Research Article

A Quantitative Comparison of Multiple Population Mortality Model on Some East Asian Countries and Regions

Ming Zhao [ID],[1] Xiaojun Wang,[2] Yu Liu,[1] and Jiali Shen[1]

[1]*School of Finance, Capital University of Economics and Business, Beijing 100070, China*
[2]*School of Statistics, Renmin University of China, Beijing 100872, China*

Correspondence should be addressed to Ming Zhao; zhaoming@cueb.edu.cn

Received 16 January 2020; Accepted 11 March 2020; Published 9 April 2020

Guest Editor: Ming Zhou

This paper reviews the progress of the multiple population mortality model and the defects in parameter estimation and proposes an effective method to improve the performance of the mortality model. We set up a multiple population group, using the data of mainland China, Hong Kong (China), and Japan, to test fitting performance and forecasting performance. Using the TSWLS and TSSVD methods in a multiple population stochastic mortality model has advantages in fitting performance and robustness. In addition, the forecasting value of mortality ratio between any two populations can converge to a fixed constant in a certain time period which obeys the regular of human biological characteristics.

1. Introduction

With the demographic dividend gradually disappearing worldwide, it is common for the elders to have fewer children. In the future, it will seriously change the age structure of the population and interfere with the country's formulation of certain strategic policies. Additionally, the acceleration of the life expectancy and the ageing of the population would lead to varying shocks to the national pension system, commercial insurance companies, and families, which makes a negative impact on the economic development for a country. Therefore, it is beneficial to take countermeasures in advance to help economic entities, using scientific methods to forecasting the population mortality and reasonably assessing the impact of longevity risk. The research on the method of forecasting of mortality has experienced the development from a single population model to multiple population model. Among them, some classical methods, such as the Lee–Carter model, APC model, and CBD model, have been verified for stability, which represent the frontier progress of the research on the stochastic mortality model.

The stochastic mortality model which is used for a single population group is first proposed by Lee and Carter in [1].

The Lee–Carter model assumes that the logarithmic mortality is composed of independent age and period effects, with fewer model parameters, simple fitting process, robust forecasting results, and other advantages, which has been widely used by scholars all over the world. The Lee–Carter model applies a two-stage method to estimate parameters. In the first stage, it uses the orthogonal least squares (OLS) method, the maximum likelihood estimate (MLE) method, or singular value decomposition (SVD) method to estimate static parameters. For the second stage, dynamic parameters are fitted by the time series model [2]. Scholars have made many improvements to the Lee–Carter model, including the improvement of the parameter estimation method [3] and model hypothesis [4]. Renshaw et al. [5] took the cohort effect of population mortality into consideration and further expanded it into the age-period-cohort (APC) model. Although the APC model is used in medicine for a long time, the idea of modelling stochastic mortality originated from the Lee–Carter model. Furthermore, due to the relatively small number of population exposures of old age, the Lee–Carter model and APC model are not well suitable for the mortality of the elders. However, it is found that the CBD model with two factors could cope with this problem [6]. The CBD mode with cohort effects is a prominent choice [7] for

fitting mortality of the elders when both BIC information criterion and robustness of the parameters are considered. With the continuous development of population mortality models, there is an increasing number of shortcomings of single population stochastic mortality model exposed and there would be unreasonable crossover or deviation in mortality forecasting in a long time [8, 9]. Because the mortality modelling is a kind of systematic work, it only considers a single population group, which will cause different population mortality violating human biological laws over time. Therefore, it is necessary to promote the forecasting performance of mortality in the long term by combining with two or more populations in one model, which can make the mortality model have better fitting goodness and forecasting performance.

Carter and Lee [10] proposed the first multiple mortality model called the Joint-k model, which assumes that the mortality of multiple populations has the common period effect factor, and the gap among population groups reflects only in the individual age effect factor. Li and Lee [11] extended the age effect factor to a common factor based on the Joint-k model and then put forward the Li–Lee model, an augmented common factor model, which has common age and period effect factors plus the additional age and period effect factors that represent the mortality in a single population. Li and Hardy [12] proved that there is a cointegration relationship on the trend of the period effect factor among multiple populations and then established a linear time-effect factor model called the cointegrated Lee–Carter model. Kleinow [13] proposed a common age effect (CAE) model, using a common principal component analysis to estimate the parameter. The model is simple and has a better fitting performance. Enchev et al. [14] compared the above models with the MLE method to select which model has the best fitting performance. According to the result, the CAE model and Li–Lee model are more suitable for modelling on multiple population mortality, but there is a problem in converging of parameter estimation. Li and Liu [15] built a logistic two-population mortality projection model for the mortality at ages 80 to 100 of both sexes, applied this model and its extensions to high-quality old-age mortality data of Belgium, Sweden, Switzerland, and the UK, and produced a decent model performance in both mortality fitting and forecasting. Tsai and Zhang [16] proposed a nonparametric method to forecast the mortality of a multiple populations of the United States, the United Kingdom, and Japan.

