Forecasting Malaysian mortality rates using the Lee-Carter model with fitting period variants

Nur Shatikah Mohamad Ibrahim¹, Norazliani Md Lazam² and Syazreen Niza Shair³

¹,²,³ Centre for Actuarial Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara Shah Alam, 40450, Selangor, Malaysia

³E-mail: syazreen@tmsk.uitm.edu.my

Abstract. This study aims to forecast Malaysian mortality rates by age and gender using the well-known Lee-Carter model. Data obtained from the Department of Statistics Malaysia which consists of central mortality rates by age and gender from year 1970 to 2018. Two different sets of fitting periods were determined based on the observations of changes in mortality index patterns over the years. The Set A consists of 24-year fitting period from 1970 to 1993 whereas Set B comprises 31-year fitting period from 1970 to 2000. The in-sample evaluation of the Lee-Carter model are performed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) whereas the out-sample errors are calculated using the Mean Absolute Percent Errors (MAPE) and Root Mean Square Error (RMSE). Results show that the set A is better fitted into the Lee-Carter model than that of set B by having lower values of AIC and BIC, consequently produced more accurate out-sample forecast values for females. The Lee-Carter model is a reliable model for Malaysia data, however careful attention must be given when selecting the best fitting period.

1. Introduction
Forecasting mortality rates have been crucial due to uncertainty decreasing rate in mortality rates which have resulted in an increase in human life expectancy in recent years. Major causes to these changes are due to improvement in health care innovations and lifestyles. Hence, accurate mortality forecasting model is needed to help policy maker in dealing with uncertainty to estimate future number of deaths, which necessary for social and population planning.

There are many mortality forecasting models proposed by researchers in the past years. The Lee-Carter model [1] is a prominent model that has been used as a benchmark to develop many of its extended versions. Few researchers proposed the extension versions of Lee-Carter model in order to forecast mortality rates in different countries, such as in Hungary [2], Indonesia [3] and Republic of Iran [4]. For example, the Lee-Carter model was extended by [3] to support the Artificial Neural Networks (ANN) in forecasting mortality rates of Indonesian population. These extensions were developed to improve the accuracy of mortality rate forecast values.

In Malaysia, research from [5] suggested that the Lee-Carter model as an adaptation to meet the outcome results with approach of regression due to its simplicity and reliability. Thus, the application of the Lee-Carter model to Malaysian mortality data has been growing over the last ten years. Work from [5] estimated Malaysian mortality data using Lee-Carter model and proved that the Lee-Carter model gave a good fit to the Malaysia mortality data. Research from [6] indicated that the functional data version of Lee-Carter model outperformed the original Lee-Carter model for Malaysia and other developing
countries. In addition, [7] have proved that the Lee-Carter model supported by Neural Network fits well to both Malaysian male and female mortality data by age groups and from the year 1991 to 2016.

In addition to extended version of the Lee-Carter model, fitting periods also affected the accuracy of forecast values in some ways. For example, a study from [1] showed that the Lee-Miller (LM) method fits better for Malaysian male age-specific mortality rates whereas Hyndman-Ullah 6 (HU6) method fits better into long period of data for females. When the mortality models fitted into short period of data, it shows that the Booth-Mordonald-Smith variant (BMS) method is the best good of fit for male and HU6 for females. These indicated that the length fitting period affects the forecast performance of mortality forecasting models.

Hence, it is of interest to this study to analyse the effect of fitting period variants on the mortality forecast values from the original Lee-Carter model. This research extends existing works by proposing two different fitting periods using more recent data from the year 1970 to 2018. Fitting period variant I is called as set A consists of fitting period from 1970 to 1993 while variant II referred to set B includes fitting period from 1970 to 2000. These fitting period variants were determined based on the changes in time-component of the Lee-Carter model over the full length of observation period. Thus, the aims of this study are to analyse Malaysian mortality rates over the observation years and to observe the validity of the original Lee-Carter model using current Malaysian mortality data with two different length of fitting periods and subsequently to forecast mortality rates for future years.

This research paper is organized as follows. Section 2 describes data and methods use to achieve the research objectives. Section 3 discusses in detailed the results and findings of this research and finally Section 4 concludes.

2. Research Methods

In this study, data is collected from the Department of Statistics Malaysia (DoSM). Data includes Malaysian mortality rates segregated by quinquennial age groups from 0 to 75 years old, year from 1970 until 2018 and gender, $m_{xt,i}$. The age-specific mortality rate patterns are analysed by age, year, and gender. From these mortality rates the life expectancy at births is calculated and analysed. The analysis of the data is performed using R analytics before we fit the Lee-Carter model into the observed data.

