SCIENCE FORUM (JOURNAL OF PURE AND APPLIED SCIENCES) 23 (2023) 469 - 479

Science Forum (Journal of Pure and Applied Sciences)

Journal Homepage: https://atbu.edu.ng/science-forum/

# A Gated Recurrent Unit Aided Lee-Carter Model for mortality projection

**A.U Shelleng[1], H.G Dikko[2], J. Garba[3], M. Abdulkarim[4] and B.B Alhaji[5]**

*Department of Mathematics, Gombe State University, Gombe[1]*
*Department of Statistics, Ahmadu Bello University Zaria, Kaduna[2,3]*
*Department of Computer Science, Ahmadu Bello University Zaria, Kaduna[4]*
*Department of Mathematics, Nigerian Defence Academy, Kaduna[5]*

*ABSTRACT*

*Over the past decades, new models for mortality modelling have been developed. A pioneering and one of the most influential model, among others, is the Lee-Carter (LC) model, developed by Lee and Carter in 1992. The method traditionally used in the LC model shows obvious drawback in describing the shape of future mortality. This work used a more robust approach based on a deep learning technique that differs from the literature. More precisely, a Gated recurrent unit (GRU) was integrated with the Lee–Carter model to increase its capacity for prediction. The LC-GRU model performs well in terms of predicted accuracy. The strength of the gated recurrent unit network lies in its capability to handle the non-linearity inherent in mortality data and long-term dependencies in time series data. The research results shows that the LC-GRU model has minimum errors and thus performs better than the LC model.*

ARTICLE INFO

Received 12 July 2023 Received in Revised Form 11 August 2023
Accepted 11 August 2023
Published 30 October 2023
Available online 30 June 2023

KEYWORDS

*Lee-Carter model,*
*ARIMA,*
*Singular value Decomposition (SVD),*
*Gated recurrent unit and*
*Mortality rate.*

## 1. Introduction

The need to forecast into the future and have a good understanding of what future mortality rates and life expectancy will look like is of major importance to researchers. Numerous methods have been developed recently to project mortality using stochastic models (McNown and Rogers, 1989; Lee and Carter, 1992; Alho, 1990; Currie *et al.,* 2004; Cairns *et al.*, 2006; Renshaw and Haberman, 2006). The Lee-Carter model, created by Lee and Carter in 1992, is one of the most often used models for projecting mortality rates among the several models that have been used to forecast mortality. The model has garnered a lot of interest and has come to be regarded as a benchmark for mortality modeling and projection. For instance, the Lee-Carter model was applied to Japan (Wilmoth, 1993), Australia (Booth *et al.,* 2002), Austria (Carter and Prskawetz, 2001), and Belgium (Tuljapurkar, 2008). (Brouhns *et al.,* 2002).

The Lee-Carter model (LC model) has a number of shortcomings despite its widespread acceptance (Lee, 2000). These shortcomings have been criticized by researchers such as Gutterman and Vanderhoof (1999). Researchers and actuaries have started using more advanced methods for predicting future mortality rates in order to treat longevity risk and predict future mortality rates with more accuracy. With recent improvements in computational ability, artificial intelligence (AI) are back on stage. AI techniques enclose numerous algorithms that learn from vast datasets with numerous attributes, and they are highly effective in spotting obscure and difficult-to-detect patterns.

The lack of wide used of AI approaches in demography is largely attributable to what are frequently referred to as "black boxes," which means that it is impossible to fully comprehend the inner workings of the hidden layer and that these workings are thus challenging to interpret. Additionally, the algorithms utilized in AI are based on data rather than specific hypothesis, which are of greater interest to demographers.

To correct some of the drawbacks of the existing models which includes inability to handle non-linearity inherent in mortality data and long-term dependence in time series, this work combines the Lee-Carter model with a Gated recurrent unit (GRU). The GRU makes it possible to get mortality projections that are more consistent with the variability and nonlinear trends observed in mortality. More specifically, the GRU is set up to handle lengthy data sequences, creating a memory that can hold onto important links between data. In this sense, the time series framework of the GRU enables the prediction of future mortality over time, taking into account the significant influence of the historical mortality pattern and accurately reproducing it in the trend that is predicted. The information will also be preserved over time via the GRU, preventing it from gradually disappearing throughout processing.

