birth at times 2020 and 2030. Life expectancy at x, is given in Table 6:

According to the [United Nations (2017) Revision] forecasts of life expectancy combined total population is given by Table 7:



8.png 8.png

Figure 8. Life Expectancy Forecasts

#### 4.8 Construction of an abridged life table for the Kenyan Population

From the forecasted values of  $k_t$ , the age specific death rates and probabilities of death are determined. The death rates are computed using the calculated b parameter, the death rates and the parameter  $k_t$  on the last year where data is available that is 2015. Due to limitation of reliable data to construct a complete set of tables, I will use the projected age

specific central death rates  $m_x$  to come up with an abridged life table for males, females and combined population for the period 2017-2020 using available data which are in the age groups of five years apart from age 0 to 1 year and age 1 to 4 years up to a maximum of 80+ age interval.

#### Assumptions:

The exact value for  $a_x$  can only be calculated from the full death records. Since this is estimated on forecasted data, it is impossible to have the specific number of deaths and so according to [Newell (1988)] choosing the correct  $a_x$  for young ones is critical, he states that using  $a_0$  to be 0.3 is conventional for high mortality countries which Kenya is among,  $a_1$  to be 0.4 and  $a_x$  to be 0.5 for all  $x > 5$  thereafter. The life table is to be constructed in the following matrix format:

- Age  $x$  years - representing the age groups: 0-1, 1-4, 5-10, ... 75-79, 80+
- Length of the interval of age intervals  $n$
- Average fraction of life lived in the age interval  $x$  to  $x+n$  for the deceased,  ${}_n a_x$
- Forecasted central death rates,  ${}_n m_x$
- Probability of death within  $n$  years,  ${}_n q_x$
- Starting cohort number alive radix,  $l_x$  to 100,000
- Number of deaths,  ${}_n d_x$
- Number of person years lived in the range,  ${}_n L_x$
- Aggregate number of years lived by individuals in cohort at start of age interval,  $T_x$
- Expectation of life, life expectancy,  $e_x$

The life expectancy for those aged 1-4 years is higher than for age 0 which is a typical scenario for developing countries [Canudas-Romo and Becker (2011)].

## 4.9 Applying the projected rates to life assurance and annuity contracts

Using the projected central death rates  $m_x(t)$  we use them to calculate the Actuarial Present Values (APV hereafter) of life assurance and annuity contracts by first obtaining the probabilities of dying at each period of age interval  ${}_t q_x$ .

We assume the Force of mortality  $\mu_x(t) = m_x(t)$  estimation is based on the assumption:  $\mu_{x+k_1, t+k_2} = \mu_{x,t}$ ,  $0 \leq k_1, k_2 < 1$  and let  ${}_t q_x = 1 - {}_t p_x = 1 - \exp(-\mu_x(t)) = 1 - \exp(-m_x(t))$ . We further let  $i$  be the effective rate of interest for one period preferably one year and  $v$  be the present value such that  $v = \frac{1}{1+i}$ :



Table 8. Combined life table for both male and female populations 2017-2020

years	<b>n</b>	$nq_x$	$nm_x$	$nq_x$	$l_x$	$nd_x$	$nL_x$	$T_x$	$e_x$
0-1	1	0.3	0.03187	0.03137	100000.00	3136.66	97804.34	6592489.67	65.9249
1-4	4	0.4	0.00295	0.01174	96863.34	1136.99	384724.58	6494685.33	67.0500
5-9	5	0.5	0.00235	0.01169	95726.35	1118.71	475834.96	6109960.74	63.8274
10-14	5	0.5	0.00202	0.01004	94607.63	950.22	470662.62	5634125.79	59.5525
15-19	5	0.5	0.00249	0.01239	93657.41	1160.16	465386.67	5163463.17	55.1314
20-24	5	0.5	0.00340	0.01685	92497.25	1558.30	458590.52	4698076.49	50.7915
25-29	5	0.5	0.00371	0.01836	90938.96	1669.76	450520.38	4239485.97	46.6190
30-34	5	0.5	0.00395	0.01958	89269.20	1747.68	441976.77	3788965.59	42.4443
35-39	5	0.5	0.00534	0.02636	87521.51	2307.11	431839.78	3346988.82	38.2419
40-44	5	0.5	0.00693	0.03404	85214.40	2900.93	418819.69	2915149.03	34.2096
45-49	5	0.5	0.00741	0.03637	82313.47	2993.51	404083.59	2496329.34	30.3271
50-54	5	0.5	0.00963	0.04701	79319.96	3729.06	387277.15	2092245.75	26.3773
55-59	5	0.5	0.01265	0.06129	75590.90	4632.97	366372.06	1704968.61	22.5552
60-64	5	0.5	0.01870	0.08927	70957.93	6334.44	338953.55	1338596.55	18.8647
65-69	5	0.5	0.02926	0.13612	64623.49	8796.75	301125.58	999643.00	15.4687
70-74	5	0.5	0.04728	0.21055	55826.74	11754.07	249748.53	698517.42	12.5122
75-79	5	0.5	0.07657	0.31808	44072.67	14018.44	185317.25	448768.88	10.1825
80+	5	0.5	0.11408	1	30054.23	30054.23	263451.63	263451.63	8.8