According to current research, it is found that the multiple population stochastic mortality model has become the frontier progress, and there have been studies on the quantitative comparison of different types of multiple mortality models, but the quantitative comparison of parameter estimation methods is still blank. In addition, most of the data used to test the multiple population mortality models are European countries in the human mortality database, but few studies use data from statistical institutions in developing countries. As a country with a large population and rapid economic development, the mortality rate of mainland China is rapidly decreasing, but there are few studies on mortality forecasting based on the multiple

population model. It is meaningful for mainland China to build a multiple mortality model with neighbouring countries or regions which not only have the lower values and higher quality mortality but also have much closer genes and habits that can affect mortality. Therefore, this paper selects East Asian countries and regions, including Hong kong (China) and Japan to build a multiple population model with mainland China to compare quantitatively the methods of parameter estimation.

The structure of this paper is as follows. Section 2 explains the rationale of the methods of parameter estimation for the multiple mortality model. Next, Section 3 describes the data features and research scheme. Following this, Section 4 inspects the fitting performance of the proposed methods. Section 5 examines the forecasting performance among the three population groups. Finally, Section 6 concludes the paper.

2. Mortality Models

2.1. Lee–Carter Model. Lee and Carter (1992) propose a logarithmic, linear stochastic mortality model as follows:

$$\ln m(x, t) = \alpha(x) + \beta(x) * k(t) + \varepsilon(x, t), \qquad (1)$$

in which $m(x, t)$ is the crude mortality rate at the age x in the year t for a single population, $\alpha(x)$ and $\beta(x)$ are variables about the age x, while $\alpha(x)$ is the average of the logarithmic mortality in all years, $\beta(x)$ shows the age effect parameter which stands for the slope of the logarithmic mortality, $k(t)$ explains the period effect parameter that represents the slope of the logarithmic mortality in the year t, and $\varepsilon(x, t)$ is the normal error with i.i.d. There are three types of methods to estimate the parameters in the Lee–Carter model, including the OLS, SVD, and MLE method.

2.2. Li–Lee Model. While requiring the high quality of death data, the Lee–Carter model can only be used to forecast the mortality of a single population. To make up for the defects of the Lee–Carter model, Li and Lee proposed a mortality model from the perspective of multiple populations as follows:

$$\ln m(x, t, i) = \alpha(x, i) + B(x) * K(t) + \beta(x, i) * k(t, i) + \varepsilon(x, t, i),$$
$$(2)$$

in which $m(x, t, i)$ is the crude mortality rate at the age x in the year t in the population i and $\alpha(x, i)$ explains the average of logarithmic mortality at the age x in all years in the population i. The common age effect parameter $B(x)$ represents the slope of the logarithmic mortality at the age x for all population groups, $K(t)$ is period effect parameter which shows the slope of the logarithmic mortality in the year t for all population groups, the specific age effect parameter $\beta(x, i)$ is the slope of the logarithmic mortality at the age x in the population i, $k(t, i)$ is the specific period effect parameter that represents the slope of the logarithmic mortality in the year t in the population i, and $\varepsilon(x, t, i)$ is the normal error with i.i.d. Similar to the Lee–Carter model, the Li–Lee model also

contains three methods to estimate the parameters that are SVD, OLS, and MLE.

Li and Lee used the SVD method to estimate the parameters of the multiple population mortality model. However, that experiment discovers the lack of suitability in the SVD method for certain deformation forms. Enchev et al. used the MLE method, which has an extensive range of applications to almost all types of multiple population mortality models, to estimate parameters. Nevertheless, the multiple population mortality models have more parameters, and there will be errors that the converging value is mistaken for the optimal local solution during the calculation process of the Newton–Raphson iterative algorithm. For the sake of solving the problems above, we propose two methods to estimate the parameter of the Li–Lee model, which are two-step weighted least squares (TSWLS) and two-step singular-value decomposition (TSSVD).

In what follows, we derive and implement constraints for the parameters that allow us to solve the identifiability issues. The constraints are as follows:

$$\sum_x B(x) = 1,$$
$$\sum_t K(t) = 0,$$
$$\sum_x \beta(x, i) = 1,$$
$$\sum_t k(t, i) = 0,$$

(3)

in which we normalise the sums of the common age parameter to equate to unity. Additionally, the specific age parameters should also equate to unity for every population i. Furthermore, the common period parameter should sum to zero, and finally, the specific period parameters should sum to zero for each population i as well. The steps of TSWLS and TSSVD methods are as follows:

Step 1: using the Lee–Carter model, define

$$\ln(x, t, i) = \alpha(x, i) + B(x) * K(t) + \varepsilon(x, t, i), \quad (4)$$

and then estimate the parameters $\alpha(x, i)$, $B(x)$, and $K(t)$ for the combined dataset of all populations. Both the methods, TSWLS and TSSVD, have the same Step 1.