## 2. Material and Methods

The data used in this research work were obtained from the United Nation World Population Prospect (WPP) database. Simulated data from Gompertz distribution and Age-specific mortality rates records of Nigeria male adults within the ages of 15 to 59 from 2009 to 2020 were used.

### 2.1 Gated Recurrent Unit aided Lee-Carter model (LC-GRU)

The main components of the Lee-Carter approach are the base model and the Time series component. In the first stage, $a_x$ is estimated using Ordinary least square method, $b_x$ and $k_t$ are estimated Singular value decomposition method. In the second stage, the Gated recurrent unit model is used to model and extrapolate the fitted values of $k_t$. Thus, if we let the mortality data collected at times to be $g(1),\ g(2),\ldots,g(J)$ then the LC-GRU model is given as:

$$\ln(m_{xg(t)}) = a_x + b_x k_{g(t)} + \varepsilon_{xg(t)} \tag{1}$$

$$m_{xg(t)} = \exp^{a_x + b_x k_{g(t)} + \varepsilon_{xg(t)}} \tag{2}$$

$$\text{where } \varepsilon_{xg(t)} \sim N(0,\sigma^2),\quad k_{g(t)} = f(k_{g(t-1)}, k_{g(t-2)},\ldots,k_{g(t-J)}) + \varepsilon_{g(t)} \tag{3}$$

$k_{g(t)}$ is a GRU model that approximates a function $f$ that links $k_{g(t)}$ to its time lags, and $j \in N$ represents the number of time lags considered and $\varepsilon_t$ is the error term.

$m_{xg(t)}$ represent segmented mortality that is infant mortality, adult mortality or old age mortality, respectively.

$a_x$ is the average shape of the age parameter for segmented mortality.

$b_x$ is the speed of change parameter for segmented mortality.

## 2.2 Parameter estimation

The least square solution of Equation (1) gives $\hat{a}_x$ as the mean or average of the observed $ln(m_{xt})$ while $\hat{b}_x$ and $\hat{k}_t$ are the first term of the SVD of the matrix $ln(m_{xt}) - \hat{a}_x$ .

$$\ln(m_{xg(t)}) = a_x + b_x k_{g(t)} + \varepsilon_{g(t)}$$
$$\varepsilon_{xg(t)} = \ln(m_{xg(t)}) - a_x - b_x k_{g(t)}$$
$$\varepsilon^2_{xg(t)} = (\ln(m_{xg(t)}) - a_x - b_x k_{g(t)})^2$$
$$\sum_{t=1}^{T} \varepsilon^2_{xg(t)} = \sum_{t=1}^{T} (\ln(m_{xg(t)}) - a_x - b_x k_{g(t)})^2$$
$$= \sum_{t=1}^{T} (\ln(m_{xt}) - a_x - b_x k_{g(t)})^2$$
$$-\sum_{t=1}^{T} \ln(m_{xg(t)}) + Ta_x + b_x \sum_{t=1}^{T} k_{g(t)} = 0$$

To achieve a unique solution we will impose the following restriction $\sum_{t=1}^{T} k_{g(t)} = 0$

$$\therefore T\hat{a}_x = \sum_{t=1}^{T} \ln(m_{xg(t)})$$

$$\hat{a}_x = \frac{\sum_{t=1}^{T} \ln(m_{xg(t)})}{T} \tag{4}$$

### 2.3 Singular Value Decomposition

Bell and Monsell introduced the Singular Value Decomposition (SVD) (1991). It is a method for breaking down a matrix into its several constituent matrices, revealing many of the valuable and intriguing characteristics of the original matrix. Any real $m \times n$ matrix $Z$ can be uniquely decomposed as:

$$Z = B\Sigma K^T \tag{5}$$

The matrices $B$ and $K$ are both defined to be orthogonal matrices. Matrix $\Sigma$ is a diagonal matrix. The elements along the diagonal of $\Sigma$ are known as the singular values of the matrix $Z$ . The columns of $B$ are known as the left-singular vectors. The right-singular vectors are the columns of $K$ .