Table 9. Male life table 2017-2020

years	n	$nq_x$	$nm_x$	$nq_x$	$l_x$	$nd_x$	$nL_x$	$T_x$	$e_x$
0-1	1	0.3	0.03410	0.03353	100000.00	3352.73	97653.09	6391878.60	63.9188
1-4	4	0.4	0.00323	0.01283	96647.27	1239.72	383613.76	6294225.51	65.1257
5-9	5	0.5	0.00274	0.01361	95407.55	1298.90	473790.50	5910611.75	61.9512
10-14	5	0.5	0.00197	0.00981	94108.65	922.86	468236.09	5436821.25	57.7717
15-19	5	0.5	0.00298	0.01481	93185.79	1380.16	462478.53	4968585.16	53.3191
20-24	5	0.5	0.00390	0.01930	91805.63	1771.71	454598.85	4506106.63	49.0831
25-29	5	0.5	0.00478	0.02360	90033.92	2124.90	444857.33	4051507.78	44.9998
30-34	5	0.5	0.00440	0.02178	87909.02	1914.79	434758.11	3606650.44	41.0271
35-39	5	0.5	0.00692	0.03401	85994.23	2924.24	422660.55	3171892.33	36.8849
40-44	5	0.5	0.00817	0.04001	83069.99	3323.83	407040.39	2749231.78	33.0954
45-49	5	0.5	0.00910	0.04447	79746.17	3546.09	389865.61	2342191.39	29.3706
50-54	5	0.5	0.01120	0.05447	76200.08	4150.34	370624.53	1952325.77	25.6210
55-59	5	0.5	0.01395	0.06740	72049.73	4855.83	348109.11	1581701.24	21.9529
60-64	5	0.5	0.02076	0.09859	67193.91	6624.55	319408.16	1233592.13	18.3587
65-69	5	0.5	0.03127	0.14473	60569.35	8766.28	280931.06	914183.97	15.0932
70-74	5	0.5	0.04841	0.21496	51803.07	11135.80	231175.85	633252.91	12.2242
75-79	5	0.5	0.07796	0.32280	40667.27	13127.48	170517.67	402077.06	9.8870
80+	5	0.5	0.11893	1	27539.80	27539.80	231559.39	231559.39	8.4

