Common Factor Model with Multiple Trends for Forecasting Short Term Mortality

Wan Zakiyatussariroh Wan Husin, Mohammad Said Zainol and Norazan Mohamed Ramli

***Abstract*— This article describes a common factor model to forecast mortality rates. The proposed model is an extension Lee-Carter state space (LC-SS) model by incorporating multiple common trends through application dynamic factor analysis (DFA) to reduce the dimension of the observed mortality rates in term of common trends. The original, the LC-SS model is formalized the Lee-Carter (LC) model as a statistical model accounting for all source of variability. The proposed model is actually the LC-SS incorporating DFA and being termed as LC-DFA model. The LC-DFA model is designed specifically for analyzing short and non-stationary mortality series. As in LC-SS, the parameters in the proposed LC-DFA model are estimated by maximum likelihood estimation (MLE) through an expectation-maximization (EM) algorithm. The mortality data of Peninsular Malaysia for years 1980 to 2009 were used to illustrate the performance of the proposed model. The data were split according to gender and separate LC-DFA models were each fitted for the males and female population. The LC-DFA performance in terms of the accuracy of prediction based on in-sample fitting and out-of-sample forecasts of the LC-DFA was then evaluated by comparing with the LC-SS and LC models. The most efficient forecasting model was based on lowest values of root mean square error (RMSE) and mean absolute percentage error (MAPE). The results revealed that the proposed LC-DFA model performs the best.**

***Index Terms*—Dynamic factor analysis, Expectation-maximum algorithm, Lee-Carter model, mortality, state space model.**

I. Introduction

DIMENSION reduction techniques such as principal component analysis (PCA), factor analysis (FA) or correspondence analysis (CA) have been applied by researchers to analyze sets of data that contain relatively large number of response variables. The main purpose of dimension reduction is to simplify the data without losing relevant information in the data sets. This requires that the simplified structure explains most of the variability in the data. The basic technique of data reduction in multivariate analysis is PCA. However the PCA technique does not account for temporal variation. Dynamic factor analysis (DFA) (also known as dynamic factor model-DFM) is a dimension-reduction technique that models N observed non-stationary time series in term of M common trends. The principle of DFA is the same as other dimension reduction techniques. However, DFA is designed for time series. Although it is possible to apply PCA to time series data it does not take account of time in any way. Even though it can connect consecutive points in time with each other, interpretation of the results is likely to be difficult [1].

In 1992, Lee and Carter proposed a model which used PCA technique in modeling and forecasting age-specific mortality [2]. The LC model assumes that the log-death rates time series shares one common trend that explained by the first principal component term which represented by mortality index. This first term of principal component can be estimated by the singular value decomposition (SVD). Basically, the LC model is based on a log-bilinear form for age-specific mortality involving two equations. The parameters of these two equations are estimated separately where the first equation is computed from SVD in order to extract the principal component term (the mortality index), while, the second equation is modeled the mortality index using time series methods. The strength of the LC method is in its simplicity and robustness in the context of linear trends in age-specific death rates (ASDR) [3]. The model became the leading stochastic model in the actuarial and demographic literature and was used as a benchmark model in most academic researches and practical applications of mortality forecasting [4]-[7].

The LC method works very well for most of the countries, but, not for some countries. Therefore, the LC model has undergone various extensions and modifications exemplified in the works of [8]-[12]. Most of these extensions used PCA in extracting the first component or single common trend explained by the mortality index. A exception is the work done by [11] that included the incorporation of second and higher order terms into the LC model to cater for the additional component that are not explained by the first component. Apart from this, several extended the concept of LC model by considering multiple PCA components using dynamic factor model (DFM) in forecasting mortality [13], [14].

The extension of LC model also involves a reformulation of the model as state space model as in [15]-[18]. The main reason why a state space formulation of the LC model was suggested is due to the fact that errors of the LC equations were estimated separately. The first equation is estimated by a combination of SVD while the second as a time series model. Reference [15] highlighted the fact that the

Manuscript received December 24, 2015; revised December 31, 2015. This work was supported in part by the Ministry of Education, Malaysia and Universiti Teknologi MARA, Malaysia under Fundamental Research Grant Scheme (FRGS) (FRGS/1/2014/ST06/UiTM/02/5).

W. H. W. Zakiyatussariroh is with the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia (corresponding author: phone: +6019-3409511; fax: +603-55442000; e-mail: wanzh76@gmail.com).

M.S. Zainol, is with the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia (e-mail: saidzainol@gmail.com).

M.R. Norazan is with the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam,Selangor, Malaysia (e-mail: norazan@tmsk.uitm.edu.my).

prediction of the LC model only accounts for the errors of time series model while ignoring the errors in estimating the parameters and the variance of the error term in the first equation. In state space formulation, all the parameters in the LC model were estimated simultaneously. However, most of the existing LC in state space model retained the assumption of one common trend of LC except in the work done by [16] that considered multiple common trends using DFA. DFA model is common factors model formulated in state space framework which plays a role in dimension-reduction technique by modeling N observed non-stationary time series in term of M common patterns. These patterns may represent common trends, seasonal effects or common cycle.

Most developments of mortality model for forecasting by previous researchers concentrate on methods used on long data sets has been available in developed countries, with only scanty mentions of methods suitable for the shorter data set for developing countries.

Therefore, this study proposed to adopt DFA model with multiple common trends to cater the cases of mortality data containing large number of response variables pertaining to age groups within relatively short length of the time series data. This study focuses on the development a DFA model with EM algorithm for its parameter estimation adopted from [19] and [1] by extending the works done by [17], [18].