To obtain parameters $b_x$ and $k_t$, singular value decomposition (SVD) is applied on the matrix $Z$ where, $Z$ is given as

$$Z = \ln(\hat{m}_{x,t}) - \hat{a}_x \tag{6}$$

That is;

$$Z_{xt} = \begin{pmatrix} \ln(\hat{m}_{1,1}) - \hat{a}_1 & \ln(\hat{m}_{1,2}) - \hat{a}_1 & \ln(\hat{m}_{1,3}) - \hat{a}_1 & \ldots & \ln(\hat{m}_{1,n}) - \hat{a}_1 \\ \ln(\hat{m}_{2,1}) - \hat{a}_2 & \ln(\hat{m}_{2,2}) - \hat{a}_2 & \ln(\hat{m}_{2,3}) - \hat{a}_2 & \ldots & \ln(\hat{m}_{2,n}) - \hat{a}_2 \\ \ln(\hat{m}_{3,1}) - \hat{a}_3 & \ln(\hat{m}_{3,2}) - \hat{a}_3 & \ln(\hat{m}_{3,3}) - \hat{a}_3 & \ldots & \ln(\hat{m}_{3,n}) - \hat{a}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ln(\hat{m}_{n,1}) - \hat{a}_n & \ln(\hat{m}_{n,2}) - \hat{a}_n & \ln(\hat{m}_{n,3}) - \hat{a}_n & \ldots & \ln(\hat{m}_{n,n}) - \hat{a}_n \end{pmatrix}$$

Here we decomposed matrix $Z_{xt}$ into a new set of factors that are optimal and orthogonal based on some criterion.

Thus, the real matrix $Z$ of order $m \times n$ can be uniquely decomposed as:

$$SVD(Z_{x,t}) = B\Sigma K^T \tag{7}$$

where $B$ is $m \times m$ orthogonal matrix that represents the age component.

$\Sigma$ is $m \times n$ diagonal matrix that represents the singular values.

$K$ is $n \times n$ orthogonal matrix that represents the time component.

$$Z_{xt} = B\Sigma K^T = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1m} \\ b_{21} & b_{22} & b_{32} & \dots & b_{2m} \\ b_{31} & b_{32} & b_{33} & \dots & b_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \dots & b_{mm} \end{pmatrix} \begin{pmatrix} \sigma_{11} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{22} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \sigma_{mn} \end{pmatrix} \begin{pmatrix} k_{11}^T & k_{12}^T & k_{13}^T & \dots & k_{1n}^T \\ k_{21}^T & k_{22}^T & k_{23}^T & \dots & k_{2n}^T \\ k_{31}^T & k_{32}^T & k_{33}^T & \dots & k_{3n}^T \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{n1}^T & k_{n2}^T & k_{n3}^T & \dots & k_{nn}^T \end{pmatrix}$$

Here, $b_{11},\dots,b_{mm}$ is the Eigen vectors of $Z_{xt}Z_{xt}^T$. $\sigma_1,\dots,\sigma_r$ are the singular values of $Z_{xt}Z_{xt}^T$ satisfying $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq,\dots,\sigma_r > 0$ and $k_1, k_2 \dots k_n$ are the Eigen vectors of $Z_{xt}^T Z_{xt}$ Here, Singular values are defined as the positive square roots of the Eigen values of $Z_{xt}Z_{xt}^T$

Matrix $Z_{xt}^T Z_{xt}$ of size $m \times n$ is real and of rank $r$. Also $r$ of its Eigen values $\sigma_i^2$, $i = 1(1)r$ are positive and therefore real, while the remaining $(n-r)$ are zero.

Theorem: Let $Z_{xt} = B\Sigma K^T$ for $i = 1(1)m; j = 1(1)n$ be singular value decomposition of an $m \times n$ real matrix of rank. Then

$$Z_{xt} = B\Sigma K^T = \sum_{i=1}^{r} b_i \sigma_i k_i^T$$

Proof

$Z_{xt} = B\Sigma K^T$

$$Z_{xt} = \begin{pmatrix} \sigma_1 b_{11} & \sigma_2 b_{12} & \sigma_3 b_{13} & \dots & \sigma_r b_{1r} \\ \sigma_1 b_{21} & \sigma_2 b_{22} & \sigma_3 b_{23} & \dots & \sigma_r b_{2r} \\ \sigma_1 b_{31} & \sigma_2 b_{32} & \sigma_3 b_{33} & \dots & \sigma_r b_{3r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1 b_{r1} & \sigma_2 b_{r2} & \sigma_3 b_{r3} & \dots & \sigma_r b_{rr} \end{pmatrix} \begin{pmatrix} k_{11}^T & k_{12}^T & k_{13}^T & \dots & k_{1r}^T \\ k_{21}^T & k_{22}^T & k_{23}^T & \dots & k_{2r}^T \\ k_{31}^T & k_{32}^T & k_{33}^T & \dots & k_{3r}^T \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{r1}^T & k_{r2}^T & k_{r3}^T & \dots & k_{rr}^T \end{pmatrix}$$

