A new approach to forecast Malaysian mortality rates

Sajjad Majeed Bin Iqbal Hamid [1, *], Eric Dei Ofosu-Hene [2], Raja Rajeswari Ponnusamy [1]

[1]School of Mathematics, Actuarial and Quantitative Studies, Asia Pacific University of Technology and Innovation (APU), Kuala Lumpur, Malaysia
[2]Department of Accounting and Finance, Faculty of Business and Law, De Montfort University, Leicester, UK

A R T I C L E   I N F O

A B S T R A C T

In this paper a new approach to project the time varying index k_t for the Lee-Carter (LC) model by using a machine learning technique known as Neural Network (NN) is proposed for forecasting Malaysian male and female mortality rates. To evaluate the forecasting performance of the proposed model, the conventional LC model which uses ARIMA to forecast k_t is used as a benchmark. However unlike previous studies were done in Malaysian, we employed 9 different ARIMA models and evaluated the AIC and BIC to obtain the best fit model to forecast k_t. The forecasting performance of the two methodologies were then compared using 3 performance indicators Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). In this study, findings showed that the proposed NN model outperformed the conventional ARIMA model for forecasting both Malaysian male and female mortality rates.

1. Introduction

Over the last few decades as a result of an unprecedented increase in mortality rates around the world and the potential risk this increase pose towards population size and structure, social security systems, life insurance, and pensions industry worldwide. Modelling and forecasting mortality rates accurately has become an essential research area for practitioners of actuarial science and demography in many countries.

Several mortality forecasting models have been proposed in the past. However, the most successful model was introduced by Lee and Carter (1992) which incorporated stochastic forecasting.

The Lee-Carter (LC) model has been said to be the leading statistical model of mortality forecasting (Deaton and Paxson, 2004). It has become one of the most favorable models to be used for modelling and forecasting mortality in many counties, such as US (Lee and Carter, 1992), Chile (Lee and Rofman, 1994), China (Jiang, 1995), Japan (Wilmoth, 1996), the seven most economically advanced nations (G7) (Tuljapurkar et al., 2000), the Nordic countries (Koissi et al., 2006), India (Yadav et al., 2012), Sri

Lanka (Aberathna et al., 2014), Thai (Yasungnoen and Sattayatham, 2016) and Malaysia (Ngataman et al., 2016). Furthermore, it has also become a model that is utilized by policy analysts around the world for forecasting all-cause and cause-specific mortality (Girosi and King, 2007).

Although the LC model has its advantages such as, simplicity of parameter estimation, ease of parameter interpretation, ability to acquire all values of a countries life table and also having the capability to produce probabilistic intervals. The LC model also has some disadvantages which has caused it to become a basis model for which many modifications and extension have been proposed. Wilmoth (1993) introduced two methods to improve the fitting of the LC model by employing weighted Singular Value Decomposition (SVD) and by using Maximum Likelihood Estimation (MLE). Lee and Miller (2001) proposed a variant of the LC model which restricted the fitting period, adjusted the time varying index of the LC model by matching life expectancy and eliminated the jump off errors by forecasting from the observed rates. Booth et al. (2002) introduced a method to ascertain the optimum fitting period based on the assumption of recent trends are bound to be linear, and also fitted the time varying index to the age distribution of death after finding systematic departures from linearity in the Australian mortality decline. Brouhns et al. (2002) proposed a variant where the LC model is embedded in a Poisson regression model, suited for fitting and forecasting age sex specific mortality rates. Li and Lee (2005)

proposed an extension for modelling group population. While an extension for modelling log death rates to forecast age specific mortality and fertility rates was proposed by Hyndman and Ullah (2007), using the functional data paradigm framework of Ramsay and Silverman (2005). De Jong and Tickle (2006) introduced a more flexible model using the state space framework by Harvey (1990) for modelling the log death rates. Koissi and Shapiro (2006) implemented a fuzzy formulation of Lee-Carters model to overcome the violations of the Lee-Carter assumption of constant error variance across age. Yue et al. (2008) applied a similar approach by Bell (1997) and Hyndman and Ullah (2007), proposing an age-shift model using principal component analysis, which included two second-order interaction terms to overcome an unsupported assumption of LC in various countries (i.e., parameters are dimed to be constant). Recently another extension was introduced by Neves et al. (2017), which adopted the general framework of the LC model. But, considered several competing conditional distributions for different outcome variables and a new class of time series models, known as, Generalized Autoregressive Score model (GAS) by Creal et al. (2008; 2013), for estimating, forecasting and simulating mortality rates. Furthermore, recent advances in fast computation and numerical methods have enabled a more widespread use of the Bayesian approach in many fields of application, including population forecasting (Wiśniowski et al., 2015). Wiśniowski et al. (2015) presented a fully integrated and dynamic Bayesian approach which is embedded to LC type models to forecast populations by age and sex, it also could be used to handle different data types and sources of information. For more on Bayesian LC extensions, see Wong et al. (2018), Antonio et al. (2015) and Raftery et al. (2013).

Many of the models developed by extending or modifying the LC model had one thing in common, which is the need to forecast the time varying index. Typically, this is done by using the Box Jenkins strategy for autoregressive integrated moving average (ARIMA) models (Box et al., 2015). However due to the limitation of this approach, such as, not being able to forecast over long term interval (Girosi and King, 2007). Carter (1996) proposed a state space approach to model and forecast the time varying index, however it was pointed out by Lee and Miller (2001) that the proposed model was barely distinguishable from the original model.

