Forecasting selected specific age mortality rate of Malaysia by using Lee-Carter model

Halim Shukri Kamaruddin\textsuperscript{1} and Noriszura Ismail\textsuperscript{2}

\textsuperscript{1}Taylor’s Business School, Taylor’s University, 47500, Subang Jaya, Selangor, Malaysia
\textsuperscript{2}Pusat Pengajian Sains Matematik, Fakulti Sains dan Teknologi Universiti Kebangsaan Malaysia, 43600, Selangor, Malaysia

Email: HalimShukri.Kamaruddin@taylors.edu.my

Abstract. Observing mortality pattern and trend is an important subject for any country to maintain a good social-economy in the next projection years. The declining in mortality trend gives a good impression of what a government has done towards macro citizen in one nation. Selecting a particular mortality model can be a tricky based on the approached method adapting. Lee-Carter model is adapted because of its simplicity and reliability of the outcome results with approach of regression. Implementation of Lee-Carter in finding a fitted model and hence its projection has been used worldwide in most of mortality research in developed countries. This paper studies the mortality pattern of Malaysia in the past by using original model of Lee-Carter (1992) and hence its cross-sectional observation for a single age. The data is indexed by age of death and year of death from 1984 to 2012, in which are supplied by Department of Statistics Malaysia. The results are modelled by using RStudio and the keen analysis will focus on the trend and projection of mortality rate and age specific mortality rate in the future. This paper can be extended to different variants extensions of Lee-Carter or any stochastic mortality tool by using Malaysia mortality experience as a centre of the main issue.

1. Introduction
The rise in the life expectancy and gradual decline in the mortality rate across ages in any geographical area has attracted researchers in governments or corporate sector to investigate in further consequences of this situation. Mortality forecasting has become the central focus of study among the academicians and industrial researchers across borderless worldwide. The outcome can be used in different areas mainly in planning of retirement fund, pricing of life insurance contracts, public health or in further medical research interventions. The World Health Organization (WHO) is one of the bodies in the world concerns the movement of worldwide mortality as later is the source of public health and morbidity that most countries can rely on. Hence forecasting future mortality is an important subject for any health organization without exceptional for big organization like WHO in ensuring the stability of health in one nation or specific region in general. Accuracy and reliability forecasting tool play an important role in predicting future mortality and hence estimating the morbidity in specific sub-region. A quite number of mortality models have been proposed since Gompertz model firstly published his law of mortality in 1825 [1]. The methods used in mortality forecast were relatively simple and only involved a fair degree of subjective judgment [2]. Recent methods and tools have been developed by actuaries or demographers involved sophisticated and complicated models. Some of these methods using stochastic processes in the modelling with high
outcome expectation in producing more reliable result, which in fact has better significant outcome compared to subjective opinions methods.

This paper presents the fitting of Malaysia mortality data for the past 29 years from the calendar years 1984 to 2012 using original Lee Carter model (1992) and hence from the fitted, Malaysian mortality can be forecasted up to the next ten years. From there, the two dimensional fitted and forecasted mortality data will be cross-sectioned and examine for the certain ages. In Lee Carter model, it combines a demographic model with a statistical model of times series in order to forecast mortality rates. Girosi and King [3] views Lee Carter model as a special type of a multivariate process in which the covariance matrix depends on the drift vector and the innovations are inter-temporally correlated. The study on forecasting of mortality has been in the headline of mortality researches when the mortality rate decline dramatically in all developed countries in the twentieth century [4]. According to World Health Ranking 2012 published by WHO, Malaysia’s life expectancy in 2011 is 73.9 which a huge jump compared to Malaysia’s life expectancy in 1961, which only 54.3.

Apart from concentrating government side impact issues in mortality, insurance companies are also culprit in mortality issues mentioned above. Insurance companies are exposed to uncertainties which lead to whether a surplus or deficit of holding insurance contracts in money term. Life insurance contract in general specifies a stream of payments between the insured and the insurer contingent on some predetermined insurance events. Take whole life insurance as an example, insurer will pay benefits upon the death of the insured. Judging from this situation, the mortality rate and life expectancy are two of main studies should be focused in determining the contract price. Problems have been alarmed to either government or corporate parties in regards to the life insurance issued, mainly due to the inclining of the life expectancy of individuals. Insurance company also concerns on issues regarding insurance products that offer individuals lifetime coverage which highly dependent on the period until when they are alive which in other word individual’s life expectancy. Venturing these issues, insurance companies have to make sure a proper lay out on how their funds should be adjusted according to the rapid declining in the mortality rates in Malaysia. In other word, an insurance company should be able to identify and model the sources of risk adequately such that it is be able to price the contract correctly [5]. Hence, a suitable and applicable model or tool of adjustment from the original model is essential for insurance companies so that they can have a clear view in the long run on the mortality rate and life expectancy of Malaysia in guiding them to decide the issued contracts with full accuracy.