(1) Based on $\sum_t K(t) = 0$, get the estimation of $\alpha(x, i)$:t

$$\widehat{\alpha}_{x,i} = \frac{\sum_{t=t_L}^{t_U} \ln m_{x,t,i}}{t_U - t_L + 1}. \quad (5)$$

(2) Based on $\sum_x B(x) = 1$, get the estimation of $K(t)$:

$$\widehat{K}_t = \sum_{i=1}^{r} \sum_{x=x_L}^{x_U} w_i \times \left[\ln(m_{x,t,i}) - \widehat{\alpha}_{x,i}\right]. \quad (6)$$

in which $t = t_L, \ldots, t_U$, $i = 1, \ldots, r$, $x = x_L, \ldots, x_U$, and $i = 1, \ldots, r$; w_i is the weight of the group i, that is,

$\sum_{i=1}^{r} w_i = 1$; in this article, the weight of every group is $w_i = 1/r$.

(3) Based on $[\ln(m_{x,t,i}) - \alpha_{x,i}]$, use WLS to obtain the estimation of $B(x)$:

$$\widehat{B}_x = \frac{\sum_{i=1}^{r} \sum_{t=t_L}^{t=t_U} w_i \left[\widehat{K}_t \times \left(\ln m_{x,t,i} - \widehat{\alpha}_{x,i}\right)\right]}{\sum_{t=t_L}^{t_U} \widehat{K}_t^2}. \quad (7)$$

Step 2: using the Li–Lee model, define

$$\ln m(x, t, i) = \widehat{\alpha}(x, i) + \widehat{B}(x) * \widehat{K}(t) + \beta(x, i) * k(t, i). \quad (8)$$

in which the estimated $\alpha(x, i)$, $B(x)$, and $K(t)$ are fixed as constant values.

(4) Based on $\sum_x B(x, i) = 1$, obtain the estimation of $k(t, i)$:

$$\widehat{k}_{t,i} = \sum_{x=x_L}^{x_U} \left[\ln\left(m_{x,t,i}\right) - \widehat{\alpha}_{x,i} - \widehat{B}_x \times \widehat{K}_t\right]. \quad (9)$$

(5) Based on $[\ln(m_{x,t,i}) - \widehat{\alpha}_{x,i} - \widehat{B}_x \times \widehat{K}_t]$, use OLS to obtain the estimation of $\beta(x, i)$:

$$\widehat{\beta}_{x,i} = \frac{\sum_{t=t_L}^{t_U} \left[\widehat{k}_{t,i} \times \left(\ln m_{x,t,i} - \widehat{\alpha}_{x,i}\right)\right] - \widehat{B}_x \sum_{t=t_L}^{t_U} K_t \widehat{k}_{t,i}}{\sum_{t=t_L}^{t_U} \widehat{k}_{t,i}^2}. \quad (10)$$

The above step is the parameter estimation process of the TSWLS method, and the process of the TSSVD method can be directly used by singular-value decomposition.

2.3. Mortality Graduation Method.

Although the Li–Lee model, as Kang et al. [17] reveal, is one of the classical methods among the multiple population mortality models, the fluctuation of crude mortality data defects the fitting performance. Therefore, graduating the crude mortality data before fitting the model is necessary. In this paper, we apply the two-dimensional beta kernel density method, which can guarantee the reasonable smoothing results on both age and time dimensions, to graduate the crude mortality [18]. The two-dimensional beta kernel density method requires three steps, as follows:

Step 1: modelling two-dimensional beta kernel density:

(1) Define the distribution function of the numbers of death:

$$d(x, y) \sim \text{Bin}[e(x, y), q(x, y)], \quad (11)$$

in which x is the age, y is the year, at the age x in the year y, $d(x, y)$ is the numbers of death satisfying the binomial distribution, $e(x, y)$ is the exposure, and $q(x, y)$ is the actual mortality.

(2) Define the function of two-dimensional beta kernel density:

$$k_{h_Z}(z; m_Z) = \left(z - a_Z + \frac{1}{2}\right)^{(m_Z - a_Z + (1/2))/(h_Z(c_Z + 1))}$$
$$\cdot \left(b_Z + \frac{1}{2} - z\right)^{(b_Z + (1/2) - m_Z)/(h_Z(c_Z + 1))}, \quad z \in \mathcal{Z}, \tag{12}$$

in which Z is a two-dimensional random variable that has the value space of $\mathcal{Z} = \{a_Z, b_Z\}$, $c_Z = b_Z - a_Z$, and h_z is the bandwidth. The above formula can be standardized as

$$K_{h_Z}(z; m_Z) = \frac{k_{h_Z}(z; m_Z)}{\sum_{\omega \in Z} k_{h_Z}(\omega; m_Z)}, \quad z \in \mathcal{Z}. \tag{13}$$

