Mortality Modelling Using Stochastic Mortality Models: A Study on Malaysia's Ethnic Groups

Qian Yun Ng[1] and Lay Guat Chan[1,2](✉)

[1] Department of Actuarial Science and Risk, School of Mathematical Sciences, Sunway University, 47500 Petaling Jaya, Selangor, Malaysia
layguatc@sunway.edu.my

[2] Healthy Ageing and Well-Being Research Cluster, School of Mathematical Sciences, Sunway University, 47500 Petaling Jaya, Selangor, Malaysia

Abstract. Globally, the life expectancy of the population has been continuously improving over the years due to healthcare and socioeconomic advancements. The rapid increase in life expectancy over the last few decades leads to an ageing population as the population lives longer. With the rise of elderly population in the society, insurance companies and pension funds need to deal with longevity risk, which is the risk of incurring greater pay-out ratios than projected as life expectancies exceed pricing assumptions. Hence, accurate mortality modelling and projection are of key interest to insurance companies, pension providers and government to minimize such risks. This study will focus on the modelling of mortality rates in Malaysia based on three main ethnic groups, namely Malay, Chinese and Indian using data from Abridged Life Tables for a 20-year period (2001–2020) obtained from Department of Statistics Malaysia. Mortality rates for six subpopulations (Malay male, Malay female, Chinese male, Chinese female, Indian male and Indian female) under 18 age groups will be modelled using three stochastic mortality models, i.e. Lee-Carter model, Hyndman-Ullah model and Augmented Common Factor model. We conclude that Hyndman-Ullah model has the best fit for past mortality rates with the lowest values of goodness-of-fit using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Future research can be conducted by using Hyndman-Ullah model to forecast mortality rates in Malaysia based on age, gender and ethnic groups, which can be then applied in updating pension and annuities calculations on the existing and new contracts to minimize financial losses arising from longevity risk.

Keywords: Stochastic mortality model · mortality modelling · Hyndman-Ullah model

1 Introduction

The life expectancy of the population worldwide has been improving over the years due to medical advancements and socioeconomic improvements. In Malaysia, the life expectancy at birth for females was 74.7 years in 2000 and increased to 77.4 years in 2019, whereas the life expectancy at birth for males improved from 70.0 years to

© The Author(s) 2023
N. Wahi et al. (Eds.): ICMSS 2022, ACSR 98, pp. 163–172, 2023.
https://doi.org/10.2991/978-94-6463-014-5_16

72.5 years, for 2000 and 2019 respectively [1, 2]. Malay, Chinese and Indian are the three main ethnic groups in Malaysia, where in 2019 Malays formed the majority of the population (69.3%), then Chinese (22.8%), followed by Indian (6.9%) [3]. Some studies [4, 5] conducted in Malaysia have concluded that different ethnic groups have different mortality rates, specifically, Chinese has lower mortality rates compared to Malay and Indian for all ages. Also, female mortality rates are lower than male mortality rates in general.

The increase of life expectancy in Malaysia is due to the decrease in the mortality rates, indicating that the population tends to live longer and will slowly lead to an ageing population. Insurance companies and pension funds need to deal with longevity risk when actual mortality rates exceed expectations or mortality assumptions during pricing [6]. Hence, understanding the trends and patterns of a population's mortality rates are important as it can be applied in forecasting future mortality rates, developing appropriate pension schemes, social security systems and healthcare services planning. As each gender and ethnic group in Malaysia have different mortality rates, accurate modelling and projection of mortality trends will also provide insights on how life expectancies at birth are expected to change in the future for each gender and ethnic group.

There has been an increasing number of studies on modelling and projecting mortality rates in recent years. Stochastic mortality models were developed to model and project mortality rates by considering time effect and producing probabilistic confidence intervals, allowing researchers to assess the uncertainty level involved and to study the possibility of mortality rates might vary from now to a future time point [7]. Lee-Carter model [8] is one of the significant stochastic mortality models and many extensions have been made to improve its model fitting, such as Hyndman-Ullah model [9]. The Hyndman-Ullah model produces forecasted mortality pattern that are more accurate and realistic as compared to the Lee-Carter model by capturing additional variation of mortality rates through utilizing higher order principal components and non-parametric smoothing. A few studies have compared the performance of Lee-Carter model and Hyndman-Ullah model in modelling and forecasting mortality rates in Malaysia. Hyndman-Ullah model was found to be more performing than Lee-Carter model for Malaysian mortality rates [10, 11]. However, another study found that Lee-Carter model fitted and forecasted Malaysia mortality rates better than Hyndman-Ullah model [12]. Further study can be conducted to compare the performance of Lee-Carter model and Hyndman-Ullah model due to the contradictory results of the above studies.

