ASM Sc. J., 18, 2023

https://doi.org/10.32802/asmscj.2023.1465

Mortality Index Simulation for Forecasting Malaysian Mortality Rates

N. Redzwan, P. Sivasundaram and R. Ramli*

Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

Mortality studies are very important in demography and actuarial areas because they assist policymakers and life insurers in managing longevity and mortality risks. In recent decades, many extrapolative mortality models have been developed following the Lee-Carter model. Despite the widely used Lee-Carter model for projecting mortality rates, the literature that has a thorough explanation of it is limited. In this study, we aim to provide a comprehensive explanation of the model with a focus on its fitting and simulation forecasting techniques. We fitted the mortality rates of the Malaysian population for the years 1991 to 2012 using the Lee-Carter model. We then projected the mortality rates for the years 2013 to 2018 using an autoregressive integrated moving average (ARIMA) (0,1,0) model by using a simulation of the mortality index. Findings showed that the Lee-Carter model performs well for this dataset based on the computed standard accuracy measures. The estimated age parameters exhibited a high mortality rate in the age group of 0-4 years, while the estimated time-varying parameter indicated a decreasing trend. This study presents a thorough interpretation of the Lee-Carter model and a detailed simulation of the ARIMA (0,1,0) model and hence provides a comprehensive reference for beginners in mortality studies.

Keywords: forecasting; simulation; mortality; Lee-Carter; Malaysia

I. INTRODUCTION

The global population is experiencing an increase in life expectancies. According to the United Nations (2019), the Malaysian life expectancy at birth in 2019 was 76.2 years, as compared to 70.9 years in 1990. The increase in life expectancies is a challenge, especially for life insurers and pension funds, in managing longevity risks. Following this, much scientific literature provides insights on fitting and forecasting mortality rates. Mortality models can be categorised into three types: expectation, explanatory, and extrapolative (Booth & Tickle, 2008). The expectation model is based on the opinions of experts; the explanatory model is based on structural models of certain causes of death; and the extrapolative model is based on the past mortality trend. The extrapolative model can minimise the problem of bias in judgements faced by the expectation model. It is also suitable to be used for long-term forecasting, as compared to the explanatory model, which is usually limited to short-term forecasting.

Much literature on mortality models in actuarial and demography fields are based on the extrapolative model. The development of these models can be traced back to works by Heligman and Pollard (1980), McNown and Rogers (1989), and Lee and Carter (1992). Lee and Carter (1992) proposed a mortality model with a bilinear factor and has since become a prominent stochastic mortality model. Following its development, many researchers have included some modifications and extensions to the Lee-Carter model. These include modifications to its statistical foundations and the development of new models (Cairns *et al.*, 2011). For example, Brouhns *et al.* (2002), Booth *et al.* (2002), and Delwarde *et al.* (2007) modified the statistical foundations of the Lee-Carter model. Extensions of the Lee-Carter model as proposed by Renshaw and Haberman (2006), Cairns *et al.* (2006), and Plat (2009) included other factors such as the

*Corresponding author's e-mail: itaramli@ukm.edu.my

cohort effect and the age-period effect. Further details on the development of the Lee-Carter model are described in the next section.

During the last decades, studies on stochastic mortality modelling have gained the interest of researchers. However, most of this literature is not straightforward, especially for young researchers and students. For example, previous studies provide insufficient explanation on methods such as singular value decomposition (SVD) and the autoregressive integrated moving average (ARIMA) forecasting model. This often leads to confusion in understanding the technique of the Lee-Carter model (Lee & Carter, 1992; Booth *et al.*, 2002; Brouhns *et al.*, 2002; Haberman & Russolillo, 2005; Renshaw & Haberman, 2006; Tsai & Lin, 2017). Furthermore, not many studies focused on forecasting by using the simulation process (Lee & Carter, 1992; Haberman & Russolillo, 2005; Tsai & Lin, 2017). Since the Lee-Carter model is the base model for many mortality models, this study is aimed at providing a comprehensive explanation on the model with a focus on its fitting and simulation forecasting techniques. The thorough interpretation of the Lee-Carter model and ARIMA (0,1,0) simulation in this work provides a good reference for beginners in mortality studies. It is important to understand the fundamentals of this model, so that they can comprehend its modifications and extensions. In addition to that, a practical application of Lee-Carter in modelling and forecasting is shown in Section 3.

This paper is organised as follows: Section 2 describes the fundamentals of the Lee-Carter model with its modifications and extensions; Section 3 shows the practical application of the Lee-Carter model to the Malaysian population's mortality data; and Section 4 provides the conclusions of this paper.

II. EXTRAPOLATIVE MORTALITY MODELS

A. Lee-Carter Model

Mortality data is an example of panel data in which the mortality rates for each age group (individual ages in insurance) are observed over time. Ages (age groups) and years are represented by positive integers to simplify the notations. While the first age and the first year are both represented by 1, the last age and the last year are represented by X and T, respectively. This data can be displayed in a matrix with ages and years as rows and columns, respectively. Let $\boldsymbol{M}$ be an $X \times T$ original matrix whose element is the mortality rate $m_{x,t}$ given by:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}} \tag{1}$$

where $D_{x,t}$ is the total number of deaths for age x during period t and $E_{x,t}$ is the total exposure (number of people alive) for age x at the beginning of period t.