This study aims to propose the use of an alternative method to model and forecast time varying index for LC type models by using a machine learning model known as Neural Network (NN). Neural network models have been prevailing as a forecasting method in both theoretical and empirical works (Kourentzes and Crone, 2010a). Over the years, NN models have attracted vast amount of attention in the Time Series and Forecasting (TSF) community due to its superior performance in classification and regression problems in machine learning (Yan, 2012). In contrast to statistics-based forecasting techniques the neural network approaches have several unique characteristics such as: 1) being non-parametric, hence, not having the need to rely on any underlying model, 2) being both nonlinear and data driven and 3) being more applicable to complicated models due to its flexibility (Zhang et al., 1998). As a result, NN models have been acknowledged by many experts to be a rising technology in TSF (Yan, 2012).

Prior to this study, as per our knowledge only one paper has been published concerning the forecasting of the time varying index using NN, which was by Safitri et al. (2018), and it was found in their study that the NN model performed well in long term forecasting with low errors.

The rest of this paper will be structured as followed. Section 2 will provide information on the dataset used. Section 3 will give a brief overview on their LC model. Section 4 will discuss the fitting and forecasting methods for the time varying index. Section 5 will present information on the performance measures. In Section 6 numerical results will be presented and in section 7 the conclusion.

2. Data

For both male and female five-year age group life-tables are acquired from Department of Statistics Malaysia (DOSM) separated into 18 age groups: (<1), (1-4), (5-14), (15–24), so on till (75–84) and then (>85) for each year from 1991 till 2016. The collected lifetables are broken down into single age groups (e.g., 0, 1, 2,...>89) by employing the methodology provided by Wilmoth et al. (2007). This is done by utilizing R program package MortHump by Remund et al. (2018).

The acquired single age dataset will be divided into two, where one portion will be the "training set" spanning from 1991 to 2012 and the other portion will be the "testing set", spanning from 2013 to 2016.

3. Lee-Carter model

The basic assertion of the LC model is that, there is linear relationship between the explanatory variables age x at time t and the logarithm of the age-specific death rate for age x at time t, $\ln(m_{x,t})$. Lee and Carter (1992) proposed model is as follows:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \tag{1}$$

where, $m_{x,t}$ is the central death rates relative to age x at time t, a_x represents the general shape of the of the mortality curve, k_t is the time varying index at time t, b_x is the deviation from the age profile as k_t varies for each age x and $\varepsilon_{x,t}$ is the error term which is assumed to be homoscedastic. Due to the LC model being over parameterized, as a result of being invariant to the following transformations:

$$\{ a_x, b_x, k_t \} \rightarrow \{ a_x, b_x/c, ck_t \} \tag{2}$$

$$\{ a_x, b_x, k_t \} \rightarrow \{ a_x - cb_x, b_x, \ k_t + c \} \qquad (3)$$

The model is not able to be fitted using ordinary regression methods. However, it is able to be identified using a two-step procedure proposed by Lee and Carter (1992). In the first step the parameters b_x and k_t are evaluated by imposing the following constraint, which is normalizing b_x to sum to unity, $\sum b_x = 1$ and k_t to sum to nought, $\sum k_t = 0$, then a_x can be acquired by averaging $ln(m_{x,t})$ over time, $a_x = \frac{1}{n} \sum_{t=1}^{n} ln(m_{x,t})$. Following with the use of SVD to estimate b_x and k_t. To do so, SVD is applied to the following matrix Z:

$$Z = \sum_{t=1}^{n}(\ln(m_{x,t}) - \hat{a}_x) \qquad (4)$$

$$Z$$

$$\begin{bmatrix} \sum_{t=1}^{n}(ln(m_{0,1991}) - \hat{a}_0) & \cdots & \sum_{t=1}^{n}(ln(m_{0,2016}) - \hat{a}_0) \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^{n}(ln(m_{85+,1991}) - \hat{a}_{85+}) & \cdots & \sum_{t=1}^{n}(ln(m_{85+,2016}) - \hat{a}_{85+}) \end{bmatrix}$$

In the second step, k_t is refitted to the total observed death rates to give more weight to ages which have higher deaths. As a result, it compensates the effects of using the logarithm of rates in the LC model. Once k_t is refitted we would have acquired the appropriate k_t values to be forecasted.

4. Forecasting

Once the model has been fitted and estimates of the parameters are acquired. To forecast using the LC model, all that has to be done is to forecast k_t. In this section a brief overview of the methods used is provided.

4.1. ARIMA

The time varying index is generally forecasted by employing a random walk with drift ARIMA model, which was originally used by Lee and Carter (1992). However, it is explicitly mentioned in Lee and Carter (1992) that different population may have to use different ARIMA models to fit the mortality index. On the contrary, studies done in Malaysia did not consider other ARIMA models. For this reason, in this study we adopt the methodology suggested by Chavhan and Shinde, (2016), where 9 different possible choices of ARIMA models are fitted to the mortality index and compared to acquire the best fitted model using two most commonly used model selection criterions, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Burnham and Anderson, 2004).

