Forecasting the mortality rates using Lee-Carter model and Heligman-Pollard model

R I Ibrahim, N Ngataman and W N A Wan Mohd Abrisam

Faculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia
E-mail: rose.irnawaty@usim.edu.my

Abstract. Improvement in life expectancies has driven further declines in mortality. The sustained reduction in mortality rates and its systematic underestimation has been attracting the significant interest of researchers in recent years because of its potential impact on population size and structure, social security systems, and (from an actuarial perspective) the life insurance and pensions industry worldwide. Among all forecasting methods, the Lee-Carter model has been widely accepted by the actuarial community and Heligman-Pollard model has been widely used by researchers in modelling and forecasting future mortality. Therefore, this paper only focuses on Lee-Carter model and Heligman-Pollard model. The main objective of this paper is to investigate how accurately these two models will perform using Malaysian data. Since these models involve nonlinear equations that are explicitly difficult to solve, the Matrix Laboratory Version 8.0 (MATLAB 8.0) software will be used to estimate the parameters of the models. Autoregressive Integrated Moving Average (ARIMA) procedure is applied to acquire the forecasted parameters for both models as the forecasted mortality rates are obtained by using all the values of forecasted parameters. To investigate the accuracy of the estimation, the forecasted results will be compared against actual data of mortality rates. The results indicate that both models provide better results for male population. However, for the elderly female population, Heligman-Pollard model seems to underestimate to the mortality rates while Lee-Carter model seems to overestimate to the mortality rates.

1. Introduction
Human life expectancies have shown remarkable improvement and this leads to significant reductions in mortality rates across all age groups, genders and countries. In Malaysia, mortality rates have been declined significantly due to the increases of life expectancies of an individual. The increase in life expectancies will increase the demand and expenses on health care and social welfare services. Consequently, several models have been proposed to forecast the mortality rates. Instead of other models, Lee-Carter model and Heligman-Pollard model are among the well-known model that have been suggested by most researchers in forecasting future mortality rates. An adequate model should be used in order to acquire the most accurate results. The Lee-Carter model is a demographic model with statistical time series techniques while Heligman-Pollard model is a parameterisation function.
Lee and Carter [1] proposed a base model of age-specific death rates with a dominant time component and a fixed relative age component, and a time series model (ARIMA) of the time component whereby the model was firstly applied to United State death rates data for year 1933 to 1987. The results show that the model performs well in generating the output. Lee-Carter model has been widely used by researchers such as Bravo et al. [2] has applied Lee-Carter model and its extension for forecasting human mortality rates for Portugal, Ozeki [3] has compared Lee-Carter model and Heligman-Pollard model with two other methods using Japan Life Table. In addition, the study by Ngataman et al. [4] has shown that the fitted log-age specific death rates from Lee-Carter model fits Malaysian population data quite well as the estimated rates move along with the actual data for 30 years period of study, starting from year 1981 to 2010.

On the other hand, Heligman and Pollard [5] introduced Heligman-Pollard model to fit the Australian mortality rates. Van Wei [6] has concluded that Heligman-Pollard model seems to fit better compared to Lee-Carter model even though the results show that big errors had occurred at age 0 and 110+ compared with Lee-Carter model which only have errors at adolescent ages when applied to Netherlands, Denmark and Norway data. Similar with Gaille and Sherris [7], the Heligman-Pollard model shows an improvement compared with Lee-Carter model because of its smoothness in forecasting mortality rates. Furthermore, it is stated that Heligman-Pollard model capacity has adequate adaptability to take into consideration of the full age scope of mortality and having parameters that are deciphered for their impact on death rates. However, the forecasting of mortality rates using this model has its limitation due to unstable and large number of parameters.

2. Methodology
Nowadays, many methods have been proposed to forecast mortality rates. Among the many methods, Lee-Carter model and Heligman-Pollard model are the well-known model that have been suggested by most researchers in forecasting future mortality rates.

2.1. Lee-Carter Model
The Lee-Carter model is probably the best known method for mortality forecasting these days. Lee-Carter model is a time series model for the time component defined as follows:

\[
\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}
\]

(1)

where \( m_{x,t} \) represents the central mortality rate of age \( x \) at time \( t \); \( a_x \) is the average log-mortality of age \( x \); \( b_x \) measures the response of age \( x \) to change in \( k_t \); \( k_t \) represents the overall level of mortality at time \( t \) and \( \epsilon_{x,t} \) is the error term or residual.

The value of \( a_x \) is estimated by taking the average log-mortality at age \( x \) while Singular Value Decomposition (SVD) is applied to obtain the estimated values of \( b_x \) and \( k_t \) respectively by using MATLAB software. Then, the parameter of \( k_t \) is re-estimated using the following equation:

\[
\sum_x D_{x,t} = \sum_x E_{x,t} \exp\left(\hat{a}_x + \hat{b}_x \hat{k}_t\right)
\]

(2)

where \( D_{x,t} \) is total number of deaths at age \( x \) in time \( t \); \( E_{x,t} \) is the population (exposure to risk) at age \( x \) in time \( t \) and \( \hat{a}_x, \hat{b}_x, \hat{k}_t \) are the estimated values of \( a_x, b_x, k_t \). ARIMA model is applied to forecast values of \( k_t \). Then, the forecasted mortality rates are obtained by inserting the values of \( \hat{a}_x, \hat{b} \) and forecasted \( k_t \) into equation (1).
2.2. Heligman-Pollard Model

The Heligman-Pollard model is a parameterisation function that consists of eight parameters, defined as follows:

\[
\frac{q_x}{p_x} = A(x + B)^C + D \exp \left[ -E \left( \ln \left( \frac{x}{F} \right) \right)^2 \right] + GH^x
\]  

(3)

where \( q_x \) is the probability that an individual who has reached age \( x \) will die before reaching age \( x+1 \); \( p_x = 1 - q_x \) and \( A, B, C, D, E, F, G, H \) are the eight positive parameters to be estimated by using non-linear weighted least squares methods.

