Coherent Mortality Model in A State-Space Approach
(Model Kemortalan Koheren dalam Pendekatan Keadaan-Ruang)

Siti Rohani Mohd Nor*, Fadhilah Yusof & Siti Mariam Norrulashikin

ABSTRACT

Mortality improvements that have recently become apparent in most developing countries have significantly shaped queries on forecast divergent between populations in recent years. Therefore, to ensure a more coherent way of forecasting, previous researchers have proposed multi-population mortality model in the form of independent estimation procedures. However, similar to single-population mortality model, such independent approaches might lead to inaccurate prediction interval. As a result of this inaccurate mortality forecasts, the life expectancies and the life annuities that the mortality model aims to generate is underestimated. In this study, we propose another new extension of the multi-population mortality model in a joint estimation approach by recasting the model into a state-space framework. A combination of augmented Li-Lee and O'Hare-Li methods are employed, before we transform the proposed model into a state-space formulation. In addition, this study incorporates the quadratic age effect parameter to the proposed model to better capture the younger ages mortality. We apply the method to gender and age-specific data for Malaysia. The results show that our latter framework brings a significant contribution to the multi-population mortality model due to the incorporation of joint-estimate and quadratic age effect parameters into the model's structure. Consequently, the proposed model improves the mortality forecast accuracy.

Keywords: Coherent mortality model; multi-population; state-space

ABSTRAK

Kadar kematian yang semakin menurun di kebanyakan negara membangun telah menimbulkan beberapa persoalan penting terhadap perbezaan jurang ramalan antara populasi bagi tahun-tahun kebelakangan ini. Oleh itu, untuk memastikan hasil ramalan yang lebih koheren, penyelidik sebelum ini telah mengemukakan model kemortalan berbilang penduduk dalam bentuk prosedur anggaran yang dibuat secara berasingan antara populasi. Walau bagaimanapun, sebagaimana model kemortalan penduduk tunggal, pendekatan berasingan mungkin menyebabkan ramalan yang tidak tepat. Akibat ramalan kemortalan yang tidak tepat ini, jangkaan hayat dan anuiti hayat yang dihasilkan oleh model kemortalan akan menjadi lebih rendah daripada yang sepatutnya. Dalam kajian ini, kami mencadangkan satu lagi model kemortalan yang mengintegrasikan maklumat antara populasi dengan cara menggabungkan model tersebut dalam rangka keadaan-ruang. Gabungan kaedah Li-Lee dan O'Hare-Li digunakan dan kemudian kami mengubah model yang dicadangkan ke dalam formulasi keadaan-ruang. Di samping itu, kajian ini menggabungkan parameter kesan usia kuadratik kepada model yang dicadangkan untuk menganggar kematian yang berlaku pada usia muda dengan lebih baik. Kami menggunakan kaedah tersebut ke atas data jantina dan data khusus umur bagi Malaysia. Keputusan menunjukkan bahawa rangka kerja ini membawa sumbangan penting kepada model kemortalan pelbagai penduduk kerana menggabungkan parameter kesan umur dan kuadratik parameter ke dalam struktur model. Hasilnya, model yang dicadangkan dapat meningkatkan lagi ketepatan ramalan kematian.

Kata kunci: Keadaan-ruang; kepelbagaian penduduk; model kemortalan koheren

INTRODUCTION

Recent decades have shown a growing number of interests in extensively developing a sophisticated mortality model in order to overcome two challenges associated with developed countries, which are longevity risk and increased correlation between closely related populations such as gender, race, states, and countries. These two challenges are important information to be incorporated

into a modelling structure in order to obtain proper forecast reading. Without accurate forecast reading, life expectancy is prone to be underestimated and hence will cause inadequate amount of retirement income, low allocation of health plan, and ageing population (Nor et al. 2018). Realising the significant impact that these two challenges may have on the aforementioned aspects, Li and Lee (2005) and Li (2013) suggested the need for joint modelling of the multi-population study to hedge the risk and to create coherent forecast between populations.

The Li and Lee (2005) model has contributed significantly to the line of multi-population mortality model. However, it still has limited flexibility because the model's built-up is based on a model developed by Lee and Carter (1992), which does not allow for variances between ages and does not have good historical fit and forecast performances as compared to the other single-population mortality models (Booth et al. 2006). Motivated by the limitations of Li and Lee (2005) model, Wan and Bertschi (2015) modified it by employing the Plat (2009) framework into the regression modelling structure. However, the model by Wan and Bertschi (2015) still have limitation in a way that it is unable to capture the nonlinear pattern observed at the lower ages. This is due to the incorporation of Plat (2009) model, which is only suitable to be fitted to the mortality rates above age 20 (O'Hare & Li 2012). Hauser and Weir (2016) and Weir (2010) stated that the incorporation of younger ages is essential since it works as a preliminary analysis for future life. In addition, O'Hare and Li (2012) and Scherbov and Ediev (2016) added that the mortality changes between all ages have significant impact on the estimation of average life expectancy. Hence, the inability of the stochastic mortality model to capture the young mortality would eventually cause an inaccurate estimate of annuity pricing. To close the gap in the model by Wan and Bertschi (2015), this study extends the Li and Lee (2005) model by incorporating the quadratic effect parameter proposed by O'Hare and Li (2012).

