

RESEARCH ARTICLE

Open Access



A multi-country comparison of stochastic models of breast cancer mortality with P-splines smoothing approach

Sumaira Mubarik¹, Ying Hu¹ and Chuanhua Yu^{1,2*} 

Abstract

Background: Precise predictions of incidence and mortality rates due to breast cancer (BC) are required for planning of public health programs as well as for clinical services. A number of approaches has been established for prediction of mortality using stochastic models. The performance of these models intensely depends on different patterns shown by mortality data in different countries.

Methods: The BC mortality data is retrieved from the Global burden of disease (GBD) study 2017 database. This study include BC mortality rates from 1990 to 2017, with ages 20 to 80+ years old women, for different Asian countries. Our study extend the current literature on Asian BC mortality data, on both the number of considered stochastic mortality models and their rigorous evaluation using multivariate Diebold-Marino test and by range of graphical analysis for multiple countries.

Results: Study findings reveal that stochastic smoothed mortality models based on functional data analysis generally outperform on quadratic structure of BC mortality rates than the other lee-carter models, both in term of goodness of fit and on forecast accuracy. Besides, smoothed lee carter (SLC) model outperform the functional demographic model (FDM) in case of symmetric structure of BC mortality rates, and provides almost comparable results to FDM in within and outside data forecast accuracy for heterogeneous set of BC mortality rates.

Conclusion: Considering the SLC model in comparison to the other can be obliging to forecast BC mortality and life expectancy at birth, since it provides even better results in some cases. In the current situation, we can assume that there is no single model, which can truly outperform all the others on every population. Therefore, we also suggest generating BC mortality forecasts using multiple models rather than relying upon any single model.

Keywords: Breast cancer mortality, Goodness of fits, Lee carter model, Multivariate Diebold-Marino-test, P-splines

Background

The modeling and projections of future cancer related incidence and mortality rates are essential for development of public health programs and clinical amenities [1]. Breast cancer (BC) ranked among the top global burden of diseases, and it threaten the health all over the world.

Among Asian women, the BC is consider the second leading cause of cancer related morbidity and mortality. According to previous studies, the BC among Asian women constitute approximately 40% of all BC diagnosed worldwide [2–4]. Asia has much higher mortality to incidence ratio of BC than western countries [2, 3]. A study conducted on BC differences in Asian regions, reported that BC increasing among Asian women. Cause of this increasing rate are associated with higher prevalence of BC risk factors like, delayed childbirth, increased obesity [5].

* Correspondence: yuchua@whu.edu.cn

¹Department of Epidemiology and Biostatistics, School of Health Sciences, Wuhan University, 185 Donghu Road, Wuhan 430071, Hubei, China

²Global Health Institute, Wuhan University, Wuhan 430071, Hubei, China



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Other risk factors that may contribute to increase mortality risk due to BC are include aging process, high body mass index, alcohol consumption and low physical activity [2–4, 6, 7]. Though, a number of cumulative exposures contribute to raise the BC mortality rates, therefore it is expected that these rates follow a smooth curvilinear pattern and future mortality rates are predictable based on these pattern. The rapid change pattern of BC rates over the historical time is remain a challenge for its prediction. The Asian countries selected in this study are suffering from increase BC burden and also having similar circumstances related to over population, poverty, socio-cultural background, and insufficient access to diagnosis, advanced screening and proper treatment [4]. Moreover, due to lack of proper statistical registries system in the developing countries, statisticians always encounter the problem of insufficient and unsatisfying Asian demographic and disease registration data sets. This data deficiency raise the problems when it used in stochastic mortality models. Solving those problems bring lots of research passion in this field. In light of limited data access combined with poor quality, less technically refined methods are available for Asian countries compared with developed countries. One of the methods which could be applied on Asian data is the stochastic population approach. This approach has been used in various studies to generate the future mortality forecasts for different western countries [8].