Forecasting mortality rate has been developed as early as at the beginning of 1980s which can be sub-divided under three broad approaches: expectation, extrapolation and explanation. In the 1980s and early 1990s, actuaries have relied heavily on the subjective judgments for mortality forecast [2]. Evolution of research in mortality forecasting has drastically changed in the beginning of 1990s, researchers turn to sophisticated extrapolative methods and this forecasting method fully uses the element of statistical method [6]. This approach consists of the traditional and simple extrapolation of aggregate measures for example life expectancy and also more complex methods like the Lee-Carter model. Extrapolative methods assume future trends will essentially be a continuation of the past. In mortality forecasting, this is usually a reasonable assumption because of historical regularities.

Explanation approach makes use of the structural or epidemiological models of certain causes of death involving disease processes and known risk factors, example of this approach is the dependence of lung cancer on tobacco smoking. The main advantage of the explanatory approach is the feedback mechanisms and limiting factors can be taken into account in the modeling. However, this approach is not widely applied in mortality forecasting due to the insufficient study of the relationships between risk factors and mortality, which make them less reliable in forecasting. Many models used in explanatory forecasting are regression-based and therefore they are fit to be used in the GLM methods.

Some situations extrapolation method will not be useful for instance due to certain unpredictable factors, such as fluctuation and stagnation of mortality rate, the researchers began to develop alternative models instead of using simple linear extrapolation. Method that always used for this kind of case is time series due to their stochastic properties and capability to do calculation of probabilistic prediction interval for the estimated value. Time series model can have zero-factor or multiple factors such as time, age and cohort (factor that include correlation between time and age). For time series
models with a zero factor, a simple random walk with drift model is developed and applied to age-specific rate. However, the model has a disadvantage which it will create a doubtful age pattern, so the researchers use iterated form of the seemingly unrelated regression estimator to forecast first differences of age-specific log mortality rates based on cross-correlated time series models [1]. For one-factor time series models, some researchers use parameterization functions and Brass relational model, which widely used when encountering poor data situation. Parameterization functions (laws of mortality) are one-factor models signifying the age model of mortality with the advantage of forecasting smoothness across age, familiarity and interpretability. Parameterization method requires repetitious fitting procedure from annual data to obtain a time series for each model parameter using the time series methods [1]. For two-factor time series model, e.g. Lee-Carter model and bivariate regression model which use generalized linear model (GLM) to discover the relationship of mortality rate between age and time (calendar year). Lee-Carter model uses matrix decomposition method to identify independent mortality factors such as age, time and other factors whereas GLM used appropriate time value for forecasting. As a result, time in GLM is considered as explicit covariate while time in Lee-Carter is an independent mortality factor [1]. In conclusion from past research certain extrapolative models and methods will be chosen to allow the discovery of the structural changes which happened among Malaysians as well as the impact of time and age factors towards in constructing fitting mortality rate and hence forecasting the mortality rate in the future which our main concerned focus in Malaysia.

2. Methodology
The Malaysia mortality data for this research are from year 1984 to year 2012 which obtained from Department of Statistics of Malaysia. For each calendar year, it consists of each age group from 0 to 80 year-old, number of Malaysia population as the exposure and number of death for each age. The data obtained are used to fit into the Lee-Carter model for both fitted mortality model and forecasted model. There are three ages selected, 5, 25 and 60 for mortality rate cross-sectional observation. These three ages are chosen to represent mortality pattern for three different groups of ages, i.e. children, young adult and elderly. The fitted mortality rate produced from Lee-Carter model will be cut cross and the trend of mortality will be analysed and observed.