2.1. The Lee-Carter model

The equation in [8] and [9] describing the original Lee-Carter model as below:

$$\ln(m_{xt,i}) = a_{x,i} + b_{x,i}k_{t,i} + \varepsilon_{x,t,i}$$  \hspace{1cm} (1)

where

$m_{xt,i}$ is the central death rate for the age group $x$, year $t$ and $i$ refers to male or female;
$a_{x,i}$ is the average log-mortality at age $x$ for gender $i$;
$b_{x,i}$ indicates the rate of improvement of the level of mortality at age $x$ for gender $i$;
$k_{t,i}$ represents the mortality index in the year $t$ for gender $i$;
$\varepsilon_{x,t,i}$ is the error term for age $x$ and year $t$ for gender $i$.

$b_{x,i}$ represents the tendency of mortality rates at each age go against the reaction changes in $k_{t,i}$. The $k_{t,i}$ is mortality index that reduces over the year. There is a linear relationship between the parameters and explanatory factors of the age $x$, and the time $t$ in the Lee-Carter model. The estimation of parameters $b_{x,i}$ and $k_{t,i}$ can be derived using the Singular Value Decomposition (SVD) method with an assumption that all the errors have the same finite variance. Moreover, the constraints $\sum x b_{x,i} = 1$ and $\sum t k_{t,i} = 0$ are applied, and the estimation of parameter $a_{x,i}$, such that $a_{x,i} = \frac{1}{t} \sum_t \ln (m_{xt,i})$. 


2.2. Evaluation on goodness fit model.

The goodness fit model is capable to evaluate the goodness of fit for the Lee-Carter model. Basically, this is to check how fit is the model towards the data provided. The selection of fitting periods is based on mortality index patterns. Set A represents fitting period from year 1970 to 1993 and Set B represents fitting period from year 1970 to 2000. The set A fitting period is chosen because the pattern matches the Lee-Carter model assumption that the $k_{t,1}$ pattern should be linearly decreasing over time (see Figure 5). However, if we take the set B fitting period, there is slightly an increasing pattern in recent years (See Figure 6). In this study, there are two criterion is used which are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) that are calculated using StMoMo package and gnm package in R. The lower the value of AIC and BIC is preferable. The equation (2) and (3) are expressed as:

$$AIC = 2v - 2L$$  \hspace{1cm} (2)

$$BIC = v \log K - 2L$$  \hspace{1cm} (3)

where

- $L$ is the maximized log-likelihood
- $v$ is the number of parameters (so assuming that the dispersion is known)

2.3. Forecast Error Measurement.

The forecast error is measured as the actual value minus the predicted value. Mean Absolute Percent Error (MAPE) and the Root Mean Square Error (RMSE) is used to determine the error of forecast values. The formula (4) and (5) represent Mean Absolute Percent Error (MAPE) and the Root Mean Square Error (RMSE) respectively:

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - P_i}{Y_i} \right| \times 100$$  \hspace{1cm} (4)

$$RMSE = \left[ \frac{1}{N} \sum_{i=1}^{n} (P_i - Y_i)^2 \right]^{1/2}$$  \hspace{1cm} (5)

where

- $N$ is the number of data
- $Y_i$ is an actual value
- $P_i$ is the predicted value of the i-th data obtained

2.4. Estimation of Life expectancy at Births

Life expectancy at birth, $e_0$ is the number of years a person born today can be expected to live use the current age-specific mortality rates. Life expectancy usually used to measure health status of population. The formula (6) below illustrates the formula and notation used in [10] to estimate life expectancy using the actuarial life table formula

$$e_x = \frac{7x}{nLx}$$  \hspace{1cm} (6)

Where $Tx$ indicates the total number of person-years lived by the cohort from age $x$ until all members of the cohort have died while $nLx$ is the total number of persons who survive by the cohort from age $x$ to $x + n$. 


3. Results and Discussion of Analysis
Before we fit the Lee-Carter model into the data, we perform data analysis of Malaysian mortality rates by age and gender and life expectancy at births from year 1970 to 2018 as below.

3.1. Analysis of mortality rates and life expectancy in Malaysia.

![Figure 1](image.png)

Figure 1. The rainbow plots of log age-specific death rates for Malaysian males (left) and females (right). Colors indicates years in which the first few years are shown in red, followed by orange, yellow, green, blue, indigo and violet represents the last few years.

Figure 1 above describes the log death rates for both gender which are male and female using Malaysian observation mortality data from the year 1970 to 2018. Based on the figure 1, it generally shows the mortality rates for both genders have the same pattern that is increasing along with the age. The pattern interprets that both gender’s mortality in Malaysia formed a roughly J pattern between the age 0 to 75. The mortality rate for infants is higher compared to other age groups. However, infants’ mortality has improved over the years indicating that the government has put much effort to reduce number of deaths among newborns. Malaysia’s government have successfully reducing the infant mortality by formulation of policies, strategies, public health facilities and programs that have addressed access to services through socio-economic, cultural, educational, gender and poverty dimensions based on the United Nations Children's Fund (UNICEF) [11].