Till now, variety of forecasting techniques for prediction of mortality trends have been tested and proposed consistently [9, 10]. In this regard, the work of Lee and Carter [11] is incontrovertible contribution in literature of mortality forecasting. Their contributions based on generalization of different stochastic mortality models and its extended versions. The basic Lee-Carter (LC) model uses age and time parameter to estimate the future mortality trends by using principal component approach. This initial model now considered a milestone in the stochastic modeling and forecasting of trends in mortality data. After this, various extensions and modifications were proposed. These extensions vary in number of fundamental features, for example, smoothness assumption, causes of randomness, different estimation methods and addition of cohort effects. Now these methodologies has become widely used to attain a broader interpretation and to capture the main structures of vigorous of mortality intensity [12–17]. However, in context of different stochastic mortality models comparisons, many studies confirm that there is no single model among them which clearly dominates the others according to measured evaluation standards [18–20]. Lee and Tuljapurkar [8] proposed a new method in case of few observations at uneven intervals, and they applied it to China and South Korea data. Hyndman and Ullah [15] developed a

more general method by treating the underlying demographic process as functional data, employing the functional principal components to extract more than one explaining components and providing robust estimation and forecast. Previously, functional demographic models on breast cancer mortality data has been used to estimate the future trends in breast cancer mortality for United State and England-Wales [21].

This study apply the three stochastic mortality models belong to family of generalized LC with smoothing p-splines approach to four Asian countries' data sets, evaluate the performance of these three methods, compare the similarities of different Asian regions and accordingly propose potential improvements. The models that apply in current study named as smoothedleecarter (SLC) model, Functional demographic model (FDM) and Booth-Maindonald-Smith (BMS) model. Though, the prediction of mortality for future by different stochastic methods have been extensively reviewed elsewhere [22, 23]. This paper covers in-sample and out-sample evaluation of different stochastic smoothed mortality methods by applying them to BC mortality of four Asian countries.

Reference to current study data sets, the parameter estimation of LC models is based on least square method using singular value decomposition (SVD) algorithm of the matrix of the log age-specific observed BC death rates. The BC mortality data in form of death counts and exposures to risk have to fill a rectangular matrix. Hereafter, it will be denoted as $M_{x, t}$ that is observed BC mortality rates at age x during calendar year t . it is achieved by the ratio between the number of deaths at age x during year t , ($D_{x, t}$), from an exposure to risk, that is, the number of person years from which $D_{x, t}$ happened ($E_{x, t}$). As regards the Asian women population data set on the basis of the BC mortality rates from 20 to 80+ year of age, presented by countries and individual year, it shows curvilinear mortality behavior and suggest the smoothing is suitable for this data (see Fig. 1). We can see the random variations in the data, especially for ages below 50, where the reductions in the BC mortality rates are solid. Besides, we can also observe the BC mortality irregularities for older ages. Therefore, the model estimation on four Asian countries' data set follow the smoothing technique for model fitting to control the random variations in the data, otherwise we could not get the resulting BC mortality rates much reliable.

The aim of current study based in achieving the following specific objectives: (a) To estimate the BC mortality trends by using three stochastic mortality models with smoothing P-splines approach for four Asian countries' data sets. (b) To evaluate both the