The development of single-population and multi-population mortality models are a huge influence in the aim to obtain accurate mortality forecast reading. However, the performances of both methods are still limited by the fact that the models designed are composed of two stages of independent estimation procedures which are regression model estimation procedure and time series model estimation procedure. Such separate estimation leads to problems including unreliable forecasts (Fung et al. 2018), and potential hidden errors that are being ignored in the first step of the estimation procedure (Pedroza 2006). Those issues related to independent estimation have led

to new interests in recasting the existing mortality model into a single estimate of the state-space framework. Fung et al. (2017, 2015), Husin et al. (2015), Liu and Li (2016a, 2016b) and Pedroza (2006) have found that the mortality model reformulated in a state-space framework has outperformed the independent estimation technique of mortality model in terms of the overall forecast accuracy. However, since most of the existing state-space mortality models are represented as Lee and Carter (1992) model, which has a lot of limitations that need to be addressed since it is the very first model introduced for stochastic mortality model, this study aims to reformulate the combination of Li and Lee (2005) and O'Hare and Li (2012) mortality models into a state-space framework.

Realising the significant contribution that the multi-population mortality and the state-space mortality model has brought into, there are two primary aims of this study, which are to propose new extension of multi-population mortality model by extending the models by Li and Lee (2005) and O'Hare and Li (2012), and to reformulate the proposed method into a state-space framework. The proposed models are able to capture the information of the mortality rates more accurately as compared to the existing models since the proposed models are able to capture the variances for the whole age ranges and the two estimation processes of the model are integrated into a single estimate only. This article is structured as follows. The Materials and Methods section describe the mortality data and the methods used in this article. The Results and Discussion section evaluate and compares the performances of the methods used in this article in terms of in sample fit, standardised residuals and forecasted mortality rates. Finally, the Conclusion section presents the conclusion of overall article.

MATERIALS AND METHODS

THE DATA SET

The new extended models as proposed in this study are applied on Malaysia mortality dataset, collected from the Department of Statistics Malaysia (DSM). The dataset comprises of the number of deaths, and the number of exposures to populations in Malaysia since the beginning of 1980 until 2015, available for the five-year age span ranges from 0 to 80. Two groups of populations are involved which are male and female. The formula of the mortality rate is:

$$m_{x,t,i} = \frac{D_{x,t,i}}{E_{x,t,i}}$$

where $D_{x,t,i}$ is the number of deaths; $E_{x,t,i}$ is the number of exposures for age x; time t and population i.

EXISTING MORTALITY MODELS

In this study, the abbreviation (LC, Plat, OH, Li-Lee, Wan-Bertschi) in Table 1 will be used to refer to the methods proposed by Lee and Carter (1992), Plat (2009), O'Hare and Li (2012), Li and Lee (2005) and Wan and Bertschi (2015), respectively. LC, Plat, and OH models are the single-population mortality models, whereas Li-Lee and Wan-Bertschi models are the multi-population mortality models. The models listed in Table 1 are estimated by using maximum likelihood estimation (MLE) method. The models employed are listed in Table 1.

TABLE 1. The six candidate models' structures under consideration

Model	Formula
LC	$\ln(m_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$
Plat	$\ln(m_{x,t}) = \alpha_x + k_t^1 + k_t^2(\bar{x} - x) + k_t^3(\bar{x} - x)^+ + \varepsilon_{x,t}$
OH	$\ln(m_{x,t}) = \alpha_x + k_t^1 + k_t^2(\bar{x} - x) + k_t^3((\bar{x} - x)^+ + [(\bar{x} - x)^+]^2) + \varepsilon_{x,t}$
Li-Lee	$\ln(m_{x,t,i}) = \alpha_{x,i} + \beta_x k_t + b_{x,i} k_{t,i} + \varepsilon_{x,t}$
Wan-Bertschi	$\ln(m_{x,t,i}) = \alpha_x + k_t^1 + k_t^2(x - \bar{x}) + k_t^3(\bar{x} - x)^+ + \alpha_{x,i} + \sum_{j=1}^{L} b_{x,i,j} k_{t,i,j} + \varepsilon_{x,t}$

The symbol $m_{x,t,i}$ is referred to central death rate at age x, year t and population i. The parameter α_x describes the overall mortality rate across ages, β_x is the additional age-specific component that represents the speed of mortality rate that responds to the change of time-varying mortality index k_t, k_t^1, k_t^2, k_t^3 and $k_{t,i}$. Parameter $\bar{x}$ indicates the sample average of the age groups x, where $(\bar{x} - x)^+ = \max(\bar{x} - x, 0)$ as such when $(\bar{x} - x)^+$ is positive, it takes the value of $\bar{x} - x$, whereas when $(\bar{x} - x)^+$ is negative, it takes the value of zero. $\varepsilon_{x,t}$ is the error term used in Lee-Carter model with mean zero and variance σ_ε

THE PROPOSED MODEL

There are two research objectives in this study, which are to propose multi-population mortality model that is able to capture accurate historical mortality data for the full age ranges and to reformulate the independent estimation of multi-population mortality model into a unified estimation. In order to avoid unreliable forecasts due to the separate estimation method, this study reformulates the independent estimation into a unified one by using a state-space representation framework. These two proposed mortality models are denoted as Augmented O'Hare (Augmented OH) and Augmented O'Hare State-Space (Augmented OH SS) models. The programming codes for the existing models are available in MARSS package, written in R (Holmes et al. 2012).

AUGMENTED OH MODEL

The first proposed mortality model in this study combines O'Hare and Li method into the first equation of Li-Lee method which is $\ln(m_{x,t,i}) = \alpha_{x,i} + \beta_x k_t$. The developed model is denoted in (1) and is named as Augmented O'Hare and Li (Augmented OH) method:

$$\text{Ln}(m_{x,t,i}) = \alpha_x + k_t^1 + k_t^2(\bar{x} - x) + k_t^3((\bar{x} - x)^+ +$$
$$[(\bar{x} - x)^+]^2) + \alpha_{x,i} + \sum_{j=1}^{L} \beta_{x,i,j} k_{t,i,j} + \varepsilon_{x,t,i} \qquad (1)$$

The indicated Augmented OH framework is designed to capture the central tendency between a group of populations and population-specific fluctuations, respectively. Following Hyndman et al. (2013), the weighted average of mortality rates is calculated by using geometric mean $(\prod_{i=1}^{M} m_{x,t,i})^{1/M}$ for the ith population M where the number of population is and for the jth principal components where the number of principal components is L. The first part of $\alpha_x + k_t^1 + k_t^2(\bar{x} - x) + k_t^3[(\bar{x} - x)^+ + [(\bar{x} - x)^+]^2]$ is estimated by using maximum likelihood estimation (MLE) Poisson distribution technique. The second part of the fitting procedure is to estimate the remaining information of the mortality model. The Singular Value Decomposition

(SVD) approach is employed to estimate population-specific parameters $\beta_{x,i,j}$ and $k_{t,i,j}$.