2.1. Lee-Carter model
The basic form of Lee-Carter model is as following:

\[ \ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t} \]  

where

\( m_{x,t} \): the age-specific death rate for age \( x \) at time \( t \)
\( a_x \): the average age-specific mortality
\( b_x \): the rate of improvement of the level of mortality at age \( x \)
\( k_t \): the mortality index at time \( t \)
\( \varepsilon_{x,t} \): the random error for age \( x \) at time \( t \)

In Lee-Carter model, there is a linear relationship among the parameters and explanatory factors which are the age \( x \), and the time \( t \). The parameters \( b_x \) and \( k_t \) are estimated by applying the Singular Value Decomposition (SVD) method, assuming the errors are homoscedastic (all the errors have the same finite variance). Using this method, parameters \( b_x \) and \( k_t \) are subject to the constraints of:

\[ \sum_x b_x = 1 \]  
\[ \sum_t k_t = 0 \]

Whereas, the \( a_x \) is simply the average values over time of the \( \ln m_{x,t} \) values for each \( x \). Therefore, \( a_x \) is then can be estimated by the following formula:
\[ \hat{a}_x = \frac{1}{n} \sum_{t=1}^{n} \ln(m_{x,t}) \]  

(4)

At the same time, \( k_t \) is estimated by

\[ \bar{k}_t = \sum_x (\ln m_{x,t} - a_x) \]  

(5)

These two estimated parameters will be used to estimate the \( b_x \) using the least squares estimation which the formula as follows:

\[ (\ln m_{x,t} - \bar{a}_x) = b_x \bar{k}_t + \varepsilon_{x,t} \]  

(6)

\[ \sum_{x,t} (\ln m_{x,t} - \hat{a}_x - \hat{b}_x \hat{k}_t)^2 \Rightarrow \hat{b}_x = \frac{\sum_{t=1}^{n} k_t (\ln m_{x,t} - a_x)}{\sum_{t=1}^{n} k_t^2} \]  

(7)

The estimated value of \( \alpha_x \) or \( \bar{a}_x \) will give the general pattern of the age-specific death rates, \( m_{x,t} \), and estimated value of \( k_t \) or \( \bar{k}_t \) is the index describes the variation in the level of mortality at \( t \). The \( \hat{b}_x \) will describe the tendency of changes of mortality at the specific age \( x \) when the general level of mortality \( (\bar{k}_t) \) changes. There is high possibility lacking of similarity between the predicted and the actual deaths at the beginning stage of estimation due to the estimation is calculated according to the log of death rates instead of the death rates. In order to overcome the issue and hence to reduce the discrepancies between the fitted death rates and the actual number of deaths, the \( \hat{k}_t \) is re-estimated, by taking the same value of \( \hat{a}_x \) and \( \hat{b}_x \) which estimated from the first step. According to [7], the new \( \hat{k}_t \) denotes as \( \bar{k}_t \), which estimated by an iterative procedure. When the fitted death rates are equal to the actual death rates, then following equation will be fulfilled:

Let

\[ q_{x,t} = e^{(\hat{a}_x + b_x \bar{k}_t)} \]  

(8)

Then

\[ \sum_{t} d_{x,t} = \sum_{t} N_{x,t} q_{x,t} \]  

(9)

where

\[ d_{x,t} \] : the actual observed death rate

\[ N_{x,t} \] : the population aged \( x \) at time \( t \)

The iterative method is done as follows:

1. First, compare the total expected deaths (TE) to the total actual death (TA) in each period.
2. There are three possible states of these comparison:
   (i) If \( TE > TA \), decrease in \( TE \) is needed, thus adjusting the \( \hat{k}_t \) so that \( \hat{k}_t \) will be:
      1. \( \hat{k}_t = \bar{k}_t (1 - d) \), if \( \bar{k}_t > 0 \)
      2. \( \hat{k}_t = \bar{k}_t (1 + d) \), if \( \bar{k}_t < 0 \)
      the \( d \) is a small number.
   (ii) If \( TE = TA \), stop here the iteration.
   (iii) If \( TE < TA \), increase in \( TE \) is needed, thus adjusting the \( \hat{k}_t \) so that \( \hat{k}_t \) will be:
      1. \( \hat{k}_t = \bar{k}_t (1 + d) \), if \( \bar{k}_t > 0 \)
      2. \( \hat{k}_t = \bar{k}_t (1 - d) \), if \( \bar{k}_t < 0 \)
3. Go back to Step 1.
As the $k_t$ index is obtained by the two stages of estimation, it is then able to forecast the index by using ARIMA. Lee and Carter (1992) describes $k_t$ as a random walk with a drift term model such that:

$$k_t = k_{t-1} + d + \varepsilon_t$$  \hspace{2cm} (10)

where

- $d$: the drift parameter
- $\varepsilon_t$: the uncorrelated term

The term $d$ is estimated by the least square estimator which equal to

$$\hat{d} = \frac{1}{n-1} (k_n - k_1)$$  \hspace{2cm} (11)

After the $k_t$ index has been forecasted, mortality rate of the projected years is then possible to be obtained. From Andreozzi, Blaconá, & Arnesi [7], the forecast values of $k_{n+h}$ are substituted in the formula:

$$\hat{m}_{x,n+h} = \hat{m}_{x,n} e^{\hat{b}_x (k_{n+h} - \hat{k}_n)}$$  \hspace{2cm} (12)

where

- $n$: the final year from which data are available
- $h$: the forecast horizon
- $x$: the age

As mention at the beginning of this section the procedure will be run using R-programming and the result and outcome of the data will be discussed in the analysis section.