$$Z_{xt} = \begin{pmatrix} (\sigma_1 b_{11} k_{11}^T + \ldots + \sigma_r b_{1r} k_{r1}^T) & (\sigma_1 b_{11} k_{12}^T + \ldots + \sigma_r b_{1r} k_{r2}^T) & \ldots & (\sigma_r b_{11} k_{1r}^T + \ldots + \sigma_r b_{1r} k_{rr}^T) \\ (\sigma_1 b_{21} k_{11}^T + \ldots + \sigma_r b_{2r} k_{r1}^T) & (\sigma_1 b_{21} k_{12}^T + \ldots + \sigma_r b_{2r} k_{r2}^T) & \ldots & (\sigma_r b_{21} k_{1r}^T + \ldots + \sigma_r b_{2r} k_{rr}^T) \\ \vdots & \vdots & \ddots & \vdots \\ (\sigma_1 b_{r1} k_{11}^T + \ldots + \sigma_r b_{rr} k_{r1}^T) & (\sigma_1 b_{r1} k_{r2}^T + \ldots + \sigma_r b_{rr} k_{r2}^T) & \ldots & (\sigma_1 b_{11} k_{1r}^T + \ldots + \sigma_r b_{rr} k_{rr}^T) \end{pmatrix}$$

$$= \sigma_1 b_1 k_1^T + \begin{pmatrix} (\sigma_2 b_{12} k_{21}^T + \ldots + \sigma_r b_{1r} k_{r1}^T) & (\sigma_2 b_{12} k_{22}^T + \ldots + \sigma_r b_{1r} k_{r2}^T) & \ldots & (\sigma_2 b_{12} k_{2r}^T + \ldots + \sigma_r b_{1r} k_{rr}^T) \\ (\sigma_2 b_{22} k_{21}^T + \ldots + \sigma_r b_{2r} k_{r1}^T) & (\sigma_2 b_{22} k_{22}^T + \ldots + \sigma_r b_{2r} k_{r2}^T) & \ldots & (\sigma_2 b_{22} k_{2r}^T + \ldots + \sigma_r b_{2r} k_{rr}^T) \\ \vdots & \vdots & \ddots & \vdots \\ (\sigma_2 b_{r2} k_{21}^T + \ldots + \sigma_r b_{rr} k_{r1}^T) & (\sigma_2 b_{r2} k_{22}^T + \ldots + \sigma_r b_{rr} k_{r2}^T) & \ldots & (\sigma_2 b_{r2} k_{2r}^T + \ldots + \sigma_r b_{rr} k_{rr}^T) \end{pmatrix}$$

$$= \boldsymbol{\sigma_1 b_1 k_1^T + \sigma_2 b_2 k_2^T + \ldots + \sigma_r b_r k_r^T}$$

$$= \boldsymbol{\sigma_1 b_1 k_1^T + \sigma_2 b_2 k_2^T + \ldots + \sigma_r b_r k_r^T = \sum_{i=1}^{r} \sigma_i b_i k_i^T}$$

$$\boldsymbol{Z_{xt} = \sum_{i=1}^{r} \sigma_i b_i k_i^T}$$

## 2.4 Merits of the Gate recurrent unit aided LC Model

i. The number of parameters employed is modest.
ii. It can handle both the linear and non-linear nature of mortality trends, thus adding confidence to extrapolations.
iii. The model can handle the long-term dependence associated with time series data.
iv. The model performs better than other models in terms of its prediction errors and generates reasonable estimates of mortality forecast.

## 3. Results and Discussion

The mortality index $k_t$ is modelled using a gated recurrent unit in order to take care of the non-linearity inherent in mortality data and long-term dependence in time series data. As usual, the model parameters $b_x$ and $k_t$ were estimated using SVD.