AUGMENTED OH SS MODEL

Multivariate State-Space model is innovated to deal with the parameter uncertainty estimation, missing data problem, and high-dimensional complex estimation problem (Holmes et al. 2012). The proposed stochastic mortality model in (1) with a state space framework, can be expressed as the Augmented O'Hare State-Space (Augmented OH SS) as denoted in (2) as follows:

$$m_t = \alpha + \beta k_t + \varepsilon_t \tag{2}$$

where $m_t = \left(\ln(m_{x_1,t,1}), \dots, \ln(m_{x_N,t,1}), \dots, \ln(m_{x_1,t,M}), \dots, \ln(m_{x_N,t,M}) \right)$ with age age $x = x_1, \dots, x_N$, time $t = 1,2,\dots,T$ and population, $i = 1, 2, \dots, M$, α, is the intercept parameter, and β is the population parameter. Parameter k_t is for the period effect parameter at time t and ε_t is denoted as the residual of the model at time t.

The proposed model design in (2) is composed of: common population parameters; and population-specific parameters with more than one single factor of multiple dynamic factor analysis (DFA) component. In order to have a similar structure as (1), the matrix β in (2) must fulfil the following condition:

$$\beta = \begin{bmatrix} 1 & \bar{x}-x_1 & (\bar{x}-x_1)^+ & +[(\bar{x}-x_1)^+]^2 & \beta_{x_1,1,1} \dots \beta_{x_1,1,F} & 0 & 0 & 0 & 0 \dots 0 \\ 1 & \bar{x}-x_2 & (\bar{x}-x_2)^+ & +[(\bar{x}-x_2)^+]^2 & \beta_{x_2,1,1} \dots \beta_{x_2,1,F} & 0 & 0 & 0 & 0 \vdots 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots \dots \vdots \\ 1 & \bar{x}-x_N & (\bar{x}-x_N)^+ & +[(\bar{x}-x_N)^+]^2 & \beta_{x_N,1,1} \dots \beta_{x_N,1,F} & 0 & 0 & 0 & 0 \dots 0 \\ 1 & \bar{x}-x_1 & (\bar{x}-x_1)^+ & +[(\bar{x}-x_1)^+]^2 & 0 & 0 & 0 & \beta_{x_1,2,1} \dots \beta_{x_1,2,F} & 0 \dots 0 \\ 1 & \bar{x}-x_2 & (\bar{x}-x_2)^+ & +[(\bar{x}-x_2)^+]^2 & 0 & 0 & 0 & \beta_{x_2,2,1} \dots \beta_{x_2,2,F} & 0 \dots 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots \vdots \vdots \\ 1 & \bar{x}-x_N & (\bar{x}-x_N)^+ & +[(\bar{x}-x_N)^+]^2 & 0 & 0 & 0 & \beta_{x_N,2,1} \dots \beta_{x_N,2,F} & 0 \dots 0 \end{bmatrix}$$

where the first four columns of matrix β are the common population parameters, whereas the latter columns are the population-specific parameters $\beta_{x,i,f}$ for age-group $x_1,..,x_N$, at a specific population $i = 1,2$ and $f = 1,2,3\dots,F$ is the number of multiple components included in the state space model.

The residual of the model in (2) is denoted as follow:

$$\varepsilon_t = \left(\varepsilon_{x_1,t}, \varepsilon_{x_2,t}, \dots, \varepsilon_{x_N,t} \right)', \text{ where } \varepsilon_t \sim \text{MVN}(0, R)$$

The identity matrix of is:

$$R = \begin{bmatrix} r & 0 & \cdots & 0 & 0 \\ 0 & r & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & r & 0 \\ 0 & 0 & \dots & 0 & r \end{bmatrix}$$

The latent states vector symbolized by k_t takes the form as in:

$$k_t = D k_{t-1} + \epsilon_t; \text{ where } D = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 1 & 0 & \dots & 1 & 0 \\ 0 & 0 & 1 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

$$k_t = \left(k_t^c, k_{t,i,f}, U_t^c \right)', \text{ where } \epsilon_t \sim MVN(0, Q)$$

where MVN is the multivariate normal distribution, k_t^c is denoted as the common parameters of the model and U_t^c is the stochastic drift with c = 1,2,3. $k_{t,i,f}$ is denoted as the population-specific parameters of the considered model where t is the number of years, $i = 1,2,\dots, M$ is the number of populations and $f = 1,2,3\dots, F$ is the number of multiple components included in the state space model. The states assumption error is $\epsilon_t \sim MVN(0, Q)$, where Q is the unequal variances and must be positive-definite, and p is the number of column matrix for parameter β.

MEASUREMENT ERRORS

The comparison of the model performances is measured by employing Li et al. (2015a) method in percentage form. The three measurement errors are denoted in (3)-(5).