3. Analysis

The fitted mortality rate from 1984 to 2012 using Lee-Carter portrayed in Figure 1(a) which has been generated by using RStudio. Each coloured lines represent different years and the colours of the lines follow the rainbow colour scheme: red, orange, yellow, green, blue, indigo and violet. The earliest year will be shown in red colour while the latest year will be in violet. Based on the Figure 1(a), the plotted lines generally show the mortality rates increase along with the age in any observed calendar year. However, between the age 0 and 20, the lines formed a rough U pattern. In between these ages, the lines are smoother and less jagged. The lines in Figure 1(a) also show the mortality rate for infants is higher compared to any children aged, young adults in between 15 and 34 and adults before age 40. These findings proved Malaysia has a high birth death rate. But across the calendar year from 1984 to 2012, the infant mortality rate from age 0 to age 5 has sharp decreasing dive with approximately 63.7% in average for these twenty-nine years observation. Government and WHO play an important role in realisation of this improvement. WHO has long term target to reduce two thirds of under five-year-old mortality from 1990 to 2015, which this agenda called Millennium Development Goal (MDA) [8]. From Figure 1(a) mortality rate under five-year-old is declining with average of 8.8% per annum. This statistically implying Malaysia has improved significantly in maternal and child health throughout the years. This drop in the infant mortality rate may due to several different factors. This remarkable improvement in minimizing the child and maternal mortality are the result of hard works from different factor of reasons mainly in Malaysia’s policies, strategies and programs that involving socio-economic, cultural, educational, gender, and poverty dimensions [9]. Study in the past also found the mother’s educational level has influenced her child’s health in Malaysia. Thus, children’s chances of surviving have positive correlation with the mother’s education [8]. Observation of 29 years in Figure 1(a) also spot a trend of decreasing for mortality rate between age 5 to 10, age 15 to 40 and age 80 and above. However, there is slight increasing in rate of mortality for age in between 45 and 55 throughout 29 years. Moreover, roughly starting from age 30, it can be observed the rate of mortality increase begins to amplify. The Figure 1(a) also can be concluded in general the mortality rate of Malaysia have been lower from one calendar year to another. This may be due to the improvement on the hospital facilities or generally healthcare in Malaysia [10]. Besides that,
Malaysian citizen also pretty much aware over the years the importance of healthy lifestyle, and adopt it in such balance diets and exercises to maintain the healthiness.

![Graph of Fitted Line Using Lee-Carter Model](image1.png)  
**Figure 1(a). **Graph of Fitted Line Using Lee-Carter Model

![Graph of Forecasted Malaysia Mortality from 2013 to 2022.](image2.png)  
**Figure 1(b). **Graph of Forecasted Malaysia Mortality from 2013 to 2022.

The Figure 1(b) shows the forecasted mortality rate from year 2013 to 2022. There is no different in colour representing compared to The Figure 1(a). Mortality rate of Malaysia has obvious decreasing trend for infants and teenagers meanwhile the rest of the ages’ trends has small increasing tendencies. In general in the future in whatever forecasted year, there is a sharp decrement of mortality rate at the infant age and also slightly decrease in teenager ages. However, a small increment of mortality rate from age 10 to 20. The elder group has the highest log mortality rates throughout the forecasted calendar years as compared to other age groups.

