Pertanika J. Sci. & Technol. 27 (2): 797 - 814 (2019)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

# A Functional Data Approach to the Estimation of Mortality and Life Expectancy at Birth in Developing Countries

**Syazreen Niza Shair*, Nur Azhani Zolkifli, Nur Faiqah Zulkefli and Azizah Murad**

*Department of Actuarial Science, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40500 UiTM, Shah Alam, Selangor, Malaysia*

## ABSTRACT

The functional data model has recently received increasing attention particularly in their application to mortality forecasting. The advantage of this method over the well-known Lee-Carter model is the ability to treat the underlying process as functional, and provide estimations that are robust to outliers. This research investigates the accuracy of functional data approach in estimating the mortality rates and life expectancy at births in developing countries including Malaysia, Indonesia, Thailand and Singapore. The functional data method was applied to these countries' mortality data, and the out-sample forecast errors showed that, in terms of overall, the functional data model was more accurate than that of the original Lee-Carter model for males and females. The results provide evidence that the functional model is accurate to forecast the life expectancy at births for developing countries.

*Keywords:* Functional data model, Lee-Carter model, life expectancy at births, mortality forecasts

## INTRODUCTION

Throughout the second half of the 20th century, majority of developing countries experienced a continuation of declining mortality rates, following the trends that had occurred in many developed nations in the past. This mortality decline had resulted in an increase in the average human's survival. There is no sign that the average human life span will approach a certain limit. Hence, life expectancy is expected to continue to increase in the future. As a result, an accurate mortality forecasting model which accounts for changes in mortality trends is important for insurance companies as well as government agencies. For illustration, the

ARTICLE INFO

*Article history:*
Received: 09 February 2018
Accepted: 26 November 2018
Published: 25 April 2019

*E-mail addresses*:
syazreen@tmsk.uitm.edu.my (Syazreen Niza Shair)
zolkiflinurazhani@yahoo.com (Nur Azhani Zolkifli)
faiqahz22@yahoo.com (Nur Faiqah Zulkefli)
azizah@tmsk.uitm.edu.my (Azizah Murad)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

calculation of the expected present values of benefits in pricing requires accurate mortality projections to avoid underestimation of future costs. Moreover, accurate estimation of deaths is necessary for government planning particularly to ensure the social security benefits and medical care costs are sufficient for a growing number of elderly.

Actuaries and demographers put extensive efforts to develop methods to estimate mortality rates. Initially, several human mortality and survival functions were proposed including DeMoivers, Gompertz and Makeham laws (Bowers et al., 1997). Nonetheless, these functions do not include the change in mortality over time. The understanding of mortality improvement over time is important in more recent decades, in particular for social policy planning. Thus, the development of stochastic approaches based on time series methods received an increasing recognition. Example works are from McNown and Rogers (1989, 1992), Bell and Monsell (1991), and Lee and Carter (1992). The Lee and Carter model became one of the most influential stochastic method in the field of mortality forecasts. The model involves two parameters which are known as age-component and time-component. In order to estimate the two parameters, singular value decomposition (SVD) method is applied to decompose the log of age and year specific mortality rates matrices. It is noteworthy that, only the time-component is forecasted using a non-stationary time series model. Due to its parsimony in research, the Lee-Carter model received recognition and has been applied worldwide.

In recent years, we have seen a significant development of the Lee-Carter model purposely to improve the accuracy of forecast values. These extensions often use the original approach as a reference with some additional statistical techniques included in the models. Booth et al. (2005) provided reviews on some of these extensions. One of the Lee-Carter extensions is the functional data model from Hyndman and Ullah (2007). This model combines the idea from the functional data analysis, non-parametric smoothing and robust statistics to model the age specific mortality rates. It was proven that the functional model was more accurate than that of the original Lee-Carter model when forecasting both, mortality rates and life expectancy at births of fourteen selected developed countries (Shang et al., 2011). Moreover, the functional data model had successfully enhanced the accuracy of the age-mortality predictions for breast cancer patients in United States and England-Wales (Erbas et al., 2010). In addition, the model's applications were extended to other demographic components in Australia including fertility and migration (Hyndman & Booth, 2008).

Nonetheless, the application of the functional data model to developing countries is rather limited. Although a few researches showed the functional data model outperformed the Lee-Carter model for Malaysian mortality rates (Shair et al., 2017; Husin et al., 2015), the performance of the model in other developing countries is unknown. Thus, it is of interest to this research to extend the application of the functional data model to other developing countries.

This research aims to evaluate two mortality forecasting models: the Lee-Carter and its extended version, the functional data. Each method is applied to mortality data by age and gender in four developing countries including Malaysia, Indonesia, Thailand and Singapore. The evaluation of the models involves fitting different methods to data from 1960 to 2001, then forecast within the period of 2002 to 2015. To evaluate the forecast accuracies of both models, the forecast values of mortality rates and life expectancies were then compared with observations.

This paper is organized as follows. Section 2 describes data that we use in this study and explains the Lee-Carter and the functional data models and error measurements for evaluation. Section 3 reports the out-sample forecast errors of both models. Finally, concluding comments with suggestions for future works are presented in Section 4.

## MATERIALS AND METHODS

### Data

We utilized mortality data by age and gender in developing countries including Malaysia, Indonesia, Thailand and Singapore. The length of data is consistent for each country, from 1960 to 2015 (56-year data). For Malaysia, data were collected from the Department of Statistics Malaysia (DoSM), whereas for other developing countries, data were obtained from the United Nation (2015). The oldest age groups were excluded due to high mortality variations and some missing values, giving the age range was from 0 to 85 years old. It should be noted that, from the collected mortality data, the observed life expectancy at birth was calculated using the standard actuarial life table approach.

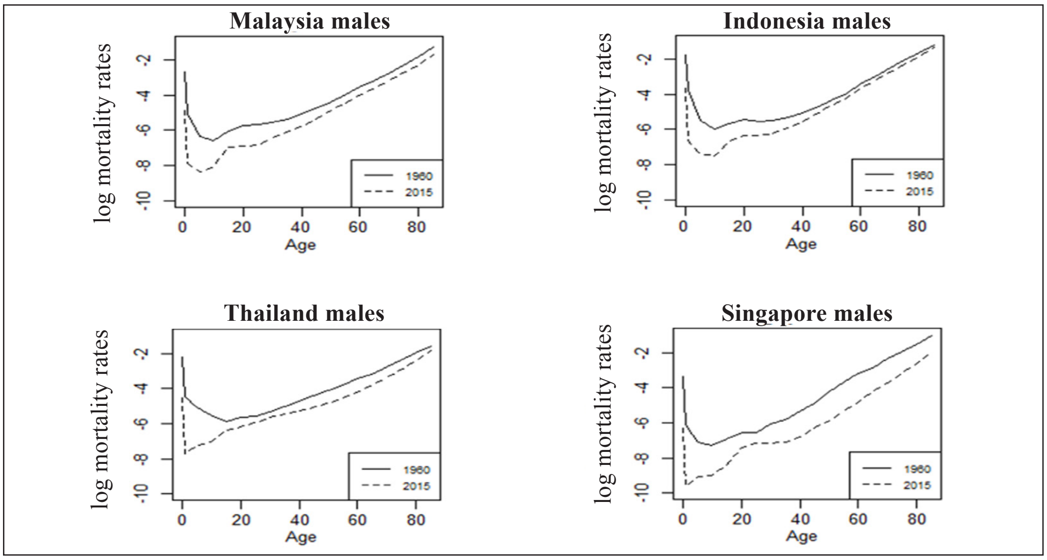


*Figure 1*. Age-specific mortality rates of Malaysia Indonesia, Thailand and Singapore males from year 1960 (——) to 2015 (----)