The average measurement error (AE) measures the uncertainty between observed and predicted values:

$$AE = \left(\frac{1}{NT} \sum_x \sum_t \frac{\hat{m}_{x,t} - m_{x,t}}{m_{x,t}} \right) \times 100 \tag{3}$$

where $\hat{m}_{t,x}$ is the predicted value of mortality rates; $m_{x,t}$ is the observed mortality rates; N is the length of age-groups x; and T is the length of period t.

The Mean Average Percentage Error (MAPE) measures the magnitude of average error:

$$MAPE = \frac{1}{NT} \sum_x \sum_t \frac{|\hat{m}_{x,t} - m_{x,t}|}{m_{x,t}} \times 100 \tag{4}$$

The root mean square error (RMSE) value is significantly influenced by the existence of outliers. The higher error reported by the RMSE value indicates a greater deviance value:

$$RMSE = \left(\sqrt{\frac{1}{NT} \sum_x \sum_t \left(\frac{\hat{m}_{x,t} - m_{x,t}}{m_{x,t}} \right)^2} \right) \times 100 \tag{5}$$

Other than AE, MAPE and RMSE measurement errors, Cairns et al. (2009) stated that it is natural for a

model with a higher number of parameters to outperform a model with fewer number of parameters. Haberman and Renshaw (2011) stated that to address this concern, the models with a large number of parameters need to be penalized by including a penalty function, which is Akaike Information Criterion (AIC). Hence as suggested, to evaluate the model fairly, regardless of how many numbers of parameters the model has, Akaike Information Criterion (AIC) is considered in selecting the best model. The model with the lowest value of AIC indicates the model is selected as the best model since it is the most parsimonious.

$$AIC = 2\delta - 2 \log(W)$$

where W is the maximum value of the likelihood; and δ is the number of model's parameters.

RESULTS AND DISCUSSION

Table 2 summarizes the comparison of the fitting performances of each model fitted to Malaysia dataset in terms of AE, MAPE, and RMSE. The model that gives the lowest values of the error measurements is the best model.

TABLE 2. In-sample fit error measurements for males and females

	Male			Female		
	AE	MAPE	RMSE	AE	MAPE	RMSE
LC	0.0088	0.0763	0.0369	0.0065	0.0625	0.0214
Plat	0.0237	0.1870	0.0826	0.0236	0.1960	0.1091
OH	0.0270	0.1994	0.0923	0.0218	0.1596	0.0809
Li-Lee	0.0100	0.0731	0.0395	0.0067	0.0622	0.0215
Wan-Bertschi	0.0033	0.0338	0.0093	0.0033	0.0288	0.0098
Augmented OH	0.0030	0.0314	0.0094	0.0024	0.0249	0.0094
Augmented OH SS	0.0021	0.0196	0.0068	0.0015	0.0172	0.0045

According to Table 2, the single-population mortality models which are LC, Plat, and OH have the highest AE, MAPE, and RMSE for both genders. This shows that the single-population mortality models could not give best fitting results as compared to the multi-population mortality models since they did not incorporate the parameter that could captured the correlation between population. The proposed multi-population mortality model which is Augmented OH SS outperform the existing multi-population mortality models by giving the lowest values of AE which are 0.0021 for males and 0.0015 for females, MAPE which are 0.0196 for males and 0.0172 for females and RMSE which are 0.0068 for males and 0.0045 for females. This indicates that the multi-population mortality model is improved by incorporating variations for all age groups and by recasting the separate estimation into single estimate of the state-space framework.

TABLE 3. The approximate AIC for Malaysia dataset

	AIC	
Model	Male	Female
LC	159780.20	87537.50
Plat	373555.40	124519.50
OH	380443.90	124386.60
Li-Lee	204658.90	131974.90
Wan-Bertschi	125644.70	55318.67
Augmented OH	96801.93	46136.31
Augmented OH SS	54752.89	36502.01

Next, the quality of fit of the different mortality models is evaluated using AIC presented in Table 3. The model with the smallest value of AIC is selected as the best model. In Table 3, the proposed models Augmented OH and Augmented OH SS has smaller value of AIC with an approximate 96801.93 and 54752.89 for males, and 46136.31 and 36502.01 for females. These results indicate that although our models are considered to have more number of parameters as compared to the others, the models are still considered to be the best-selected model since the lower AIC values imply that the model is not overfitting and has a better fit.

Overall, the results in this section indicate that the extended multi-population mortality model, which incorporates the dependency between ages coherently would eventually improve the fitting performance of the historical Malaysia mortality rates for both males and females. Other than that, the multi-population mortality model in a state-space formulation could explain the historical mortality data better than the multi-population mortality model with the independent estimation. Therefore, we can conclude from here that both of our proposed models provide a good quality fit to the mortality dataset. This is one of the desirable characteristics any good mortality model should have (Cairns et al. 2009).