(3) Estimate the crude mortality:

$$\hat{q}(x, y) = \sum_{u \in \mathcal{X}} \sum_{v \in \mathcal{Y}} K_{h_X, h_Y}(u, v; m_X = x, m_Y = y)\dot{q}(u, v),$$
$$(x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{14}$$

in which $\dot{q}(x, y)$ is defined as the crude mortality that is accord with the $q(x, y)$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$. Otherwise,

$$K_{h_X, h_Y}(x, y; m_X, m_Y) = K_{h_X}(x; m_X) K_{h_Y}(y; m_Y),$$
$$(x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{15}$$

Step 2: setting the adaptive bandwidth:

The formula of bandwidth adaptation is

$$h_Z(z; s_Z) = h_Z[l_Z(z)]^{s_Z}, \quad z \in \mathcal{Z}, \tag{16}$$

in which h_Z stands for the global bandwidth factor and s is a sensitive parameter as the local bandwidth factor, $s \in [0, 1]$. The reliability function $l_Z(z)$ can determine the numerical ranges of the local bandwidth factor, and at the same time, reliability function is limited by the local bandwidth factor so that extreme values do not occur.

Mazza and Punzo used variation coefficient (VC) to measure $l_Z(z)$, which is as follows:

$$VC(z) = \frac{\sqrt{e(z)\bar{q}(z)[1 - \bar{q}(z)]}}{e(z)\bar{q}(z)}, \quad z \in \mathcal{Z}. \tag{17}$$

When the reliability function satisfies the condition of $[l_Z(z)]^{s_Z} \in [0, 1]$,

$$[l_Z(z)] = \frac{VC(z)}{\sum_{\omega \in \mathcal{Z}} VC(\omega)}, \quad z \in \mathcal{Z}. \tag{18}$$

Step 3: selection of sensitive parameters:

In the process of mortality graduation of the two-dimensional beta kernel, the selection of bandwidth is realised by minimizing CV statistics, whose formula is

$$CV(h_Z) = \sum_{z \in \mathcal{Z}} \text{res}^2[\dot{q}(z), \hat{q}_{-z}(z)], \tag{19}$$

in which $\text{res}[\dot{q}(z), \hat{q}_{-z}(z)]$ is the residual in z and $\dot{q}(z)$ is the rude mortality in z. The formula of $\hat{q}_{-z}(z)$ can be expressed as

$$\hat{q}_{-z}(z) = \sum_{\substack{v \in Z \\ v \neq zZ}} \frac{k_{h_Z}(v; m_Z = z)}{\sum_{\substack{w \in Z \\ w \neq z}} k_{h_Z}(w; m_Z = z)} \dot{q}(z). \tag{20}$$

Furthermore, the proportional difference form of residual in mortality graduation is commonly used, that is,

$$\text{res}[\dot{q}(z), \hat{q}_{-z}(z)] = \frac{\hat{q}_{-z}(z)}{\dot{q}(z)} - 1. \tag{21}$$

2.4. Fitting and Forecasting. Before forecasting the mortality out-of-sample, it is necessary to test the fitting performance. The average absolute percentage error (MAPE) is used frequently [19] whose expression is

$$MAPE_i = \frac{1}{x_U - x_L + 1} \frac{1}{t_U - t_L + 1} \sum_{x = x_L}^{x_U} \sum_{t = t_L}^{t_U} \left|\frac{\hat{m}_{x,t,i} - m_{x,t,i}}{m_{x,t,i}}\right|. \tag{22}$$

Forecasting the mortality should model the time-effect factors first. The Li–Lee model includes both the common time-effect factor, $K(t)$, and the specific time-effect factor, $k(t, i)$. They are modeled as

$$\hat{K}_t = \theta + \hat{K}_{t-1} + \varepsilon_t,$$
$$\hat{k}_{t,i} = \phi_0 + \phi_1 \hat{k}_{t-1,i} + \varepsilon_{t,i}. \tag{23}$$

in which $K(t)$ is the processes of a random walk with drift and $k(t, i)$ explains the autoregressive process with one order. On this basis, we get the following mortality prediction formula:

$$\hat{m}_{x,t,i} = \dot{m}_{x,s,i} \exp\left[\hat{B}_x(\hat{K}_t - K_s) + \hat{\beta}_{x,i}(\hat{k}_{t,i} - k_{s,i})\right]. \tag{24}$$

The forecasting results can not only reflect the trend of mortality decline but also ensure the long-term consistency of mortality among different population groups.

3. Data and Research Scheme

We perform our analysis based on data obtained from public resources of the National Bureau of Statistics of China (BSC)

and Human Mortality Database (HMD). The typical dataset consists of the numbers of deaths and the central exposure. The age period range considered is from 0 up to 100 years (101 consecutive ages in total) and from 1994 up to the year 2014 (21 years in total), respectively.