II. Lee-Carter Model

Let be $m_{x,t}$ the mortality rates for a group of x ages in year t with $x = 1, \dots\dots, N_{(Age)}$ and $t = 1, \dots\dots, T_{(Year)}$. LC model analyzes the linear relationship between the logarithm of original $m_{x,t}$ and two factors, namely are age x and year t. The model is represented as

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

where α_x is the age pattern of log mortality rates averaged across year; β_x is the first principal component reflecting relative change in the $log(m_{x,t})$ at each age; k_t is the first set of principal score by year t representing mortality index that measures the general level of the $log(m_{x,t})$ and $\varepsilon_{x,t}$ is the error term, assumed homoskedastic. The model was estimated using SVD with two constrains to ensure it is identifiable. The two constrains are $\sum_{t=1}^{T} k_t = 0$ and $\sum_{x=1}^{N} \beta_x = 1$. In addition, the LC method adjusts k_t by refitting it to the total number of deaths. The purpose of this adjustment is to give more weight to high rates [2]. The adjustment is as follows:

$$d_t = \sum_{x=1}^{A} [p_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x \tilde{k}_t)],$$

where d_t is the number of deaths for a group of ages x, $p_{x,t}$ is the average number of people living for a group of x ages in year t and $\tilde{k}_t$ is the adjusted estimated mortality index. The adjusted estimated k_t is then extrapolated using autoregressive integrated moving average (ARIMA) method [20],[21], specifically, ARIMA (0,1,0) as in the original paper of [2].

The ARIMA (0,1,0) which represent a random walk with drift model is expressed as follow,

$$k_t = k_{t-1} + \theta + u_t \qquad t = 1, \dots\dots, T, \tag{2}$$

where θ is a drift parameter representing the constant annual change in the series of k_t and u_t is the error term. The procedures of the LC method are summarizing as follow.

1) Estimate α_x, β_x and k_t using historical age specific mortality rates.
2) Adjust the estimated k_t to ensure equality between the observed and estimated number of deaths in a certain period.
3) Extrapolate the series of adjusted k_t using ARIMA.
4) Forecast log ASDR using extrapolated adjusted k_t with fixed values of estimated α_x and β_x. Here, the forecasted values of adjusted k_t and the estimated α_x and β_x are substituted into (1), then convey back the estimated $log(m_{x,t})$ to the original scale in order to get forecast value for the ASDR. Thus, the h-step forecast of $m_{x,T+h}$ is:

$$\hat{m}_{x,T+1} = \exp(\hat{\alpha}_x + \hat{\beta}_x \tilde{k}_{T+h}).$$

III. Lee-Carter as Common Factor Model

A. *Lee-Carter State Space Model*

In order to improve the quality of the forecast, the LC model was set up in a standard class of stochastic model based on the concept of common factor model. Previously, the LC model was set up as a state space model (SSM) as in [15-18]. The model contains N observations for N age groups and a single state equation that explains the dynamics of the unobservable variable. Let $\boldsymbol{m}_t$ be the vector of N log ASDR for year t, that is $\boldsymbol{m}_t = (m_{1t}, m_{2t}, \dots, m_{Nt})'$ where m_{it} is the value of the i[th] ASDR at time t ($i = 1, \dots, N$ and $t = 1, \dots, T$). Thus, the LC model as a SSM is represented by

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{\alpha} + \boldsymbol{\beta} k_t + \boldsymbol{\varepsilon}_t \qquad && \boldsymbol{\varepsilon}_t \sim MVN(\boldsymbol{0}, \boldsymbol{R}) \\ k_t &= k_{t-1} + \theta + u_t && u_t \sim N(0, q) \end{aligned} \tag{3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ and u_t are error terms that are assumed to be independent. A univariate random walk with drift model is assumed for the state vector. The LC in SSM (hereafter LC-SS model) model is the simplest common factor model which contains only one common trend. This model assumes all the age groups follow the same pattern by one common trend. It also assumes the error for the state u_t is independent, identically and normally distributed, with zero mean and variance q. The LC-SS model for the log ASDR provides a joint distribution for the N age groups at any given time and allows us to estimate all the parameters simultaneously. It assumes that the error is independent across time where $\boldsymbol{\varepsilon}_t$ are independent identically and normally distributed with $N \times 1$ variance vector $\boldsymbol{R}$. The errors $\boldsymbol{\varepsilon}_t$ and u_t are uncorrelated.

B. Extended Lee-Carter State Space Model into Dynamic Factor Analysis Model

The LC-SS model (3) with one common trend has been further extended with multiple common trends. The LC-SS model with multiple common trends is similar to common trend model or also known as DFA model. The LC-SS model is the simplest DFA model which contains only one common trend. The model with one common trend assumes that all the age groups follow the same pattern. The proposed model here is taking account of the dynamic structure of mortality data by extending the LC-SS using DFA model which is also based on state space formulation. In this case, the LC-SS model is considered as a special case of DFA model with the number of common trends being 1, i.e. $M = 1$. The mathematical representation for this extended LC-SS through DFA model (hereafter LC-DFA model) with M commons trends together with a level parameter and a noise component is

$$m_{xt} = \beta_{x1}k_{1t} + \beta_{x2}k_{2t} + \cdots + \beta_{xM}k_{Mt} + \alpha_x + \varepsilon_{xt}, \qquad (4)$$

where $\varepsilon_{xt} \sim MVN(0, R)$, m_{xt} is the value of x^{th} log ASDR of the 17 age group series at time t, k_{jt} is the j$^{\text{th}}$ common trend, β_{xi} is the factor loading, α_x constant level parameter and ε_{xt} is noise. In matrix notation, the model is as

$$\boldsymbol{m_t} = \boldsymbol{\beta k_t} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon_t} \text{ where } \boldsymbol{\varepsilon_t} \sim MVN(0, \boldsymbol{R}), \qquad (5)$$

where $\boldsymbol{m_t}$ is a $N \times 1$ vector containing the values of the 17 logged death rates series for at time t, $\boldsymbol{k_t}$ represent the values of M common trends at time t, $\boldsymbol{\alpha}$ is a $N \times 1$ vector of constant level parameter that allow each linear combination of common trends to move up or down, and $\boldsymbol{\varepsilon_t}$ is a $N \times 1$ noise component which is assumed to be normally distributed with mean 0 and covariance matrix $\boldsymbol{R}$. The $N \times 1$ vector $\boldsymbol{\beta}$ contains the values of factor loading which determines the linear combinations of common trends. The factor loading identifies which common trends are important to a particular age groups and which set of age groups are related to the same common trend. The trends represent the underlying common patterns over time.