Coherent mortality models, sometimes also referred to multi-population models, are part of the stochastic mortality models, were developed to project the mortality rates of at least two subpopulations simultaneously and to produce non-divergent forecasts of subpopulations in the long run. For instance, one of the coherent mortality models, i.e. Augmented Common Factor model [13] extends the Lee-Carter model by incorporating a mortality reference in the base model to maintain historic relationships between groups and to restrict the time component of subpopulations, ensuring non-divergent forecasts indefinitely over time. Limited studies in Malaysia compared coherent mortality models with Lee-Carter model. For instance, one of the studies concluded that Augmented

Common Factor model performed better in fitting and forecasting mortality rates for both genders in Malaysia as compared to Lee-Carter model [14].

Based on the literatures, none of the research compared the accuracy in fitting between Lee-Carter, Hyndman-Ullah and Augmented Common Factor models in Malaysian context based on ages, genders and ethnic groups. Hence, the main objective of this study is to fit mortality rates of years 2001 to 2020 based on ages, genders and ethnic groups in Malaysia using three stochastic mortality models, i.e. Lee-Carter model, Hyndman-Ullah model and Augmented Common Factor model. The performance of each model in fitting mortality rates will be evaluated using two goodness-of-fit techniques, i.e., Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

2 Methodology

Mortality data in Malaysia from years 2001 to 2020 are obtained from Abridged Life Tables by age groups, gender and ethnic groups, provided by Department of Statistics Malaysia [15]. There are 18 age groups (<1, 1–4, 5–9, 10–14 until 80+), 2 genders (male and female) and 3 ethnic groups (Malay, Chinese and Indian) under the Abridged Life Table. The central mortality rate of each gender and ethnic group at age x and time t, $m_{x,t}$, will be used in estimating the parameters of three stochastic models, i.e. Lee-Carter model, Hyndman-Ullah model and Augmented Common Factor model. The equations and parameters of each stochastic mortality model will be discussed next. Data from 2001 to 2020 will be used to fit six subpopulations data – Malay males, Malay females, Chinese males, Chinese females, Indian males and Indian females, under 18 age groups using three stochastic models. Lastly, goodness-of-fit of the three stochastic models will be performed to evaluate the models' performance in fitting mortality rates.

2.1 Lee-Carter Model

The Lee-Carter model expresses the log central mortality rate at age x in year t as:

$$ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t} \tag{1}$$

α_x refers to the mortality level at age x, β_x estimates the response at age x to change in the overall mortality level across the years, κ_t is the mortality index in year t, and $\varepsilon_{x,t}$ is the normally distributed error term. The original Lee-Carter method adjusts the mortality index κ_t to fit observed total number of deaths [8]. However, some researchers skip this procedure and forecast the original κ_t directly [13]. Hence, κ_t will not be adjusted in this study. Under the Lee-Carter model, constraints $\sum_{x=1}^{X} \beta_x = 1$ and $\sum_{t=1}^{T} \kappa_t = 0$ are introduced to ensure uniqueness of solution. α_x is estimated by taking the mean of log central mortality rates across years. Singular Value Decomposition (SVD) method is used to find β_x and κ_t when applied to the matrix $\mathbf{M} = \ln(m_{x,t}) - \alpha_x$. The estimates of α_x, β_x and κ_t are fitted into Eq. (1), in order to obtain the fitted values of $\ln(m_{x,t})$.