To compare advantages of methods used in this paper, we make the following designs: (1) due to the lower number of deaths, we select male mortality as the research samples; (2) in order to keep up with the standard of HMD data, we use two-dimensional beta kernel density method to graduate the mortality in mainland China; (3) we select the data from East Asian countries or regions of mainland China, Hong Kong (China), and Japan to compare our method with the Lee–Carter model; and (4) we test the robustness of our method by replacing the data of crude mortality with the graduated mortality.

4. Quantitative Comparison of Fitting Performance

In this paper, we compare models between the multiple population stochastic mortality model and single population Lee–Carter model, so do methods between the TSWLS method and the TSSVD method in the Li–Lee model, from the perspective of horizontal comparison. Because major research on mortality modelling of China uses the crude data, this paper first applies the same mortality data for analysis. Also, the MLE method cannot converge to a stable value all the time in the Li–Lee model, so it is not included in our study. Based on methods of the second part in this paper, we can calculate the MAPE values.

Table 1 shows the MAPE values of the three methods in mainland China, Hong Kong (China), and Japan, in which a smaller MAPE value indicates a better fitting performance. In general, the Li–Lee model with the TSSVD method has the lowest MAPE value of 20.22% in all the years from 1995 to 2014, while the Lee–Carter model has the highest MAPE value of 21.18%. Therefore, we can infer that the fitting performance of multiple population mortality model under the two methods for parameter estimation is better than that under the single population mortality model. Next, we plan to analyse the fitting performance of different population groups below. For mainland China, the Lee–Carter model gets the lowest MAPE value, while the Li–Lee model under the TSWLS method has the highest in all the years from 1995 to 2014. However, in most of the five-year periods, the Li–Lee model does better than the Lee–Carter model on fitting performance. For example, the Li–Lee model with TSSVD method has the best performance in the years of 1995~2000 and 2006~2010. By contrast, the Lee–Carter model only shows a surprisingly lowest MAPE value in the years of 2011~2014, which brings about the best fitting performance in the whole historical period. In Hong Kong (China), the Li–Lee model with TSSVD method always displays the lowest MAPE value during 1995 and 2014, and only in 1995~2000, it gets slightly higher value than the TSWLS method. As for Japan, the last population group, the Li–Lee model with TSSVD method is the one that has the best performance among any periods of years. In terms of numerical values, Japan's MAPE value ranges from 5% to 9%, while that in mainland China is from 27% to 49% and 18% to 25% in

Hong Kong (China). From the above analysis, we can assume a preliminary conclusion that when the higher the smoothness is, the better fitting performance the mortality model will have.

Then, we use the two-dimensional beta kernel density method, aiming to derive mortality with higher smoothness in the dimension of the period, to graduate the data in mainland China. Although the mortality from the HMD database is graduated, the method is constrained within the dimension of age, which ignores the impact of the period trend. To not only test the robustness of the estimation but also explain whether the smoother population mortality data can improve the fitting performance of the model, we use the two-dimensional smoothing method for the population mortality in mainland China.

Table 2 shows the MAPE values of three methods in mainland China, Hong Kong (China), and Japan, respectively. From Table 2, we notice that the fitting mortality model based on the graduated data makes a considerable positive impact on mainland China, which reduces the MAPE value range from the 27%~49% to 9%~15% but has less effect on the fitting performance of Japan and Hong Kong (China). It means that the smoothness of the data can significantly improve the fitting effect of the Li–lee model, with both methods, and the Lee–Carter model. At the same time, we can observe that the three methods show a stable estimation effect in all populations. Consistent with Table 1, the Li–Lee model with TSSVD method gets the lowest MAPE value, while the Lee–Carter model has the highest MAPE value. Consequently, we can make advancement in model fitting performance by using the graduated mortality data and solving the multiple population mortality models with the TSSVD method, especially for forecasting of mortality in mainland China.

5. Quantitative Comparison of Forecasting Performance

According to equation (11), we can obtain the numerical values of out-of-sample mortality from 2015 to 2050. In this paper, we take the 30-year-old population as an example and then display the values in Table 3. All three types of methods could show a decreasing trend of population mortality in the future, but there are some differences in detail. From the perspective of different population groups, the mortality of the Lee–Carter model in mainland China declines rapidly from 0.97‰ in 2015 to 0.28‰ in 2050, while the forecasting values of the other two methods are very close to each other which decline from 1.00‰ in 2015 to 0.50‰ in 2050 together. For Hong Kong (China), the mortality forecasting results by all methods are similar, except the Li–Lee model with TSWLS, which has slightly higher values. Compared with that in mainland China, the future values of population mortality in Japan is also close under the two methods of the Li–Lee model, but the Lee–Carter model has higher forecasting values. The reason for the above situation is that the multiple population mortality model is systematic, which can consider the interrelationship between different population groups and make the expected results reasonable in the long term. Because the single population of the Lee–Carter model assumes that the mortality decreases with a fixed constant, the forecasting values of mortality among different population groups will cross or

TABLE 1: MAPE values of all the models on crude mortality data (%).