The common trend is modeled as a multivariate random walk with drift, i.e.,

$$\boldsymbol{k_t} = \boldsymbol{c} + \boldsymbol{k_{t-1}} + \boldsymbol{u_t} \qquad (6)$$

where $\boldsymbol{u_t} \sim MVN(0, \boldsymbol{Q})$, $\boldsymbol{c}$ **is** a drift parameter and $\boldsymbol{u_t}$ is the error vector that assumed to be normally distributed with mean 0 and diagonal covariance matrix $\boldsymbol{Q}$. The error $\boldsymbol{u_t}$ is independent of $\boldsymbol{\varepsilon_t}$. Hence, the j$^{\text{th}}$ trend at time t is equal to the j$^{\text{th}}$ trend at time $(t-1)$ plus a contribution of the noise component. If the diagonal element of $\boldsymbol{Q}$ is relatively small, then the contribution of the error component is likely to be small for all t and the j$^{\text{th}}$ trend will be a small curve. If it is large, then the j$^{\text{th}}$ trend will show more variation. Hence the trends are smoothing functions over time and are independent of each other. To complete the LC-DFA model, the initial condition of the state is assumed to be normally distributed with mean $\boldsymbol{a_0}$ and variance $\boldsymbol{P_0}$ such that $\boldsymbol{k}_0 \sim MVN(\boldsymbol{a_0}, \boldsymbol{P_0})$.

C. Parameter Estimation for LC-DFA Model

In this study, the LC-DFA model in (5) takes account of the dynamic structure of mortality data using a model based on a SSM formulation with the unknown parameters denoted as $\boldsymbol{\delta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{R}, \mathbf{Q}, \boldsymbol{\theta}, \boldsymbol{a}_0 \text{ and } \boldsymbol{P}_0\}$**.** These parameters are estimated using MLE under the assumption that the initial state is normal that is $k_0 \sim MVN(a_0, \boldsymbol{P}_0)$. The joint log-likelihood function of the observations $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots\ldots, \boldsymbol{m}_T$ and the trend component $\boldsymbol{k}_0, \boldsymbol{k}_1, \ldots\ldots, \boldsymbol{k}_T$ are given as follows:

$$\begin{aligned}
&\log L(\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots\ldots, \boldsymbol{m}_T, \boldsymbol{k}_0, \boldsymbol{k}_1, \ldots\ldots, \boldsymbol{k}_T) = \\
&-\frac{1}{2}\log|\boldsymbol{P_0}| - \frac{1}{2}(\boldsymbol{k}_0 - \boldsymbol{a_0})'\boldsymbol{P_0}^{-1}(\boldsymbol{k}_0 - \boldsymbol{a_0}) - \frac{\mathrm{T}}{2}\log|\boldsymbol{Q}| \\
&-\frac{1}{2}\sum_{t=1}^{\mathrm{T}}(\boldsymbol{k}_t - \boldsymbol{k}_{t-1} - \boldsymbol{\theta})'\boldsymbol{Q}^{-1}(\boldsymbol{k}_t - \boldsymbol{k}_{t-1} - \boldsymbol{\theta}) \\
&-\frac{\mathrm{T}}{2}\log|\boldsymbol{R}| + \text{constant} \\
&-\frac{1}{2}\sum_{t=1}^{\mathrm{T}}(\boldsymbol{m}_t - \boldsymbol{\beta k}_t - \boldsymbol{\alpha})'\boldsymbol{R}^{-1}(\boldsymbol{m}_t - \boldsymbol{\beta k}_t - \boldsymbol{\alpha})
\end{aligned}$$

This likelihood function is maximized numerically using the method of numerical maximizing, the EM algorithm as in [17] and [18] which originally adopted from [1], [19]. The EM algorithm is an iterative method for finding the MLEs of all parameters in (5) and (6). This is done by successively maximizing the conditional expectation of the likelihood function. The EM algorithm involves procedure with expectation step (*E*-step) and maximization step (*M*-step). Starting with an initial set of parameter that denoted $\widehat{\boldsymbol{\delta}}_1$, and updated parameter set $\widehat{\boldsymbol{\delta}}_2$ is obtained by finding the $\widehat{\boldsymbol{\delta}}_2$ which maximizes the expected value of the likelihood over the distribution of the state condition on $\widehat{\boldsymbol{\delta}}_1$. Then, using $\widehat{\boldsymbol{\delta}}_2$, an updated parameter set $\widehat{\boldsymbol{\delta}}_3$ is calculated. These steps are repeated until the expected log-likelihood stops increasing. Then, the varimax rotation is applied to the factor loading after which an inverse factor rotation is applied to the common trends and the corresponding covariance matrix is modified. The overall procedure has been regarded as simply alternating between the Kalman filtering (KF) and smoothing (KS) recursions. The following are the steps involved in maximizing the likelihood via EM algorithm. It was maximized using the KF [22] and KS [23]. The EM algorithm involves the following steps and detail can be found in [24], [25].

1) Set an initial estimates of parameters, on iteration *j*, (*j* = 0,1,2,…..):
2) Compute the incomplete-data likelihood and perform the *E-Step*. Based on initial parameters $\boldsymbol{\delta^j}$, the expected values of k_t conditioned on all the observed data m_1^T are calculated.
3) Perform the *M-Step*. A new set $\boldsymbol{\delta^{j+1}}$ was computed by finding the parameters that maximize the expected log-likelihood function with respect to $\boldsymbol{\delta}$.

4) Repeat step 2 and 3 for convergence. New expectations are computed using $\boldsymbol{\delta}^{j}$, then a new set of parameters $\boldsymbol{\delta}^{j+1}$ is generated. This process is continued until the log likelihood stops increasing at a specified tolerance level.
5) Apply a varimax rotation to the factor loadings and apply an inverse factor rotation to the common trends and modify the corresponding covariance matrix.

Here, a number of models have been estimated with different number of common trend with a diagonal and non-diagonal error covariance matrix $\boldsymbol{R}$. To decide which model has the best fit, corrected Akaike's information criterion (AICc) is was applied. The DFA model with the smallest AICc value is chosen.

D. Kalman Filter and Smoother

The overall procedures involved in EM may be regarded as simply alternating between the KF and KS recursions. LC-DFA forecasts the observables using KF recursion based on the MLEs that were obtained from the recursive procedure. The unknown parameters in the model have been estimated using KF algorithm that solves for the expected value of the hidden state $(\boldsymbol{k_t})$ at time *t* conditioned on the observed data up to time *t*. The KF gives the optimal (lowest mean square error) estimates of the unobserved $\boldsymbol{k_t}$ based on the observed data up to time *t*. While, the KS (solves the expected value of the hidden state conditioned on all the data. The estimators from the KF and KS yielded the maximum-likelihood estimates.