Generally, this model is separated into three distinct terms, which describes the entire lifetime. The first term consists of three parameters \( A, B \) and \( C \) which represents a rapidly declining exponential and decrease in mortality the early childhood years for ages less than 10. The second term is a lognormal function which consists of three parameters \( D, E \) and \( F \) that represents incidental mortality for both genders and maternal mortality for female in the middle age. The last term of parameters \( G \) and \( H \) reflects the Gompertz mortality law and exponential growth of mortality at older ages.

When applying the model to the Malaysian Life Table, all the parameters are determined by minimizing the sum of square,

\[
S^2 = \sum_s^{25} \left( \frac{q_s}{\hat{q}_s} - 1 \right)^2
\]

(4)

where \( q_s \) is the fitted mortality at age \( s \) and \( \hat{q}_s \) is the observe mortality. In this study, the Matrix Laboratory Version 8.0 (MATLAB 8.0) software will be used to estimate the parameters and the details on the calculation can be referred to Ibrahim [8]. Then, the forecasted parameters are obtained using ARIMA procedure. After forecasting each of the parameters, all forecasted parameters are plugged back into equation (3) to obtain the forecasted mortality rates.

3. Results and Discussion

The empirical data sets of Malaysian population for period of 1981 to 2015 for both genders will be considered, which the period of 1981 to 2010 is used as “training set” and the period of 2011 to 2015 as “testing set”. The results of both methods will be analysed and compared as in Figure 1 and Figure 2.

Figure 1 (a) shows the mortality rates for male in year 2011. It was found that Lee-Carter model fits very well to the actual data and Heligman-Pollard model seems to underestimate the actual data at the older ages. However, Heligman-Pollard model seems to fit very well to the actual data in year 2012 until 2015, while Lee-Carter model tends to overestimate the actual data at the older ages.

Figure 2 shows the mortality rates for female in which the trends are almost same for the five years’ period of study. The forecasted mortality rates for both models fit very well to the actual data for ages less than 65 years. However, Heligman-Pollard model seems to continue to underestimate the actual data starting from ages 65 years and above, while Lee-Carter model seems to overestimate the actual data at the older ages.
Figure 1. Forecasted Mortality Rates for Male.

Figure 2. Forecasted Mortality Rates for Female.
Comparing their overall fit, both models seem to fit the actual data very well for ages less than 65 years as the forecasted rates move along with the actual data at almost all age groups for both male and female. For the older ages, Lee-Carter model seems to overestimate the actual data for both male and female and Heligman-Pollard model seems to underestimate the actual data for female. However, it was found that the relative difference values, which is the errors between actual mortality and forecasted mortality for both models in year 2011 to 2015 is below 10 percent range. It is concluded that both models having the small error values for almost all age groups and both models are fit well to the Malaysian data.

4. Conclusion

In order to forecast mortality rates for Malaysian population, this paper have been applied two mortality models namely Lee-Carter model and Heligman-Pollard model. The Matrix Laboratory Version 8.0 (MATLAB 8.0) software is used to estimate the parameters for both models. Then, ARIMA is used to estimate all the forecasted parameters for both models.

The results showed that the forecasted mortality rates for both models move along with the actual data for ages less than 65 years for both male and female. The results also showed that the Heligman-Pollard model seemed to fit better to the male data than the the Lee-Carter model for all ages. Whereas for female, the results showed that Heligman-Pollard model seemed to underestimate the actual data and Lee-Carter model seemed to overestimate the actual data for ages greater than 65 years. However, it can be concluded that both models seemed to fit very well to the Malaysian data for ages less than 65 years and the result is fair for the older ages because the relative difference between actual and forecasted mortality is below the 10 percent range in these ages.

Acknowledgments

The authors would like to thank the Department of Statistics, Malaysia for providing the data and Universiti Sains Islam Malaysia for their financial support (PPT/GP/FST/30/14615).

References

[1] Lee R and Carter L 1992 *J. Am. Stat. Assoc.* 87 pp 659-71.
[2] Bravo B, Coelho E, and Magalhães M 2010. *Mortality projections in Portugal.* (Luxemberg: European Union) p 241.
[3] Ozeki M 2005. *Application of Mortality Models to Japan.* (Orlando, Florida: Society of Actuary) p 1.
[4] Ngataman N, Ibrahim R and Md Yusuf M 2016. *AIP Conf. Proc.* 1750 0200091–96.
[5] Heligman L and Pollard J 1980. *J. Inst. Actuaries* 107 pp 49-80.
[6] Van W 2015 *Quantitative Methods in Business and Economics* 1 pp 89-113.
[7] Gaille S and Sherris M 2010. Improving Longevity and Mortality Risk Models with Common Stochastic Long-Run Trends. *UNSW. Australian School of Business Research Paper 2010ACTL13.*
[8] Ibrahim R 2008 *Matematika* 24 pp 1-10.