To have an idea of mortality evolution in developing countries, Figures 1 and 2 display the decreasing trends in log mortality rates from 1960 (solid lines) to 2015 (dashed lines), for males and females respectively. In comparison to other age groups, infants and children who were below five years old experienced remarkable decreases in mortality over the years for both genders. This trend is shown by the biggest gaps between the two lines for each country and gender. Factors contributing the declines in infant deaths among developing countries were the decrease in the number of children who were malnourished and poor also the improvement in mother's education background (Rutstein, 2000).

Conversely, it is noticeable that the gap between the two lines for people aged fifty-five years and above is diminutive for Malaysia, Thailand and Indonesia, indicating the curative medical technologies in recent years had little effects on reducing deaths of old-age people of these countries. This trend however is not applicable to Singapore as both figures show that the gaps for older male and female Singaporean are wider than that of other three countries and they are consistent for all age groups. This result suggests that healthcare services in Singapore has successfully reduced deaths among the elderly. According to Tan (2017), Singapore has adopted one of the best health care system so called the hybrid health care model that combines government subsidies with patient co-payments.

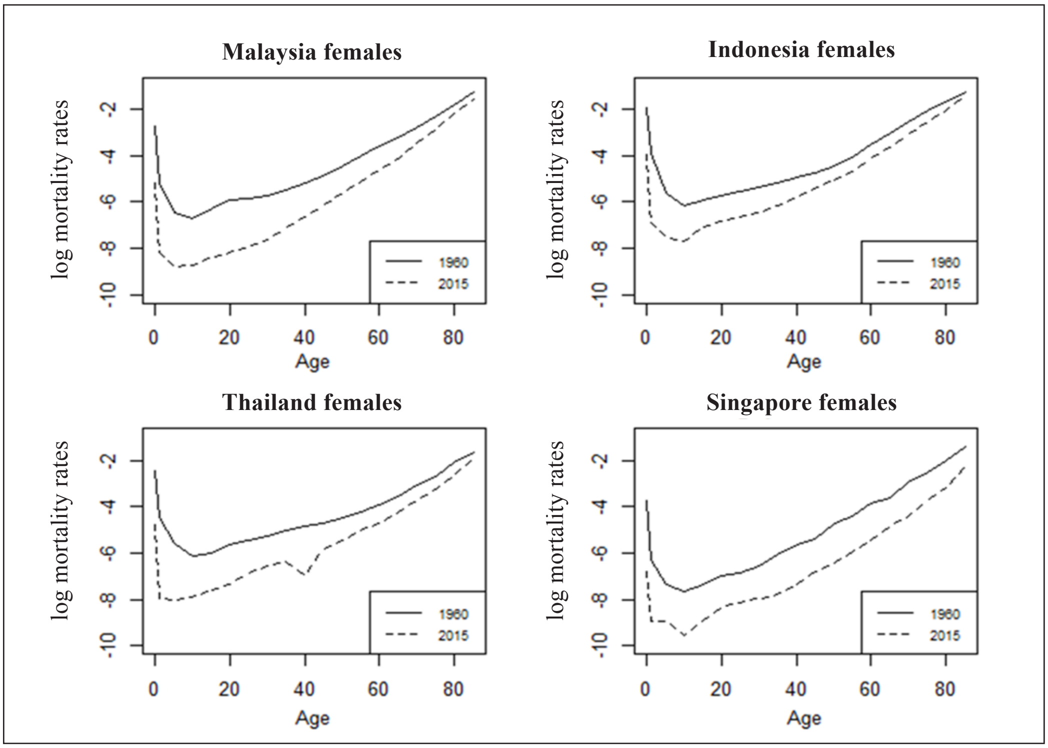


*Figure 2.* Age-specific mortality rates of Malaysia Indonesia, Thailand and Singapore females from year 1960 (——) to 2015 (----)

Due to a decline in mortality over time, life expectancy at births of all four developing countries have increased significantly for both genders. From the Figure 3, it can be seen that the differences in life expectancy between genders are generally wider over the years. The sex differentials in mortality and life expectancy are normally attributed to trends in behavioral and social risk factors such as cigarette smoking, heavy drinking, violence and occupational hazards (Villegas & Haberman, 2014).

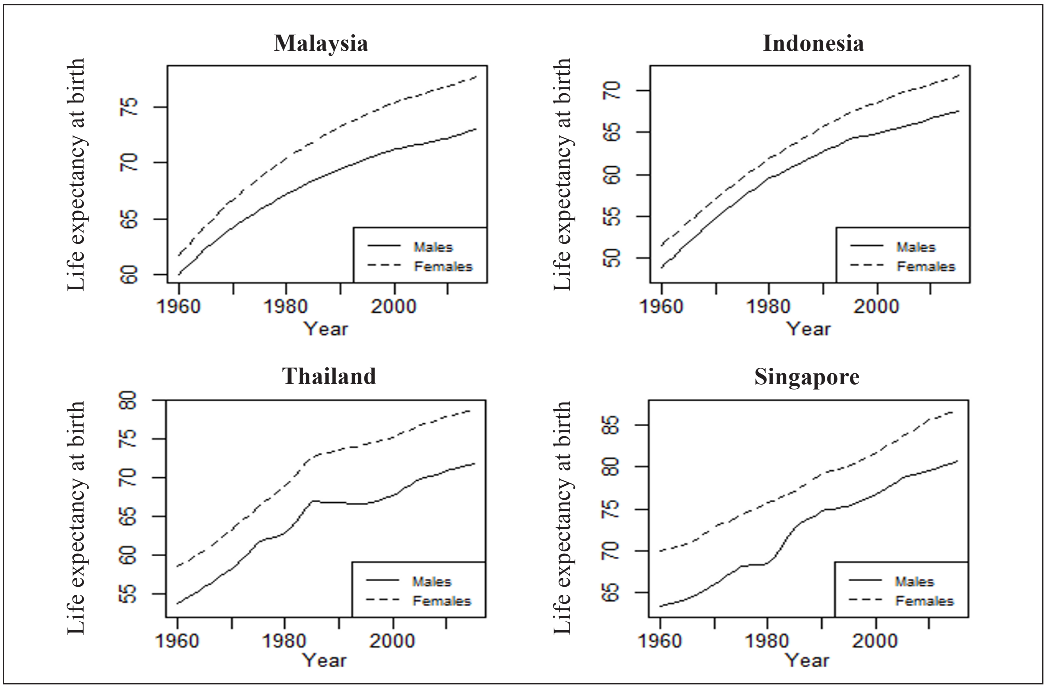


*Figure 3.* Life expectancy at births for Malaysia, Indonesia, Thailand and Singapore females (---) and males (——) from 1960 to 2015