Table 10. Female life table 2017-2020

years	n	$nq_x$	$nm_x$	$nQ_x$	$l_x$	$nd_x$	$nL_x$	$T_x$	$e_x$
0-1	1	0.3	0.03097	0.03050	100000.00	3049.82	97865.12	6784549.87	67.8455
1-4	4	0.4	0.00278	0.01104	96950.18	1070.48	385231.54	6686684.75	68.9703
5-9	5	0.5	0.00194	0.00965	95879.69	925.18	477085.50	6301453.21	65.7225
10-14	5	0.5	0.00203	0.01008	94954.51	957.08	472379.83	5824367.70	61.3385
15-19	5	0.5	0.00200	0.00997	93997.42	936.83	467645.05	5351987.87	56.9376
20-24	5	0.5	0.00295	0.01463	93060.60	1361.61	461898.97	4884342.82	52.4856
25-29	5	0.5	0.00280	0.01392	91698.99	1276.72	455303.15	4422443.84	48.2278
30-34	5	0.5	0.00367	0.01819	90422.27	1645.16	447998.43	3967140.70	43.8735
35-39	5	0.5	0.00462	0.02281	88777.10	2025.29	438822.31	3519142.27	39.6402
40-44	5	0.5	0.00571	0.02817	86751.82	2443.82	427649.55	3080319.96	35.5073
45-49	5	0.5	0.00574	0.02828	84308.00	2384.00	415580.00	2652670.41	31.4640
50-54	5	0.5	0.00790	0.03875	81924.00	3174.70	401683.24	2237090.42	27.3069
55-59	5	0.5	0.01103	0.05367	78749.30	4226.61	383179.97	1835407.18	23.3070
60-64	5	0.5	0.01613	0.07749	74522.69	5774.92	358176.15	1452227.21	19.4870
65-69	5	0.5	0.02636	0.12347	68747.77	8488.19	322518.37	1094051.07	15.9140
70-74	5	0.5	0.04470	0.20029	60259.58	12069.40	271124.40	771532.70	12.8035
75-79	5	0.5	0.07298	0.30572	48190.18	14732.69	204119.17	500408.30	10.3840
80+	5	0.5	0.11292	1	33457.49	33457.49	296289.14	296289.14	8.9

### 4.9.1 Life Assurance Contracts

The key variable of interest is the time until death random variable for a person aged  $x$  and is denoted by  $T(x)$ , the distribution function of  $T(x)$  is given by:

$$F_{T(x)}(t) = P[(x) \leq t] = P[X - x \leq t | X - x] = {}_tq_x \quad (27)$$

Where  ${}_tq_x$  is the probability that a person aged  $x$  dies in the age interval  $(x, x+t)$ . A discrete case is a random variable denoted by  $K(x)$  and is linked to the continuous case  $T(x)$  however it is defined as the largest integer strictly smaller than  $T(x)$  Chavan et al (2016). For  $x=0,1,2,3,\dots$ . The probability mass function of  $K(x)$  is given by

$$P[K(x) = k] = P[k < T(x) \leq k + 1] = {}_k p_x * q_{x+k} \quad (28)$$

where  ${}_k p_x = 1 - {}_k q_x$

Therefore the APV of a whole life assurance contract payable at the end of the year of death of person aged  $x$  years, denoted by  $A_x$  and its expression is given by:

$$A_x = \sum_{k=0}^{\infty} v^{k+1} * {}_k p_x * q_{x+k} \quad (29)$$

The APV of a term life assurance contract payable at the end of the year of death of person aged  $x$  years and limited for a period of  $n$  years, denoted by  $A_{x:1:\overline{n}|}$  its expressed as

$$A_{x:1:\overline{n}|} = \sum_{k=0}^{n-1} v^{k+1} * {}_k p_x * q_{x+k} \quad (30)$$

The tables below represent the values of  $A_x$  and  $A_{x:1:\overline{n}|}$  calculated at 10 percent effective rate of interest p.a. 10 percent represents the rate of Central Bank of Kenya rate as at 1st of July 2017.

There is a reduction of assurance factors over the period 2020-2030 to depict improvement in mortality, however, there is a reduction of assurance factors over the age to show increase in risk due to mortality which naturally increases with age. It is possible to

**Table 11. Whole life and Term Assurance factors**

APV	Gender	Year	Age (x)			
			20	30	40	50
Whole life Assurance	Male	2020	0.0661	0.0907	0.1343	0.1970
		2030	0.0600	0.0801	0.1182	0.1793
	Female	2020	0.0486	0.0703	0.1051	0.1678
		2030	0.0451	0.0646	0.1017	0.1682
Temporary Life Assurance	Male	2020	0.0360	0.0758		
		2030	0.0333	0.0661		
	Female	2020	0.0248	0.0568		
		2030	0.0217	0.0509		

compute other types of assurance factors like deferred assurance, endowment and pure endowment using combinations of whole life and term assurance factors. Additionally one can also interpolate to obtain assurance factors of ages between 20-30, 30-40 and 40-50.