2.2 Hyndman-Ullah Model

The Hyndman-Ullah model is a generalization of the Lee-Carter model, where it combines functional data method, nonparametric smoothing and robust statistics [9]. Let $y_t(x_i)$ be the log observed mortality rates for

$$y_t(x_i) = ln\, m_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \tag{2}$$

where $f_t(x_i)$ is the smooth mortality function of age x, $\sigma_t(x_i)$ is the smooth volatility function that varies with age x in year t and $\varepsilon_{t,i}$ is the normally distributed error term for every $i = 1, \ldots, p$ and $t = 1, \ldots, n$. For $x \in [x_1, x_p]$,

$$f_t(x) = \alpha(x) + \sum\nolimits_{j=1}^{J} \beta_j(x) k_{t,j} + e_t(x) \tag{3}$$

$\alpha(x)$ refers to the average mortality pattern by age over time. $\beta_j(x)$ is a set of orthonormal basis functions that is a set of first J principal components reflecting the sensitivity to time-varying index over age, which is similar to β_x in Lee-Carter model. Besides, $k_{t,j}$ is a set of uncorrelated principal component scores which are also explained as time series coefficients. It is similar to κ_t in Lee-Carter model that represents the time-varying index. Next, $e_t(x)$ is the error of the model, whereas $J < n$ is the number of principal components extracted. The number of principal components (pair of basis functions and coefficients) used is set to $J = 6$, as presented in the original paper of [9]. To fit the model, the data for each t are smoothed using a non-parametric smoothing method to estimate $f_t(x)$ for $x \in [x_1, x_p]$. The smoothed data is estimated using a one dimensional (function of age only) non-parametric approach, based on weighted penalized regression splines with a monotonicity increasing for $x \geq c$ to reduce noise in the estimated curves at older ages. In this study, c is set to 40, as done in [16]. For mortality rates, observations in year t are given by $y = m_{x,t}$, the weights ω are taken as the inverse of the estimated variances of y in which y is assumed to have a binomial distribution. The details of this smoothing technique may refer to [9]. Using functional principal component analysis (FPCA), a set of curves in Eq. (3) is decomposed into orthogonal functional principal component $\{\beta_j(x)\}$ and their uncorrelated principal component scores $\{k_{t,j}\}$. $\alpha(x)$ is the mean function, estimated by $\alpha(x) = \frac{1}{n}\sum_{t=1}^{n} f_t(x)$. The coefficients obtained in the previous step are implemented to get the $f_t(x)$ as in Eq. (2). From Eq. (2), we can see that $y_t(x)$ is fitted.

2.3 Augmented Common Factor Model

An additional common factor has been introduced in the Augmented Common Factor model to extend Lee-Carter model [13]. The log central mortality rate at age x in year t for gender i is modelled as:

$$ln(m_{x,t,i}) = \alpha_{x,i} + B_x K_t + \beta_{x,i}\kappa_{t,i} + \varepsilon_{x,t,i} \tag{4}$$

$\alpha_{x,i}$ refers to the mortality level at age x for gender i, $B_x K_t$ represents the aggregated subpopulation's common factor for both genders, $\beta_{x,i}\kappa_{t,i}$ is the gender-specific factor for

gender i, and $\varepsilon_{x,t,i}$ is the error term normally distributed. K_t represents the aggregated subpopulation's overall time trend, whereas B_x estimates the sensitivity to decrease in mortality at age x. The shared component by subpopulations, $B_x K_t$, is a required and adequate condition for avoiding divergence in central forecast of subpopulations. Equivalently, $\kappa_{t,i}$ represents the gender-specific mortality time index, and $\beta_{x,i}$ is the measurement of age sensitivity. Hence, the mortality trend of each specific gender i on top of the aggregated subpopulation's overall trend is captured by the component $\beta_{x,i}\kappa_{t,i}$. As the mortality index in Lee-Carter model is not adjusted, the common trend K_t in Augmented Common Factor model will not be adjusted to ensure comparability of the results. Similar to Lee-Carter model, constraints $\sum_{x=1}^{X} B_x = 1$, $\sum_{t=1}^{T} K_t = 0$, $\sum_{x=1}^{X} \beta_{x,i} = 1$ and $\sum_{t=1}^{T} \kappa_{t,i} = 0$ are introduced to ensure uniqueness of solution. $\alpha_{x,i}$ is estimated by taking the mean of log central mortality rates across years for each gender. SVD method is applied to $\mathbf{M} = \sum_{i=1}^{r} w_i[\ln(m_{x,t,i}) - \alpha_{x,i}]$ to find B_x and K_t where r is the number of groups, w_i is the weight of the group i, that is, $\sum_{i=1}^{r} w_i = 1$. To obtain the fitted values of $\ln(m_{x,t,i})$, the estimates of $\alpha_{x,i}$, B_x, K_t, $\beta_{x,i}$ and $\kappa_{t,i}$ are fitted into Eq. (4).