	1995~2000	2001~2005	2006~2010	2011~2014	All years
Mainland China					
Li–Lee model (TSWLS)	28.76	**27.45**	40.76	47.19	35.17
Li–Lee model (TSSVD)	**28.75**	27.61	40.03	45.91	34.77
Lee–Carter model	29.02	27.99	**39.32**	**44.24**	**34.42**
Hong Kong (China)					
Li–Lee model (TSWLS)	**23.61**	19.21	20.15	23.03	21.05
Li–Lee model (TSSVD)	23.64	**18.66**	**18.75**	**22.57**	**20.49**
Lee–Carter model	23.13	19.15	20.21	24.10	21.16
Japan					
Li–Lee model (TSWLS)	5.52	4.82	5.10	6.93	5.49
Li–Lee model (TSSVD)	**5.49**	**4.70**	**4.81**	**6.91**	**5.39**
Lee–Carter model	6.99	7.59	8.46	8.65	7.95
Sum of above					
Li–Lee model (TSWLS)	19.30	17.16	22.00	25.72	20.57
Li–Lee model (TSSVD)	**19.29**	**16.99**	**21.20**	**25.13**	**20.22**
Lee–Carter model	19.71	18.24	22.66	25.66	21.18

TABLE 2: MAPE values of all the models on graduated mortality data (%).

	1995~2000	2001~2005	2006~2010	2011~2014	All years
Mainland China					
Li–Lee model (TSWLS)	7.36	**6.26**	6.96	13.55	8.39
Li–Lee model (TSSVD)	7.76	6.62	7.41	**9.13**	**7.86**
Lee–Carter model	**7.35**	6.28	**6.89**	14.39	8.54
Hong Kong (China)					
Li–Lee model (TSWLS)	23.39	19.25	20.03	23.30	21.04
Li–Lee model (TSSVD)	23.53	**18.68**	**18.70**	**22.23**	**20.38**
Lee–Carter model	**23.13**	19.15	20.21	24.10	21.16
Japan					
Li–Lee model (TSWLS)	**5.50**	4.68	4.71	**6.88**	**5.36**
Li–Lee model (TSSVD)	5.55	**4.68**	**4.70**	6.98	5.38
Lee–Carter model	6.99	7.59	8.46	8.65	7.95
Sum of above					
Li–Lee model (TSWLS)	**12.08**	10.06	10.57	14.58	11.60
Li–Lee model (TSSVD)	12.28	**9.99**	**10.27**	**12.78**	**11.21**
Lee–Carter model	12.49	11.01	11.85	15.71	12.55

TABLE 3: Numerical values of out-of-sample mortality from 2015 to 2050 (‰).

	2015	2020	2025	2030	2035	2040	2045	2050
Mainland China								
Li–Lee model (TSWLS)	0.98	0.89	0.80	0.72	0.65	0.58	0.52	0.47
Li–Lee model (TSSVD)	1.00	0.91	0.82	0.73	0.66	0.59	0.53	0.48
Lee–Carter model	0.97	0.81	0.68	0.57	0.48	0.40	0.34	0.28
Hong Kong (China)								
Li–Lee model (TSWLS)	0.59	0.54	0.49	0.44	0.39	0.35	0.32	0.28
Li–Lee model (TSSVD)	0.58	0.52	0.46	0.42	0.37	0.33	0.30	0.27
Lee–Carter model	0.57	0.51	0.46	0.41	0.37	0.33	0.30	0.27
Japan								
Li–Lee model (TSWLS)	0.65	0.56	0.50	0.45	0.40	0.36	0.32	0.29
Li–Lee model (TSSVD)	0.67	0.60	0.53	0.48	0.43	0.39	0.35	0.31
Lee–Carter model	0.68	0.66	0.63	0.61	0.59	0.57	0.56	0.54

deviate abnormally in the long term. Therefore, the multiple population mortality model is more suitable for forecasting mortality than the single population mortality model in the long term, also effectively improves the reliability of the prediction value.