The Lee–Carter model (Lee & Carter, 1992) is used to model mortality rates as follows:

$$ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \tag{2}$$

$$\text{for all } x = 1,2,\dots,X \text{ and } t = 1,2,\dots,T.$$

For all x and t, the error term $\varepsilon_{x,t}$ is independent and identically distributed normal random variable with mean 0 and variance σ^2. The parameter a_x and k_t in (2) denote the general mortality rates for age x and year t, respectively. The parameter b_x in (2) denotes the change in mortality at age x for a unit change in total mortality at time t as shown in (3).

$$\frac{d}{dt} ln(m_{x,t}) = b_x \frac{d}{dt} k_t \tag{3}$$

Equation (3) implies that:

$$\frac{\frac{d}{dt} ln(m_{x,t})}{\frac{d}{dt} k_t} = \frac{d\, ln(m_{x,t})}{d\, k_t} = b_x \tag{4}$$

Therefore, b_x measures the change in $ln(m_{x,t})$ as k_t changes with respect to time. The estimations are subject to constraints $\sum_{x=1}^{X} b_x = 1$ and $\sum_{t=1}^{T} k_t = 0$. First, matrix $\boldsymbol{Y}$ is formed for estimating parameter a_x. For a given age x, parameter a_x is estimated by minimising the error terms for the age group. The constraint $\sum_{t=1}^{T} k_t = 0$ and minimising the errors imply:

$$\begin{aligned}\sum_{t=1}^{T} ln(m_{x,t}) &= \sum_{t=1}^{T} (a_x + b_x k_t + \varepsilon_{x,t}) \\ &= a_x T + b_x \sum_{t=1}^{T} k_t + \sum_{t=1}^{T} \varepsilon_{x,t} \\ &= a_x T.\end{aligned}$$

Thus, the estimate $\hat{a}_x$ is given by:

$$\hat{a}_x = \frac{\sum_{t=1}^{T} ln(m_{x,t})}{T} \tag{5}$$

Note that (5) is a row mean of matrix $\boldsymbol{Y}$. Next, matrix $\boldsymbol{Z}$ is formed by subtracting $\hat{a}_x$ from each of the elements of row x

of matrix $\boldsymbol{Y}$ for estimating parameter b_x and k_t. Note that $\boldsymbol{M}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are all $X \times T$ matrices. The element of matrix $\boldsymbol{Z}$ is $z_{x,t} = ln(m_{x,t}) - a_x$. By the definition of $z_{x,t}$, it is apparent that the constraint $\sum_{t=1}^{T} k_t = 0$ is satisfied (the row sum of matrix $\boldsymbol{Z}$ equals 0 since each element is a mean-adjusted value).

Since X and T are usually different, reduced singular value decomposition (RSVD) can be used to estimate b_x and k_t for matrix $\boldsymbol{Z}$. Since the rank of matrix $\boldsymbol{Z}$ is less than or equal to the minimum between X and T, the maximum possible rank is $l = min\{X, T\}$. The SVD on matrix $\boldsymbol{Z}$ produces matrices $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{V}$ such that $\boldsymbol{Z}$ can be decomposed as follows:

$$\boldsymbol{Z} = \boldsymbol{U\Sigma V^T} \tag{6}$$

The dimension of each matrix is labelled in the subscript as follows:

$$\boldsymbol{Z} \xrightarrow{\text{RSVD}} \boldsymbol{U}_{X\times l}, \boldsymbol{\Sigma}_{l\times l}, \boldsymbol{V}_{T\times l}$$

Both matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ contain orthogonal orthonormal singular vectors. $\boldsymbol{\Sigma}$ is a diagonal matrix whose diagonal entries are $\sigma_1, \sigma_2, \ldots, \sigma_{l-1}, \sigma_l$. The squared of the sigma values are eigenvalues to the respective eigenvectors in $\boldsymbol{U}$ and $\boldsymbol{V}$ which are ordered such that $\sigma_1 > \sigma_2 > \cdots > \sigma_{l-1} > \sigma_l$. Letting $u_{i,j}$ and $v_{i,j}$ be the i^{th} row and j^{th} column element of matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively, the full-rank RSVD enables each element in $\boldsymbol{Z}$ to be written as follows:

$$z_{x,t} = \sigma_1 u_{x,1} v_{t,1} + \sigma_2 u_{x,2} v_{t,2} + \cdots + \sigma_{l-1} u_{x,l-1} v_{t,l-1} + \sigma_l u_{x,l} v_{t,l} \tag{7}$$

Equation (7) produces one element of matrix $\boldsymbol{Z}$ and it is noticeable that equation (6) gives the complete matrix $\boldsymbol{Z}$. The Lee–Carter model uses the first-rank RSVD, hence each of its elements is estimated as follows:

$$\hat{z}_{x,t} = \sigma_1 u_{x,1} v_{t,1} \tag{8}$$

The left-hand side of equation (8) is the product of $\hat{b}_x$ and $\hat{k}_t$; hence, the constraint $\sum_{x=1}^{X} b_x = 1$ is used to characterise each of the estimated parameters.

$$\hat{b}_x = \frac{u_{x,1}}{\sum_{x=1}^{X} u_{x,1}} \tag{9}$$

$$\hat{k}_t = \sigma_1 v_{t,1} \sum_{x=1}^{X} u_{x,1} \tag{10}$$

We want equation (8) to be written in the form of $\hat{z}_{x,t} = \hat{b}_x \hat{k}_t$. Since $\hat{z}_{x,t} = \sigma_1 u_{x,1} v_{t,1}$, there is no unique representation of $\hat{b}_x$ and $\hat{k}_t$. For example, two possible representations of the parameters are $\hat{b}_x = \sigma_1 u_{x,1}$ with $\hat{k}_t = v_{t,1}$ and $\hat{b}_x = u_{x,1}$ with $\hat{k}_t = \sigma_1 v_{t,1}$.