When comparing between genders, there is a significance difference between both gender’s pattern for the age referring to the accident hump age. There is less pronounced accident hump can be seen from the female mortality plot, indicating that less number of death due to accidents for females compared to males. According to the Malaysian Institute of Road Safety Research [12], the number of women drivers is less compared to men drivers over the last few years leading to less road accidents among females. In addition, females’ mortality rates show significant improvement due to substantial gap between red and violet lines.
Figure 2. Malaysian life expectancy at birth for males (solid line) and females (dotted line).

Figure 2 shows the life expectancy at birth for Malaysia calculated from the observation mortality data from the year 1970 to 2018 for male and female, using the life table approach. Malaysian new-born males and females are expected to live longer than previous years new-born. The graph shows that the life expectancy of male population is consistently lower than female population over the years. This proved that Malaysian female tend to live longer than Malaysian male. Significant increase in life expectancy is a great achievement for Malaysia. However, there is possibility that Malaysian life expectancy may continue increasing to catch up the life expectancies of developed countries.

3.2. Fitting Malaysian mortality data into the Lee-Carter model

3.2.1. Analysis on each estimated parameter of the Lee-Carter model. Figures 3 and 4 indicate the values of $a_{x,i}$ and $b_{x,i}$ for the Lee-carter model from two fitting periods which are Set A and Set B. There is not much different in $a_{x,i}$ between those two different sets of fitting periods and between genders. Parameter values of $b_{x,i}$ tells us which rates declined rapidly and which rates decline slowly in response to changes in mortality index $k_{t,i}$. The negative value of $b_{x,i}$ for Malaysian male in set B shows that mortality at those ages tends to rise as mortality goes down over the years.
Figure 3. The estimated $a_{x,i}$ values from the Lee-Carter model for both set A and B fitting periods.

Figure 4. The estimated $b_{x,i}$ values from the Lee-Carter model for both set A and B fitting periods.
Moreover, Figures 5 and 6 shows changes in age-specific mortality rates over the years as driven by a scalar time. The $k_{t,l}$ value is estimated to minimize the errors in logs of death rates. Figure 5 concludes that $k_{t,l}$ value from set A is declining linearly at almost constant rate over the year. However in Figure 6, $k_{t,l}$ values display non-linear decreasing pattern. The mortality index using set B show a decrease pattern from year 1970 to 1992 and then started to increase from year 1993 to 2000. Different outcomes from
different length of fitting periods affect the accuracy of forecast values. The evaluation process is performed in the following sub-sections to determine the performance of the Lee-Carter model using fitting period variants.

3.2.2. Evaluation on goodness fit model. The value of AIC and BIC are calculated to indicate the fitness of the Lee-Carter model in predictive performance for Malaysian in-sample data. The lower the value of AIC and BIC, the better the model fits into the data.

Table 1. The value of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for Malaysian data male and female using Lee-Carter model.

|                | Set A Male | Set A Female | Set B Male | Set B Female |
|----------------|------------|--------------|------------|--------------|
| Number of parameters | 56         | 56           | 63         | 63           |
| Log-likelihood    | -2.43      | -1.51        | -2.97      | -1.87        |
| AIC              | 116.8615   | 115.0186     | 131.9384   | 129.7345     |
| BIC              | 341.4925   | 339.6496     | 400.7721   | 398.5681     |

Table 1 presents the calculated value of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) of the Lee-Carter model for both gender, male and female by applying the Malaysian in-sample data for set A from the year 1970 to 1993 and set B from the year 1970-2000. By comparing in the terms of gender, we can conclude that the Lee-Carter model is best fit for Malaysian females than Malaysian males for both set A and B because lower value of AIC and BIC. Furthermore, by comparing different sets of in-sample data, set A shows a better result than set B by having lower AIC and BIC values for both males and females. This research proves that set A is better fit for Malaysia mortality data to perform mortality forecasting. In other words, linearly decreasing pattern of mortality index $k_{t,i}$ improves the performance fit of the Lee-Carter model. This is in line with research from [8] that indicated the original Lee-Carter model should portray a linearly decreasing pattern of $k_{t,i}$ value.

3.2.3. Forecast Error Measurement. Table 2 presents the estimated value for Mean Absolute Percent Errors (MAPEs) and Root Mean Square Error (RMSEs) for male and female for the Lee-Carter model. The lower the error measurement value, the better the model to predict future values. Set A represents the out-sample forecast error from the year 1994 to 2018 when we fit the mortality data from 1970-1993 into the Lee-Carter model whereas set B performs the out-sample forecast error from the year 2001 to 2018 when we fit the mortality data from 1970-2000 into the Lee-Carter model.