STANDARDISED RESIDUALS ANALYSES

Figures 1 and 2 show the deviance of residuals plot in terms of heat-maps for the considered models. Based on our analysis, most of the residual patterns for single-population mortality model for both genders as well as Li-Lee multi-population mortality model suggest that there is a need for additional factors to be included into the model structure in order to adequately capture the mortality changes in Malaysia. This is because of the apparent observed diagonal and horizontal residual pattern plots that the models produced, which show that the models are unable to capture the cohort and age-effect. More than that, LC, Plat and OH models show a strong clustering which imply that these models overestimated and underestimated the mortality rates, respectively. On the other hand, there exists strong negative clustering of residuals in Wan-Bertschi model which assumes that the model is inadequate to allow for time-varying mortality changes with age (Villegas 2015) and the mortality rate is overestimated by the model (Coelho 2013). Augmented OH and Augmented OH SS display the most random pattern which indicate that our models are able to capture all the important time-varying effect and the age effect of the data. Thus, of the eight models considered, our model has outperformed the rest for males and females in terms of residual's visual inspection.
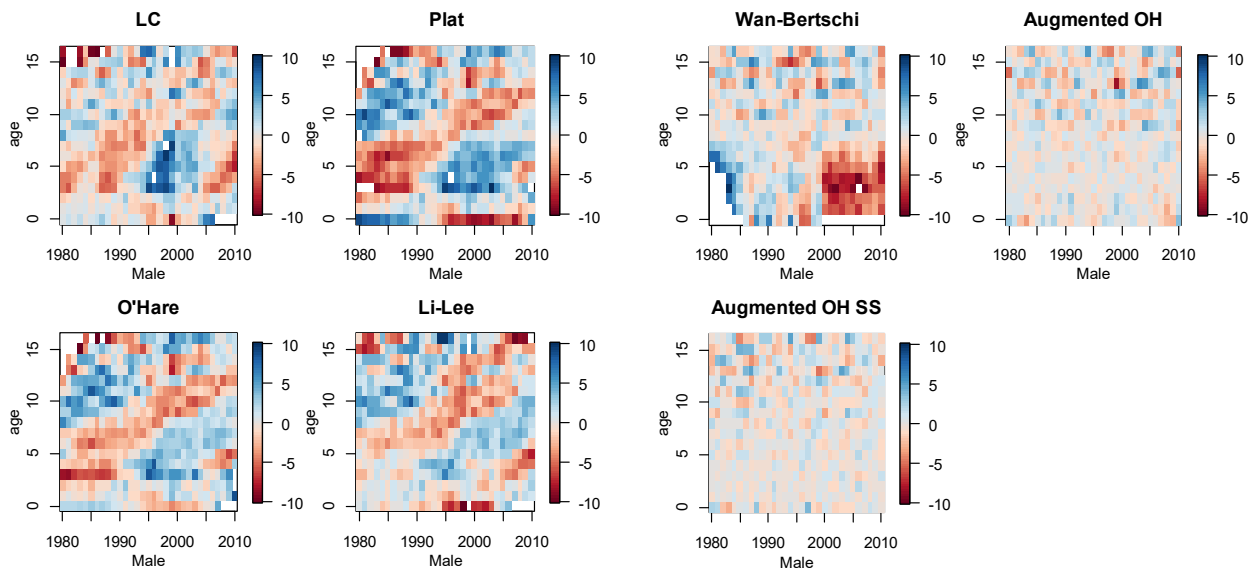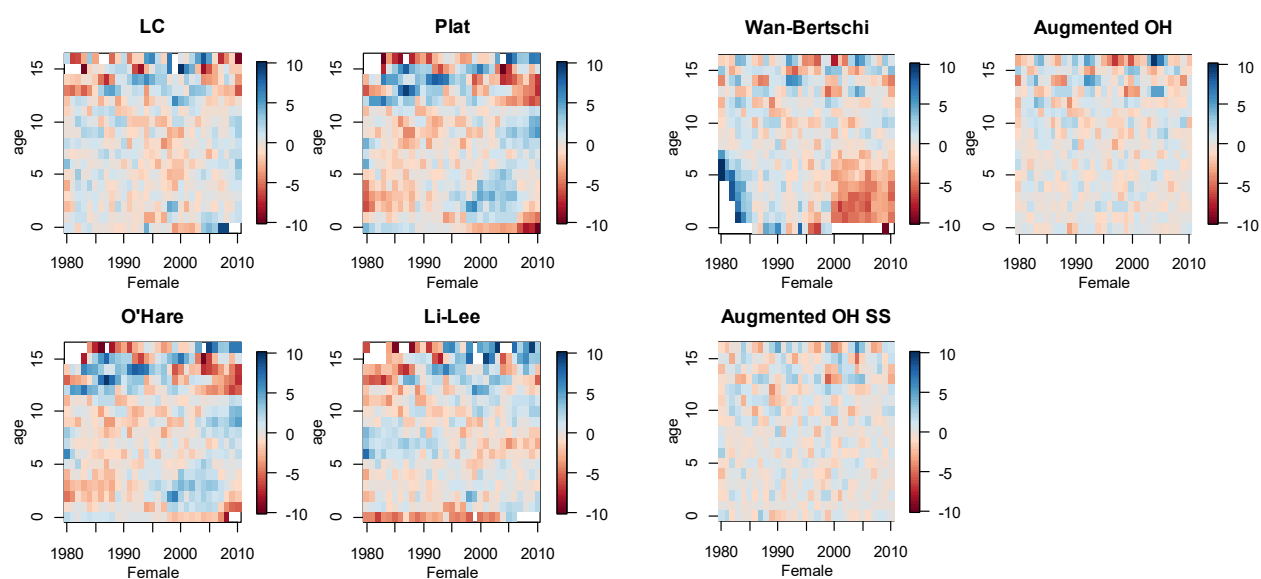

FIGURE 1. Residual plots in terms of heat map for male

FIGURE 2. Residual plots in terms of heat map for female

Figures 1 and 2 also show that the standardized residuals proposed by the other authors such as LC, Plat, OH, Li-Lee and Wan-Bertschi models do not meet the requirement of the models' assumptions random errors for males and females. It can be seen that the residuals of the models are not random since the blue and red colours are strongly visible. On the other hand, our proposed models which are the Augmented OH and the Augmented OH SS for males and females, are normal since the colours indicate that the residuals of the models are close to zero.