Writing $\hat{z}_{x,t} = \sigma_1 u_{x,1} v_{t,1} \frac{\sum_{x=1}^{X} u_{x,1}}{\sum_{x=1}^{X} u_{x,1}}$ and letting $\hat{b}_x$ be a weighted average parameter as seen in equation (9) guarantees a unique characterisation of parameterisation of equation (8). Hence, it follows that $\hat{k}_t$ is obtained by equation (10). Therefore, equation (2) implies:

$$\widehat{m}_{x,t} = e^{(\hat{a}_x + \hat{b}_x \hat{k}_t)} \tag{11}$$

From (11), $\widehat{m}_{x,t}$ represents the value of fitted mortality rate for age (age group) x and year t.

B. Modifications and Extensions to the Lee-Carter Model

According to Cairns *et al.* (2011), different versions of the Lee-Carter model can be categorised into two types. The first type is modifications to its statistical foundations. A study by Lee and Miller (2001) reported that forecasts of mortality rates performed better after adjustments to jump-off rates. Brouhns *et al.* (2002) proposed to model the number of deaths in a Poisson setting for Belgian mortality data. They concluded that their method allows for applications in life insurance. Meanwhile, both studies by Booth *et al.* (2002) and De Jong and Tickle (2006) proposed an improved model of k_t to Australian mortality data. In terms of age effects, Delwarde *et al.* (2007) applied the p-splines method to overcome a lack of smoothness in the estimated b_x's.

The second type of Lee-Carter variant is in terms of its extensions and development of new models. Some of these extensions include inclusions of cohort effects, age-period and cohort effects, and multi-population factor. For example, Renshaw and Haberman (2006) included a cohort effect, denoted by γ_{t-x}. Cohort effect is also known as the year-of-birth effect, which means that different individuals experience different mortality improvements determined by their year of birth.

Besides age and period effects, mortality rates in some countries, such as England and Wales, were also determined by cohort effects (Cairns *et al.*, 2008). Currie (2006) simplified the method proposed by Renshaw and Haberman (2006) and removed the robustness problem. Cairns *et al.* (2008) also incorporated cohort effects into their multifactor age-period model for the England and Wales old-aged males mortality data. Another type of age-period-cohort model is proposed by Plat (2009). The model combines the good features proposed by Lee-Carter (1992), Renshaw and Haberman (2006), Currie (2006), and Cairns *et al.* (2006, 2009). The model showed a better fit to U.S. male mortality data.

In 1992, Carter and Lee (1992) suggested a joint-k model to be applied to multi-population data. This means that the model would have the same value of k_t for all populations. However, Li and Lee (2005) argued that the Lee-Carter model works well for single populations because modelling multi-population mortality using the model will lead to divergent in forecasts. To overcome this, they proposed a multi-population model based on the Lee-Carter approach to provide a coherent forecast of mortality for a group of populations.

Figure 1 illustrates the graphical representation of the summary of the Lee-Carter model and its variants as discussed above.

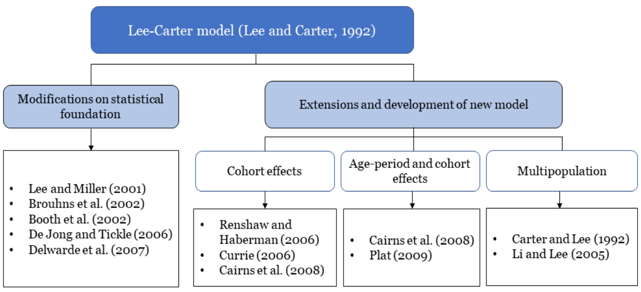


Figure 1. Lee-Carter model and its modifications and extentions.

III. METHODOLOGY

As a supplementary, this paper provides a practical application of the Lee-Carter model in forecasting Malaysian age-specific mortality rates.

A. Data

The data for this study was obtained from the Department of Statistics Malaysia for a period of 28 years, from 1991 to 2018. The data includes the number of people in 17 five-year age groups ranging from 0 to above 80 years old, as well as the number of deaths in those groups. Mortality rates were then obtained by using equation (1). The data set was then divided into two periods: fitting and forecasting. The fitting period spanned from 1991 to 2012, while the forecasting period spanned from 2012 to 2018. By comparing the forecasts to actual out-of-sample data, the fitting period was used to compute one-year, three-year, and six-year forecast horizons and determine the forecast errors.

B. Forecast of Mortality Index

The estimation of the parameters of the Lee-Carter model were conducted by methods explained in Section II (A). Given that the age groups are fixed, a set of $\{k_t | t > T\}$ was determined using an autoregressive integrated moving average (ARIMA) model on time series $\{k_t | t \leq T\}$ to forecast $m_{x,t}$ for $t > T$. An ARIMA model of order p, d, and q, denoted as ARIMA(p, d, q), implies d times difference on the original time series $\{k_t | t \leq T\}$ to obtain a stationary time series $\{k_t^d\}$ for the ARIMA model, p autoregressive terms and q moving average terms as follows:

$$k_t^d = \mu + \varphi_1 k_{t-1}^d + \varphi_2 k_{t-2}^d + \cdots + \varphi_p k_{t-p}^d + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \quad (12)$$

For $d = 0$, the stationary time series k_t^d is equal to k_t. For $d > 0$, $k_t^d = k_t^{d^-} - k_{t-1}^{d^-}$, where $d^- = d$ (see Appendix for details). Once parameters $\{\hat{\mu}, \hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p, \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q\}$ have been obtained, a simulation is run using equation (12) and the assumption that the error terms $\{\varepsilon_t\}$ are independent and identically normal-distributed with mean 0 and constant variance σ^2 (estimated by the ARIMA model) to give $\{\hat{k}_t | t > T\}$ for forecasting the mortality rates beyond time T, $\{m_{x,t} | t > T\}$.