2.4 Goodness-of-Fit Techniques

Goodness-of-fit for a model is the measure of deviation between fitted and observed data. Two techniques will be used in our study, which are Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The smaller the RMSE and MAPE, the better performance of a model in fitting the data.

Root Mean Square Error (RMSE)
RMSE measures the standard deviation between observed and predicted values [17].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f_i - o_i)^2}{n}} \tag{5}$$

where f_i is the forecasted value, whereas o_i is the observed value and n is the number of observations.

Mean Absolute Percentage Error (MAPE)
MAPE computes the percentage difference between actual values and forecasted values [18].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{o_i - f_i}{o_i}\right| \times 100 \tag{6}$$

Similar to Eq. (5), f_i is the forecasted value, o_i is the observed value and n is the number of observations.

3 Results and Discussion

3.1 Goodness-of-Fit Results

The parameters of Lee-Carter model, Hyndman-Ullah model and Augmented Common Factor model are estimated using the steps given in methodology section. The estimation

Table 1. Measures of Fits of Models

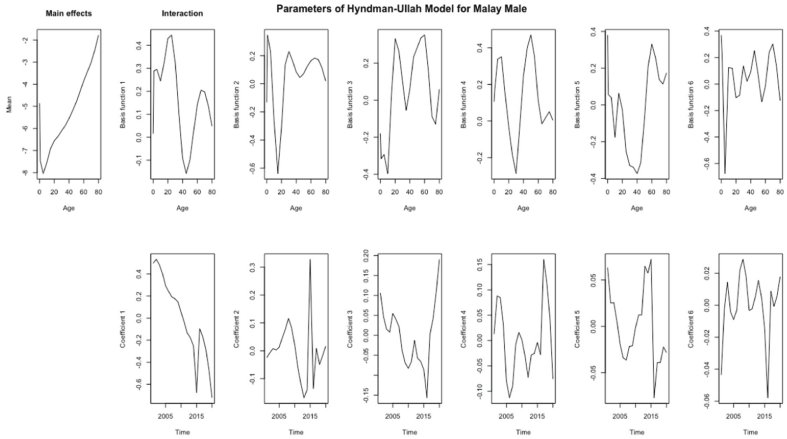
Model	Lee-Carter Model		Hyndman-Ullah Model		Augmented Common Factor Model	
Subpopulations	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Malay Male	0.05187	0.740	0.00894	0.148	0.04728	0.721
Malay Female	0.03592	0.550	0.00837	0.120	0.03616	0.561
Chinese Male	0.04550	0.574	0.01581	0.202	0.04239	0.589
Chinese Female	0.05523	0.674	0.01643	0.189	0.04830	0.622
Indian Male	0.06863	1.079	0.01581	0.252	0.05677	0.886
Indian Female	0.07382	1.089	0.02757	0.362	0.06191	0.804

of parameters for each model are performed using RStudio. The performance of the models will be assessed using RMSE and MAPE, as shown in Eqs. (5) and (6). Lower RMSE and MAPE values indicate a better fitting model.