TABLE 4. Average cross-validation error measures in short and long-term forecast

Gender	Male				Female			
Normality Tests	Variance ratio p-value	Skewness p-value	Kurtosis p-value	Royston p-value	Variance ratio p-value	Skewness p-value	Kurtosis p-value	Royston p-value
LC	0.020	0.988	1.000	0.000	0.814	0.966	1.000	0.000
Plat	0.000	0.644	0.948	0.000	0.069	0.899	1.000	0.023
OH	0.064	0.600	0.909	0.000	0.136	0.952	1.000	0.000
Li-Lee	0.687	0.970	1.000	0.001	0.902	0.953	1.000	0.000
Wan-Bertschi	0.000	0.881	1.000	0.000	0.000	0.932	1.000	0.000
Augmented OH	0.620	0.978	1.000	0.732	0.486	0.976	1.000	0.016
Augmented OH SS	0.502	0.997	1.000	0.623	0.116	0.949	1.000	0.276

In order to validate the assumptions, the standardised residuals illustrated from Figures 1-2 are then numerically supported in Table 4. The variance-ratio test, Jarque-Bera test and Royston test are employed to check for the constant variance, skewness and kurtosis predictions and multivariate normality test, respectively. Table 4 presents the p-values of the corresponding tests for both genders. The results in Table 4 shows that LC, Plat, OH, Li-Lee and Wan-Bertschi models perform quite poorly as compared to the others since the p-values of Royston test for LC, Plat,

OH, Li-Lee and Wan-Bertschi model are approximately 0.000 for both genders. Thus, the residuals of these models do not conform to standard normal distribution since the *p*-values are very low giving the strong evidence to reject the null hypothesis of the normality test at 1% significance level. On the other hand, the *p*-values of the Variance-ratio test for Plat, and Wan-Bertschi models indicate that the residuals' variances of these models are not equal. This is because, the *p*-values are very low giving the strong evidence to reject null hypothesis of the equality of the residuals' variances at 1% significance level. This is expected since Plat and Wan-Bertschi models could not capture full ages of mortality data.

The analyses reported in Table 4 suggests that only our proposed models meet the specifications needed for the model assumption random error. For example, the null hypothesis of constant variance for Augmented OH SS was failed to be rejected since the *p*-values of variance-ratio test are 0.502 and 0.116, for males and females, which are more than 1% significance level. In addition, for Augmented OH SS, the null hypothesis of the normality prediction was failed to be rejected at 1% significance level because the *p*-values for Royston test for both male and female are 0.623 and 0.276. Overall, the *p*-values for every test for Augmented OH and Augmented OH SS were found to be more than the significant value 1% which eventually indicates that the residual of the considered models are approximately normal.

MORTALITY RATES FORECASTS

One of the essential features of an accurate stochastic mortality model is to have a good forecasting ability. Inaccurate estimate of future mortality rates could lead to inaccurate measurement of annuity product pricing. Therefore, in this section, we assess the goodness of fit of stochastic mortality models in terms of forecast mortality.

TABLE 5. Average cross-validation error measures in short and long-term forecast

Gender	Male					Female				
Forecast Horizon	1-year ahead	5-year ahead	10-year ahead AE	15-years ahead	20-years ahead	1-year ahead	5-year ahead	10-year ahead AE	15-years ahead	20-years ahead
LC	0.0313	0.0439	0.0461	0.0409	0.0210	0.0135	0.0145	0.0142	0.0108	0.0108
Plat	0.0250	0.0273	0.0400	0.0406	0.0266	0.0141	0.0144	0.0193	0.0171	0.0168
OH	0.0239	0.0242	0.0325	0.0313	0.0359	0.0156	0.0161	0.0215	0.0193	0.0224
Li-Lee	0.0121	0.0157	0.0306	0.0222	0.0195	0.0083	0.0136	0.0210	0.0128	0.0132
Wan-Bertschi	0.0139	0.0161	0.0174	0.0194	0.0252	0.0230	0.0393	0.0274	0.0593	0.0269
Augmented OH	0.0074	0.0158	0.0305	0.0174	0.0157	0.0070	0.0107	0.0237	0.0104	0.0094
Augmented OH SS	0.0057	0.0097	0.0168	0.0187	0.0181	0.0035	0.0163	0.0111	0.0094	0.0097
	MAPE						MAPE			
LC	0.1566	0.2078	0.2079	0.2030	0.1602	0.1011	0.1043	0.1075	0.1718	0.1382
Plat	0.3571	0.3532	0.8384	1.0845	0.5529	0.3210	0.3416	0.5545	0.7639	0.3188
OH	0.4851	0.3917	0.6215	0.5655	0.4715	0.4631	0.4136	0.5808	0.4940	0.2614
Li-Lee	0.0739	0.0975	0.1523	0.1315	0.2221	0.0589	0.0689	0.1189	0.0935	0.1421
Wan-Bertschi	0.1040	0.1044	0.1293	0.3261	0.3809	0.8964	0.9934	1.0558	1.6636	0.8751
Augmented OH	0.0730	0.1030	0.1259	0.1514	0.2176	0.0606	0.0630	0.0844	0.1183	0.1371
Augmented OH SS	0.0766	0.0719	0.1326	0.1647	0.2228	0.0697	0.0690	0.1027	0.1261	0.1409
	RMSE						RMSE			
LC	0.0955	0.1621	0.1779	0.1769	0.0875	0.0365	0.0455	0.0476	0.0360	0.0349
Plat	0.0674	0.0758	0.1038	0.1094	0.0879	0.0464	0.0469	0.0581	0.0560	0.0512
OH	0.0650	0.0686	0.0916	0.0937	0.0901	0.0497	0.0508	0.0635	0.0616	0.0697
Li-Lee	0.0265	0.0505	0.1067	0.1023	0.0544	0.0211	0.0530	0.0638	0.0384	0.0442
Wan-Bertschi	0.0352	0.0463	0.0670	0.0687	0.0912	0.0387	0.0856	0.0716	0.1560	0.0621
Augmented OH	0.0204	0.0504	0.1302	0.0656	0.0509	0.0217	0.0417	0.1039	0.0276	0.0332
Augmented OH SS	0.0141	0.0370	0.0665	0.0691	0.0544	0.0094	0.0762	0.0379	0.0260	0.0331

The models' comparisons are determined based on AE, MAPE, and RMSE measurements. Li et al. (2015a) stated that in actuarial perspective, AE, MAPE, and RMSE are vital to avoid bias, to ensure overall forecast accuracy, and to avoid anomalies. Therefore, in Table 5, we are going to have a better look at which models give the best projection future mortality rates estimates. The results of this out-sample accuracy for males and females are tabulated in Table 5. The light green colour represents the first best performance, while the dark green colour represents the second-best performance.