The mortality index in this study was forecasted for six years from 2013 to 2018 with three forecasting horizons: one, three, and six years. The mortality index data has to be stationary prior to applying the ARIMA model (Haberman & Russolillo, 2005), and we employed Lee and Carter's (1992) ARIMA (0,1,0) model for forecasting k_t for $t > T$ as given by equation (13).

$$k_t^f = k_{t-1}^f + \hat{\theta} + \varepsilon_t \text{ for } t = T+1, T+2, \dots, T+r \quad (13))$$

where,

$k_t^f = \hat{k}_t$ estimated by ARIMA (0,1,0)

$\hat{\theta}$ = Drift parameter estimated by ARIMA (0,1,0)

ε_t = Error term at time t

r = forecasting horizon

The following represents the algorithm to obtain k_t^f through simulation of ε_t. Since the mortality index is a function of time, we need it to forecast $m_{x,t}$ by using equation (11). For each $t = T+1, T+2, \ldots, T+r$, 100 random values were generated for ε_t from a normal distribution with mean 0 and standard deviation σ. The latter was estimated by the ARIMA (0,1,0) standard error. The number of generated values is arbitrary as long as a set of values can be obtained to represent the distribution of random variable ε_t. In this study, 100 was used for simplicity.

We substituted the first 100 set of ε_t for $t = T+1$ in the equation (13) to give 100 values of k_{T+1}^f. These 100 values were ordered and the 50th percentile was chosen as the estimate of k_{T+1}^f. This follows from the fact that the 50th percentile coincides the mean of a normal distribution random variable.

The previous step was repeated for $t = T+2$ until $t = T + r$ for obtaining the complete estimate of forecasted mortality index, k_t^f for $t = T+1, T+2, \ldots, T+r$.

C. Measures of Accuracy

The fitting and projection accuracies were evaluated using standard accuracy measures, namely mean squared error (MSE), mean absolute percentage error (MAPE), and mean error (ME). The forecast horizons used in this study are one-year, three-year, and six-year. The forecast performance was evaluated using the out-of-sample data from the year 2013 to 2018. The three measures can be written as follows:

$$MSE = \frac{\sum_{t=T+1}^{T+r}\sum_{x=1}^{X}\left(m_{x,t} - \tilde{m}_{x,t}\right)^2}{rX} \quad (14)$$

$$MAPE = \frac{\sum_{t=T+1}^{T+r}\sum_{x=1}^{X}\left|\frac{m_{x,t} - \tilde{m}_{x,t}}{m_{x,t}}\right| \times 100}{rX} \quad (15)$$

$$ME = \frac{\sum_{t=T+1}^{T+r}\sum_{x=1}^{X}\left(m_{x,t} - \tilde{m}_{x,t}\right)}{rX} \quad (16)$$

where,

$$\tilde{m}_{x,t} = \begin{cases} \hat{m}_{x,t} & \text{for in-sample accuracy measure} \\ m_{x,t}^f & \text{for out-sample accuracy measure} \end{cases}$$

IV. RESULT AND DISCUSSION

A. Mortality Trend in Malaysia

This section discusses the trend of mortality rates in Malaysia from 1991 to 2018.

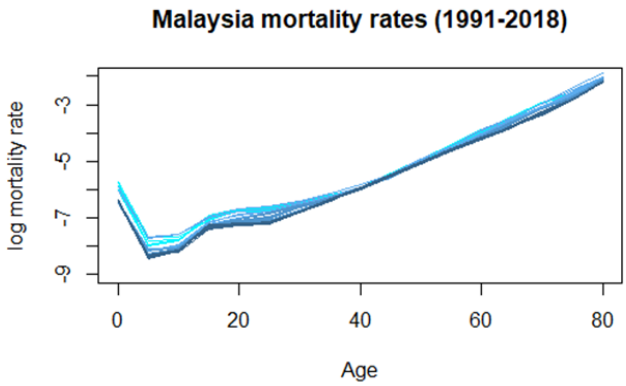


Figure 2. Malaysia mortality rates from 1991 to 2018 (Note: Lighter coloured lines represent the early years, while darker coloured lines represent later years).

Figure 2 shows that mortality rates in Malaysia followed a similar pattern every year and decreased over the 28-year observation period. Malaysia has seen an increase in mortality with respect to age. The country's mortality rate is highest between the ages of 0 and 4, then rapidly declines between the ages of 15 and 19. At the age of 20, mortality rates began to rise and continued to rise until the age of 80 and above. From one calendar year to the next, Malaysia's mortality rate decreased for all age groups. Malaysia has significantly increased its healthcare expenditure over the years (Ministry of Health Malaysia, 2019), which has an impact on the Malaysian mortality rate, as healthcare expenditure allocation affects people's life expectancy (Linden & Ray, 2017; Babenko *et al.*, 2019).

Child mortality has improved globally, with UNICEF (2015) reporting a 53% reduction in child mortality under the age of five between 1990 and 2015. Figure 1 shows that Malaysia is on the right track in this regard, as evidenced by the decreasing trend in child mortality from 1991 to 2018.

Another point to note is that the 'accident hump' that occurs between the ages of 10 and 29 is more prominent in later years. The term 'accident hump' refers to a noticeable hump on the mortality curve that occurs most commonly between the ages of 10 and 40, and is more prominent in males than females (Heligman & Pollard, 1980; Haldrup & Rosenskjold, 2019). The 'accident hump' in Figure 2 may have been caused by male mortality rates, which had a more distinct 'accident hump' than female mortality rates (Ibrahim *et al.*, 2021).

B. Parameters Estimation of the Lee-Carter Model

The age and time parameters of the Lee–Carter model were estimated for the years 1991 to 2012. Figure 3 depicts the plots of the estimated parameters.