**4.9.2 Life annuity Contracts**

The APV of a whole life annuity contract of one unit payable at the beginning of each year until death of an individual aged x is given by:

$$\ddot{a}_x = \sum_{k=0}^{\infty} v^k * {}_k p_x \tag{31}$$

The APV of a Term life annuity contract payable at the beginning of each year until death or survivor to n-year term of person aged x years is given by  $\ddot{a}_{x:\overline{n}|}$  is expressed as

$$\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^k * {}_k p_x \tag{32}$$

The table below represents the values of  $\ddot{a}_x$  and  $\ddot{a}_{x:\overline{n}|}$  calculated at 10 percent effective rate of interest p.a. 10 percent represents the rate of Central Bank of Kenya rate as at 1st of

**Table 12. Whole life and Term Annuity Factors**

APV	Gender	Year	Age (x)			
			50	55	60	65
Whole life life annuity	Male	2020	2.989	2.839	2.630	2.397
		2030	3.030	2.883	2.674	2.413
	Female	2020	3.063	2.907	2.698	2.420
		2030	3.062	2.901	2.689	2.408
Temporary Life Annuity	Male	2020	2.714	2.647	2.532	2.370
		2030	2.738	2.676	2.566	2.413
	Female	2020	2.762	2.697	2.590	2.4197
		2030	2.763	2.695	2.583	2.408

July 2017. From table 11 and 12 we see that the annuity factors increases with period 2020-2030 to show that there is an increase in longevity risk, people living more years than expected. To counter this risk the annuity factors are increasing over the periods. Additionally, using the same argument as for the assurance factors, I can also obtain annuity contracts such as deferred annuities and also annuities payable more than once in a year by Woolhouse's approximation formula.

---

## 5 Conclusions and Recommendations

### 5.1 Conclusions

This study computed the forecasts of the log mortality rates of the corresponding age-specific death rates for male, female and total population. The computation showed that the first interval for infants, the mortality observed is significantly higher than the rest of the age groups. The second interval is for early childhood. The log rates of the age specific death rates for male and female from 1979 up to 2015 were used in obtaining matrix  $Z$  as a precondition in performing the Singular Value Decomposition. The parameter  $ax$  was obtained by taking the average over time for the natural logarithm of the age specific mortality rates. The significant difference of the mortality changes between the male and female was also shown in the study and they seem to converge at some point in the future. It is observed that at most age intervals of adulthood, male deaths generally occur more than female deaths however old age mortality is uncertain because of extreme limitation of data this phenomenon cuts across both developed and developing nations. Looking at the general mortality account of the male, female and aggregate population, the arrangement of these general index tend to diminish, despite the fact that not monotonically, after some time. The Singular Value Decomposition (SVD) was used to determine the parameters of  $bx$  and  $kt$ . For the parameter  $bx$ , higher values appeared in the 0-4 year interval, which meant that, in such interval, mortality varies significantly when the general mortality index  $kt$  changes. The older ages showed lower parameters which mean that mortality slightly varies during that period. New age specific death rates were computed using RWD because of the assumption of steady decline of mortality which was evident in the data and the forecasts were up to the year 2030, this maximum period was selected because, if the forecast period is prolonged uncertainty would increase and according to [Alho and Spencer (2005)] data time frame should always be lengthy than the predicted bounds. The life expectancy forecasts were compared with the United Nations forecasts revised in 2017 and the obtained rates from my research and a close resemblance is noticeable. From the forecasted values we note that the life expectancy at age 0 is lower than at age 1 which is a common occurrence among third world countries however this trend seems to be disappearing probably because of reduction of infant mortality rates and improvement of healthcare systems in the country. Most developed countries passed this stage towards the end of the twentieth century. From my study the Kenyan life table was created up to a maximum age interval of 80+ years for the period 2017 to 2020 and the forecasted mortality rates were applied as a mortality basis in conjunction with a 10 percent Central Bank of Kenya Rate to obtain life assurance and annuity factors useful for the life insurance and pension providers. Mortality and

longevity risk which form parts of demographic risk are catered for from the new rates developed and other than this rates we expect the implementers to employ conservatism because of the continuous improvement that was not captured properly by this model.