Although fitting period set A consistently has better fit than set B for both genders, in terms of out-sample forecasts, the set A is outperforming set B only for females (see Table 2). This concludes that the Lee-Carter model has better forecast performance for female mortality data than males when using set A fitting variant, vice versa when using set B.

Table 2. The average Mean Absolute Percent Errors (MAPEs) and Root Mean Square Error (RMSEs) for Malaysian males and females using the Lee-Carter model. Average is taking over age groups and years.

| Fitting Period | MAPEs A | MAPEs B | RMSEs A | RMSEs B |
|----------------|---------|---------|---------|---------|
| Male           | 0.2361  | 0.1870  | 0.0035  | 0.0021  |
| Female         | 0.2493  | 0.4887  | 0.0021  | 0.0045  |
3.3. Forecasting for future mortality rate using the best fitting period

![Graphs showing forecasted log death rates for Malaysian males and females.](image)

**Figure 7.** The rainbow plots of forecasted log death rate for Malaysian males and females.

Figure 7 describes the observed and future log death rates for Malaysian males and females using the best fitting period determined previously. It is clearly showed that the mortality patterns by age for future years have a similar pattern with previous years such that it increases by age for both genders. In addition, the mortality rates for females show a continue decreasing pattern in future years for all age indicating mortality improvement. However, mortality rates for males in accident hump age range show the opposite pattern in which it increases in future years reaching towards the mortality level of earliest years. Efforts should put focus on the targeted group that is to reduce hazards among the young adult males.

4. Conclusion

This study investigated the accuracy of out-sample forecasts of the Lee-Carter model using two different set of fitting periods. Fitting period that produced more linearly decreasing pattern of mortality index tends to result in lower out-sample forecast errors of MAPE and RMSE for females and a smaller in-sample values of AIC and BIC for both males and females. The Lee-Carter model is still valid for Malaysia mortality data, however choosing the right fitting period may improve accuracy of the forecast values significantly. For future works, method to select fitting period may improve using optimization error procedure which help to broaden the selection criteria.

**Acknowledgements**

We would like to express our deepest gratitude to the Ministry of Education (MOE) and Research Management Centre, Universiti Teknologi MARA (UiTM) for the financial supports through Fundamental Research Grant Scheme with file number of 600-IRMI/FRGS 5/3 (125/2019) and for the permission to publish this journal.
References

[1] Husin W Z W and Abidin N S Z 2020 Point forecasts of mortality rates in Malaysia: A comparison of principal component methods *Indonesian. J. Elec. Eng & Comp. Sci* 19 p 1661-1669

[2] Baran S, Gall J, Ispany M and Pap M 2007 Forecasting Hungarian Mortality Rates Using The Lee-Carter Method *Acta Oeconomica*, 57(1), 21-34

[3] Safitri L, Mardiyati1 S and Rahim H 2018 Forecasting the mortality rates of Indonesian population by using neural network *J. Phys.: Conf. Ser.* 974 012030

[4] Aghamohamadi S, Jahangiri K, Kavousi A and Mofazali A S 2018 Analysis and Forecasting the Accident Mortality Trends in the Islamic Republic of Iran Applying Lee-Carter Model During the Years 2006 to 2035 *Bull. Emerg. Traum.* 6(4):341-348

[5] Kamaruddin H S and Ismail N 2018 Forecasting selected specific age mortality rate of Malaysia by using Lee-Carter model network *J. Phys.: Conf. Ser.* 974 012003 p 7

[6] Shair S N, Zolkifli N A, Zulkefl N F and Murad A 2019 A functional data approach to the estimation of mortality and life expectancy at birth in developing countries *Pertanika J. Sci. & Technol.* 27 (2): 797 - 814

[7] Hamid S M I, Ofosu-Hene E D and Ponnusamy R R 2019 A new approach to forecast Malaysian mortality rates *International J. Adv. & Appl. Sci.* 6 (10) 2019 p 53-61

[8] Lee R D and Carter L R 1992 Modeling and forecasting US mortality *J. Am. Stat. Assoc.* 87 p 659-671

[9] Booth H, Maindonald J and Smith L 2002 *Population Studies: A journal of Demography* 56(3) Applying Lee-Carter under conditions of variable mortality decline pp 325-336

[10] Ibrahim N S, Shair S N and Yusof A Y 2020 Mortality Rates and Life Expectancy Improvements among Malaysian Elderlies *Indonesian. J. Elec. Eng & Comp. Sci* 19 p 139

[11] Siah A K L and Lee G H Y 2015 Female Labour Force Participation, Infant Mortality and Fertility in Malaysia *J. Asia. Pac. Econ.* p 25

[12] Jabatan Keselamatan Jalan Raya (JKJR) Malaysia 2019 *Buku Statistik Kemalangan Jalan Raya* Kementerian Pengangkutan Malaysia p 10