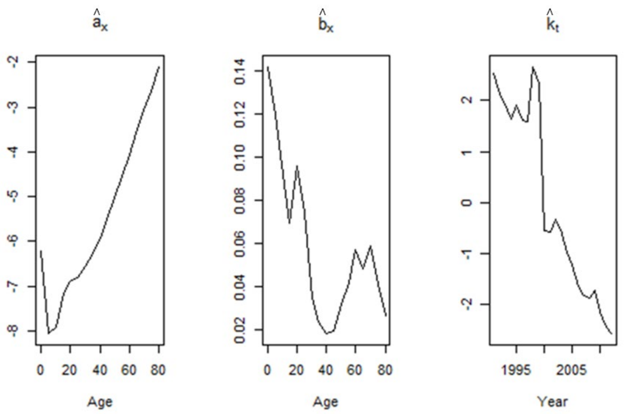


Figure 3. Estimated values of parameters a_x, b_x and k_t obtained from the Lee–Carter model.

The plot of a_x which represents the average log mortality rate, shows a high mortality rate in the age group 0 to 4 years old, followed by a sharp decline in the age group after that. The mortality rates of the remaining age groups increased as they grew older.

Aside from the age parameter, the Lee–Carter model includes a time-varying parameter or mortality index denoted by k_t , while b_x captures the rate of change in mortality as k_t changes. The plot of k_t shows a decreasing trend from 1991 to 2012. Similarly, the plot of b_x shows a decreasing trend for all age groups. Similar patterns of mortality improvement were visible in other countries, with the majority of the global population experiencing increased life expectancy (Chavhan & Shinde, 2016; Bozikas & Pitselis, 2018).

C. Fitted Mortality Rates

Fitted log mortality rates for the fitting period of 1991 to 2012 were obtained after parameter estimation. Figure 4 compares the actual and fitted mortality rates for the years 1991, 2002, and 2012 to denote the beginning, middle, and end of the fitting period, respectively. Actual mortality rates are represented by solid lines, and fitted mortality rates are represented by dotted lines.

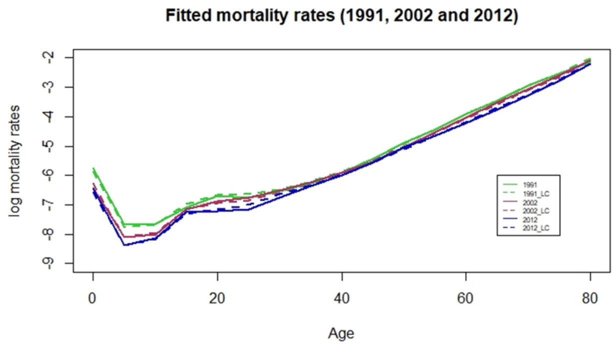


Figure 4. Fitted log mortality rates for the years 1991, 2002, and 2012.

Figure 4 shows that the Lee–Carter model fits well into Malaysian mortality data by age, with little deviation from actual data for the three years studied. The model also fits other populations well, such as Indian mortality (Chavhan & Shinde, 2016) and cancer incidence rates (Yue *et al.*, 2018). Table 1 presents the results of in-sample errors for the years 1991 to 2012 which shows that the Lee–Carter model fits Malaysian mortality rate well since the error values are very small. This indicates a small difference between the actual observed mortality rates and the fitted mortality rates. It also supported the findings by Gylys and Šiaulys (2020) that the Lee–Carter model fits data with a stable mortality trend better than populations with volatile mortality trends and outliers. Then, fitted log mortality rates by year were calculated for each of the 17 age groups. Table 2 shows the 17 age groups that were divided into four categories (Sully *et al.*, 2020; Ricci *et al.*, 2010; United Nations, 2019).

Table 1. In-sample errors

	MSE	MAPE (%)	ME
Accuracy measure	0.00000421	4.055862	0.00002399

Table 2. Classification of age groups.

Category	Age groups
Child and Adolescence	0–4, 5–9, 10–14, 15–19
Adult	20–24, 25–29, 30–34, 35–39
Middle Age	40–44, 45–49, 50–54, 55–59, 60–64
Elderly	65–69, 70–74, 75–80, 80+

As shown in Figure 5, the Lee–Carter model did not fit well for age-specific mortality by year across all ages. The mortality rates for the age group 0 to 4, as well as all age groups in the Adult and Elderly categories, are significantly different from the actual rates.

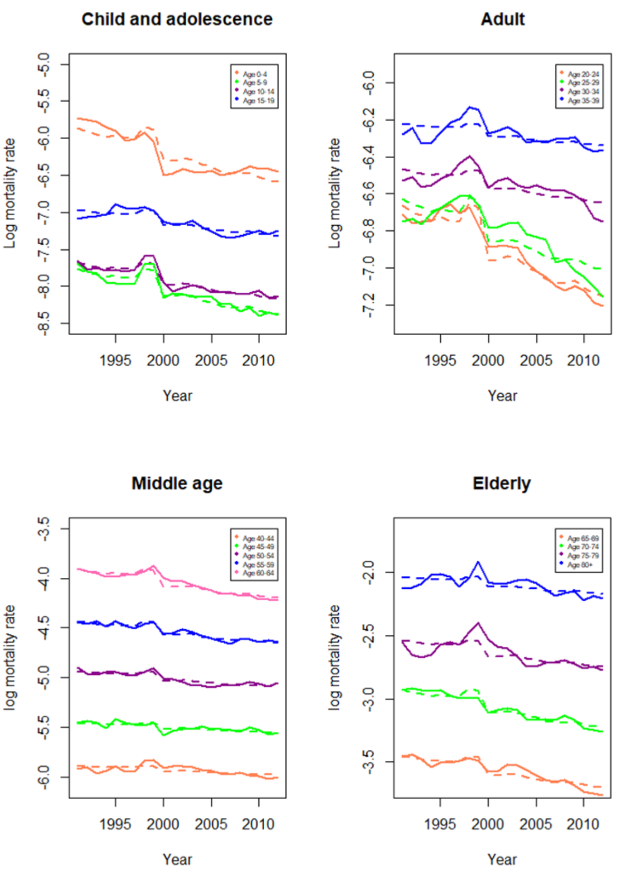


Figure 5. Fitted log mortality rates by age groups (Note: Solid lines represent the actual log mortality rate, whereas dotted lines represent the fitted log mortality rate).