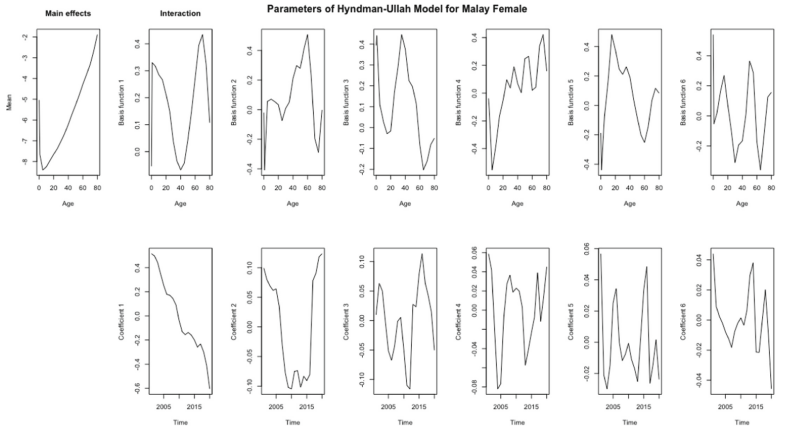
Table 1 summarizes the three stochastic models' RMSE and MAPE of log mortality rates of six subpopulations. Lee-Carter model and Augmented Common Factor model have higher RMSE and MAPE compared to Hyndman-Ullah model, indicating that these two models have a lower performance of fitting mortality rates for each gender and ethnic group in Malaysia. The Hyndman-Ullah model displays the best historical fit for all subpopulations as the RMSE and MAPE values are the lowest among all three models.

3.2 Models Fitting

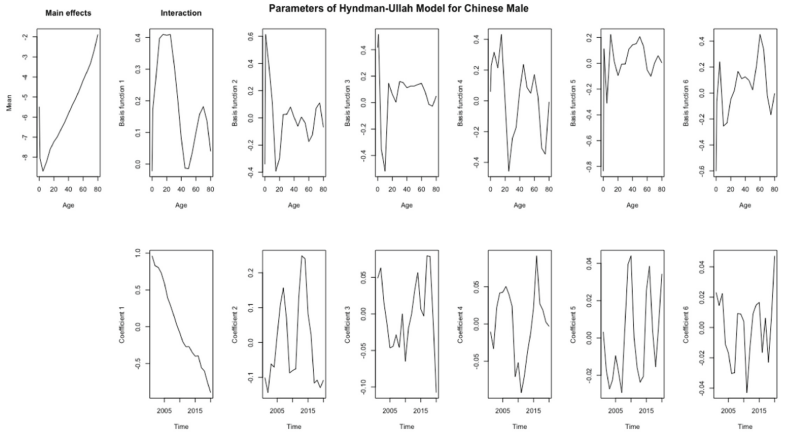
The graphs of the parameters Hyndman-Ullah model are presented and will be discussed in this section since Hyndman-Ullah model is considered as the best model based on the goodness-of-fit results in Table 1. Refer to Fig. 1, it can be observed that the values of mortality index $k_{t,1}$ in Hyndman-Ullah model have declined throughout the years, which indicates that the mortality rates have declined over the years. There is a spike in years 2017 and 2018 for the mortality indices, which could be caused by a surge in deaths due to ischemic heart diseases [19]. The main effect $\alpha(x)$ in Hyndman-Ullah model displays the average pattern of mortality across the ages. The values of main effect for all genders and ethnic groups are high at age <1, reach a minimum value at age 5, and increase until the end. Lastly, $\beta_1(x)$ in Hyndman-Ullah model reflects the sensitivity to mortality rate changes over ages. Individuals at younger and older ages are affected by the changes of mortality rate the most as compared to other age groups, displaying large values of $\beta_1(x)$.