4.2. Neural network

To model and forecast the time varying index we will be using a feed forward multilayer perceptron (MLP) neural network. The MLP network is made up of three portions interconnected by connection weights as shown in Fig. 1. The first portion consist of a set of input nodes, which represents the input layer, this is where the inputs are supplied to the network, in general for a time series forecasting neural network the inputs used are either lagged observations or other predictor variable. The second portion is one or more layers of computational nodes known as hidden layers, this layers help the network learn complex task by extracting progressively more meaningful features from the inputs as they are passed on through each layer and the last portion consist of the output layer.
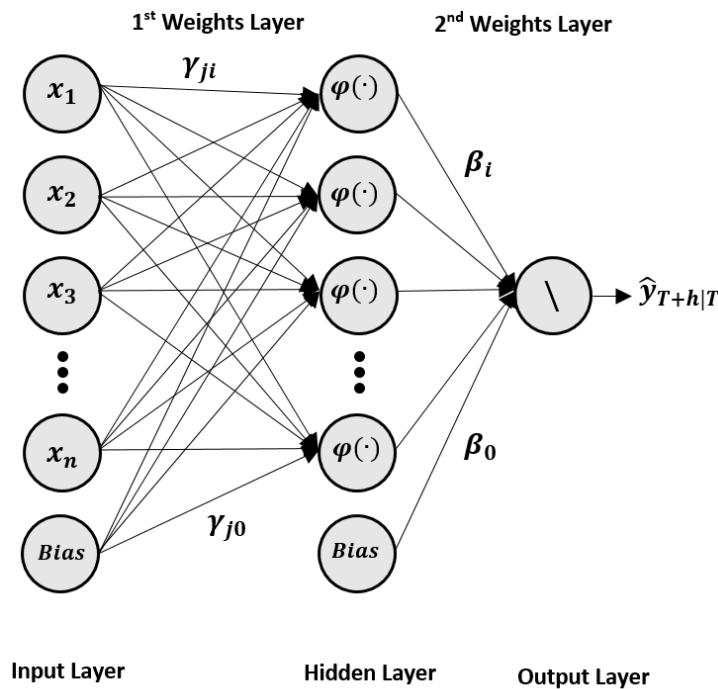

Fig. 1: General MLP neural network

$$\hat{y}_{T+h|T} = \beta_0 + \sum_{j=1}^{J} \beta_j \, \varphi\left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{ji} p_i\right) \qquad (5)$$

To compute a one step ahead forecast $\hat{y}_{T+1|T}$, equation 5 is used. The biases of each neuron which act as an intercept in a regression are denoted as β_0 and γ_0. While the network weights are denoted as $w = (\beta, \gamma)$, where β and γ are the output of the hidden layers. I denotes the number of input p_i of the network and the activation function which can be in the form of a sigmoid, bipolar sigmoid or a hyperbolic tangent, is denoted by $\varphi(.)$. J represents the number of hidden nodes in the network (Kourentzes et al., 2014). To model and forecast the time varying index, we will be using an automated fitting and forecasting neural network algorithm by Kourentzes and Crone (2010b). This will be done by employing R package "nnfor" by Kourentzes (2017). For detailed methodology see Kourentzes and Crone (2010a; 2010b) and Kourentzes et al. (2014).

5. Performance measures

After acquiring the best fitted forecasting models for k_t, of both ARIMA and Neural Network. The time varying index can be now forecasted on to the testing set.

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T} \qquad (6)$$

To evaluate the forecast performance for both models, the "errors" will be acquired using equation 6. Which is the difference between the forecasted values of the training set $\hat{y}_{T+h|T}$ and the testing set y_{T+h}. Once the errors are obtained, several different performance measures can be used to assess the forecast. However only three measures will be used in this study, which are the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) as they provide sufficient information to evaluate the forecast accuracy.

5.1. Root mean square error (RMSE)

Root Mean Square Error measures how spread forecasted data is from the actual values.

$$\sqrt{\text{mean}(e_{T+h})^2} \qquad (7)$$

RMSE close to 0 shows that the forecasted data is closer and concentrated near the actual line indicating high performance forecast. Therefore, low RMSE values are more desired.

5.2. Mean absolute error (MAE)

MAE measures the average magnitude of the errors in a set of predictions, without considering their direction.

$$\text{mean}(\,|e_{T+h}|\,) \qquad (8)$$

It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. As a result, a smaller MAE is preferred.

5.3. Mean absolute percentage error (MAPE)

Mean absolute percentage error is unit free measure of forecast error. Which measures the relative overall fit.

$$\text{mean}\left(\left|\frac{e_{T+h}}{y_t}\right|\right) \qquad (9)$$

It measures the accuracy of forecast error in percentage. Hence, making it easier to interpret. A more desired MAPE when comparing forecast is the one closer to 0%.

6. Numerical results and discussion

6.1. Components of Lee-Carter model

After unabridging, 26 male and female abridged lifetable spanning from 1991 to 2016. LC model was fitted to both male and female data. Once the model was fitted, the SVD analysis showed that 86.7% and 92.6% variation was explained by fitted LC models for Malaysian male and female mortality data respectively.

SVD analysis showed that 86.7% and 92.6% variation was explained by fitted LC models for Malaysian male and female mortality data respectively.

Fig. 2a shows the plot for the average values of the central death rate over time a_x. It can be observed that for both male and female a_x has the similar general pattern, mortality tends to decrease at lower ages, following with an exponential increase as the age increases.