Let $\boldsymbol{k}_{t|s} = E(\boldsymbol{k}_t|\boldsymbol{m}_s)$ and $\boldsymbol{P}_{t|s} = E\{(\boldsymbol{k}_t - \boldsymbol{k}_{t|s})(\boldsymbol{k}_t - \boldsymbol{k}_{t|s})'\}$. The KF algorithm for LC-DFA model in (5) is started with initial conditions $\boldsymbol{k}_{0|0}$ and $\boldsymbol{P}_{0|0}$, for $t = 1, \dots, T$ that used together with the state equation to compute

$$\boldsymbol{k}_{t|t-1} = \boldsymbol{k}_{t-1|t-1}$$
$$\boldsymbol{P}_{t|t-1} = \boldsymbol{P}_{t-1|t-1} + \boldsymbol{Q}.$$

Then, having observed the mortality data $\boldsymbol{m}_t$, the predictions $\boldsymbol{k}_{t|t-1}$ and $\boldsymbol{P}_{t|t-1}$ have been updated (correct) as

$$\boldsymbol{k}_{t|t} = \boldsymbol{k}_{t|t-1} + \boldsymbol{C}_t(\boldsymbol{m}_t - \boldsymbol{\beta}_t\boldsymbol{k}_t)$$
$$\boldsymbol{P}_{t|t} = [I - \boldsymbol{C}_t\boldsymbol{\beta}_t]\boldsymbol{P}_{t|t-1},$$

where $\boldsymbol{C}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{\beta}_t'[\boldsymbol{\beta}_t\boldsymbol{P}_{t|t-1}\boldsymbol{\beta}_t' + \boldsymbol{R}]^{-1}$ is called the Kalman gain, while the algorithm of the KS with initial conditions $\boldsymbol{k}_{T|T}$ and $\boldsymbol{P}_{T|T}$that are obtained via KF for $t = T, T-1, \dots, 1$,

$$\boldsymbol{k}_{t-1|T} = \boldsymbol{k}_{t-1|t-1} + \boldsymbol{J}_{t-1}(\boldsymbol{k}_{t|T} - \boldsymbol{k}_{t|t-1})$$
$$\boldsymbol{P}_{t-1|T} = \boldsymbol{P}_{t-1|t-1} - \boldsymbol{J}_{t-1}(\boldsymbol{P}_{t|T} - \boldsymbol{P}_{t|t-1})\boldsymbol{J}_{t-1}'$$

where $\boldsymbol{J}_{t-1} = \boldsymbol{P}_{t-1|t-1}\boldsymbol{G}'\boldsymbol{P}_{t|t-1}{}^{-1}$.

E. Forecasting with Lee-Carter State Space

LC-SS model forecasts as a standard state space model based on the MLEs obtained from the recursive procedure. In particular, the final state predictor $k_{T|T}$ implied by the MLEs together with the observation and state equation construct $\boldsymbol{m}_{T+h|T}$ for $h = 1, 2, \dots.$ according to

$$\boldsymbol{m}_{T+h|T} = \boldsymbol{\alpha} + \boldsymbol{\beta}^h\boldsymbol{k}_{T|T}.$$

IV. Application

Data set for Peninsular Malaysia's male and female all-cause mortality was used to demonstrate the use LC-DFA model in forecasting mortality specifically for the case of mortality with short time series. The data were provided by the Department of Statistics, Malaysia (DOSM) and consist of annual number of deaths and populations for 17 age groups for years 1980 to 2009. The age groups are 0 - 4, 5 - 9, ..., 75 - 79, 80+. Deaths and population with unknown age groups are not included in the analysis. Since, mortality was measured by age-specific death rates (ASDR), then the ASDR in a single calendar year were calculated based on $m_{x,t} = \left(\frac{d_{x,t}}{p_{x,t}}\right)$ where $d_{x,t}$ is the number of deaths for a group of x ages in year t and $p_{x,t}$ is the observed population for a group of x ages in year t. The observed population used the mid-year population for a group of x ages. To illustrate the methodology and to assess the performance of the models, out-of-sample for the last four years of the data was performed. That is, the data from years 1980 to 2005 were used to fit the models and then forecasts were made for years 2006 to 2009. The forecasts values were then compared to the observed values.

Time plots of the ASDR over time for male population and female population for year 1980 to 2009 are presented as Figure 1 and Fig. 2 respectively. The fluctuations in the ASDR over this period reflect the fact that mortality has decreased considerably in almost all age groups during the past 30 years and is much lower for females than for males. The decline in the female ASDR is steadier than those of the males. The males ASDR appears much noisier than for females. These data were then transformed to the logarithm (natural logarithm) because of the exponential nature in ASDR trend. In addition, it is necessary to transform the raw data by taking logarithms in order to stabilize the high variance associated with high age-specific rates.

In this study, the ASDR were fitted and forecasted using the original LC, LC-SS model with single common trend and LC-SS model with multiple common trends (LC-DFA). In estimating the LC-SS and LC-DFA, the KF with initial mean 0 and variance 5 for initial setting of the state vector $\boldsymbol{k}_0$ was used; following [17, 18]. The discussion of the results first describes the LC-DFA in estimating the underlying common trends in Malaysia's ASDR series. A total of 17 ASDR series by age group are modeled as a constant plus a linear combination of M common trends and noise term which is assumed to be normally distributed with mean 0 and two different assumptions on covariance matrix $\boldsymbol{R}$ which equal variance and unequal variance assumes.