$$\hat{a}_x = \begin{pmatrix} -1.51258 \\ -1.42149 \\ -1.31180 \\ -1.24766 \\ -1.04828 \\ -1.20401 \\ -1.39770 \\ -1.61356 \\ -1.54164 \end{pmatrix} \qquad \hat{b}_x = \begin{pmatrix} -0.22072 \\ -0.58489 \\ -0.07869 \\ 0.34606 \\ 0.27998 \\ -0.06534 \\ -0.15025 \\ -0.37868 \\ -0.46810 \end{pmatrix} \qquad \hat{k}_t = \begin{pmatrix} -2.21144 \\ 1.91964 \\ -0.35911 \\ 0.64749 \\ -0.23533 \\ 0.20401 \\ 2.01923 \\ 1.02975 \\ 0.07101 \\ -1.68809 \\ -2.31428 \\ 0.91710 \end{pmatrix}$$

The column matrices $a_x$, $b_x$ and $k_t$ above are the estimated values of the model parameters using SVD. The time component $k_t$ is them modelled using both Autoregressive Integrated Moving Average *(ARIMA)* as used by the Original LC model and a Gated Recurrent Unit (GRU).

**Table 1. Forecasted index $\hat{k}_t$ using ARIMA model and GRU for Simulated data**

| Year | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 |
|---|---|---|---|---|---|---|---|
| $\hat{k}_t(Arima)$ | 1.768146 | 1.908826 | 2.333063 | 2.527571 | 2.794256 | 2.989057 | 3.198530 |
| $\hat{k}_t(GRU)$ | -0.006259 | -0.014376 | -0.023553 | -0.037314 | -0.039273 | -0.041222 | -0.042646 |

The Result in table 1 shows the forecasted values of the mortality index $k_t$ using both ARIMA and gated recurrent unit.

**Table 2. Comparison between $\hat{k}_t$ (ARIMA) and $\hat{k}_t$ (GRU) for simulated data**

| Method | RMSE | MAE | MAPE |
|---|---|---|---|
| $\hat{k}_t(ARIMA)$ | 0.3125211 | 0.2149233 | 1.279095 |
| $\hat{k}_t(GRU)$ | **0.2818981** | **0.2083371** | **1.168281** |

It can be seen from table 2, that the GRU models the mortality index $k_t$ better than the traditional ARIMA model used by Lee and Carter in 1992. Projected $\hat{k}_t$ using GRU will then be used to predict the future central death rate $m_x$.

**Table 3 Comparison between LC-ARIMA and LC-GRU for simulated data**

| Method | RMSE | MAE | MAPE |
|---|---|---|---|
| LC-ARIMA | 0.1960865 | 0.1613628 | 1.185264 |
| LC-GRU | **0.1924219** | **0.1588562** | **1.190274** |

It is observed from table 3 that using the simulated data the (LC-GRU) model performs better than the traditional LC model used by Lee and Carter in 1992. In general, for all the error measures LC-GRU model tend to have lower values compared to the Original LC model.

Real data was also used to model mortality rates using both original LC model the GRU assisted LC model. The model parameters $a_x,\ b_x$ and $k_t$ were estimated using the SVD.

$$\hat{a}_x = \begin{pmatrix} -5.0984 \\ -4.9295 \\ -4.8682 \\ -4.8247 \\ -4.71038 \\ -4.5094 \\ -4.3111 \\ -4.0274 \\ -3.7094 \end{pmatrix} \qquad \hat{b}_x = \begin{pmatrix} 0.1112 \\ 0.13889 \\ 0.1219 \\ 0.1096 \\ 0.1102 \\ 0.1084 \\ 0.1055 \\ 0.1006 \\ 0.0933 \end{pmatrix} \qquad \hat{k}_t = \begin{pmatrix} 0.71673 \\ 1.47743 \\ 0.71673 \\ 0.11885 \\ 0.13909 \\ 0.16763 \\ 0.32585 \\ 0.63814 \\ -0.19014 \\ -0.67994 \\ -1.20958 \\ -2.22082 \end{pmatrix}$$

The above column matrices are the estimated values of the parameters $a_x,\ b_x$ and $k_t$ using SVD. The time component $k_t$ is them modelled using both Autoregressive Integrated Moving Average *(ARIMA)* as used by the original LC model and a Gated Recurrent Unit (GRU).

Table 4. Forecasted index $\hat{k}_t$ using ARIMA model and GRU for real data

| Year | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 |
|---|---|---|---|---|---|---|---|
| $\hat{k}_t(Arima)$ | -2.552026 | -2.660501 | -2.696029 | -2.707666 | -2.711477 | -2.712725 | -2.71313 |
| $\hat{k}_t(GRU)$ | -0.9610381 | -1.0214071 | -0.951606 | -0.7492634 | -0.708148 | -0.6479220 | -0.594411 |

The Result in table 4 shows the forecasted values of the mortality index $k_t$ using both ARIMA and gated recurrent unit.