Fig. 2b shows the plot for age-specific constant b_x describing the relative speed of mortality changes across ages. It can be seen that male mortality is much more sensitive compared to female mortality, especially for ages less than 16, ages between 20–34, and higher ages more than 77. Furthermore it can be observed, b_x for male are almost zero for age 36 to 45, this shows that during this ages mortality rate for male is not responsive to change in time.

Fig. 2c shows the plot for time varying index k_t, describing the general mortality for different times and captures the most important trends in death rates across all ages. In general, k_t is supposed to be decreasing as mortality is a decreasing factor and this is seen for both male and female.

6.2. Evaluating goodness of fit

To evaluate how well the LC model fitted to both male and female mortality data. The actual values and the fitted values were plotted for the following age groups 0, 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, and 81-84+.

Analyzing Fig. 3 and Fig. 4, both male and female log mortality rate fitted values follow closely to the actual values across all age groups and years. There are slight over estimation for age groups 1-10 and 11-20. However, it can be inferred that the LC model fits well to both Malaysian male and female mortality data.
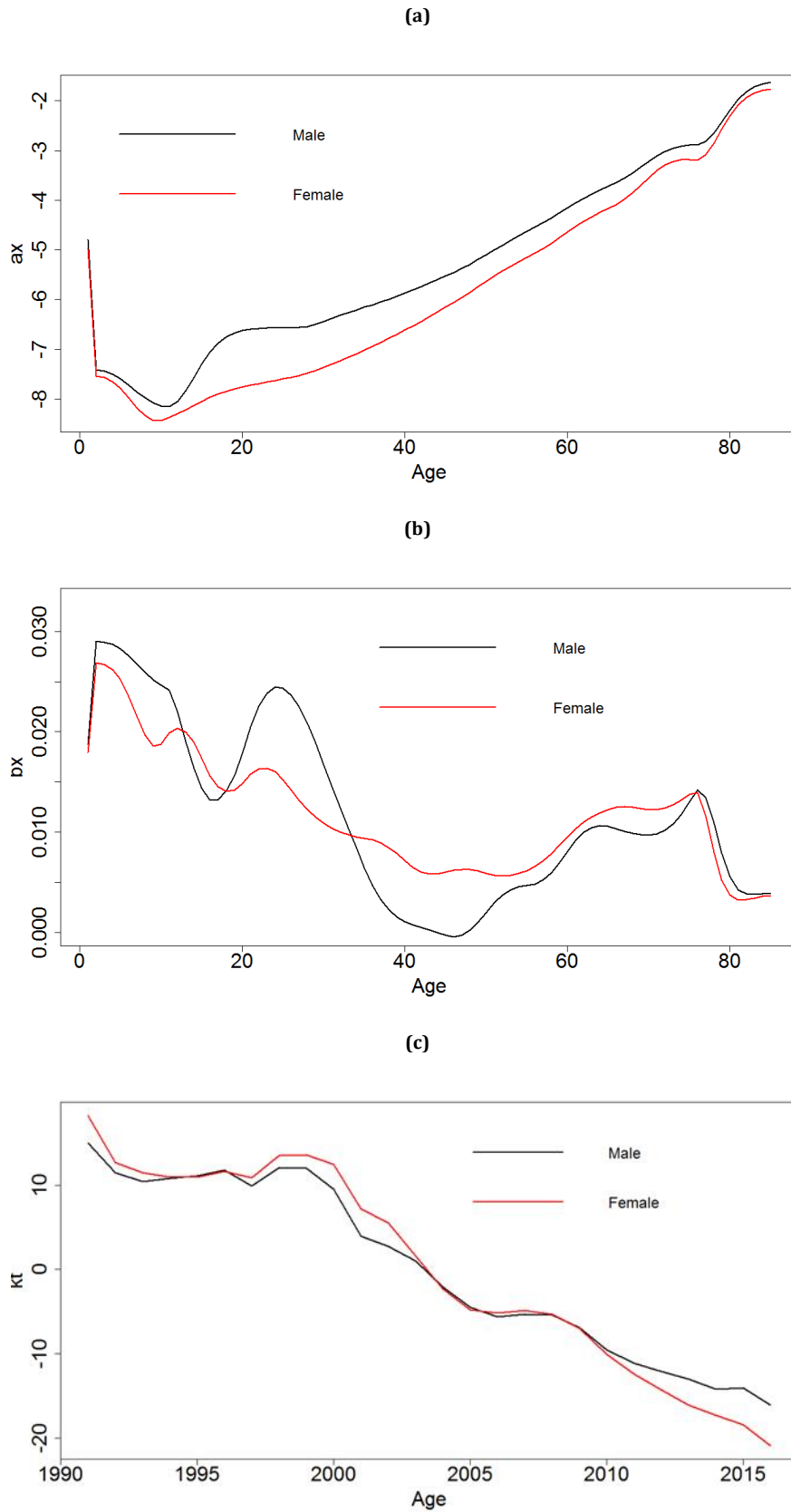
(a)

(b)

(c)