Next, we use three figures to show the forecasting performance vividly. From the perspective of different methods, we further analyse the evidence of rationality of the population mortality model. Figure 1 shows the trend of mortality in different populations under the Lee–Carter
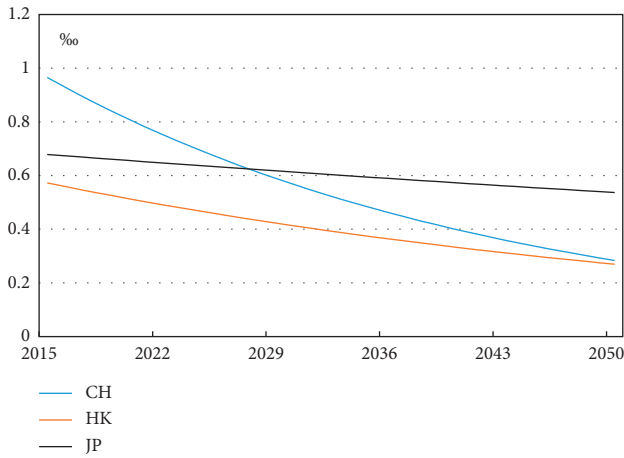
Figure 1: Trends in mortality based on the Lee–Carter model from 2015 to 2050.

Figure 2: Trends in mortality based on the Li–Lee model with TSWLS from 2015 to 2050.

Figure 3: Trends in mortality based on the Li–Lee model with TSSVD from 2015 to 2050.

model in 2015–2050. We can find two unreasonable problems in Figure 1. Firstly, the curve of mortality trend of mainland China intersects with Japan in 2038. On the other hand, it keeps below that of Japan in the following years. The reason is that the Lee–Carter model only considers the velocity of single population mortality decline in historical data. Due to the higher level of mortality in mainland China, the velocity of mortality decline rate is faster than that of Japan, which has finished the historical stage of rapid decline of mortality. If the mortality rates of the two population groups continue to fall at the current speed, the life expectancy in mainland China will surpass that in Japan soon after, though it is contradicting to the biological law of human beings. As time goes by, the gap of mortality curves between Hong Kong (China) and Japan is gradually widening with a trumpet shape, which is also attributed to the defects in the hypothesis of a single mortality model.

Figure 2 demonstrates the forecasting values of mortality based on the Li–Lee model with the TSWLS method, the above problems in the single mortality model can be fixed through the multiple mortality model. The mortality of the three groups shows a consistent trend from 2015 to 2050 in Figure 2, similar to the historical experience of decline. Mortality of mainland China maintains a higher level than the others, and the gap is narrowing. However, there will be no crossing in the future. We can tell from Figure 2 that the mortality gap between Hong Kong (China) and Japan is relatively small and even overlapped in a short period, which is not accord with the current mortality relationship between the two populations.

How to address this issue can be found in Figure 3. Figure 3 witnesses forecasting values of mortality based on the Li–Lee model with the TSSVD method from 2015 to 2050, in which the forecasting values of mortality on mainland China is close to the values in Figure 2. Yet, the TSSVD method for forecasting the mortality of two population groups in Hong Kong (China) and Japan has more advantages, as reflecting the mortality gap decreasing over time but no crossing in the short term. Overall, we conclude that the multiple population model is better than the single
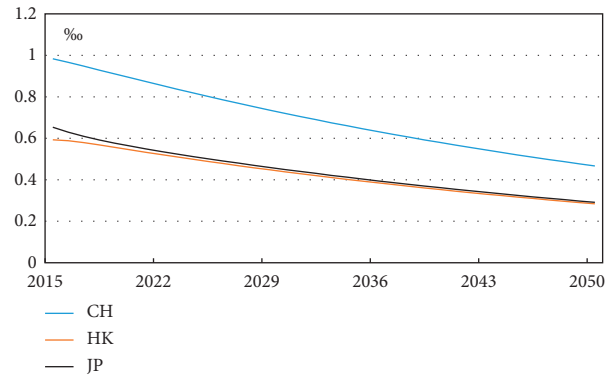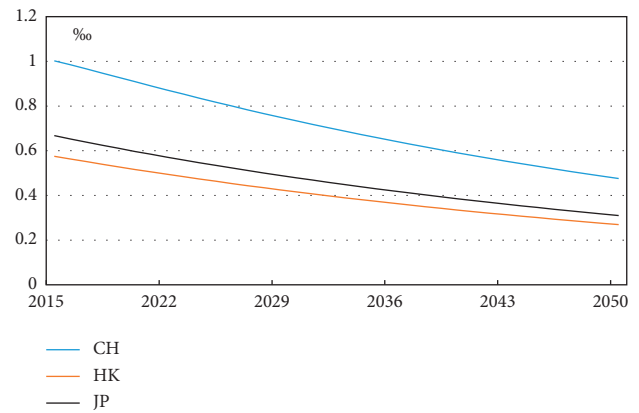
population model in terms of forecasting of mortality, and the TSSVD method is more suitable for the multiple population model than the TSWLS method.