Figure 5 also shows a decreasing age-specific mortality trend by year and fitted log mortality rates in the Middle Age category, which is consistent with the findings of Ngataman *et al.* (2016). The other three categories, on the other hand, are inconsistent in terms of deviations from actual mortality rates. This seemingly contradictory result could be due to the different samples used in both studies. While this study examined the Malaysian population as a whole, Ngataman *et al.* (2016) examined age-specific mortality rates by gender.

Figure 6 depicts the results of a mortality index forecast at the 5th, 50th, and 95th percentiles. Dotted lines represent the forecast mortality index at the 5th and 95th percentiles, while the line represents the forecast mortality index at the 50th percentile.

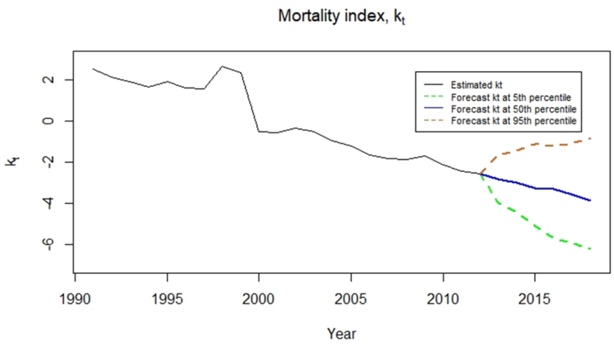
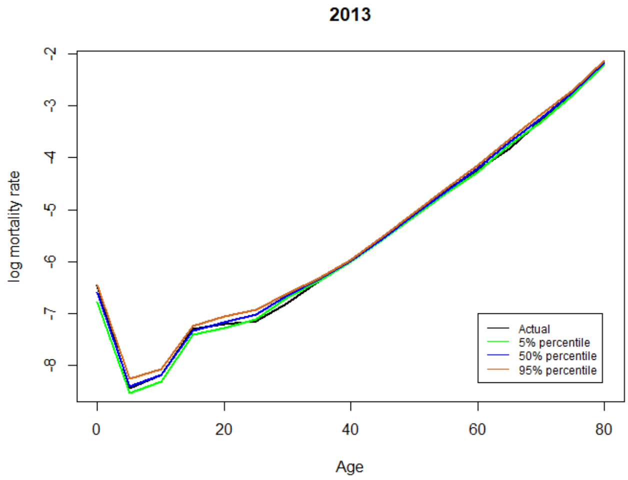


Figure 6. Estimated and forecast mortality index for the years 1991 to 2018.

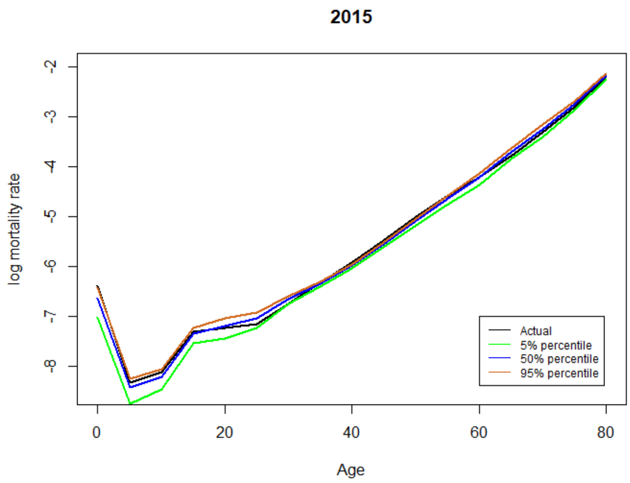
According to Figure 6, the forecast mortality index at the 95th percentile increased from 2013 to 2018, whereas the forecast mortality index at the 5th and 50th percentile decreased. It can be suggested that the Malaysian population is projected to experience mortality improvement. These findings are consistent with those of Zulkifle *et al.* (2019), Bozikas and Pitselis (2018), and Maccheroni and Nocito (2017), which emphasised the declining mortality index as a measure of mortality improvement. Also, their findings indicated that females experienced greater mortality improvements than males.

D. Forecasting Period Variants

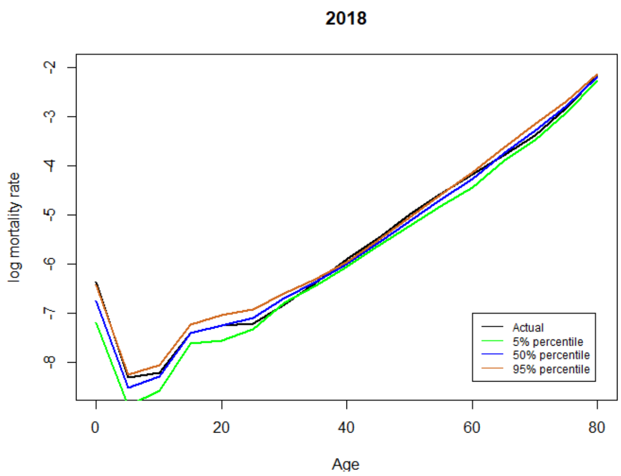
This study also focused on forecasting period variants, with out-of-sample log mortality rates that were forecasted for one-year, three-year and six-year periods.



(a) 1 year



(b) 3 years



(c) 6 years

Figure 7. Different forecast horizons of log mortality rates for: (a) 1 year, (b) 3 years, and (c) 6 years.