As mentioned in the methodology, further detail analysis of mortality rate can be observed in deferent angle which cross-section of the two dimensional mortality rate at certain selected ages. Figure 2(a), Figure 2(b) and Figure 2(c) are cross-section from different ages, which combination of fitted and forecasted mortality rate from Figure 1(a) and Figure 1(b). Figure 2(a) shows the cross-section of the log mortality rate for age 5. In general the rate has downward trend. The log mortality rate at age 5 declined from -7.51 at year 1984 to -8.35 at year 2012 approximately. However, there are minimum fluctuations within this period of time. Influenced by the modelling tool and historical data, the forecasted log mortality rate is continue to decline constantly from -8.35 to -8.75 approximately at year 2022. In overall, there is no large change on log death rate at age 5. Figure 2(b) show the log mortality rate for this age declined from -6.8 at year 1984 to -7.10 at year 1990. The log mortality rate increases inconsistently from -7.10 at year 1990 to peak, -6.58 at year 1996. The log mortality rate then drops to -7.15 at year 2000. Based on this historical data, the forecasted log mortality rate is continue to decline smoothly from -7.15 to -7.38 approximately at year 2022. The higher mortality rate among teenage age is due to smoking related illness result in public health problems. In all three National Health and Morbidity surveys, the prevalence of smoking among adults age 18 and above in Malaysia was 24.8% in 1996 as compared to 21.5% in 1986. It was reported that approximately 10000 deaths due to these illnesses within those few years which making it the primary cause of death in Malaysia [11]. Figure 2(c) shows the log mortality rate for age 60 and declining from -3.75 at year 1984 to -4.37 at year 2012 approximately. However, there are few fluctuations within this period time. Based on this historical data, the forecasted log mortality rate is continue to decline smoothly from -4.37 to -4.63 approximately at year 2022.
4. Conclusion
Construction of mortality rate for certain ages by cross-section of combination of the fitted and forecasted mortality rate are still not widely done in Malaysia studies. The development of interest in mortality analysis among researcher in Malaysia especially in two-dimensional analysis are still in the beginning phase. The focus of conclusion and discussion are not supposed about the concentration of the models or improvement of the models which most researcher and reviewers lost interest about. The focal point in mortality research is supposed more toward improvement of socio-economics by examining at statistical analysis. Government and semi-government will definitely reap the extra benefits of expending the research matter. Past study by Haberman & Russolillo [7] stated Lee-Carter model is one of the most significant recent developments in the field of mortality forecasting which one of the reasons using this method as the main tool in this research. Besides that, the important feature of Lee-Carter model is for a precision of value of the time index $k$ which enables in defining a complete set of death probabilities in order to eliminate the inconstancies in constructing the life table. Other important feature of the Lee-Carter model is it allows uncertainty in forecasting which is called longevity risk. In addition, the strengths of this method are its simplicity and robustness in the context of linear trends in age-specific death rates. Thus, these features eventually gives Lee-Carter model a good fit to the Malaysia mortality data and hence the keen analysis. This research definitely can be extended to variant different of Lee-Carter extensions or any stochastic mortality tool like Cairns-Blake-Dowd model and Poisson Log-Bilinear model by using Malaysia mortality experience as a centre of the main issue.

5. References
[1] Booth H and Tickle L 2008 Annals of Actuarial Science 3 issue 1-2 Mortality modelling and forecasting: A Review of Methods pp 3-43
[2] Pollard J H 1987 *Population Bulletin of the United Nations* 21-22 Projection of age-specific mortality rates pp 55-69

[3] Girosi F and King G 2008 Demographic Forecasting (United Kingdom: Princeton University Press)

[4] 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

[5] Dahl M 2005 On Mortality and Investment Risk in Life Insurance (Denmark: University of Copenhagen)

[6] *Continuous Mortality Investigation Report* Number 21 2004 (United Kingdom: Institute of Actuaries and Faculty of Actuaries)

[7] Haberman S and Russolillo M 2005 *City University – Actuarial Research Paper* No. 167 Lee-Carter Mortality Forecasting: Application to the Italian Population.

[8] Bujang M A, Hamid A M A, Zolkepali N A, Hamedon N M, Lazim S S M and Haniff J 2012 *Journal of Health Informatics in Developing Countries* 6(2) Mortality rates by specific age group and gender in Malaysia: Trend of 16 years, 1995–2010 pp 521-522

[9] UNICEF Malaysia - State of the World’s Children - Situation in Malaysia. (n.d.) 2014

[10] Efforts aim to improve quality of health service – BorneoPost Online | Borneo, Malaysia, Sarawak Daily News | Largest English Daily In Borneo. (n.d.). 2014

[11] Lim K H, Sumarni M G, Amal N M, Hanjeet K, Rozita W W and Norhamimah A 2009 *Trop Biomed*, 26(1) Tobacco use, knowledge and attitude among Malaysians age 18 and above pp 92-99

**Acknowledgment**

The authors would like to thank the Department of Statistics Malaysia for their support in supplying the death mortality data.