The increase in life expectancy has occurred unevenly between countries. It is reported in the Table 1 that although life expectancy at births of Indonesian are the lowest compared to other countries, the rate of increase from 1960 to 2015 is the highest, 39.03%. On the other hand, Malaysian males' life expectancy at births are ranked in the second highest among male populations, however the rate of increase is the slowest which is 21.48% over the same period. There are many factors that could affect the rate of improvement in life expectancy such as lifestyles, accident rates, government policies etc. It is also noteworthy that, the life expectancy at births of developing countries has increased at a substantially more rapid rate compared to majority of developed countries. This issue becomes a main concern to policy makers recently. Efforts should be done to ensure retirement benefits and other social benefits which are sufficient for the growing number of elderly.

Table 1
*Life expectancy at births from 1960 to 2015 for males and females in four developing countries (*MAL=*Malaysia,* IND=*Indonesia,* THA=*Thailand and* SIG=*Singapore)*

| Year | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|
| | IND | THA | MAL | SIG | IND | THA | MAL | SIG |
| 1960 | 48.91 | 53.74 | 60.09 | 63.41 | 51.57 | 58.52 | 61.76 | 70.04 |
| 1970 | 54.84 | 58.18 | 64.24 | 65.96 | 57.09 | 63.32 | 66.76 | 72.76 |
| 1980 | 59.53 | 62.81 | 67.24 | 68.44 | 61.85 | 68.88 | 70.46 | 75.67 |
| 1990 | 62.68 | 66.79 | 69.49 | 74.75 | 65.68 | 73.67 | 73.25 | 79.22 |
| 2000 | 64.88 | 67.71 | 71.26 | 76.69 | 68.54 | 75.17 | 75.42 | 81.74 |
| 2010 | 66.62 | 70.84 | 72.21 | 79.58 | 70.70 | 77.81 | 76.87 | 85.59 |
| 2015 | 67.43 | 71.83 | 73.00 | 80.62 | 71.70 | 78.68 | 77.67 | 86.63 |
| **(Δ%)** | **37.87** | **33.66** | **21.48** | **27.14** | **39.03** | **34.45** | **25.76** | **23.69** |

## The Lee and Carter Model

The Lee-Carter model is one the most popular model for projecting future mortality rates. Although there are various extensions and modifications suggested by many authors, the Lee and Carter model is often used as the performance benchmark. The Lee and Carter model has the following form:

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

where $m_{x,t}$ is the central mortality rate for persons age $x$ in year $t$. The $a_x$ variable represents the average age-pattern of mortality rates. The $k_t$ variable is the time-varying parameter which is known as mortality index, describes the level of mortality at time $t$ and the $b_x$ variable is the age-specific component, measures the change of mortality at each age $x$.

In order to estimates $k_t$ and $b_x$ parameters, following the Lee and Carter (1992) research, we use the singular value decomposition (SVD) that assumes the errors are homoscedastic. Constraints are imposed such that $\sum_x b_x = 1$ and $\sum_t k_t = 0$. Under these assumptions, the $a_x$ parameter is estimated by averaging $ln(m_{x,t})$ over time. The Lee-Carter model also assumes that the $k_t$ values portray a linearly decreasing pattern hence a random walk with drift is an appropriate model to forecast the values. The random walk with drift model is described as the following:

$$k_t = k_{t-1} + d + e_t \quad (2)$$

where $d$ is the average annual change in $k_t$ values, and $e_t$ are its respective errors. It is noteworthy that only variable $k_t$ is forecasted. Hence, the equation of age specific mortality rate forecasts for males and females in each country are calculated using the forecasted $\hat{k}_t$ values and the estimated $a_x$ and $b_x$ as follows:

$$ln(\widehat{m}_{x,t}) = a_x + b_x\ \widehat{k}_t\ _t + \mathcal{E}_{x,t} \qquad (3)$$

From the equation (3), the values of the fitted parameters $a_x$ and $b_x$ for each country (Malaysia, Indonesia, Thailand and Singapore) and gender (male and females) can be referred to Table 2 and Table 3. Also, the respective forecasted mortality index $\widehat{k}_t$ are listed in Table 4.

Table 2
*The estimated ($a_x$) of the Lee-Carter model for males and females in developing countries including Malaysia (MAL), Indonesia (IND), Thailand (THA) and Singapore (SIN)*

| Age | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|
| ($x$) | MAL ($a_x$) | IND ($a_x$) | THA ($a_x$) | SIN ($a_x$) | MAL ($a_x$) | IND ($a_x$) | THA ($a_x$) | SIN ($a_x$) |
| 0 | -3.820 | -2.429 | -3.114 | -4.624 | -3.921 | -2.637 | -3.324 | -4.831 |
| 1 | -6.358 | -4.855 | -5.856 | -7.380 | -6.433 | -4.944 | -5.865 | -7.459 |
| 5 | -7.248 | -6.205 | -6.040 | -8.205 | -7.423 | -6.354 | -6.566 | -8.279 |
| 10 | -7.224 | -6.569 | -6.320 | -7.942 | -7.457 | -6.770 | -6.940 | -8.295 |
| 15 | -6.423 | -6.115 | -6.065 | -7.314 | -7.053 | -6.436 | -6.605 | -7.875 |
| 20 | -6.115 | -5.867 | -5.702 | -6.857 | -6.786 | -6.215 | -6.315 | -7.627 |
| 25 | -6.048 | -5.918 | -5.519 | -6.930 | -6.675 | -6.038 | -6.147 | -7.468 |
| 30 | -5.863 | -5.840 | -5.376 | -6.748 | -6.450 | -5.830 | -5.959 | -7.218 |
| 35 | -5.645 | -5.624 | -5.215 | -6.426 | -6.158 | -5.577 | -5.735 | -6.792 |
| 40 | -5.385 | -5.327 | -4.999 | -5.880 | -5.810 | -5.311 | -5.514 | -6.354 |
| 45 | -5.060 | -4.952 | -4.730 | -5.376 | -5.398 | -5.035 | -5.159 | -5.905 |
| 50 | -4.637 | -4.530 | -4.416 | -4.786 | -4.939 | -4.691 | -4.812 | -5.322 |
| 55 | -4.215 | -4.096 | -4.120 | -4.248 | -4.472 | -4.330 | -4.499 | -4.829 |
| 60 | -3.739 | -3.546 | -3.739 | -3.741 | -3.987 | -3.744 | -4.149 | -4.342 |
| 65 | -3.310 | -3.094 | -3.381 | -3.145 | -3.518 | -3.271 | -3.754 | -3.834 |
| 70 | -2.872 | -2.622 | -2.966 | -2.802 | -3.034 | -2.745 | -3.276 | -3.345 |
| 75 | -2.389 | -2.159 | -2.588 | -2.361 | -2.490 | -2.250 | -2.865 | -2.815 |
| 80 | -1.912 | -1.722 | -2.090 | -1.947 | -1.958 | -1.801 | -2.286 | -2.366 |
| 85 | -1.359 | -1.230 | -1.629 | -1.433 | -1.351 | -1.301 | -1.746 | -1.700 |