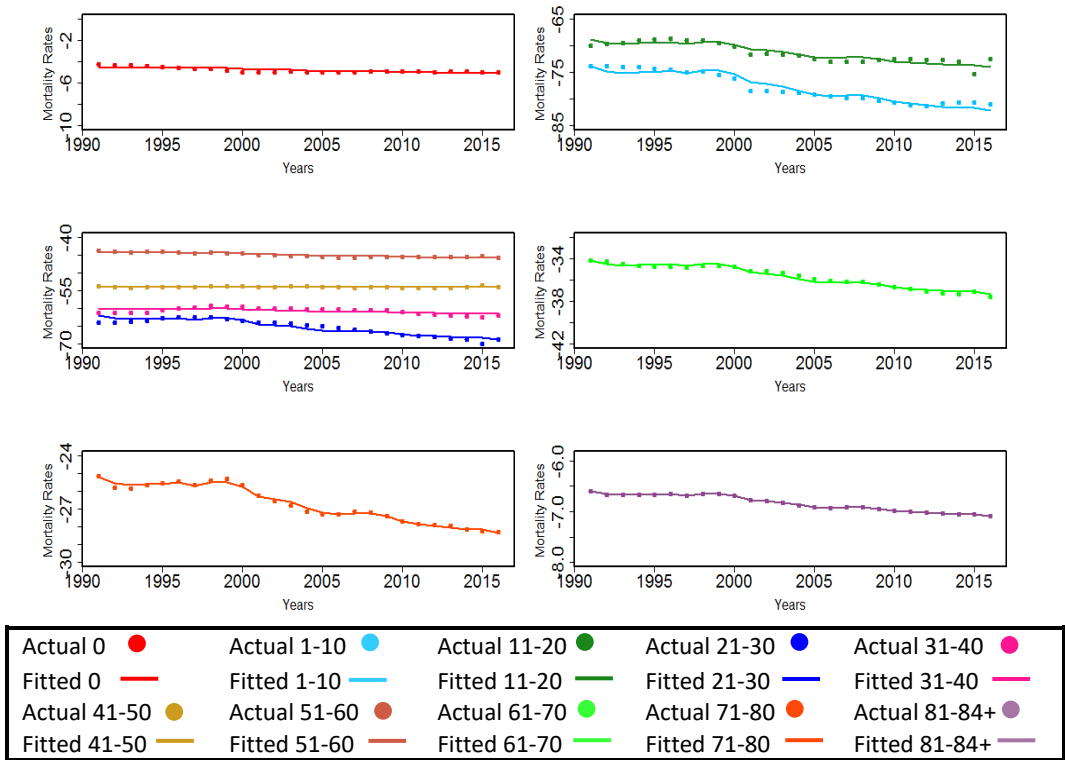**Fig. 2:** Components of fitted male and female LC model
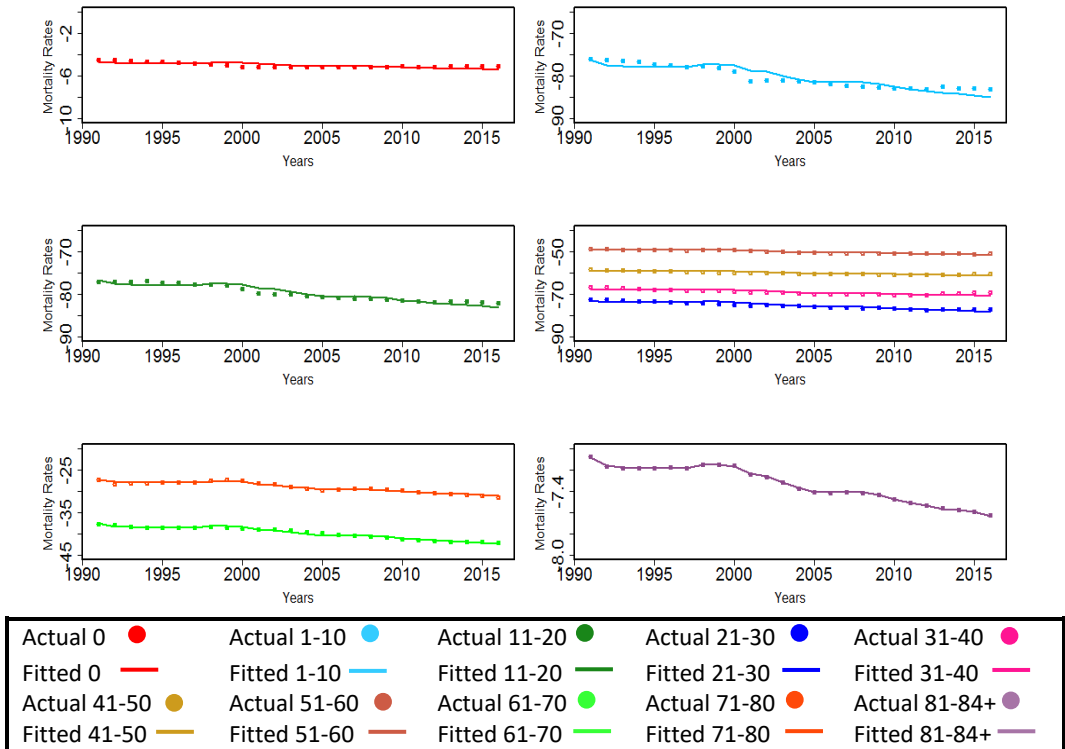
Fig. 3: Male fitted and actual plots

Fig. 4: Female fitted and actual plots

This finding is supported by Ngataman et al. (2016) who fitted Malaysian male and female mortality data to LC model from year 1981 to 2010.

6.3. Fitting ARIMA model

We now refit the entire dataset to the training set and obtain the time varying index k_t for both male and female, and further use this estimates to acquire the best forecasting models using ARIMA and Neural Network.