According to Figure 7, the one-year forecast, as opposed to the three- and six-year forecasts, remains close to the actual log mortality rates. Figures 7b and 7c show that there are significant differences between actual and forecast values for both the 5th and 95th percentiles. However, the forecast values at the 50th percentile (median) in both periods remain close to actual log mortality rates. The results also show that a longer forecasting period results in a wider gap, particularly between the ages of 0 and 24. These findings imply that different forecasting periods will produce different results. Interestingly, in the context of different fitting periods, Ibrahim *et al.* (2021) discovered that a Lee–Carter model with a shorter fitting period outperformed a model with a longer fitting period in modelling Malaysian mortality rate.

E. Forecasts Accuracy

The forecast accuracy was evaluated for each forecast horizon period using MSE, MAPE, and ME. Table 3 shows the forecast accuracy results in terms of the measure of errors between the actual mortality rates and the forecast mortality rates at the 50th percentile.

Table 3. Forecast accuracy

Forecast horizon	**MSE**	**MAPE (%)**	**ME**
1-year	0.00000176	5.305624	-0.00063
3-year	0.00000131	6.262131	-0.00045
6-year	0.00000130	7.421837	-0.00022

According to Table 3, the Lee–Carter model fits Malaysian mortality rate fairly well since the error values are very small. This indicates a small difference between the actual observed mortality rates and the forecasted mortality rates. The MAPE and ME increase as the forecast horizon increases, whereas the MSE decreases. MAPE results are consistent with a study on Malaysian fertility rates conducted by Hanafiah and Jemain (2013), in which the MAPE increases as the forecast horizon increases. As a result, this study found that longer forecasting periods result in higher error measures than shorter forecasting periods. In conclusion, a shorter forecast period fits Malaysian mortality rate better. Also, the Lee–Carter model fits data with a stable mortality trend better than populations with volatile mortality trends and outliers (Gylys & Šiaulys, 2020).

V. CONCLUSION

The purpose of this study is to guide beginners, particularly students and young researchers, in understanding the basics of the Lee-Carter model**.** As a supplementary note, this paper also provides practical application to forecast Malaysian mortality rates using the model.

Overall, the Malaysian mortality rate has shown a decreasing trend from 1990 to 2018. The log mortality curve shows a noticeable 'accident hump' as the year progresses, particularly between the ages of 10 and 29. This observation is consistent with Heligman and Pollard (1980), which state that an 'accident hump' usually occurs between the ages of 10 and 40. While the model fits the Malaysian mortality rate by age well, it does not fit age-specific log mortality rates by year across all age groups. The ARIMA (0,1,0) model was also used in this study to forecast the mortality index at the 5th, 50th, and 95th percentiles. Further research was conducted on forecasting using period variants, and error measures revealed that using a longer forecast horizon resulted in a higher error measure. The Lee–Carter model fits the Malaysian mortality rate rather well because the error values are very small.

Further research could compare Lee-Carter's forecasting performance by using the first estimation and re-estimation of the mortality index. The results of this study can also be compared to other stochastic mortality forecasting models.

VI. ACKNOWLEDGEMENT

The authors would like to thank Universiti Kebangsaan Malaysia (Research Grant: GGPM-2019-066) for their financial support and the Department of Statistics, Malaysia for providing the data for our research.

VII. REFERENCES

Babenko, V, Nehrey, M, Gaponova, E, Ryzliikova, N & Zaporozhets, E 2019, 'Life expectancy of population of the country: The role of health services effectiveness', *Research in World Economy',* vol. 10, no. 4, pp. 86-91.

Booth, H, Maindonald, J & Smith, L 2002, 'Applying Lee-Carter under conditions of variable mortality decline', *Population Studies,* vol. 56, pp. 325-336.

Booth, H & Tickle, L 2008, 'Mortality modelling and forecasting: A review of methods', *Annals of Actuarial Science,* vol. 3, pp. 3-43.

Bozikas, A & Pitselis, G 2018, 'An empirical study on stochastic mortality modelling under the age-period-cohort framework: The case of Greece with applications to insurance pricing', *Risks,* vol. 6, no. 44, pp. 1- 34

Brouhns, N, Denuit, M & Vermunt, J 2002, 'A Poisson log-bilinear regression approach to the construction of projected lifetables', *Insurance: Mathematics and Economics,* vol. 31, pp. 373-393.

Cairns, AJG, Blake, D & Dowd, K 2006, 'A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration', *Journal of Risk and Insurance,* vol. 73, no. 4, pp. 687-718.

Cairns, AJG, Blake, D & Dowd, K 2008, 'Modelling and management of mortality risk: A review', *Scandinavian Actuarial Journal,* vol. 2008, pp. 79-113.

Cairns, AJG, Blake, D, Dowd, K, Coughlan, GD, Epstein, D, Ong, A & Balevich, I 2009, 'A quantitative comparison of stochastic mortality models using data from England and Wales and the United States', *North American Actuarial Journal,* vol. 13, pp. 1-35.

Cairns, AJG, Blake, D, Dowd, K, Coughlan, GD, Epstein, D & Khalaf-Allah, M 2011, 'Mortality density forecasts: An analysis of six stochastic mortality models', *Insurance: Mathematics and Economics,* vol. 48, pp. 355-367.

Carter, LR & Lee, RD 1992, 'Modeling and forecasting US sex differentials in mortality', *International Journal of Forecasting,* vol. 8, pp. 292-411.

Chavhan, RN & Shinde, RL 2016, 'Modeling and forecasting mortality using the Lee-Carter model for Indian population based on decade-wise data', *Sri Lankan Journal of Applied Statistics,* vol. 17, no. 1, pp. 51-68.

Currie, ID 2006, *Smoothing and forecasting mortality rates with p-splines*, Conference presentation, Institute and Faculty of Actuaries, viewed 26 July 2022, < http://www.actuaries.org.uk/documents/smoothing-and-forecasting-mortality-rates-p-splines-slides>

De Jong, P & Tickle, L 2006, 'Extending Lee-Carter mortality forecasting', *Mathematical Population Studies',* vol. 13, pp. 1-18.