## **5.2 Recommendations**

My research involved application of Lee Carter model with limited data and high epidemic (HIV/AIDS) situation to the Kenyan population only. As a further study one can take up this model to model and forecast mortality rates for multi-population such as for similar countries that face similar mortality experience like the sub-Saharan region by models proposed by Cairns et al. Smoothness and graduation test for our life table should be carried out to ensure premiums for assurance and annuities of subsequent ages are priced correctly and develop select period mortality rates ours have only taken into account ultimate rates.



## Bibliography

- [Alho and Spencer (2005)] Alho, J. M., and Spencer, B. D., Statistical Demography and Forecasting, *Springer*, (2005).
- [Bell (1997)] Bell, W. R., Comparing and Assessing Time Series Methods for Forecasting Age Specific Demographic Rates, *Journal of Official Statistics*, 13, (1997), 279-303.
- [Benjamin and Soliman (1993)] Benjamin, B., and Soliman, A. S., Mortality on the move: methods of mortality projection, (1993).
- [Bongaarts and Feeney (2002)] Bongaarts, J., and Feeney, G., “How long do we live?”, *Population and Development Review*, 28(1), (2002), 13-29.
- [Booth et al (2002)] Booth, H., Maindonald, J., and Smith, L., Applying Lee-Carter under conditions of variable mortality decline, *Population Studies*, 56(3), (2002), 325–336.
- [Brass (1971)] Brass, W., On the scale of mortality. In W. Brass (Ed.), Biological Aspects of Demography, *Taylor and Francis*, (1971), 69-110.
- [Brouhns et al (2002)] Brouhns, N., Denuit, M., and Vermunt, J. K., A Poisson log-bilinear regression approach to the construction of projected lifetables, *Insurance: Mathematics and Economics*, 31(3), (2002), 373–393.
- [Cairns et al (2009)] Andrew, J.G., Cairns, Blake, D., Dowd, k., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I., A Qualitative Comparison of Stochastic Mortality Models using Data from England and Wales and the United States, (2009).
- [Cairns et al (2006)] Cairns, A.J., Blake, D., and Dowd, K., A tw-factor model for stochastic mortality with parameter uncertainty: theory and calibration, *Journal of Risk and Insurance*, 73(4), (2006), 687-718.
- [Canudas-Romo and Becker (2011)] Canudas-Romo, V., and Becker, S., The crossover between life expectancies at birth and at age one: The imbalance in the life table, *Demographic Research*, 24(4), (2011), 113-144.
- [Chiang (1984)] Chiang, C. L., The Life Table and its Applications, *Malabar, Robert E. Krieger Publishing Company*, (1984).
- [Chow (1960)] Chow, G. C., Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, (1960), 591-605.
- [Coale and Demeny (1966)] Coale, A.J., and Demeny, P. G., Regional model life tables and stable populations, *Princeton University Press*, (1966).

- 
- [Currie et al (2004)] Currie, I.D., Durban, M., and Eilers, P.H.C., Smoothing and Forecasting Mortality Rates, *Statistical Modelling*, 4, 2004, 279-298.
- [De Moivre (1725)] De Moivre, A., Annuities upon Lives, or, the Valuation of Annuities upon any Number of Lives, as also, of Reversions to which is added, an Appendix Concerning the Expectations of Life, and Probabilities of Survivorship, (1725).
- [Dellaportas et al (2001)] Dellaportas, P., Smith, A. F., and Stavropoulos, P., Bayesian analysis of mortality data, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(2), (2001), 275-291.
- ["Forecasting"] In Wikipedia, "Forecasting", *The Free Encyclopedia*, (2017, August 23).
- [Forfar et al (1988)] Forfar, Cheon, M. A., and Wilkie, A. D., On graduation by mathematical formula, *Journal of the Institute of Actuaries*, 115(1),(1988), 97-245.
- [Gavrilov (1991)] Gavrilov, L. A., and Gavrilova, N. S., The biology of life span: a quantitative approach,(1991).
- [Global Aids Epidemic Report (2000)] Joint United Nations Programme on HIV/AIDS, Report on the global HIV/AIDS epidemic, June 2000. Joint United Nations Programme on HIV/AIDS, UNAIDS.
- [Global mortality experience (2011)] Purushotham, M., Valdez, E., and Wu, H., Global mortality improvement experience and projection techniques, *Society of Actuaries Report*, (2011).
- [Gompertz (1825)] Gompertz, B., 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,(1825), 513-583.
- [Heligman and Pollard (1980)] Heligman, L., and Pollard, J. H., The age pattern of mortality, *Journal of the Institute of Actuaries*, 107(1, No 434), (1980), 49-80.
- [Irwin (1988)] Irwin, R., 1980-1986 Intercensal population estimates by race, sex, and age, (1988).
- [Koissi and Shapiro (2006)] Koissi, M. C., Shapiro, A., and Högnäs, G., Fitting and Forecasting Mortality Rates for Nordic Countries Using the Lee-Carter method, *Actuarial Research Clearing House*, Vol.1(21), (2005).
- [Koissi et al (2005)] Koissi, M.C., Shapiro, A.F., and Hognas, G., Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Interval, *Insurance: Mathematics and Economics*, (2005).
- [Lee (2000)] Lee, R., The Lee-Carter method for forecasting mortality with various extensions and applications, *North American actuarial journal*, 4(1), (2000), 80-91.