Table 3
*The estimated age-component ($b_x$) of the Lee-Carter model for males and females in developing countries including Malaysia (MAL), Indonesia (IND), Thailand (THA) and Singapore (SIN)*

| Age | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|
| ($x$) | MAL ($b_x$) | IND ($b_x$) | THA ($b_x$) | SIN ($b_x$) | MAL ($b_x$) | IND ($b_x$) | THA ($b_x$) | SIN ($b_x$) |
| 0 | 0.150 | 0.116 | 0.129 | 0.103 | 0.102 | 0.091 | 0.077 | 0.098 |
| 1 | 0.167 | 0.179 | 0.207 | 0.097 | 0.108 | 0.142 | 0.131 | 0.092 |
| 5 | 0.118 | 0.126 | 0.141 | 0.087 | 0.085 | 0.101 | 0.099 | 0.077 |
| 10 | 0.083 | 0.100 | 0.117 | 0.056 | 0.067 | 0.082 | 0.082 | 0.049 |
| 15 | 0.048 | 0.067 | 0.027 | 0.038 | 0.070 | 0.066 | 0.067 | 0.053 |

Table 3 *(continue)*

| Age ($x$) | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|
| | MAL ($b_x$) | IND ($b_x$) | THA ($b_x$) | SIN ($b_x$) | MAL ($b_x$) | IND ($b_x$) | THA ($b_x$) | SIN ($b_x$) |
| 20 | 0.052 | 0.064 | 0.013 | 0.040 | 0.076 | 0.064 | 0.062 | 0.051 |
| 25 | 0.051 | 0.060 | -0.019 | 0.051 | 0.071 | 0.063 | 0.047 | 0.056 |
| 30 | 0.042 | 0.054 | -0.013 | 0.047 | 0.065 | 0.060 | 0.055 | 0.064 |
| 35 | 0.040 | 0.048 | 0.013 | 0.048 | 0.059 | 0.054 | 0.061 | 0.054 |
| 40 | 0.042 | 0.040 | 0.035 | 0.049 | 0.054 | 0.047 | 0.071 | 0.060 |
| 45 | 0.038 | 0.030 | 0.046 | 0.048 | 0.047 | 0.037 | 0.046 | 0.049 |
| 50 | 0.034 | 0.022 | 0.049 | 0.048 | 0.042 | 0.033 | 0.040 | 0.043 |
| 55 | 0.031 | 0.015 | 0.048 | 0.045 | 0.038 | 0.029 | 0.033 | 0.047 |
| 60 | 0.024 | 0.014 | 0.051 | 0.043 | 0.032 | 0.031 | 0.032 | 0.041 |
| 65 | 0.021 | 0.013 | 0.042 | 0.039 | 0.028 | 0.028 | 0.025 | 0.036 |
| 70 | 0.017 | 0.015 | 0.044 | 0.043 | 0.023 | 0.027 | 0.027 | 0.039 |
| 75 | 0.015 | 0.014 | 0.038 | 0.040 | 0.016 | 0.022 | 0.021 | 0.032 |
| 80 | 0.014 | 0.013 | 0.019 | 0.044 | 0.011 | 0.017 | 0.017 | 0.034 |
| 85 | 0.012 | 0.009 | 0.012 | 0.034 | 0.005 | 0.008 | 0.006 | 0.024 |

Table 4

*The out-sample forecasted mortality index ($\hat{k}_t$) of the Lee-Carter model for males and females in developing countries including Malaysia (MAL), Indonesia (IND), Thailand (THA) and Singapore (SIN)*

| Year ($t$) | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|
| | MAL ($\hat{k}_t$) | IND ($\hat{k}_t$) | THA ($\hat{k}_t$) | SIN ($\hat{k}_t$) | MAL ($\hat{k}_t$) | IND ($\hat{k}_t$) | THA ($\hat{k}_t$) | SIN ($\hat{k}_t$) |
| 2002 | -0.401 | -0.305 | -0.355 | -0.543 | -0.580 | -0.414 | -0.528 | -0.548 |
| 2003 | -0.801 | -0.610 | -0.710 | -1.086 | -1.160 | -0.828 | -1.055 | -1.097 |
| 2004 | -1.202 | -0.915 | -1.065 | -1.629 | -1.740 | -1.242 | -1.583 | -1.645 |
| 2005 | -1.603 | -1.220 | -1.420 | -2.172 | -2.319 | -1.655 | -2.110 | -2.193 |
| 2006 | -2.004 | -1.525 | -1.775 | -2.715 | -2.899 | -2.069 | -2.638 | -2.742 |
| 2007 | -2.404 | -1.830 | -2.130 | -3.258 | -3.479 | -2.483 | -3.165 | -3.290 |
| 2008 | -2.805 | -2.135 | -2.485 | -3.802 | -4.059 | -2.897 | -3.693 | -3.839 |
| 2009 | -3.206 | -2.440 | -2.840 | -4.345 | -4.639 | -3.311 | -4.220 | -4.387 |
| 2010 | -3.607 | -2.745 | -3.195 | -4.888 | -5.219 | -3.725 | -4.748 | -4.935 |
| 2011 | -4.007 | -3.049 | -3.550 | -5.431 | -5.799 | -4.139 | -5.275 | -5.484 |
| 2012 | -4.408 | -3.354 | -3.905 | -5.974 | -6.378 | -4.552 | -5.803 | -6.032 |
| 2013 | -4.809 | -3.659 | -4.259 | -6.517 | -6.958 | -4.966 | -6.330 | -6.580 |
| 2014 | -5.210 | -3.964 | -4.614 | -7.060 | -7.538 | -5.380 | -6.858 | -7.129 |
| 2015 | -5.610 | -4.269 | -4.969 | -7.603 | -8.118 | -5.794 | -7.386 | -7.677 |

## The Functional Data Model

The model of Hyndman and Ullah (2007) uses the functional data to forecast the age specific mortality rates. The fundamental principle of the functional data analysis is to transform discrete data into a function or curve (Ramsay et al., 2009; Shaadan et al., 2014). In order to estimate mortality rates, the functional data model from Hyndman and Ullah (2007) extended the Lee-Carter approach to include the following additional statistical procedures:

1. The model assumes there is underlying smooth function of age for each year, $f_t(x)$ in which the rates are smoothed using a non-parametric method such that:

   $$y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_{t,x} \tag{4}$$

   where $y_t(x)$ is observation rate for person age $x$ = 0,1, .....85 in year $t$ = 1,2 .... $n$.. The $f_t(x)$ is the corresponding smoothed rate, $\varepsilon_{t,x}$ are iid standard normal error in year t with and $\sigma_t(x)$ allows the error to vary by age.
2. The model used a robust principal component analysis to estimate more than one set of mortality index, $k_t$ and age component, $b_x$. Otherwise, the Lee-Carter model only estimated one set of the two variables using the SVD method. Multiple sets of ($k_t$, $b_x$) are important to comprise higher percentage of data variations. The number of set of parameters ($H$) are estimated such that it minimizes the mean integrated squared errors.
3. In order to forecast mortality indices, the model uses more general univariate time series methods, $ARIMA(p,d,q)$ than random walk with drift.