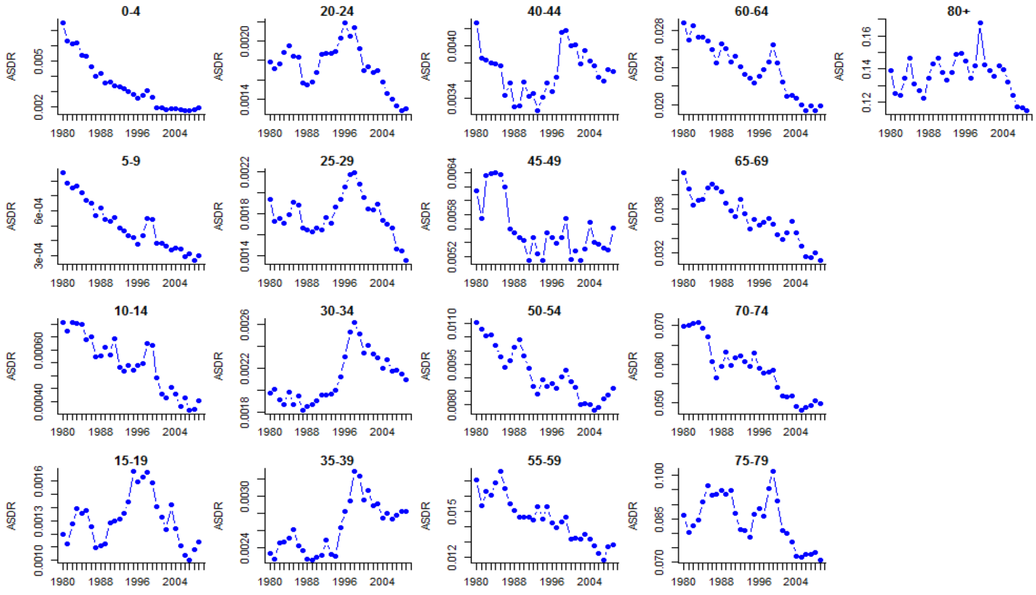

Fig. 1. Malaysian males ASDR trends.

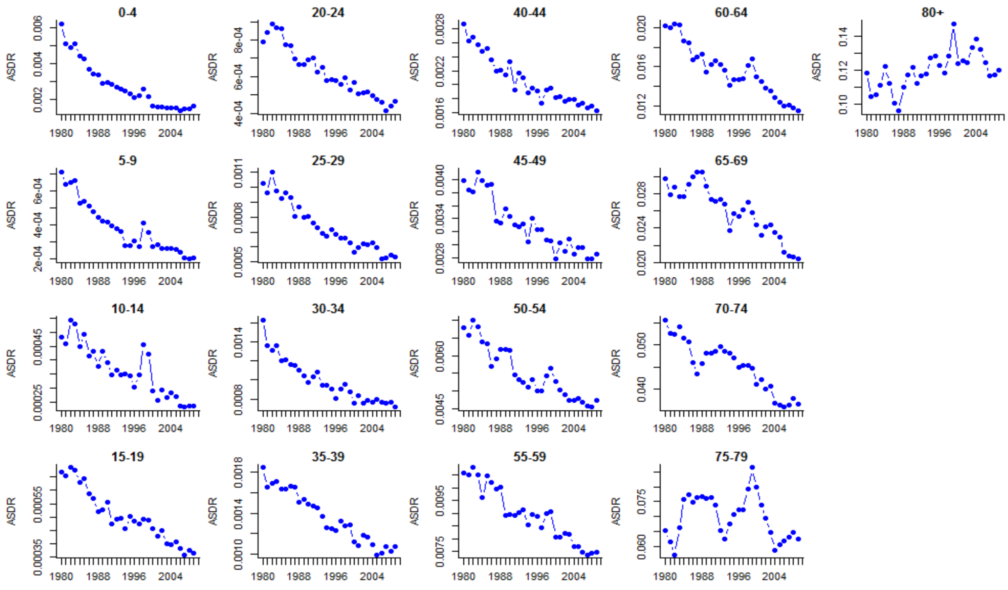

Fig. 2. Malaysian females ASDR trends

Two different assumptions on covariance matrix $\boldsymbol{R}$ were used. They are diagonal and equal covariance (equal variance) assumed for each age groups that were denoted as Model A1-M – A16-M for males and Model A1-F – A16-F for female, while, model with diagonal and unequal covariance (unequal variance) assumed for each age groups were denoted as Model B1-M – B16-M for males and B1-F – B16-F for females respectively. With different M values used, a total of 16 models were estimated for each of the two different assumptions on the variance.

Model selection results for all 16 possible models for the two error variance assumed for both males and females are shown in Table I and Table II. The results indicate that Model B5-M with M = 5 common trends with diagonal and unequal model variances was the best model for males and Model B3-F with M = 3 common trends with diagonal and unequal model variances was the best model for females. These best models were based on the lowest value of corrected Akaike information criteria (AICc).

The benefits of using the LC-DFA can be seen one compares the results with those from related mortality model; the LC model as in (1). Referring to Table I and Table II, Model A1-M and A1-F respectively is LC-DFA with M = 1 (LC-SS) with covariance matrix of observation error $\boldsymbol{R}$ is diagonal and equal is equivalent with the assumption of original LC model [6], [7]. In the LC-DFA, the DFA model is used in extracting the number of common trends and model parameters have been estimated within state space framework with EM algorithm. However, the original LC model used PCA in extracting one common factor and used ordinary least square (OLS) via SVD in estimating model parameters. Comparing the AICc values (see Table I and Table II) between A1-M (AICc = 1021.167) and B5-M (AICc = 600.419) for males and A1-F (AICc = 617.7207) and B3-F (AICc = 432.3728) for females respectively, clearly shows that LC-DFA models with multiple common trends (5 common trends for males and 3 common trends for females) are the best models, i.e., has lower AICc.