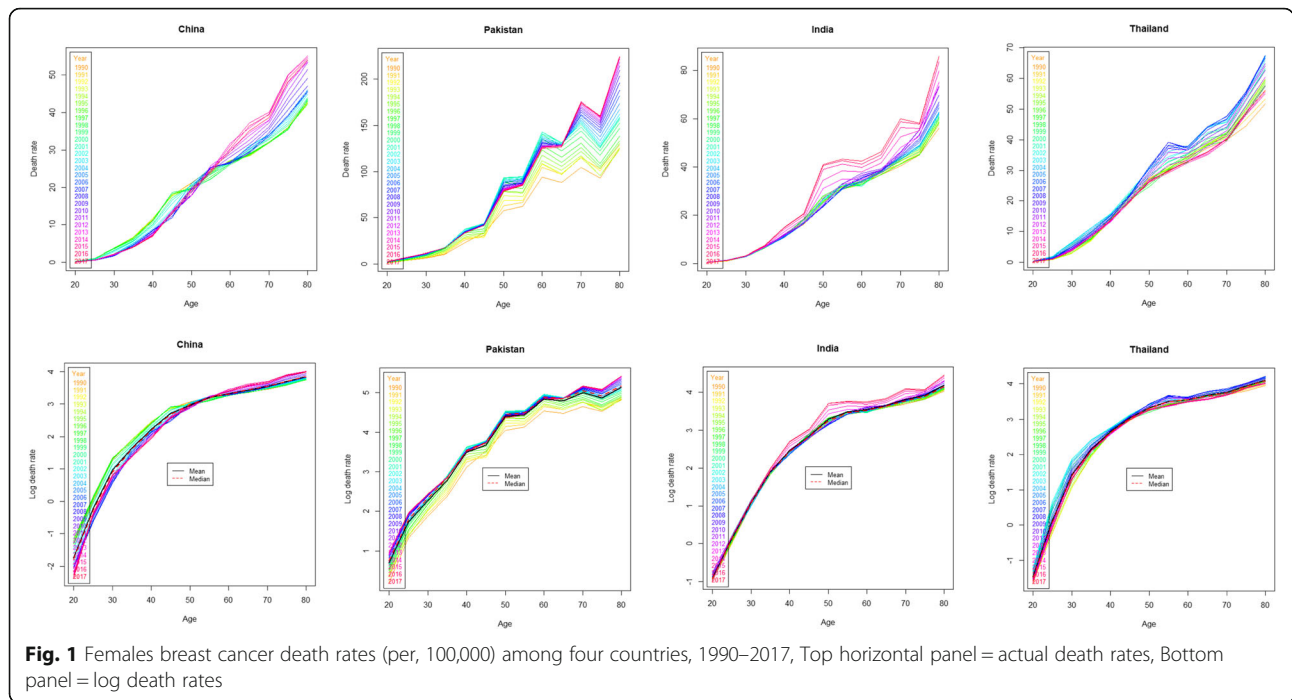


Fig. 1 Females breast cancer death rates (per, 100,000) among four countries, 1990–2017, Top horizontal panel = actual death rates, Bottom panel = log death rates

goodness of fit and predictive ability (In-sample and Out-sample) of the three models using advance quantitative forecast accuracy test as well as by range of graphical display on each Asian country’s data set. To the best of our knowledge, this is the first study to extend the current literature on Asian mortality data due to breast cancer, on both the number of considered stochastic mortality models and their rigorous evaluation using multivariate Diebold-Marino test and by range of graphical analysis for multiple countries under consideration. Our research to some extent validates the existing results but also presents a more comprehensive view. The rest of the current study is organized in the following sections. The methodology is addressed in section 2 and results of empirical analysis is presented in section 3. Discussion and conclusion is merged in Section 4.

Methods

In this section, analysis methods of BC mortality modeling is discussed. P-splines approach for Smoothing BC mortality rates were employed first. Then, three stochastic mortality models were fitted to the smoothed BC mortality data. Goodness of fits tests included graphical displays and multivariate Diebold–Mariano (DM) test was used to evaluate the model performance.

Smoothing data using P-splines approach

The violation of homogeneity assumption due to presence of outliers in mortality data may often under or overestimate the actual mortality estimates. Particularly, the mortality rates in older ages have high variability

because of small number of cases in the population. This high variability can create the problems in estimation of mortality rates for older ages. Therefore, use of smoothing techniques for such a data can avoid this deficiency of facts. Otherwise, the heavy difference at older ages influences the fitting of mortality models and can over or under fit the model [24]. To overcome this situation various studies suggested the Penalized splines which also known P-splines is now widely used method of smoothing in stochastic mortality models and generalized linear models [9, 25–27]. Concerning to its application, the following main features of the methodology is adopted: Firstly, the B-splines used as the basis for the regression modeling; secondly, the regression coefficients with different penalties is used for modifying the log-likelihood. In line with the previous literature, the intensity of mortality is decomposed as follow: [26, 28].

$$\ln \mu_x(t) = \sum_{i,j} \Theta_{i,j} B_{i,j}(x, t) \tag{1}$$

The two-dimensional B-splines $B_{i, j}$ for specific age and calendar year x and t respectively with regularly spaced knots is used. The parameters $\Theta_{i, j}$ ’s, to be estimated based on the data set. As suggested by a study, the penalty can be used based on finite differences of the coefficients of the adjacent B-spline to limit the influence of the knots on the fitted value and this approach is called P-spline [25]. The calculation of penalties for both dimensions age x and calendar year t , depend on sums of following components $(\Theta_{i, j} - 2\Theta_{i-1, j} + \Theta_{i-2,$