Overall, as indicated by the measurement errors marked in green colours, the proposed mortality models which are Augmented OH and Augmented OH SS could give better forecasts. Since the models have the lowest values of AE, RMSE and MAPE for all years. For example, Augmented OH SS and Augmented OH perform relatively well with an approximate 1-year AE of 0.0074 and 0.0057 for males and 0.0070 and 0.0035 for females. In addition, the proposed models outperform the others in terms of long-term forecast with an approximate 20-year AE of 0.0157 and 0.0181 for males and 0.0094 and 0.0097 for females. Similar to single-population mortality models, Li-Lee and Wan-Bertschi models did not perform very well across time. This is because, from previous section, these models did not meet the specifications needed for the model assumption random error, therefore, the forecasts results produced are not accurate.

The extended models which are the Augmented OH and the Augmented OH SS demonstrate an important aspect of good mortality model, which are good in-sample historical fit performance and good out-sample forecast. As can be seen from Table 5, the extended models are marked with green colours which show that they have good forecasting ability in the short-term and long-term projection. This feature allows the models to produce more stable forecasts mortality for both genders.

Next, we visually inspected the coherence ratio of the predicted mortality rates for both males and females in Malaysia. The coherence assumption is proposed based on the narrowing gap of gender-differential on the mortality improvements. Therefore, it is expected that the projection of future mortality rates of males and females would be converging to the same value in which eventually the ratio of the male-to-female should be constant as the time increases. In this section, the eight models considered in this study are analysed to see if they are following the coherence assumption as mentioned by Li (2013) and Li et al. (2015b). The coherence of male and female performances from different models were compared for 20-years of forecast horizons in Figure 3. The dotted colours represent the respective Malaysian age span ranges from 0 to 80. There are 31 observation periods, with the most recent year is 2010.
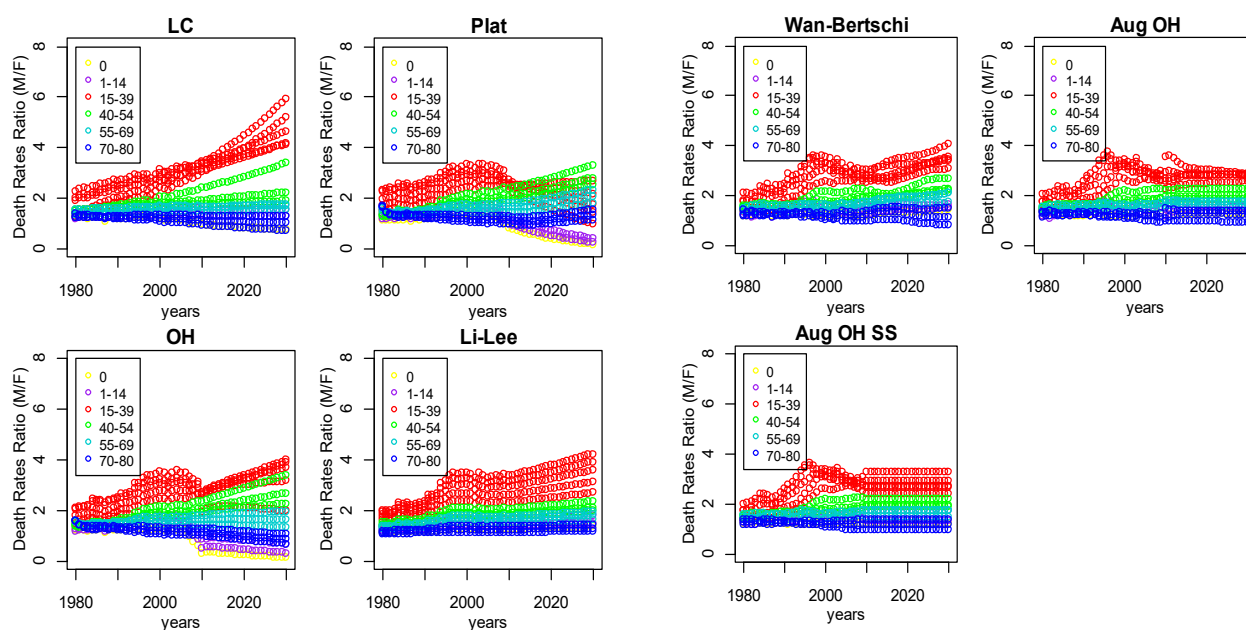

FIGURE 3. Observation on Malaysia death rates for male-to-female ratio 1980-2010, and forecast Malaysia death rates for male-to-female ratio 2011-2030 for stochastic mortality models

Based on Figure 3, our proposed models which are the Augmented OH and the Augmented OH SS, pose one of the main criteria in building the mortality model which is to produce non-diverge mortality rates forecast in the long run for all age-groups. Wan and Bertschi model performs the worst in terms of in-sample fit and out-sample forecast since it does not incorporate full age-ranges in their model's structures. According to the in-sample results, LC model outperformed the Plat and OH model. However, the coherence plots indicate that LC model gives diverge mortality rates forecast in the future and therefore violated the coherence assumption between populations. On the other hand, the projected mortality rates of certain age groups of Wan-Bertschi, Plat, LC, and OH mortality models tend to diverge and crossover between each other. These undesirable conditions can be eliminated by assimilating the relationship between the populations simultaneously. Other than that, Li-Lee mortality model produced diverge forecasts at the age 15 to 39. This shows that joint parameter included in the modelling structure of Li-Lee failed to capture the variances across ages, which consequently leads to forecast divergence.