6.3.1. ARIMA

Table 1 shows all the nine considered models, together with their Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) to assess and obtain the best fit model for forecasting the k_t.

According to the values in Table 1, it was found that, for male ARIMA(0,1,0) with drift and for female ARIMA(0,2,0) without drift are the best fitted models for the k_t.

This results, together with the results of studies done on the Peruvian population and Indian population by Cerda-Hernández and Sikov (2018) and Chavhan and Shinde (2016) respectively, confirm the importance of finding an appropriate model for modelling the motility index before forecasting.

6.4. Neural network

After employing the "nnfor" package to acquire the best fitted MLP network. Fig. 5 and Fig. 6 were produced, which show the best fitted NN structures for the training set data of both male and female respectively.

It was found that a 1 input, 20 hidden node and 1 univariate lag combined forecast using a median operator was the best neural network fitted for male k_t. While for female k_t, the best fit model had 3 input, 20 hidden node and 1, 3, 4 univariate lags combined forecast using a median operator was the best neural network fitted for male k_t. Both fitted models were acquired at a MSE's of 2.7556 and 2.8088 for male and female respectively.
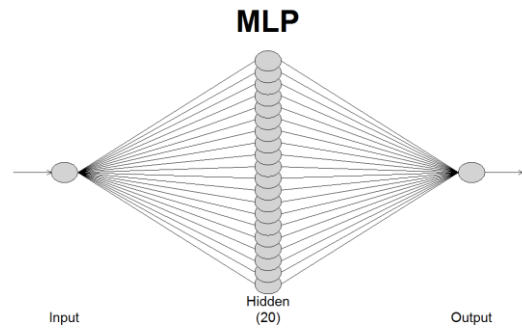

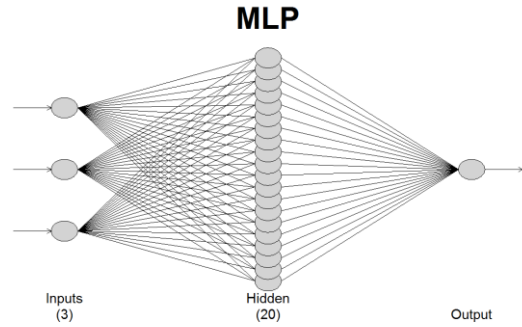**Fig. 5:** Male neural network structure

Fig. 6: Female neural network structure

Table 1: Male and female AIC & BIC

Model	Male		Female	
	AIC	BIC	AIC	BIC
ARIMA(2,1,2) with drift	89.3	95.56	98.43	104.70
ARIMA(2,1,1) with drift	86.73	91.96	96.48	101.70
ARIMA(1,1,2) with drift	87.61	92.83	96.53	101.75
ARIMA(1,1,1) with drift	87.69	91.86	98.59	102.77
ARIMA(1,1,0) with drift	86.13	89.26	96.59	99.73
ARIMA(0,1,1) with drift	85.69	88.82	97.53	100.66
ARIMA(0,1,0) with drift	**85.69**	**87.78**	97.91	100.00
ARIMA(1,2,0) without drift	88.61	90.60	95.05	97.04
ARIMA(0,2,0) without drift	87.33	88.33	**94.36**	**95.35**

6.5. Evaluating forecasting model accuracy

Now that the best fit forecasting models have been acquired. The time varying index can now be forecasted onto the testing set and errors between the forecast mortality rates and actual mortality values may be evaluated.

Table 2 show the three error measures used to evaluate the forecast between the two models. Looking across the years for both male and female error tables it can be observed, as the time horizon increases the size of errors increases for both models. However, for the Neural Network model the size of the errors increase at a lower rate, showing better performance over a larger time horizon. Similar results were reported by Safitri et al. (2018) for the Indonesian population.

Furthermore, it can also be noted that the Neural Network model out performs the ARIMA method by consistently showing lower RMSE, MAE and MAPE values for both male and female over the years. As a result this confirms that the NN model is a better

forecasting method for the time varying index k_t compared to the conventional ARIMA method.

7. Conclusion

In this paper, LC model was fitted to Malaysian male and female mortality data, following with the use of two models to forecast the time varying index (k_t). The two models used were the conventional ARIMA model, and a new approach which has been prevailing as a forecasting method known as Neural Network. Unlike previous studies done in Malaysian, to obtain the best fit ARIMA model to forecast k_t, 9 different ARIMA models were compared and the best fitted model was obtained through evaluating the AIC and BIC of the 9 models. The best fit ARIMA model obtained was then used as a benchmark model, and its forecast performance was compared with the Neural Network models forecast using 3 performance measures; Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

Table 2: Error for male ARIMA and neural network models (top) and error for male ARIMA and neural network models (bottom)

FEMALE	2013		2014		2015		2016	
	ARIMA	NN	ARIMA	NN	ARIMA	NN	ARIMA	NN
RMSE	0.14202	0.13442	0.16016	0.14520	0.18785	0.16435	0.20991	0.17878
MAE	0.11074	0.10446	0.13061	0.11835	0.15438	0.13606	0.17348	0.14835
MAPE	1.71005	1.62468	2.02643	1.85786	2.37918	2.13245	2.65132	2.31726
MALE	2013		2014		2015		2016	
	ARIMA	NN	ARIMA	NN	ARIMA	NN	ARIMA	NN
RMSE	0.09755	0.09412	0.11561	0.10714	0.18091	0.17803	0.13448	0.10857
MAE	0.07444	0.07136	0.08542	0.07931	0.13325	0.12875	0.10087	0.08265
MAPE	1.22508	1.16879	1.38515	1.28442	2.10533	2.01810	1.62148	1.35175

The finding in this paper firstly showed that the LC model fitted well for both Malaysian male and female data with 86.7% and 92.6% variation being explained by both male and female fitted models respectively. This results were in line with Ngataman et al. (2016) who fitted Malaysian male and female mortality data to LC model from year 1981 to 2010.