(a) Malay Male

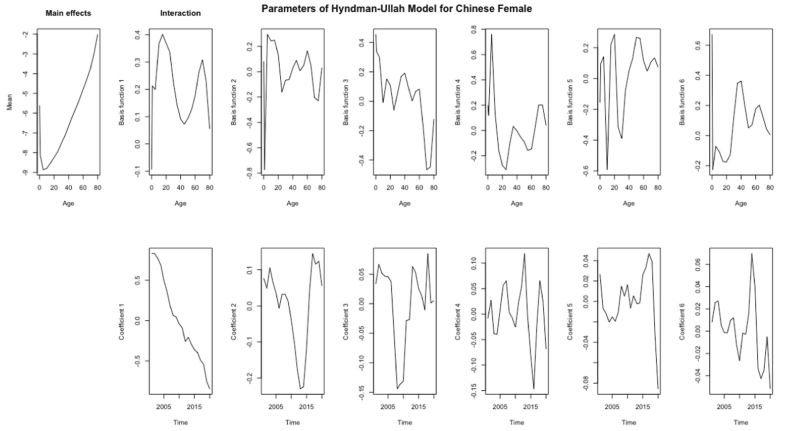


(b) Malay Female

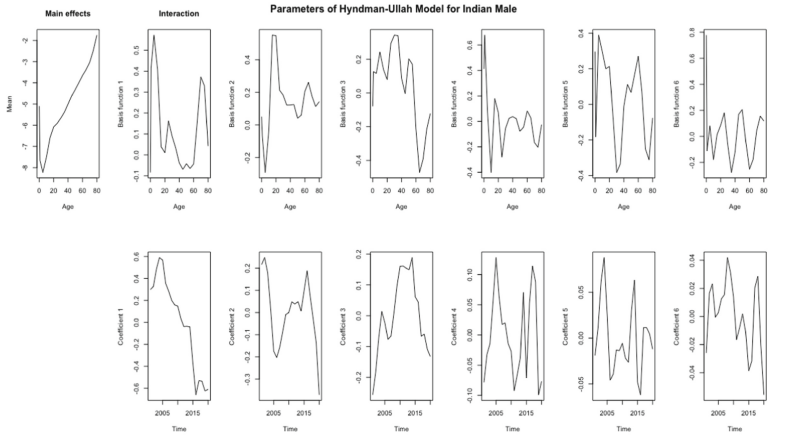


(c) Chinese Male

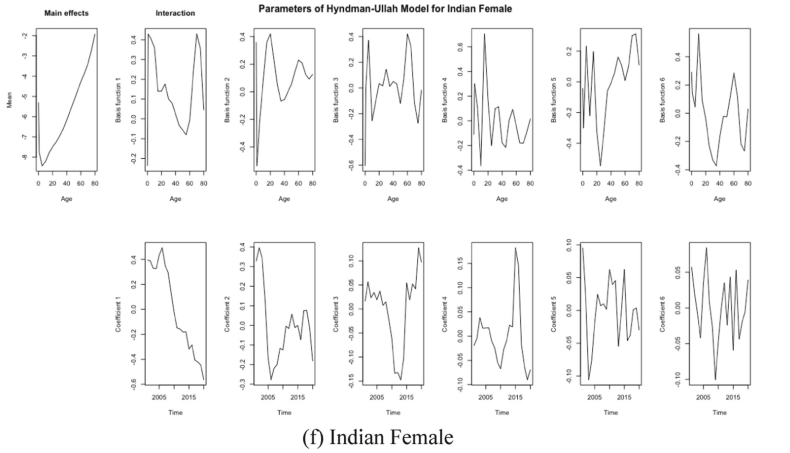
Fig. 1. Parameters of Hyndman-Ullah Model for Six Subpopulations.



(d) Chinese Female



(e) Indian Male



(f) Indian Female

Fig. 1. (*continued*)

4 Conclusion

There is an urgent need to understand the mortality rates as Malaysia is projected to be an ageing population in less than a decade from now. The implications of this mortality trend are worrying as it may negatively impact Malaysia's economic growth and productivity as well as put a strain on the healthcare system, public and private funds for pensions. In our study, we have modelled mortality rates of years 2001 to 2020 using three stochastic mortality models, i.e., Lee-Carter model, Hyndman-Ullah model and Augmented Common Factor model. Among the three stochastic mortality models, Hyndman-Ullah model has the lowest values of RMSE and MAPE, concluding that it is the best fitted model for Malaysia's ethnic groups mortality rates for years 2001 to 2020.

Future studies can be conducted to apply Hyndman-Ullah model to forecast mortality rates based on age, gender, and ethnic groups in Malaysia. By having the forecasted mortality rates using Hyndman-Ullah model, it will allow the government and insurance companies to update pension calculations on the existing and new contracts to reduce financial losses due to longevity risk. A limitation of this research is only three stochastic mortality models are used to model mortality rates of Malaysian ethnic groups. Future research can include other stochastic mortality models such as Cairns-Blake-Dowd model, Renshaw-Haberman model and Lee-Miller model to fit mortality rates based on age, gender and three main ethnic groups in Malaysia.