Overall, our proposed models are found to be superior in terms of historical fit as well as predictive accuracy for male and female mortality. The models also exhibit the criteria needed in selecting the best stochastic mortality model, which is to ensure non-diverging projection mortality rates in the long run.

CONCLUSION

We proposed two new stochastic multi-population mortality models which are Augmented OH and Augmented OH SS. The models established are then numerically and visually compared with the models by Lee and Carter (1992), Cairns et al. (2009), Plat (2009), O'Hare and Li (2012), Li and Lee (2005), and Wan and Bertschi (2015). The selected model fits the data well if it is able to follow the criteria needed for a good mortality model which is, the model should have a: good quality of fit, normal standardized residuals error assumption, good forecasting ability in terms of short term and long term, and non-diverge male-to-female projection ratio. The analyses output demonstrates that Lee and Carter (1992), Plat (2009), O'Hare and Li (2012), Li and Lee (2005), Wan and Bertschi (2015) models do not follow all of these four sets criteria of a good model, thus are not considered to be a good mortality model for our datasets. Taken together, the results of this study have shown that our two latter frameworks have consistently performed the best for all four sets criteria of a good mortality model. In addition, the proposed mortality model brings significant contribution to the multi-population mortality model

because the model could capture the correlation between population and correlation between full ages mortality, simultaneously. Other than that, we have shown that the model with a unified estimation of a state-space framework gives more accurate predictions as compared to the model with independent estimation procedures.

REFERENCES

Booth, H., Hyndman, R.J., Tickle, L. & De Jong, P. 2006. Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289-310.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. & Balevich, I. 2009. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal* 13(1): 1-35.

Coelho, E. 2013. Modelling and forecasting mortality patterns. Ph.D. Thesis, Nova University of Lisbon, Portugal (Unpublished).

Fung, M.C., Peters, G.W. & Shevchenko, P.V. 2018. Cohort effects in mortality modelling: A Bayesian state-space approach. *Annals of Actuarial Science* 13(1): 109-144.

Fung, M.C., Peters, G.W. & Shevchenko, P.V. 2017. A unified approach to mortality modelling using state-space framework: Characterisation, identification, estimation and forecasting. *Annals of Actuarial Science* 11(2): 343-389.

Fung, M.C., Peters, G.W. & Shevchenko, P.V. 2015. A state-space estimation of the Lee-Carter Mortality Model and implications for annuity pricing. In *MODSIM2015 - 21st International Congress on Modelling and Simulation*, edited by Weber, T., McPhee, M.J. & Anderssen, R.S. pp. 952-958. http://www.mssanz.org.au/modsim2015/E1/fung.pdf.

Haberman, S. & Renshaw, A. 2011. A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics* 48(1): 35-55.

Hauser, R.M. & Weir, D. 2016. Recent developments in longitudinal studies of aging. *Demography* 23(5): 1079-1084.

Holmes, E.E., Ward, E.J. & Wills, K. 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. *The R Journal* 4(1): 11-19.

Husin, W.Z.W., Zainol, M.S. & Ramli, N.M. 2015. Performance of the Lee-Carter State Space Model in forecasting mortality. *Proceedings of the World Congress on Engineering*. pp. 39-52.

Hyndman, R.J., Booth, H. & Yasmeen, F. 2013. Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* 50(1): 261-283.

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

Li, H., O'Hare, C. & Zhang, X. 2015a. A semiparametric panel approach to mortality modeling. *Insurance: Mathematics and Economics* 61: 264-270.

Li, J. 2013. A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies* 67(1): 111-126.

Li, J.S., Zhou, R. & Hardy, M. 2015b. A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics* 63: 121-134.

Li, N. & Lee, R. 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42(3): 575-594.

Liu, Y. & Li, J.S.H. 2016a. The locally linear Cairns-BlakeDowd Model: A note on Delta-Nuga hedging of longevity risk. *ASTIN Bulletin* 47(1): 79-151.

Liu, Y. & Li, J.S.H. 2016b. It's all in the hidden states: A longevity hedging strategy with an explicit measure of population basis risk. *Insurance: Mathematics and Economics* 70: 301-319.

Nor, S.R.M., Yusof, F. & Bahar, A. 2018. Multi-Population mortality model: A practical approach. *Sains Malaysiana* 47(6): 1337-1347.

O'Hare, C. & Li, Y. 2012. Explaining young mortality. *Insurance: Mathematics and Economics* 50(1): 12-25.

Pedroza, C. 2006. A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics* 7(4): 530-550.

Plat, R. 2009. On stochastic mortality modelling. *Insurance: Mathematics and Economics* 45(3): 393-404.

Scherbov, S. & Ediev, D. 2016. Does selection of mortality model make a difference in projecting population ageing? *Demographic Research* 34(2): 39-62.

Villegas, A.M. 2015. Mortality: Modelling, socio-economic differences and basis risk. Ph.D. Thesis, Cass Business School, London (Unpublished).

Wan, C. & Bertschi, L. 2015. Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach. *Insurance: Mathematics and Economics* 63: 66-75.

Weir, D.R. 2010. Grand challenges for the scientific study of ageing. *American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*. Institute for Social Research, Ann Arbor: University of Michigan.

Department of Mathematical Sciences
Faculty of Science
Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Johor Darul Takzim
Malaysia

*Corresponding author; email: sitirohani@utm.my