   Therefore, the functional data model is mathematically expressed as follows.

   $$f_t(x) = \mu(x) + \sum_{h=1}^{H} b_h(x)\, K_{t,h} + e_t(x) + \sigma_t(x)\varepsilon_{t,x} \tag{5}$$

   where $f_t(x)$ is the smoothed mortality rates across ages in each year. Following Hyndman and Ullah (2007), 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 applies after age c. In this research, we set $c$ = 40 as the log mortality plot in Figure 1 and Figure 2 show that the rates constantly increase after age 40 for all countries.

According to Dokumentov and Hyndman (2013), the weighted penalized regression spline involves calculating a vector β which minimizes the following expression:

$$||\omega(y-X\beta)||^2 + \lambda^2\beta^T D\beta \tag{6}$$

where $y$ is a vector of observations, $X$ is a matrix representing linear spline bases, $D$ is a diagonal matrix, $\omega$ is a vector of weights and $\lambda$ is a parameter. In the case of mortality rates, observations in year $t$ are given by $y = m_{x,t}$ the weights $\omega$ are taken as the inverse of the estimated variances of $y$ in which $y$ is assumed to have a binomially distribution. For details of this smoothing technique, see Hyndman and Ullah (2007). In this research, the weighted penalized regression spline smoothing procedure is performed using the *demography* package for R developed by Hyndman (2013).

The $\mu(x)$ from equation (5) refers to the average of mortality rates for each age across years. The $b_h(x)$ and $K_{t,h}$ are the $h^{th}$ age component and mortality index respectively. The optimum number $H$ is determined using the mean integrated squared errors (MISE):

$$MISE = \frac{1}{n}\sum_{t=1}^{n}\int e_t(x)^2 d_x \tag{7}$$

The $b_h(x)$ and $K_{t,h}$ variables are estimated by decomposing the matrix of $|f_t(x) - \mu(x)|$ using the principal component analysis. The $e_t(x)$ is the error term assumed to be homoscedastic such that $e_t(x) \sim (0, v(x))$. Next is to forecast the variables $K_{t,h}$ using the fitted univariate time series models to get $\widehat{K}_{t,h}$. Finally, the forecasts of $f_t(x)$ is obtained as the following:

$$\hat{f}_t(x) = \mu(x) + \sum_{h=1}^{6} b_h(x)\widehat{K}_{t,h} + e_t(x) \tag{8}$$

From the equation (8), the estimated age parameters or bases of principle components for Malaysian males: $\mu(x)$, $b_1(x)$, $b_2(x)$, $b_3(x)$, $b_4(x)$, $b_5(x)$ and $b_6(x)$ are summarized in Table 5 whereas the respective forecasted coefficient values: $\widehat{K}_{t,1}$, $\widehat{K}_{t,2}$, $\widehat{K}_{t,3}$, $\widehat{K}_{t,4}$, $\widehat{K}_{t,5}$, and $\widehat{K}_{t,6}$ re presented in Table 6. It should be noted that, we provide the parameter values only for Malaysian males. The parameter values for the remaining seven sub-populations including Malaysian females, Indonesian males, Indonesian females, Thailand males, Thailand females, Singaporean males and Singaporean females are not included in this paper, however they are available by contacting the correspondence author of this paper.

Table 5
*The estimated of principle component bases of the functional data model for Malaysian males*

| $x$ | $\mu(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | -3.823 | 0.390 | 0.214 | -0.176 | -0.069 | 0.393 | -0.948 |
| 1 | -5.842 | 0.199 | -0.304 | -0.011 | -0.416 | -0.058 | -0.088 |
| 5 | -6.689 | 0.123 | -0.295 | 0.013 | 0.054 | 0.042 | 0.054 |
| 10 | -6.700 | 0.121 | -0.111 | 0.009 | 0.234 | -0.045 | -0.009 |
| 15 | -6.512 | 0.131 | 0.026 | 0.011 | 0.163 | -0.177 | -0.033 |
| 20 | -6.304 | 0.135 | 0.092 | 0.023 | 0.022 | -0.191 | -0.012 |
| 25 | -6.100 | 0.131 | 0.107 | 0.023 | -0.116 | -0.133 | 0.036 |
| 30 | -5.899 | 0.123 | 0.090 | 0.009 | -0.120 | -0.002 | 0.062 |
| 35 | -5.646 | 0.113 | 0.072 | -0.042 | -0.104 | 0.039 | 0.072 |
| 40 | -5.371 | 0.105 | 0.048 | -0.079 | 0.032 | 0.139 | 0.048 |

Table 5 *(continue)*

| $x$ | $\mu(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ | $b_1(x)$ |
|---|---|---|---|---|---|---|---|
| 45 | -5.026 | 0.096 | 0.037 | -0.116 | 0.089 | 0.156 | 0.030 |
| 50 | -4.620 | 0.086 | 0.042 | -0.114 | 0.034 | 0.064 | 0.027 |
| 55 | -4.203 | 0.076 | 0.041 | -0.047 | 0.026 | 0.038 | 0.015 |
| 60 | -3.765 | 0.065 | 0.033 | 0.020 | 0.039 | 0.077 | 0.004 |
| 65 | -3.306 | 0.053 | 0.029 | 0.061 | 0.003 | 0.080 | 0.008 |
| 70 | -2.858 | 0.045 | 0.027 | 0.117 | -0.020 | 0.061 | 0.010 |
| 75 | -2.406 | 0.040 | 0.024 | 0.188 | 0.009 | 0.065 | 0.002 |
| 80 | -1.905 | 0.036 | 0.020 | 0.228 | 0.031 | 0.071 | -0.004 |
| 85 | -1.360 | 0.031 | 0.018 | 0.225 | 0.012 | 0.058 | -0.001 |