TABLE I
DYNAMIC FACTOR SELECTION RESULTS FOR EQUAL VARIANCE

	Males		Females	
M	Model	AICc	Model	AICc
1	**A1-M**	**1021.167**	**A1-F**	**617.7207**
2	A2-M	855.6871	A2-F	521.9101
3	A3-M	771.8541	A3-F	457.3038
4	A4-M	718.9853	A4-F	452.4413
5	A5-M	674.2541	A5-F	464.1094
6	A6-M	630.4468	A6-F	469.9247
7	A7-M	619.2778	A7-F	483.8946
8	A8-M	649.956	A8-F	504.6387
9	A9-M	681.4692	A9-F	530.6317
10	A10-M	719.6707	A10-F	558.5977
11	A11-M	763.9309	A11-F	590.047
12	A12-M	782.5936	A12-F	625.762
13	A13-M	807.1731	A13-F	643.1001
14	A14-M	826.8509	A14-F	663.3278
15	A15-M	842.2377	A15-F	678.8856
16	A16-M	852.7498	A16-F	689.4343

TABLE II
DYNAMIC FACTOR SELECTION RESULTS FOR UNEQUAL VARIANCE

	Males		Females	
M	Model	AICc	Model	AICc
1	B1-M	881.6153	B1-F	445.8471
2	B2-M	747.1274	B2-F	446.595
3	B3-M	650.5703	**B3-F**	**432.3728**
4	B4-M	630.2677	B4-F	455.398
5	**B5-M**	**600.419**	B5-F	454.2373
6	B6-M	603.7697	B6-F	490.1099
7	B7-M	635.5987	B7-F	515.9922
8	B8-M	677.0028	B8-F	549.9217
9	B9-M	716.4836	B9-F	587.9814
10	B10-M	755.8986	B10-F	630.7168
11	B11-M	798.5131	B11-F	666.0875
12	B12-M	826.2154	B12-F	707.9504
13	B13-M	852.4364	B13-F	735.1113
14	B14-M	580.82	B14-F	759.1648
15	B15-M	898.531	B15-F	776.7707
16	B16-M	911.2616	B16-F	787.0268

As seen in Table I, Model A1-M is the worst-performing model with the largest AICc compared to other models in the case of males ASDR. However, for females, Model A1-F is not comparatively worse than other models.

The models were then evaluated based on the goodness of fit for in-sample fitting and out-sample forecast of the ASDR. Two different error measures were used; the root mean square error (RMSE) and mean absolute percentage error (MAPE). The evaluations of these models were carried out for each age group and overall performance for males and females respectively. In evaluating the performance in each group, these measures of accuracy were averaged over years, whereas, for overall performance, the measures were averaged over different ages and years.

Summary of the results of the goodness of fit based on in-sample fitting respectively for both males and females are presented in Table III and IV. It is obvious that the value of RMSE and MAPE for each age group for both males and females exhibit similar patterns. Among males, both measures indicate that the error for the LC-DFA model is smallest than the LC-SS and LC models in all the age groups. In the case of females, the LC-DFA most performed for almost all the age groups except for 10-14 where the MAPE value shows the LC is well fitted than LC-DFA and LC-SS and for age group 70-74 both measures show that the LC is most performed.

TABLE III
IN-SAMPLE EVALUATION FOR MALES

	LC		LC-SS		LC-DFA	
Age	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
0-4	0.00059	14.279	0.00031	6.680	**0.00009**	**1.944**
5-9	0.00005	9.019	0.00004	6.752	**0.00002**	**3.301**
10-14	**0.00002**	3.459	0.00005	6.083	**0.00002**	**3.181**
15-19	0.00018	10.593	0.00016	9.567	**0.00004**	**2.507**
20-24	0.00017	7.878	0.00018	8.267	**0.00003**	**1.177**
25-29	0.00016	7.323	0.00015	6.620	**0.00006**	**2.369**
30-34	0.00023	7.969	0.00016	5.835	**0.00005**	**1.724**
35-39	0.00031	9.129	0.00022	6.598	**0.00004**	**1.388**
40-44	0.00028	6.429	0.00028	6.284	**0.00010**	**2.232**
45-49	0.00029	4.170	0.00030	4.412	**0.00015**	**2.302**
50-54	0.00044	4.066	0.00038	3.277	**0.00017**	**1.460**
55-59	0.00062	3.478	0.00060	3.111	**0.00022**	**1.240**
60-64	0.00080	2.584	0.00119	3.751	**0.00053**	**1.739**
65-69	0.00163	3.744	0.00117	2.577	**0.00096**	**2.062**
70-74	0.00267	3.719	0.00273	3.572	**0.00140**	**1.955**
75-79	0.00611	6.040	0.00706	6.940	**0.00013**	**0.130**
80+	0.00981	5.272	0.00864	4.693	**0.00703**	**4.147**
Overall	0.00292	6.421	0.00282	5.590	**0.00176**	**2.050**

TABLE IV
IN-SAMPLE EVALUATION FOR FEMALES

	LC		LC-SS		LC-DFA	
Age	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
0-4	0.00032	9.008	0.00021	4.556	**0.00016**	**3.736**
5-9	0.00004	9.362	0.00003	6.228	**0.00002**	**5.012**
10-14	**0.00003**	7.053	0.00004	7.003	**0.00003**	**6.435**
15-19	0.00003	4.707	0.00002	3.351	**0.00002**	**2.908**
20-24	0.00005	6.207	0.00003	3.460	**0.00003**	**3.377**
25-29	0.00007	6.119	0.00004	4.016	**0.00003**	**2.833**
30-34	0.00008	6.58	0.00006	4.062	**0.00006**	**3.725**
35-39	0.00009	5.122	0.00006	3.462	**0.00004**	**2.345**
40-44	0.00012	4.837	0.00009	3.166	**0.00007**	**2.561**
45-49	0.00018	4.485	0.00014	3.603	**0.00013**	**3.356**
50-54	0.00024	3.876	0.00023	3.307	**0.00019**	**2.679**
55-59	0.00035	3.226	0.00028	2.665	**0.00024**	**2.224**
60-64	0.00066	3.402	0.00072	3.373	**0.00032**	**1.642**
65-69	0.00133	3.656	0.00123	3.480	**0.00020**	**0.614**
70-74	**0.00209**	**3.546**	0.00240	4.229	0.00227	4.053
75-79	0.00728	8.992	0.00723	8.945	**0.00439**	**5.302**
80+	0.00980	6.043	0.00781	4.928	**0.00665**	**4.594**
Overall	0.00303	5.66	0.00267	4.343	**0.00201**	**3.376**

Generally, based on this comparison, it can be concluded that the proposed LC-DFA is preferred than the LC-SS and LC in almost all of the age groups in fitting historical data, while, the performance of the LC-SS and LC models are almost identical. However, there is no clear pattern of over or underestimation from these three models. This indicates that the estimated ASDR produced by all models vary.