Authors' Contributions. Qian Yun Ng and Lay Guat Chan both cooperatively designed the research. Qian Yun Ng implemented the research and performed the analysis of results. Both Qian Yun Ng and Lay Guat Chan contributed to the final version of the manuscript. Lay Guat Chan supervised the research.

References

1. Department of Statistics Malaysia, Population & Demography, Abridged Life Tables, Malaysia, 2008–2010. Department of Statistics Malaysia, Putrajaya (2011)
2. Department of Statistics Malaysia, Population & Demography, Abridged Life Tables, Malaysia, 2019–2021. Department of Statistics Malaysia, Putrajaya (2021)
3. Department of Statistics Malaysia, Current Population Estimates, Malaysia, 2020. Department of Statistics Malaysia, Putrajaya (2020)
4. Ibrahim, R.I., Siri, Z.: Analysis of mortality trends by specific ethnic groups and age groups in Malaysia. In: AIP Conference Proceedings, vol. 1605, pp. 1002–1006 (2014). https://doi.org/10.1063/1.4887727
5. Ishak, S.A., Shair, S.N., Shukiman, W.N.A.W.A., Radzi, N.M., Rahman, N.S.A.: The trends of age and gender specific mortality rates by ethnic groups. In: Kor, L.-K., Ahmad, A.-R., Idrus, Z., Mansor, K.A. (eds.) Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017), pp. 475–480. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-7279-7_59
6. Richards, S.J., Jones, G.L.: Financial aspects of longevity risk. In: Prudential Assurance, London, pp. 5–58 (2004)
7. Booth, H., Tickle, L.: Mortality modelling and forecasting: a review of methods. Ann. Actuarial Sci. **1**, 3–43 (2008). https://doi.org/10.1017/S1748499500000440

8. Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992). https://doi.org/10.2307/2290201
9. Hyndman, R.J., Shahid Ullah, M.: Robust forecasting of mortality and fertility rates: a functional data approach. Comput. Stat. Data Anal. **51**(10), 4942–4956 (2007). https://doi.org/10.1016/j.csda.2006.07.028
10. Husin, W.Z.W., Zainol, M.S., Ramli, N.M.: Stochastic models in forecasting Malaysian mortality: the Lee-Carter model and its extension. Adv. Sci. Lett. **21**(6), 1850–1853 (2015). https://doi.org/10.1166/asl.2015.6135
11. Sapri, N.N.F.F., Ramli, N.M., Ghani, N.A.M., Husin, W.Z.W.: A comparison of mortality models based on life expectancy and log death rates for Malaysia age-specific death rates. Adv. Sci. Lett. **22**(12), 4018–4022 (2016). https://doi.org/10.1166/asl.2016.8168
12. Kamaruddin, H.S., Ismail, N.: Statistical comparison of projection Malaysia mortality rate by using Lee-Carter model and Lee-Carter extension of Hyndman-Ullah. In: AIP Conference Proceedings, vol. 2111 (2019). https://doi.org/10.1063/1.5111214
13. Li, N., Lee, R.: Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. Demography **42**(3), 575–594 (2005). https://doi.org/10.1353/dem.2005.0021
14. Nor, S.R.M., Yusof, F., Bahar, A.: Multi-population mortality model: a practical approach. Sains Malaysiana **47**(6), 1337–1347 (2018). https://doi.org/10.17576/jsm-2018-4706-31
15. Department of Statistics Malaysia, Abridged Life Tables. Department of Statistics Malaysia, Putrajaya (2021)
16. Shair, S.N., Zolkifli, N.A., Zulkefli, N.F., Murad, A.: 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 (2019)
17. Salkind, N.J.: Root mean square error. In: Encyclopaedia of Research Design, vols. 1–0 (2010)
18. de Myttenaere, A., Golden, B., Le Grand, B., Rossi, F.: Mean absolute percentage error for regression models. Neurocomputing **192**, 38–48 (2016). https://doi.org/10.1016/j.neucom.2015.12.114
19. Loh, F.F.: Heart attack leading cause of death. In: The Star, Kuala Lumpur (2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.