- 
- [Li and Chan (2005)] Li, S., and Chan, W., The Lee-Carter Model for Forecasting Mortality Revisited, (2005).
- [Lee-Carter (1992)] Lee, R., and Carter, L., Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association*, 87, (1992), 659-671.
- [Lee-Miller (2001)] Lee, R. D. and Miller, T., Evaluating the performance of the Lee-Carter method for forecasting mortality, *Demography*, 38(4), (2001), 537-549.
- [Lewis (1982)] Lewis, C.D., Industrial and business forecasting methods, *Butterworths*, (1982).
- [Makeham (1860)] Makeham, W., On the Law of Mortality and the Construction of Annuity Tables, *The Assurance Magazine and Journal of the Institute of Actuaries*, 8(6), (1860), 301-310.
- [Mazur (1969)] Mazur, D. P., Expectancy of life at birth in 36 nationalities of the Soviet Union: 1958-60, *Population Studies*, 23(2), (1969), 225-246.
- [Newell (1988)] Newell, Methods and Models in Demography, *Belhaven Press London*, (1988).
- [OECD (2014)] OECD, Mortality Assumptions and Longevity Risk: Implications for pension funds and annuity providers, *OECD Publishing*, (2014).
- [Oeppen and Vaupel (2002)] Oeppen, J., and Vaupel, J. W. Broken limits to life expectancy, *Science*, 296, (2002), 1029-31.
- [Olivieri (2001)] Olivieri, A., Uncertainty in Mortality Projections: An Actuarial Perspective, *Insurance: Mathematics and Economics*, 29.2, (2001), 231-245.
- [Renshaw and Haberman (2006)] Renshaw, A.E., and Haberman, S., A cohort-based extension to the Lee-Carter model for mortality reduction factors, *Insurance: Mathematics and Economics*, 38, (2006), 556-570.
- [Renshaw and Haberman (2003)] Renshaw, A.E., and Haberman, S., Lee-Carter Mortality Forecasting: a Parallel Generalized Linear Modelling Approach for England and Wales Mortality Projections, *Journal of the Royal Statistical Society, Series C* 52, (2003), 119-137.
- [Tuljapurkar et al (2004)] Li, N., Lee, R., and Tuljapurkar, S., Using the Lee-Carter Method to Forecast Mortality for Populations with Limited Data. *International Statistical Review*, 72(1), (2004), 19-36.
- [Tuljapurkar et al (2000)] Tuljapurkar, S., Li, N., and Boe, C., A Universal Pattern of Mortality change in the G7 Countries, *Nature*, 405, (2000), 789-92.

- [United Nations (2017) Revision] United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP/248.
- [United Nations (2015)] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects 2015 – Data Booklet (ST/ESA/SER.A/377).
- [United Nations (1982)] United Nations, Model Life Tables for Developing Countries, *Population Studies 77 New York, USA: United Nations Department of International Economic and Social Affairs*,(1982).
- [Wilmoth (1993)] Wilmoth, J.R., Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change, *University of California*, (1993).
- [Wolf (2004)] Wolf, D. A., Another variation on the Lee-Carter model. *Population Association of America*, (2004).