6. Conclusion

This paper reviews the progress of the multiple population stochastic mortality model and finds some defects in parameter estimation, so that it proposes an effective method to improve the parameter estimation. We set up multiple population groups, using the data of mainland China, Hong Kong (China), and Japan, to test fitting performance and forecasting performance, and draw the following conclusions: (1) for the parameter estimation method, TSWLS and TSSVD methods, in a multiple population stochastic mortality model, can avoid some problems, like maximum likelihood estimation is not converge, or the result is local optimal caused by the more parameters. The two methods are simple and easy to understand and are the tools for the application in the multiple population mortality model. (2) For fitting performance, the multiple population mortality model is robustness based on TSWLS and TSSVD methods. Additionally, the fitting performance can be significantly improved based on graduated data, which illustrates that the multiple population mortality model is more suitable for

smooth mortality. (3) For forecasting performance, the multiple population stochastic mortality model with additional period effect factors can obtain consistent values among different population groups, and the mortality ratio between any two populations can converge to a fixed constant in a certain period which obeys the regular of human biological characteristics.

Data Availability

The data used to support the findings of this study are from the Human Mortality Database and China National Bureau of Statistics.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] R. D. Lee and L. R. Carter, "Modeling and forecasting U.S. mortality," *Journal of the American Statistical Association*, vol. 45, no. 6, pp. 4–19, 1992.

[2] J. Bongaarts and G. Feeney, "How long do we live?" *Population and Development Review*, vol. 28, no. 1, pp. 13–29, 2002.

[3] J. R. Wilmoth, "Mortality projections for Japan: a comparison of four methods," in *Health and Mortality Among Elderly Populations*, G. Caselli and A. Lopez, Eds., pp. 266–287, Clarendon Press, Oxford, UK, 1996.

[4] N. Li, R. Lee, and S. Tuljapurkar, "Using the Lee–Carter method to forecast mortality for populations with limited data," *International Statistical Review*, vol. 72, no. 1, pp. 19–36, 2004.

[5] A. E. Renshaw, S. Haberman, J. Insmatheco et al., "A cohort-based extension to the Lee-Carter model for mortality reduction factors," *Insurance: Mathematics and Economics*, vol. 38, no. 3, pp. 556–570, 2006.

[6] A. J. G. Cairns, D. Blake, K. Dowd et al., "A quantitative comparison of stochastic mortality models using data from England and Wales and the United States," *North American Actuarial Journal*, vol. 13, no. 1, pp. 1–35, 2007.

[7] A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah, "Morality density forecasts: an analysis of six stochastic mortality models," *Insurance: Mathematics and Economics*, vol. 48, no. 3, pp. 355–367, 2011.

[8] W. Yu, Y. Yong, G. Guan, Y. Huang, W. Su, and C. Cui, "Valuing guaranteed minimum death benefits by cosine series expansion," *Mathematics*, vol. 7, no. 9, p. 835, 2019.

[9] Z. Zhang, Y. Yaodi, and W. Yu, "Valuing equity-linked death benefits in general exponential Lévy models," *Journal of Computational and Applied Mathematics*, vol. 365, Article ID 112377, 2020.

[10] L. R. Carter and R. D. Lee, "Modeling and forecasting US sex differentials in mortality," *International Journal of Forecasting*, vol. 8, no. 3, pp. 393–411, 1992.

[11] N. Li and R. D. Lee, "Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method," *Demography*, vol. 42, no. 3, pp. 575–594, 2005.

[12] J. S.-H. Li and M. R. Hardy, "Measuring basis risk in longevity hedges," *North American Actuarial Journal*, vol. 15, no. 2, pp. 177–200, 2011.

[13] T. Kleinow, "A common age effect model for the mortality of multiple populations," *Insurance: Mathematics and Economics*, vol. 63, pp. 147–152, 2015.

[14] V. Enchev, T. Kleinow, and A. J. G. Cairns, "Multi-population mortality models: fitting, forecasting and comparisons," *Scandinavian Actuarial Journal*, vol. 2017, no. 4, pp. 1–24, 2017.

[15] J. Li and J. Liu, "A logistic two-population mortality projection model for modelling mortality at advanced ages for both sexes," *Scandinavian Actuarial Journal*, vol. 2019, no. 2, pp. 97–112, 2019.

[16] C. C. L. Tsai and Y. Zhang, "A multi-dimensional Bühlmann credibility approach to modeling multiple population mortality rates," *Scandinavian Actuarial Journal*, vol. 2019, no. 1, pp. 406–443, 2019.

[17] M. Kang, Y. Liu, S. H. Li, and W. S. Chan, "Mortality forecasting for multiple populations: an augmented common factor model with a penalized log-likelihood," *Communications in Statistics: Case Studies, Data Analysis and Applications*, vol. 4, no. 3-4, pp. 118–141, 2019.

[18] A. Mazza and A. Punzo, *Graduation by Adaptive Discrete Beta Kernels, Classification and Data Mining*, Springer, Berlin, Germany, 2013.

[19] T. Lin, C.-W. Wang, and C. C.-L. Tsai, "Age-specific copula-AR-GARCH mortality models," *Insurance: Mathematics and Economics*, vol. 61, pp. 110–124, 2015.