Table 6

*The out-sample forecast values of principle component coefficients of the functional data model for Malaysian males*

| $t$ | $\widehat{K}_{t,1}$ | $\widehat{K}_{t,2}$ | $\widehat{K}_{t,3}$ | $\widehat{K}_{t,4}$ | $\widehat{K}_{t,5}$ | $\widehat{K}_{t,6}$ |
|---|---|---|---|---|---|---|
| 2002 | -2.451 | -0.547 | -0.058 | -0.059 | -0.044 | 0.066 |
| 2003 | -2.536 | -0.476 | -0.058 | -0.062 | -0.046 | 0.068 |
| 2004 | -2.623 | -0.405 | -0.058 | -0.064 | -0.049 | 0.070 |
| 2005 | -2.709 | -0.334 | -0.058 | -0.067 | -0.051 | 0.072 |
| 2006 | -2.795 | -0.263 | -0.058 | -0.070 | -0.054 | 0.074 |
| 2007 | -2.881 | -0.192 | -0.058 | -0.072 | -0.056 | 0.076 |
| 2008 | -2.967 | -0.121 | -0.058 | -0.075 | -0.059 | 0.078 |
| 2009 | -3.054 | -0.050 | -0.058 | -0.077 | -0.061 | 0.080 |
| 2010 | -3.140 | 0.021 | -0.058 | -0.080 | -0.064 | 0.082 |
| 2011 | -3.226 | 0.092 | -0.058 | -0.083 | -0.066 | 0.084 |
| 2012 | -3.312 | 0.163 | -0.058 | -0.085 | -0.069 | 0.086 |
| 2013 | -3.398 | 0.234 | -0.058 | -0.088 | -0.071 | 0.089 |
| 2014 | -3.484 | 0.305 | -0.058 | -0.091 | -0.074 | 0.091 |
| 2015 | -3.570 | 0.375 | -0.058 | -0.093 | -0.076 | 0.093 |

## The Forecast Error Measurements

In order to evaluate the performance of the Lee-Carter model and the functional data model, data from 1960 to 2001 (42 years) were fitted into the models and then forecasted the mortality rates and life expectancy at births over the 2002 and 2015 period (14 years). On the issue of the length of data series for estimation and evaluation processes, there is no clear indication on the appropriate length of data for each process. However, Lazim (2012) suggested that at least one-quarter of the full length of data was appropriate to be saved for evaluation. Furthermore, Hyndman and Athanasopoulos (2014) stated that the size of test set was typically 20% of the total data. The difference between the forecasted values of mortality rates and life expectancy at births from 1990 to 2015, and its respective actual rates in the same period are defined as the out-sample forecast errors. In this research, the out-sample forecast errors are estimated using two error measurements which are the Root Mean Squared Forecast Error (RMSFE) for mortality and the Mean Absolute Forecast Error (MAFE) for life expectancy at births.

The RMSFE and MAFE are commonly used scale-dependent measure which have been applied in many fields and the formula for each sub-population are given as follows:

$$RMSFE = \sqrt{\frac{\sum_{n=1}^{N}\sum_{j=1}^{p}\left(F_{j,n}-y_{j,n}\right)^2}{N\times P}} \tag{5}$$

$$MAFE = \frac{\sum_{n=1}^{N}|F_n-y_n|}{N} \tag{6}$$

Where

$F_{j,n}$ is the forecast value (mortality rate) for a person age $j$ in year $n$

$y_{j,n}$ is the observed value (mortality rate) for a person age $j$ in year $n$

$F_n$ is the forecast value (life expectancy at birth) in year $n$

$y_n$ is the observed value (life expectancy at birth) in year $n$

$P$ is the maximum age

$N$ is the latest observation year

## RESULTS AND DISCUSSIONS

In this section we report the forecast errors of log age and gender specific mortality rates and life expectancy at births for two different mortality forecasting models, the Lee-Carter and functional data in four developing countries including Malaysia, Indonesia, Thailand and Singapore.

Table 7 summarizes the RMSFE of log mortality rates segregated by genders and countries from both models. For males, the functional model is more accurate than that of the Lee-Carter model in all four countries including Malaysia, Indonesia,Thailand and Singapore, hence outperforming in terms of overall error (taking the average over countries). For females, although the functional model performs better than the Lee-Carter model only for two out of four countries, Indonesia and Thailand, the overall error indicates that the functional model is more accurate than the Lee-Carter model.

Figure 4 and 5 display the root mean square forecast errors of log mortality rates by age and country from two models: the Lee-Carter (solid lines) and the functional data (dashed line). Figure 4 clearly shows that the Lee-Carter model provides larger errors than the functional data model for males in certain age groups such as infants and 40 to 60 in

Malaysia, 0 to 50 in Indonesia and 0 to 19, 45 to 55 and 65 to 75 in Thailand. The forecast errors from both models are almost the same in Singapore. Furthermore, Figure 5 shows that the Lee-Carter model provides considerably large errors in 0 to 50 age groups for Indonesian females and in 10 to 35 for Thailand females, resulting the functional model out-performing the Lee-Carter model for females in these two countries. These results are consistent with the previous outcomes in the Table 7. In addition, it is noteworthy that both models successfully produce accurate results for older males and females age 50 and above as the forecast errors fluctuate approximately around zero in all countries.

Table 7
*The out-sample Root Mean Squared Forecast Errors (*RMSFEs*) of mortality rates for males and females in four developing countries using two methods (*LC=*Lee-Carter and* FD= *Functional Data)*

| Country | Males | | Females | |
|---|---|---|---|---|
| | LC | FD | LC | FD |
| Malaysia | 0.15256 | 0.12681 | 0.20340 | 0.23593 |
| Indonesia | 0.16304 | 0.05492 | 0.24213 | 0.06628 |
| Thailand | 0.21023 | 0.13055 | 0.13359 | 0.08826 |
| Singapore | 0.21797 | 0.21599 | 0.19623 | 0.22453 |
| Overall | 0.18595 | **0.13207** | 0.19383 | **0.15377** |

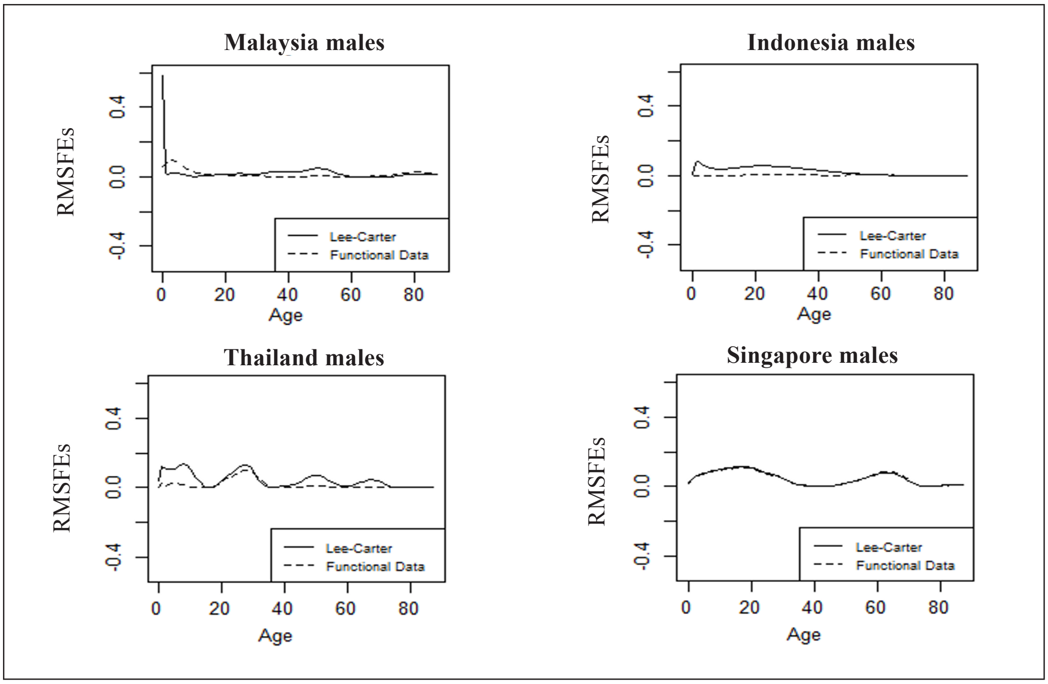


*Figure 4*. Root Mean Square Forecast Errors (RMSFEs) of mortality rates by age and methods: Lee-Carter (—–) and Functional data (----), for males in four developing countries