Secondly, moving on to fitting the time varying index, for the ARIMA model, it was found that for Malaysian male data the time varying index k_t fitted best to ARIMA (0,1,0) with drift, while for Malaysian female data the time varying index k_t fitted best with ARIMA (0,2,0) without drift. Therefore, justifying the importance of testing ARIMA models before forecasting the time varying index as mentioned by Lee and Carter (1992). On the other hand, for the neural network model it was found that 1 input, 20 hidden nodes and 1 univariate lag combined forecast using a median operator was the best neural network fitted for male k_t. While for female k_t, the best fit model had 3 input, 20 hidden nodes and 1, 3, 4 univariate lags combined forecast using a median operator was the best neural network fitted for male k_t.

Lastly, coming to the evaluation of forecast for both models, our finding showed that overall the Neural Network model outperforms the conventional ARIMA model in forecasting mortality rates. The Neural Network model consistently showed lower errors against actual values and the size of errors increased at a lower rate as time horizon increases, compared to the forecast of the ARIMA model.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aberathna W, Alles L, Wickremasinghe WN, and Hewapathirana I (2014). Modeling and forecasting mortality in Sri Lanka. Sri Lankan Journal of Applied Statistics, 15(3): 141-170. https://doi.org/10.4038/sljastats.v15i3.7794

Antonio K, Bardoutsos A, and Ouburg W (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. European Actuarial Journal, 5(2): 245-281. https://doi.org/10.1007/s13385-015-0115-6

Bell WR (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. Journal of Official Statistics, 13(3): 279-303.

Booth H, Maindonald J, and Smith L (2002). Applying Lee-Carter under conditions of variable mortality decline. Population Studies, 56(3): 325-336. https://doi.org/10.1080/00324720215935 **PMid:12553330**

Box GE, Jenkins GM, Reinsel GC, and Ljung GM (2015). Time series analysis: Forecasting and control. John Wiley and Sons, Hoboken, USA.

Brouhns N, Denuit M, and Vermunt JK (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics, 31(3): 373-393. https://doi.org/10.1016/S0167-6687(02)00185-3

Burnham KP and Anderson DR (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research, 33(2): 261-304. https://doi.org/10.1007/b97636

Carter LR (1996). Forecasting US mortality: A comparison of Box-Jenkins ARIMA and structural time series models. The Sociological Quarterly, 37(1): 127-144. https://doi.org/10.1111/j.1533-8525.1996.tb02334.x

Cerda-Hernández J and Sikov A (2018). Lee-Carter method for forecasting mortality for Peruvian population. arXiv:1811.09622. Available online at: https://bit.ly/2ORykzV

Chavhan R and Shinde R (2016). Modeling and forecasting mortality using the Lee-Carter model for Indian population based on decade-wise data. Sri Lankan Journal of Applied Statistics, 17(1): 51-68. https://doi.org/10.4038/sljastats.v17i1.7845

Creal D, Koopman SJ, and Lucas A (2008). A general framework for observation driven time-varying parameter models. Tinbergen Institute Discussion Paper No. 08-108/4, Amsterdam, Netherlands. https://doi.org/10.2139/ssrn.1297183

Creal D, Koopman SJ, and Lucas A (2013). Generalized autoregressive score models with applications. Journal of Applied Econometrics, 28(5): 777-795. https://doi.org/10.1002/jae.1279

De Jong P and Tickle L (2006). Extending Lee–Carter mortality forecasting. Mathematical Population Studies, 13(1): 1-18. https://doi.org/10.1080/08898480500452109

Deaton AS and Paxson C (2004). Mortality, income, and income inequality over time in Britain and the United States. In: Wise DA (Ed.), Perspectives on the economics of aging: 247-286. University of Chicago Press, Chicago, USA. https://doi.org/10.7208/chicago/9780226903286.003.0007

Girosi F and King G (2007). Understanding the Lee-Carter mortality forecasting method. Working Paper, Harvard University, Cambridge, USA.

Harvey AC (1990). Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge, USA. https://doi.org/10.1017/CBO9781107049994

Hyndman RJ and Ullah MS (2007). Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics and Data Analysis, 51(10): 4942-4956. https://doi.org/10.1016/j.csda.2006.07.028

Jiang L (1995). Changing kinship structure and its implications for old-age support in urban and rural China. Population Studies, 49(1): 127-145. https://doi.org/10.1080/0032472031000148286

Koissi MC and Shapiro AF (2006). Fuzzy formulation of the Lee–Carter model for mortality forecasting. Insurance: Mathematics and Economics, 39(3): 287-309. https://doi.org/10.1016/j.insmatheco.2005.11.011

Koissi MC, Shapiro AF, and Högnäs G (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. Insurance: Mathematics and Economics, 38(1): 1-20. https://doi.org/10.1016/j.insmatheco.2005.06.008