Delwarde, A, Denuit, M & Eilers, P 2007, 'Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach', *Statistical Modelling,* vol. 7, no. 1, pp. 29-48.

Gylys, R & Siaulys, J 2020, 'Estimation of uncertainty in mortality projections using state space Lee-Carter model', *Mathematics,* vol. 8, no. 1053, pp. 1-22.

Haberman, S & Russolillo, M 2005, 'Lee-Carter mortality forecasting: Application to the Italian population', Actuarial Research Paper No. 167, Cass Business School, London, UK.

Haldrup, N & Rosenskjold, CPT 2019, 'A parametric factor model of the term structure of mortality', *Econometrics,* vol. 7, pp. 1-22.

Hanafiah, H & Jemain, AA 2013, 'Structural modelling of age specific fertility curves in Peninsular Malaysia: An approach of Lee-Carter method', *AIP Conference Proceedings,* vol. 1571, pp. 1062-1068.

Heligman, L & Pollard, JH 1980, 'The age pattern of mortality', *Journal of the Institute of Actuaries,* vol. 107, pp. 49-80.

Ibrahim, NSM, Lazam, NM & Shair, SN 2021, 'Forecasting Malaysian mortality rates using the Lee-Carter method with fitting period variants', *Journal of Physics: Conference Series,* vol. 1988, pp. 1-10.

Lee, RD & Carter, LR 1992, 'Modeling and forecasting U.S. mortality', *Journal of the American Statistical Association,* vol. 87, no. 419, pp. 659-671.

Lee, R & Miller, T 2001, 'Evaluating the performance of the Lee-Carter method for forecasting mortality', *Demography,* vol. 38, no. 4, pp. 537-549.

Li, N & Lee, R 2005, 'Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method', *Demography,* vol. 42, no. 3, pp. 575-594.

Linden, M & Ray, D 2017, 'Life expectancy effects of public and private health expenditures in OECD countries 1970-2012: Panel time series approach', *Economic Analysis and Policy,* vol. 56, pp. 101-113.

Maccheroni, C & Nocito, S 2017, 'Backtesting the Lee-Carter and the Cairns-Blake-Dowd stochastic mortality models on Italian death rates', *Risks,* vol. 5, no. 34, pp. 1-23.

McNown, R & Rogers, A 1989, 'Forecasting mortality: A parameterized time series approach', *Demography*, vol. 26, no. 4, pp. 645-660.

Ministry of Health Malaysia 2019, *Malaysia National Health Accounts Health Expenditure Report 1997-2017,* < https://www.moh.gov.my/moh/resources/Penerbitan/PenerbitanUtama/MNHA/Laporan_MNHA_Health_Expenditure_Report_1997-2017_03122019.pdf>

Ngataman, N, Ibrahim, RI & Yusuf, MM 2016, 'Forecasting the mortality rates of Malaysian population using Lee-Carter method', *AIP Conference Proceedings,* vol. 1750, pp. 020009-1-6.

Plat, R 2009, 'On stochastic mortality modeling', *Insurance: Mathematics and Economics,* vol. 45, no. 3, pp.393-404.

Renshaw, AE & Haberman, S 2006, 'A cohort-based extension to the Lee-Carter model for mortality reduction factors', *Insurance: Mathematics and Economics,* vol. 38, pp. 556-570.

Ricci, NA, Aratani, MC, Doná, F, Macedo, C, Caovilla, HH & Ganança, FF 2010, 'A systematic review about the effects of the vestibular rehabilitation in middle-age and older adults', *Brazilian Journal of Physical Therapy*, vol. 14, pp. 361-371.

Sully, EA *et al*. 2020, 'Adding It Up: Investing in Sexual and Reproductive Health 2019', New York: Guttmacher Institute, <https://www.guttmacher.org/report/adding-it-upinvesting-in-sexual-reproductive-health-2019>.

Tsai, CC & Lin, T 2017, 'A Bühlmann credibility approach to modeling mortality rates', *North American Actuarial Journal*, vol. 21, no. 2, pp. 204-27.

UNICEF 2015, *Beyond averages: Learning from the MDGs,* Progress for Children no. 11, New York, USA.

United Nations, 2019, *World Population Prospects 2019 Volume II: Demographic Profiles,* Department of Economic and Social Affairs, United Nations.

Yue, JC, Wang, HC, Leong, YY & Su, WP 2018, 'Using Taiwan National Health Insurance Database to model cancer incidence and mortality rates', *Insurance: Mathematics and Economics',* vol. 78, pp. 316-324.

Zulkifle, H, Yusof, F & Nor, SRM 2019, 'Comparison of Lee Carter model and Cairns, Blake and Dowd model in forecasting Malaysian higher age mortality', *Matematika,* vol. 35, pp. 65-77.

APPENDIX

For $d = 0$, we set $k_t^d = k_t$. The following are examples when $d = 1, d = 2$ and $d = 3$.

(Note: k_t^d is the stationary time series for ARIMA model)

d	k_t^d
1	$k_t^1 = k_t^0 - k_{t-1}^0 = k_t - k_{t-1}$
2	$k_t^2 = k_t^1 - k_{t-1}^1$ $= (k_t - k_{t-1}) - (k_{t-1} - k_{t-2})$ $= k_t - 2k_{t-1} + k_{t-2}$
3	$k_t^3 = k_t^2 - k_{t-1}^2$ $= (k_t - 2k_{t-1} + k_{t-2}) - (k_{t-1} - 2k_{t-2} + k_{t-3})$ $= k_t - 3k_{t-1} + 3k_{t-2} - k_{t-3}$