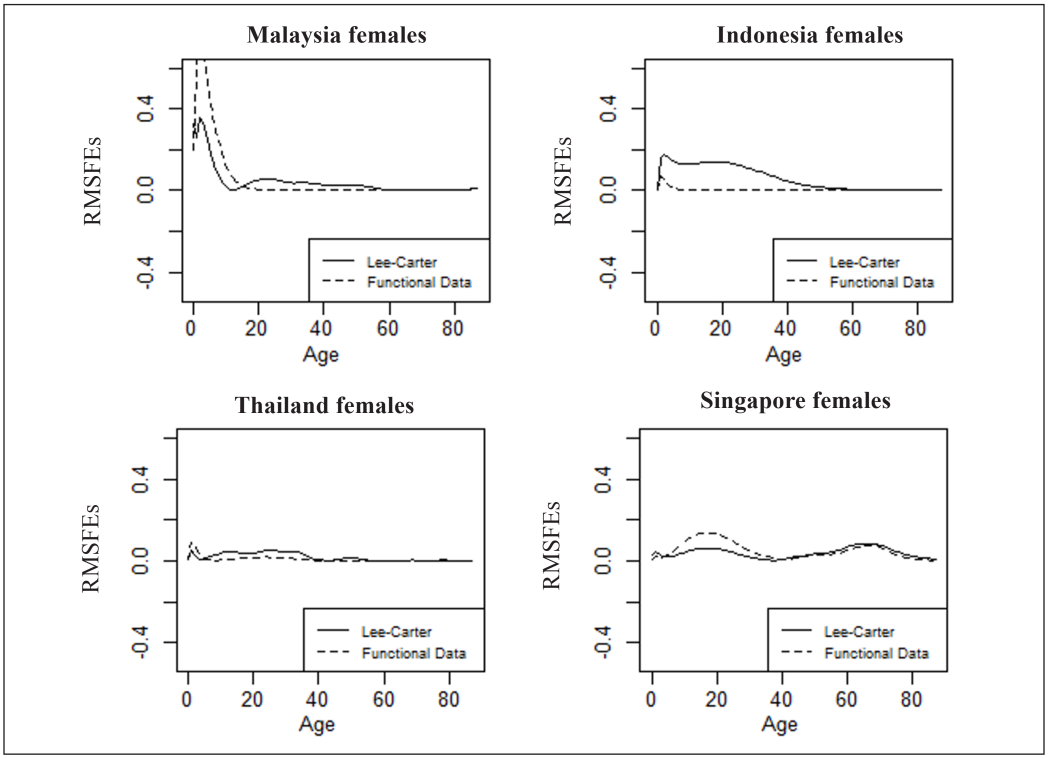


*Figure 5.* Root Mean Square Forecast Errors (RMSFEs) of mortality rates by age and methods: Lee-Carter (——) and Functional data (----), for females in four developing countries

In addition to mortality rates, we include another outcome measure which is life expectancy at births, for models' evaluation. The Table 8 summarizes the out-sample forecast errors of life expectancy at births segregated by genders and countries. Promising results can be concluded from the Table 8 in which the functional data model significantly out-performs the Lee-Carter model for females in all countries and provides better forecasts for males in three out of four countries, Malaysia, Indonesia and Thailand. For example, as we can see from the Table 8, the functional model substantially reduces the forecast error of Indonesian males from 1.00015 to 0.38785 (by 61%) and Indonesian females from 1.23932 to 0.22426 (by 82%).

The superior performance of the functional data model to forecast life expectancy at births arises for several reasons. Firstly, the model allows more complex dynamics than the original Lee-Carter model by setting $K>1$, thus allowing higher order terms to be included for estimation. Secondly, the additional statistical procedures adopted in the functional model such as smoothing technique allows the observational error to be treated separately from the time series forecast hence reducing the forecast errors. Thirdly, the use of robust method minimizing the risk of outlying years or data (Hyndman et al., 2007).

Table 8
*The out-sample Mean Absolute Forecast Errors (*MAFEs*) of Life expectancy at births for males and females in four developing countries using two methods (*LC=*Lee-Carter and* FD= *Functional Data)*

| Country | Males | | Females | |
|---|---|---|---|---|
| | LC | FD | LC | FD |
| Malaysia | 0.76023 | 0.59970 | 0.42874 | 0.35785 |
| Indonesia | 1.00015 | 0.38785 | 1.23932 | 0.22426 |
| Thailand | 0.90254 | 0.24192 | 0.35301 | 0.11989 |
| Singapore | 0.66158 | 0.81638 | 1.62805 | 1.17126 |
| Overall | 0.83112 | **0.51143** | 0.91228 | **0.46831** |

It is noteworthy that the out-sample forecast errors are not consistent between mortality and life expectancy at birth. For example, although the functional data model is less accurate than that of the Lee-Carter model to forecast the mortality of Malaysian and Singaporean females (Table 7), the model performs better than the Lee-Carter model to forecast life expectancy at births for these sub-populations (Table 8). This issue was highlighted previously with evidence that the most accurate model for mortality is not essentially the best model for life expectancy, vice versa (Booth et al., 2005).

The life expectancy at birth forecast values from different models and its respective observations are presented in the Figure 6. It can be seen clearly from the figure that both models successfully provide accurate results for Malaysian females as almost all forecasted values lie closely to the observation lines. There is no major structural change in the Malaysian females' life expectancy in the past hence simplifying predictions. In addition, the Lee-Carter model significantly overestimates the life expectancy of Indonesian males and females such that the Lee-Carter model forecast values (dashed lines) lie above the observations (full lines). The significant overestimation of the Lee-Carter model for Indonesian life expectancies maybe due to the model does not impose weight on the structural changes in the life expectancy that have occurred just before the forecast period, after 1990. On the contrary, the functional data model provides better forecasts than the Lee-Carter model for Indonesian females also for Thailand males and females. These results can be seen clearly from the Figure 6 that the forecast values from the functional model (dotted lines) lie closer to observations (full lines) than the forecast values from the Lee-Carter model (dashed lines).