$j)^2$. The selection of weight coefficients to attach with each penalty should depend on historical data. According to some studies, the P-splines are less apparent to the actuaries, particularly due to the selection of penalty parallels to the assessment of the future pattern of mortality. In order to smooth the mortality data, a study has reported the limits of using a penalized spline. As the penalty depends on the parameterization, so smoothed value is vary with respect to its choice [10]. Its solution is suggested in a study, to consider a direct smoothing and interchange the penalized differences in the contiguous coefficient with the penalized differences in the contiguous fitted values [28]. Recommended by a study, it is better to smoothing the data initial, rather than smoothing the fitted values. In this way, it become more convenient to impose the monotonic restriction on the smoothing data more simply [15].

The Lee–Carter (LC) model

The Lee-Carter (LC) model is backbone of all methodologies that used for mortality projection and it is key component of all actuarial literature related to mortality forecasting. The LC method (Lee and Carter [11]) for forecasting mortality rates uses principal components analysis to decompose the age-time matrix of central death or mortality rates into a linear combination of age and time parameters. The parameter of time is used in forecasting. LC has produced numerous variants and extensions. The two main variants of LC are Lee Miller (LM) and Booth-Maindonald-Smith (BMS) presented by lee and miller [29], and Booth et al. [12] respectively. These variants are collectively referred to as “LC methods”. A major extension of this approach uses functional data analysis. Which has presented in next sections. It further extended in various combinations of options that study variation in different dimension like age, period or cohort, and these are famous as “HU methods. Detail description of methods and extensions can be found elsewhere [23, 30].

With reference to the present study data set, the following Eq. (2) shows the LC model for BC mortality data. The logarithm of the observed BC mortality rates given for age x and year t is denoted as $M_{x,t}$. The sum of an age-specific component α_x represent the independent of time parameter. Another component that is product of a time-varying parameter k_t , and an age-specific component β_x , reflecting the general level of BC mortality and change of BC mortality at each age x with changes in general level of BC mortality respectively.

$$\ln(M_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t} \quad (2)$$

The error term is denoted as $\varepsilon_{x,t}$ and it is assume that it follow homoscedastic and normality assumption. In

order to find a least squares solution to the Eq. (2) the SVD method is used because of unavailability of regressors on right-hand side. To find the unique solution of parameter of Eq. (2) the following set of constrains is imposed on parameters of Eq. (2), firstly, the sum of the age specific coefficients is equal to one i.e. $\sum \beta_x = 1$ and secondly, the sum of the time varying parameter is equal to zero i.e. $\sum k_t = 0$. In order to forecast BC mortality by using the LC model, following two procedures were adopted in current study. First procedure is to estimate parameters α_x , β_x and k_t using historical BC mortality data. Second procedure is to estimated time-dependent parameter k_t . Model structure is based on a stochastic process which is follow autoregressive integrated moving average (ARIMA p, d, q) model scheme. It is commonly known as the standard Box and Jenkins methodology [31, 32]. Finally, the fitted ARIMA model was used to extrapolate k_t and to get a forecast of future BC mortality rates and then from these forecasts future life expectancy was derived to evaluate the model.

Booth-Maindonald-Smith (BMS) model

In LC model the age-specific component β_x , which can be defined in terms $\beta_x^{(1)}$ and $\beta_x^{(2)}$ are assumed to be constant over time and time varying parameter $k_t^{(2)}$ is assumed to be linear over time. Booth, Maindonald and Smith [12] changed these assumption in their study. In line with the prior literature, the BMS model was built by extended the LC model. The extension of model based on inclusion of more interaction terms between age x and year t . Model fitting period was restricting by improving the assumptions of homoscedastic and linearity of β_x and k_t respectively. However, in LC model only first terms of the SVD is used. While, BMS model can used “ n ” terms that allow the model to include second and higher order terms as well [22]. According to a study, any systematic variation in the residuals from fitting only the first term would be captured by the second and higher order terms [12]. With reference to current study data set, the BMS model can be presented as:

$$\ln(M_{x,t}) = \alpha_x^0 + \sum_{i=1}^n \beta_x^{(i)} k_t^{(i)} + \varepsilon_{x,t} \quad (3)$$