Overall results show that the LC-DFA is most performed than the LC-SS and the LC for both males and females with having smallest values in both measures. Additionally, these findings also revealed that in estimating the LC model, the state space framework through EM algorithm is performed better than classical LC model. This is based upon the observation, the overall performance of LC-SS and LC-DFA that performed well than original LC.

However, commonly, in time series forecasting, the performance of in-sample fit is different from the out-of-sample fit. The good forecasting model is the one that performs well in out-of-sample. Table V and VI present out sample evaluation for males and females respectively.

TABLE V
OUT-SAMPLE EVALUATION FOR MALES

	LC		LC-SS		LC-DFA	
Age	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
0-4	0.00096	51.480	0.00040	19.114	**0.00033**	**14.387**
5-9	0.00008	27.167	**0.00002**	**6.213**	**0.00002**	6.614
10-14	0.00006	13.448	**0.00003**	7.077	**0.00003**	**6.421**
15-19	0.00040	37.581	0.00037	34.984	**0.00008**	**6.367**
20-24	0.00039	29.488	0.00044	32.903	**0.00012**	**8.876**
25-29	0.00048	31.761	0.00046	30.725	**0.00025**	**15.399**
30-34	0.00052	23.814	0.00034	15.390	**0.00012**	**5.370**
35-39	0.00048	17.273	0.00028	9.966	**0.00004**	**1.192**
40-44	**0.00008**	**1.7000**	0.00013	3.248	0.00018	4.716
45-49	0.00069	12.452	0.00035	5.774	**0.00021**	**2.664**
50-54	0.00143	16.524	**0.00066**	6.624	0.00071	**7.249**
55-59	0.00105	7.102	0.00058	4.001	**0.00052**	**3.978**
60-64	0.00145	6.879	0.00105	4.870	**0.00047**	**1.941**
65-69	**0.00072**	**1.691**	0.00183	5.682	0.00234	7.341
70-74	0.00655	12.843	**0.00143**	**2.465**	0.00356	6.642
75-79	0.00397	5.386	0.00807	11.082	**0.00232**	**3.012**
80+	0.03026	25.486	0.02784	23.423	**0.01438**	**11.851**
Overall	0.00760	18.946	0.00707	13.149	**0.00369**	**6.707**

TABLE VI
OUT-SAMPLE EVALUATION FOR FEMALES

	LC		LC-SS		LC-DFA	
Age	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
0-4	0.00060	37.837	0.00031	16.275	**0.00026**	**13.943**
5-9	0.00005	22.331	**0.00001**	4.368	**0.00001**	**4.296**
10-14	0.00002	6.795	0.00002	8.907	**0.00001**	**3.169**
15-19	0.00003	8.620	**0.00001**	**2.485**	**0.00001**	2.851
20-24	0.00005	10.200	**0.00003**	**4.905**	0.00004	6.946
25-29	0.00007	12.853	**0.00002**	**3.922**	0.00003	4.980
30-34	0.00016	21.331	0.00007	8.856	**0.00005**	**5.838**
35-39	0.00015	13.307	**0.00006**	**4.880**	**0.00006**	5.289
40-44	0.000251	14.879	0.00009	5.350	**0.00005**	**2.832**
45-49	0.00031	10.720	0.00012	3.724	**0.00006**	**2.124**
50-54	0.00050	10.443	**0.00017**	**2.633**	0.00022	3.924
55-59	0.00041	5.240	0.00018	2.092	**0.00013**	**1.493**
60-64	0.00047	3.852	0.00071	5.954	**0.00034**	**2.725**
65-69	**0.00120**	5.795	0.00236	11.381	0.00147	**7.104**
70-74	**0.00137**	3.030	0.00217	**5.129**	0.00238	5.769
75-79	0.00820	12.940	0.00775	12.228	**0.00142**	**1.754**
80+	0.02276	18.835	0.01737	14.262	**0.01540**	**12.591**
Overall	0.00622	12.883	0.00468	6.9030	**0.00381**	**5.155**

Similar results are observed in out-sample prediction where both error measures for overall performance of the LC-DFA are smallest than the LC-SS and LC in which the LC-DFA performed better than the LC-SS and original LC for both males and females. However, the performance of the models for each of the age groups for both males and females, different evident is found especially for male population. Majority of the age groups show smaller values of RMSE for the LC-SS model compared to the LC-DFA. However, if we refer to MAPE values, the LC-DFA is still more dominant than LC-SS for almost all the age groups. Among females, both values of RMSE and MAPE show the LC-DFA is most performed for almost all the age groups when compared to both LC-SS and LC models.

V. CONCLUSION

This article explores the state space representation of the LC model in modeling and forecasting mortality rates. We extended LC model in state space representation called it LC-SS model by considering multiple state equation that explains the dynamics of the unobservable common trends through DFA, known as LC-DFA. As extension of the LC-SS in [17], the LC-DFA is also fitted with unequal variance assume for the each age group. This extension model considered because the observed ASDR data shows that some age groups have more variability than others. Since the aim of this paper is to introduce an alternative approach to modeling and forecasting mortality rates, we did not compare the proposed approach with other alternative methods. Here, the performance of the LC-DFA model was compared with the LC-SS model and LC model only. Evaluations were carried out using both in-sample and out-sample fit. Based on empirical results using Malaysian ASDR, as an overall performance, it is concluded that the LC-DFA model fits reasonably well for males and females in both evaluations with smallest errors measures compared to the both LC-SS and original LC model. Even though for certain age groups, the LC-SS model produced predicted ASDR with lower error, it was found that the LC-DFA model outperforms LC-SS and LC models for almost all age groups. Hence, based on empirical results, for fitting the current data set, we conclude that the performance of the LC-DFA model is the best model. We further conclude that the LC model in state space framework using EM algorithm is able to provide better estimates compared to the classical LC model. Further work in evaluating the performance of the LC-DFA model in forecasting mortality using life expectancy and other mortality indicators is currently under study.

ACKNOWLEDGMENT

We thank the Department of Statistics Malaysia (DOSM) for the data. In particular, we record our gratitude to the Ministry of Higher Education Malaysia and Universiti Teknologi MARA Malaysia for financial support under FRGS/1/2014/ST06/UiTM/02/5. The authors are grateful to the reviewers, editor, associate editor and anonymous referees for their thoughtful in-depth comments which have been very helpful in the revision of this paper.