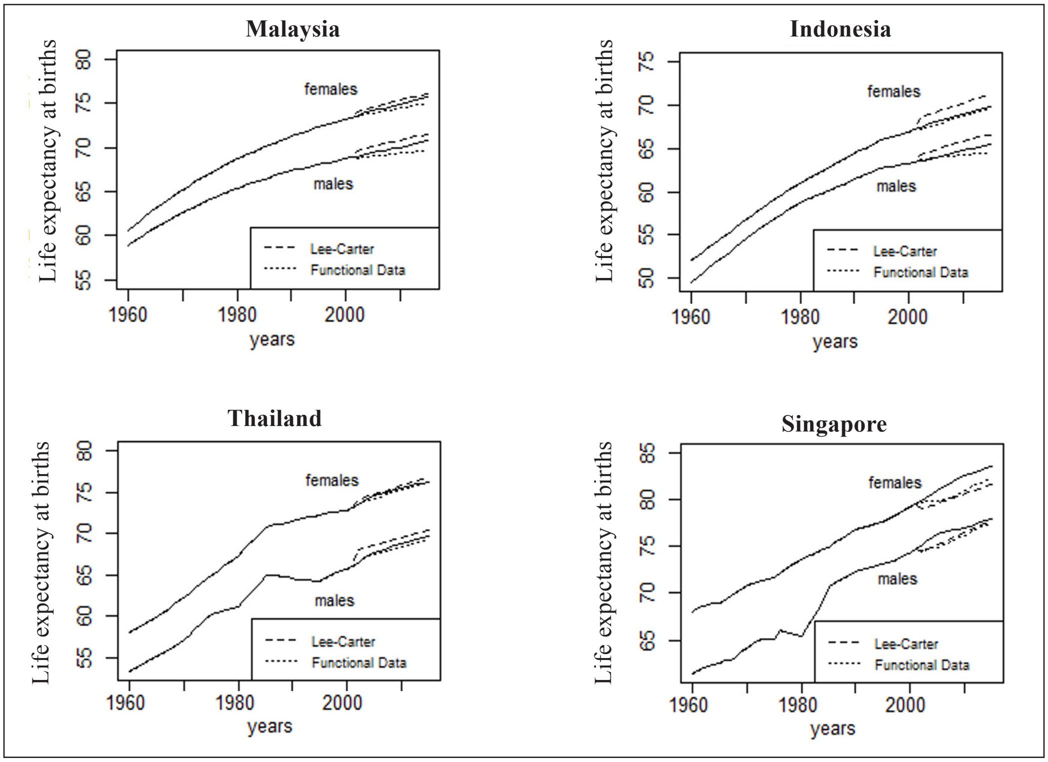


*Figure 6*. The observed and 14-year forecasts of life expectancy at births for males and females in four developing countries

## CONCLUSION

This research extends the application of two mortality forecasting models: the Lee-Carter model and its extended version, the functional data model to four developing countries including Malaysia, Indonesia, Thailand and Singapore. In overall, the out-sample mortality forecast values show that the functional model is more accurate than that of the original Lee-Carter model for males and females. Moreover, the evaluation both models using life expectancy at birth as the outcome measure indicated that the functional model significantly out-performed the Lee-Carter model for females in all countries and males in three out of four countries. It can be concluded that the functional model has the potential to accurately forecast age and gender specific mortality rates and the life expectancies of developing countries provided there is no major structural changes in recent years. It is recommended for future works to include other Lee-Carter extensions to determine most accurate model for developing countries.

## REFERENCES


Bell, W., & Monsell, B. (1991). Using principal components in time series modelling and forecasting of age specific mortality rates. In *American Statistical Association, 1991 proceedings of the Social Statistics Section* (pp. 154-159). Alexandria, Virginia: American Statistical Association.

Booth, H., Tickle, L., & Smith, L. (2005). Evaluation of the variants of Lee-Carter method of forecasting mortality: A multi-country comparison. *New Zealand Population Review, 31*(1), 13-34.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., & Nesbitt, C. J. (1997). Actuarial Mathematics (2nd Ed.). In *Survival Distribution and Life Tables* (p.51-83). Schaumburg: The Society of Actuaries.

Dokumentov, A., & Hyndman, R. J. (2013). *Two-dimensional smoothing of mortality rates* (Working paper 26/13). Department of Econometrics and Business Statistics, Monash University. Retrieved April 11, 2017, from http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Erbas, B., Akram, M., Gertig, D. M., English, D., Hopper, J. L., Kavanagh, A. M., & Hyndman, R. J. (2010). Using functional data analysis models to estimate future time trends in age-specific breast cancer mortality for the United States and England–Wales. *Journal of Epidemiology, 20*(2), 159-165.

Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis, 51*(10), 111-126.

Hyndman, R. J., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting, 24*(3), 323-342.

Hyndman, R.J. (2013). *'Demography': forecasting mortality, fertility, migration and population data, R Package version 1.14*. Retrieved March 12, 2017, from http://robjhyndman.com/software/

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. New York: Prentice Hall.

Husin, W. Z. W., Zainol, M. S., & Ramli, N. M. (2015). Stochastic models in forecasting Malaysian mortality: The lee-carter model and its extension. *Advanced Science Letters, 21*(6), 1850-1853.

Lazim, M. A. (2012). *Introductory business forecasting: a practical approach* (3rd Ed.). Shah Alam, Selangor: Penerbitan UiTM.

Lee, R. D., & Carter, L. (1992). Modeling and forecasting the time series of U.S. Mortality. *Journal of the American Statistical Association, 87*(419), 659-671.

McNown, R., & Rogers, A. (1989). Forecasting mortality: a parameterized time series approach. *Demography, 26*(4), 645-660.

McNown, R., & Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting, 8*(3), 413-432.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). Introduction to Functional Data Analysis. In J. O. Ramsay, G. Hooker & S. Graves (Eds.), *Functional Data Analysis with R and MATLAB* (p.1-19). New York: Springer-Verlag.

Rutstein, S. O. (2000). Factors associated with trends in infant and child mortality in developing countries during the 1990s. *Bulletin of the World Health Organization, 78*(10), 1256-1270.

Shaadan, N., Jemain, A. A., & Deni, S. M. (2014, July). Data preparation for functional data analysis of PM10 in Peninsular Malaysia. In *AIP Conference Proceedings* (Vol. 1605, No. 1, pp. 850-855). New York: AIP Publishing.

Shang, H. L., Booth, H., & Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research, 25*, 173-214.

Shair, S., Purcal, S., & Parr, N. (2017). Evaluating extensions to coherent mortality forecasting models. *Risks, 5*(1), 16-35.

Tan, J. (2017, January 23). The world's best health care? How Singapore's hybrid model is the envy of other countries. *The National*. Retrieved January 5, 2017, from https://www.thenational.ae/world/asia/

United Nations. (2015). *World Population Prospects: The 2015 Revision*. Department of Economic and Social Affairs, Population Division. Retrieved March 10, 2017, from https://population.un.org/wpp/ Download/Standard/Population/

Villegas, A., & Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: an application to deprivation and mortality in England. *North American Actuarial Journal, 18*(1), 168-193.