Where $M_{x,t}$ is the BC mortality rate for age x in calendar year t . α_x^0 is the intercept of the model which represent the effect of age-specific parameter. $\beta_x^{(i)}$ represent the “age interaction” parameter at age x . While, $k_t^{(i)}$ is the “time interaction” parameter that indicate the value in year t . The residual term $\varepsilon_{x,t}$ shows the value of error at age x and year t and approximation rank is denoted by n .

The purpose of inclusion of more interaction terms in the BMS model is to improve the model fit on data and to increase its capacity to explain more unexplained variations that could not account in LC model. The BMS model applied by a study on Australian data over the period 1907 to 1999. The study results reported the higher forecasted life expectancies and smaller forecast error by BMS model relative to the LC model. Another study also claim the more accurate forecasts and use of shorter fitting period by BMS model than the LC model [33, 34].

Functional Demographic Model (FDM)

Hyndman and Ullah [15] presented the modified version of the LC model, which is known as Functional Demographic Model (FDM). This methodology is based on combination of functional data analysis and nonparametric smoothing and robust statistics and it was proposed to forecast the age-specific mortality rates. Particularly, this approach allows for smooth functions of age, resulting the robust to outliers and offers a modeling framework convenient to fit to restrictions and other evidence. The methodology background of FDM based on generalization of LC method. In context of our study data set, let $y_t(x)$ represent the log of the observed BC mortality rates for specific age x and calendar year t i.e. $y_t(x) = \ln(M_{x,t})$ and we assume that there are underlying \mathbb{L}_2 continuous and smooth functions $\{f_t(x)\}$ such that

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

where, $\varepsilon_{t,i}$ is error term which is assume to be independent identically distributed (i.i.d) standard normal variable, and quantity of noise that vary with age x is denoted by $\sigma_t(x_i)$. The forecasting of log observed BC mortality rates $y_t(x)$ founded through following steps:

Step-1. First of all smoothed data set series was generated for each time t using penalized regression splines. The functions $f_t(x)$ for $x \in [x_1, x_p]$ from \mathbb{R}^2 , where $\mathbb{R}^2 = \{(x_i, y_t(x_i)) \mid (i, t) \in N_p \times N_n\}$. The functions $f_t(x)$ for $i = 1, 2, \dots, p$ was estimated for each time t i.e. $t = 1, 2, \dots, n$ by applying a nonparametric smoothing with constraint. It was assume that the function $f_t(x)$ is monotonically increasing for $x \geq c$ for some c (say 65 years), where c is reasonable threshold for mortality data. The estimated curves has less noise at older ages by imposing this condition (see Hyndman and Ullah [15], for more details).

Step 2. The smoothed curves $f_t(x)$ generated in above step-1 was decomposed by using the following basis function expansion [35].

$$f_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x), \quad x \in [x_1, x_p] \tag{5}$$

where $\mu(x)$ is a locational measurement (median curve) of $f_t(x)$, while Eigen functions $\phi_k(x)$ shows the main regions of variation or k th principal component function; $\{\beta_{t,k}\}$ are the coefficients (or corresponding principal component scores) which are uncorrelated by construction and k is the number of basis functions with $k < n$; Hyndman and Booth [36] found that $k = 6$ is sufficient to capture a substantial amount of variance in the data; $e_t(x)$ is the error function which $\sim N(0, \text{var}(x))$.

Step 3. By using Eq. (5), a univariate time series model to each principal component score $\{\beta_{t,k}\}$ was applied to obtain their future values.

Step 4. Based on fitted time series models obtained in step 3, the coefficients $\{\beta_{t,k}\}$, $k = 1, 2, \dots, K$, are forecasted for $t = n + 1, \dots, n + h$.