REFERENCES

[1] A. F. Zuur, R. J. Fryer, I. T. Jolliffe, R. Dekker, and J. J. Beukema, "Estimating common trends in multivariate time series using dynamic factor analysis," Environmetrics, vol. 14, pp. 665-685, 2003.
[2] R. D. Lee and L. R. Carter, "Modeling and Forecasting U. S. Mortality," *Journal of the American Statistical Association*, vol. 87, 1992, pp. 659-671.
[3] H. Booth, L. Tickle, and L. Smith, "Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison," New Zealand Population Review, vol. 31, pp. 13-34, 2005.
[4] R. J. Hyndman and M. Shahid Ullah, "Robust forecasting of mortality and fertility rates: A functional data approach," Computational Statistics & Data Analysis, vol. 51, pp. 4942-4956, 2007.
[5] R. J. Hyndman and H. L. Shang, "Forecasting functional time series," Journal of the Korean Statistical Society, vol. 38, pp. 199-211, 2009.
[6] S. S. Yang, J. C. Yue, and H.-C. Huang, "Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models," Insurance: Mathematics and Economics, vol. 46, pp. 254-270, 2010.
[7] H. L. Shang, H. Booth, and R. J. Hyndman, "Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods," Demographic Research, vol. 25, pp. 173-214, 2011.
[8] A. J. G. Cairns, D. Blake, and K. Dowd, "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration," Journal of Risk and Insurance, vol. 73, pp. 687-718, 2006.
[9] P. De Jong and L. Tickle, "Extending Lee–Carter Mortality Forecasting," Mathematical Population Studies, vol. 13, pp. 1-18, 2006/01/01 2006.

[10] A. Delwarde, M. Denuit, and C. Partrat, "Negative binomial version of the Lee–Carter model for mortality forecasting," Applied Stochastic Models in Business and Industry, vol. 23, pp. 385-401, 2007.
[11] [R. J. Hyndman and M. Shahid Ullah, "Robust forecasting of mortality and fertility rates: A functional data approach," Computational Statistics & Data Analysis, vol. 51, pp. 4942-4956, 2007.
[12] D. Mitchell, P. L. Brockett, R. Mendoza-Arriaga, and K. Muthuraman, "Modeling and Forecasting Mortality Rates," SSRN eLibrary, 2011.
[13] Q. Gao and C. Hu, "Dynamic mortality factor model with conditional heteroskedasticity," Insurance: Mathematics and Economics, vol. 45, pp. 410-423, 2009.
[14] D. French and C. O'Hare, "A Dynamic Factor Approach to Mortality Modeling," Journal of Forecasting, vol 32 (7), pp. 587 - 599, 2013.
[15] C. Pedroza, "A Bayesian forecasting model: predicting US male mortality," Biostatistics, vol. 7, pp. 530-550, 2006.
[16] D. Lazar and M. M. Denuit, "A multivariate time series approach to projected life tables," Applied Stochastic Models in Business and Industry, vol. 25, pp. 806-823, 2009
[17] W. H. W. Zakiyatussariroh, M.S. Zainol, and M. R. Norazan, "Lee-Carter state space modeling: Application to the Malaysia mortality data," AIP Conference Proceedings, vol. 1602, pp. 1002-1008, 2014.
[18] W. H. W. Zakiyatussariroh, M.S. Zainol, and M. R. Norazan, "Performance of the Lee-Carter State Space Model in Forecasting Mortality," Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2015, WCE 2015, vol. 1, pp. 94-99, 2015.
[19] E. E. Holmes, E. J. Ward, and K. Wills, "Marss: Multivariate autoregressive state-space models for analyzing time-series data," *The R Journal*, vol. 4, 2012, pp. 11-19.
[20] G. E. Box and G. Jenkins, "" Time Series Analysis for Forecasting and Control," Holden-D. iv, San Francisco, 1976.
[21] J. D. Hamilton, Time series analysis vol. 2: Cambridge Univ Press, 1994.
[22] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *Journal of basic Engineering* **82** (1), 35-45 (1960).
[23] H. E. Rauch, C. Striebel and F. Tung, “Maximum Likelihood Estimates of Linear Dynamic Systems”, *AIAA journal* **3** (8), 1445-1450 (1965).
[24] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications: with R examples*: Springer, 2011.
[25] E. Holmes, E. Ward, and M. Scheuerell, "Analysis of multivariate time-series using the MARSS package," 2014

W. H. W. Zakiyatussariroh received a BSc. in Statistics from UniversitiTeknologi MARA, Malaysia and MSc. in Applied Statistics from Universiti Putra Malaysia. Currently she is a doctoral candidate in UniversitiTeknologi MARA, Malaysia. Her research interests include Demography and Time Series Modeling.

She is a Senior Lecturer at the Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, Kelantan, Malaysia and has been with the department for the past 13 years.

Ms.Wan Zakiyatussariroh is a member of the Malaysian Institute of Statistics as well as the Management Science and Operations Research Society of Malaysia.

M. S. Zainol received a BSc. in Mathematics from University of Malaya, Kuala Lumpur, an MBA in Quantitative Methods from Catholic University of Leuven and a PhD from the University College of Wales in Aberystwyth. United Kingdom.

Prior to retirement in May 2015 he was a Professor in Econometrics at the Faculty of Computer Science and Mathematics, UniversitiTeknologi MARA. His research interests include Demography and Time Series Modelling.

Dr. Zainol is a member of the Malaysian Institute of Statistics, the Management Science and Operations Research Society of Malaysia as well as Malaysian Mathematics Society.

M. R. Norazan received BSc. (Math. and Statistics) from University of East London (formerly known as Polytechnic of East London), United Kingdom, and both MSc. (Statistics) and PhD (Statistics) from Universiti Putra Malaysia.

Currently she is a Senior Lecturer in Statictics, and is attached at Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia. Her research interests are Demography, Partial Least Squares Regression, SEM-PLS Modeling, Robust Regression for Image Identification / Recognition, Outliers and Missing Values and Bootstrapping Techniques

Dr. Norazan is a member of the Malaysian Institute of Statistics, the Management Science and Operations Research Society of Malaysia as well as Malaysian Mathematics Society.