**Table 5. Comparison between $\hat{k}_t$ (ARIMA) and $\hat{k}_t$ (GRU) for real data**

| Sex | Method | RMSE | MAE | MAPE |
|---|---|---|---|---|
| **Male** | $\hat{k}_t(ARIMA)$ | 2.809482 | 2.66587 | 8.866765 |
| | $\hat{k}_t(GRU)$ | **1.237017** | **1.092629** | **2.74663** |

It can be seen in table 5 that the GRU model has the minimum errors and thus, models the mortality index $k_t$ better than the traditional ARIMA model used by Lee and Carter in 1992. Projected $\hat{k}_t$ using GRU will then be used predict the future central death rate $m_x$.

Table 6. Comparison between LC-ARIMA and LC-GRU for real data

| Sex | Method | RMSE | MAE | MAPE |
|---|---|---|---|---|
| **Male** | **LC-ARIMA** | 0.00427314 | 0.003826728 | 0.2578544 |
| | LC-GRU | **0.001359604** | **0.0009497948** | **0.09406077** |

It can be seen from table 5 that the (LC-GRU) model has the minimum error measures thus performs better than the traditional LC model used by Lee and Carter in 1992. In general, for all the error measures LC-GRU model tend to have lower values compared to the LC model.

## 4. Conclusion

After obtaining all the parameters of the model, the performance of the original LC model and the GRU aided approach (GRU-LC model), was assessed. Some of the hyper parameters of the GRU used to improve the performance of the models include: epoch size (100), activation function (Swish), learning rate (0.0001), and the number of neurons in the hidden state. To assess and compare the performance of the original LC model and the LC-GRU model, error measures (RMSE, MAE, and MAPE) were used as measures of accuracy for the age-specific forecasts for the mortality rates. The forecast accuracy of the different approaches is investigated. The summaries in Tables 3, and 6 indicate that different methods show different forecast accuracy in terms of error measures for male, subpopulations and the simulated data. LC-GRU has lower errors as compared to the original LC-model proposed by Lee and Carter in 1992. Therefore LC-GRU performs better than the original LC model.

## References

Bell, W. and Monsell, B. (1991). Using principal components in time series modeling and forecasting of age-specific mortality rates. *Alexandria Virginia American Statistical Association*.

Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee-Carter under conditions of Variable mortality decline. *Population studies* 56(3), 325–336.

Box, G.E. and Jenkins, G.M. (1970). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Brouhns, N., Denuit, M. and Vermunt, J.K (2002). A Poisson log-bilinear regression approach To the construction of projected life tables. *Insurance: Mathematics and Economics* 31 (3), 373–393.

Cairns, A.J., Blake, D. and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance 73*(4), 687–718.

Carter, L.R. and Prskawetz, A. (2001). Examining Structural Shifts in Mortality Using the Lee -Carter Method. Working Paper WP2001-007, Max Planck Institute for Demographic Research, Rostock, Germany.

Chung, J., Gulcehre, C., Cho, K.H., and Bengio, Y. (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. Technical report, arXiv preprint arXiv:1412.3555, 2014 - arxiv.org

Currie, I. D., Durban, M. and Eilers, P.H (2004). Smoothing and forecasting mortality rates. *Statistical Modelling 4* (4), 279–298.

Gutterman, S., and Vanderhoof, I. (1999). Forecasting Changes in Mortality: a Search for a Law of causes and effects. *North American Actuarial Journal*, 2 (4): 135-138

Lee, R. and Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association* 87: 65-71.

Lee, R.D. (2000) The Lee-Carter method of forecasting mortality, with various extensions and applications (with discussion). *North American Actuarial Journal* 4, 80-93.

McNown, R. and Rogers, A. (1989). Forecasting mortality: A parameterized time series Approach. *Demography 26*(4), 645–660.

Renshaw, A.E., and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for Mortality reduction factors. *Insurance: Mathematics and Economics* 38(3): 556-570.

Tulijapurkar, S., (2008), Mortality declines, Longevity risk and Aging. *Asia Pacific Journal of Risk and Insurance* vol. 3(1), pp. 37-51.

Wilmoth, J.R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of Mortality change. *Technical Report, University of California, Berkeley U.S.A.*