Kourentzes N (2017). NNFOR: Time series forecasting with neural networks. GPL-3, Package. Available online at: https://bit.ly/2Z4agOk

Kourentzes N and Crone S (2010a). Feature selection for time series prediction: A combined filter and wrapper approach for neural networks. Neurocomputing, 73(10-12): 1923-1936. https://doi.org/10.1016/j.neucom.2010.01.017

Kourentzes N and Crone SF (2010b). Frequency independent automatic input variable selection for neural networks for forecasting. In The 2010 International Joint Conference on Neural Networks, IEEE, Barcelona, Spain: 1-8. https://doi.org/10.1109/IJCNN.2010.5596637

Kourentzes N, Barrow DK, and Crone SF (2014). Neural network ensemble operators for time series forecasting. Expert Systems with Applications, 41(9): 4235-4244. https://doi.org/10.1016/j.eswa.2013.12.011

Lee R and Miller T (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. Demography, 38(4): 537-549. https://doi.org/10.1353/dem.2001.0036 **PMid:11723950**

Lee RD and Carter LR (1992). Modeling and forecasting US mortality. Journal of the American Statistical Association, 87(419): 659-671. https://doi.org/10.1080/01621459.1992.10475265

Lee RD and Rofman R (1994). Modeling and projecting mortality in Chile. Notas de Poblacion, 22(59): 183-213.

Li N and Lee R (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography, 42(3): 575-594. https://doi.org/10.1353/dem.2005.0021 **PMid:16235614 PMCid:PMC1356525**

Neves C, Fernandes C, and Hoeltgebaum H (2017). Five different distributions for the Lee–Carter model of mortality forecasting: A comparison using GAS models. Insurance: Mathematics and Economics, 75: 48-57. https://doi.org/10.1016/j.insmatheco.2017.04.004

Ngataman N, Ibrahim RI, and Yusuf MM (2016). Forecasting the mortality rates of Malaysian population using Lee-Carter method. In the AIP Conference Proceedings, AIP Publishing, Melville, USA, 1750(1): 020009. https://doi.org/10.1063/1.4954522

Raftery AE, Chunn JL, Gerland P, and Ševčíková H (2013). Bayesian probabilistic projections of life expectancy for all countries. Demography, 50(3): 777-801. https://doi.org/10.1007/s13524-012-0193-x **PMid:23494599 PMCid:PMC3904289**

Ramsay JO and Silverman BW (2005). Principal components analysis for functional data. In: Ramsay JO and Silverman BW (Eds.), Functional data analysis: 147-172. Springer, New York, USA. https://doi.org/10.1007/b98888

Remund A, Camarda CG, and Riffe T (2018). Analyzing the young adult mortality hump in R with MortHump. MPIDR Technical Report No. TR-2018-2018-003, Max Planck Institute for Demographic Research, Rostock, Germany.

Safitri L, Mardiyati S, and Rahim H (2018). Forecasting the mortality rates of Indonesian population by using neural network. In the International Conference on Mathematics: Pure, Applied and Computation, Journal of Physics: IOP Publishing, 974(1): 012030. https://doi.org/10.1088/1742-6596/974/1/012030

Tuljapurkar S, Li N, and Boe C (2000). A universal pattern of mortality decline in the G7 countries. Nature, 405(6788): 789-792. https://doi.org/10.1038/35015561 **PMid:10866199**

Wilmoth JR (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical Report, Department of Demography, University of California, Berkeley, USA.

Wilmoth JR (1996). Mortality projections for Japan: A comparison of four methods. In: Caselli G and Lopez A (Eds.), Health and mortality among elderly population: 266-287. Oxford University Press, Oxford, UK.

Wilmoth JR, Andreev K, Jdanov D, Glei DA, Boe C, Bubenheim M, and Vachon P (2007). Methods protocol for the human mortality database. University of California, Berkeley, USA and Max Planck Institute for Demographic Research, Rostock, Germany.

Wiśniowski A, Smith PW, Bijak J, Raymer J, and Forster JJ (2015). Bayesian population forecasting: Extending the Lee-Carter method. Demography, 52(3): 1035-1059. https://doi.org/10.1007/s13524-015-0389-y **PMid:25962866**

Wong JS, Forster JJ, and Smith PW (2018). Bayesian mortality forecasting with overdispersion. Insurance: Mathematics and Economics, 83: 206-221. https://doi.org/10.1016/j.insmatheco.2017.09.023

Yadav A, Yadav S, and Kesarwani R (2012). Decelerating mortality rates in older ages and its prospects through Lee-Carter approach. PloS One, 7(12): e50941. https://doi.org/10.1371/journal.pone.0050941 **PMid:23236414 PMCid:PMC3516525**

Yan W (2012). Toward automatic time-series forecasting using neural networks. IEEE Transactions on Neural Networks and Learning Systems, 23(7): 1028-1039. https://doi.org/10.1109/TNNLS.2012.2198074 **PMid:24807130**

Yasungnoen N and Sattayatham P (2016). Forecasting Thai mortality by using the Lee-Carter model. Asia-Pacific Journal of Risk and Insurance, 10(1): 91-105.

Yue JC, Yang SS, and Huang HC (2008). A study of the Lee-Carter model with age-shifts. In the Living to 100 and Beyond Symposium, Orlando, USA: 1-20.

Zhang G, Patuwo BE, and Hu MY (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1): 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7