Step 5. The forecasted coefficients attained in the step 4 was applied in Eq. (5) to get the $f_t(x)$. Then $y_t(x)$ was forecasted from Eq. (4). In simple word, the $y_t(x_i)$ can be simplify by combining (4) and (5) and resulting equation was obtained as (6):

$$y_t(x_i) = \mu(x_i) + \sum_{k=1}^K \beta_{t,k} \phi_k(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \tag{6}$$

Specifically, the h -steps ahead forecasts of $y_{n+h}(x)$ can be obtained by using following formula (7):

$$y_{n+h}(x) = E[y_{n+h}(x) \mid I, \Phi] = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\phi}_k(x) \tag{7}$$

In above Eq. (7), the observed data is denoted as $I = \{y_t(x_i); t = 1, 2, \dots, n; i = 1, 2, \dots, p\}$ and Φ is the set of basis functions (or estimated set of functional principal components). The $\hat{\beta}_{n,k,h}$ corresponds to the h -step ahead forecast of $\beta_{n+h,k}$ having been estimated time series $\hat{\beta}_{1,k}, \dots, \hat{\beta}_{n,k}$, using Eq. 7.

Step 6. Lastly, the variance of error terms in Eqs. (4) and (5) can be calculated to determine confidence intervals for mortality projections. In particular, the forecast variance can be derived from Eq. (6) which is explained in prior studies [10, 30]. Due to complex theoretical calculation of interval forecast with the

fitted model, various studies suggested this equation as useful tool for interval forecast calculation [21, 37, 38].

Forecast accuracy measure

In order to measure the forecast accuracy, the difference between actual and corresponding forecast series, indicated as mean absolute forecast error (MAFE) is calculated by given formula,

$$MAFE_{\xi} = \frac{1}{J} \sum_{j=1}^J |y_{n+\xi(x_j)} - \hat{y}_{n+\xi(x_j)}| \tag{8}$$

In Eq. (8) the term $y_{n+\xi(x_j)}$ is the actual observed sample values for the j th age and ξ th curve of the projected time period, whereas $\hat{y}_{n+\xi(x_j)}$ shows the point projections for the observed sample [39].

The multivariate Diebold–Mariano (DM) test

Diebold-Mariano (DM) test is the statistical test, introduced by Diebold and Mariano in 1995. This test is used to test the null hypothesis of equal forecast accuracy or equal predicting ability between two competing models. The DM test evaluation criteria is based on its parameter loss function. This loss function may include variety of differences and skill functions such as straight, absolute or squared differences and correlation respectively. In particular, the test is free from the distributional assumption on the forecast errors. It can includes temporal autocorrelations and any other type of correlation between the two or more series. Variety of alterations to the test have been prepared for its improvement and for conveniently to use it [40–43]. In order to test the forecast accuracy among multiple models, the authors Mariano and Preve extended the DM test for multivariate forecast accuracy [44]. The general loss is used to evaluate the multiple forecast models. The framework of multivariate DM test is describes as follows. The test consist of null hypothesis that demonstrates equal predictive accuracy for the predictive models. The test procedure is as follow:

$$H_0 : E[g(e_{(1),t})] = E[g(e_{(2),t})] = \dots = E[g(e_{(k+1),t})], \tag{9}$$

$$e_{(i),t} = \{y_{it} - \hat{y}_{it}\}, i = 1, 2, \dots, k + 1. t = 1, 2, \dots, T \tag{10}$$

In the above Eq. (9) the component $E[g(e_{(1), t})]$ indicate the loss function expected value for the i th forecast. While in Eq. (10) the $e_{(i), t}$ is the $(k + 1)$ th forecast errors time series from $(k + 1)$ another model. In simple word, we can assume that these components might be the forecasts outcome that based on some subjective or objective approach, for example, surveys, judgments, extrapolation techniques, smoothing, time-series models, or any mixture of methods. The accuracy of the forecasts is to be assess

using some specified loss function that is denoted by $g(\cdot)$. As various parameter can be involved in the loss function, so usually it is assume that the loss function depends only on the forecast errors. The loss function differential d_{jt} between forecasting errors of two model is described as

$$d_{jt} = g(e_{j,t}) - g(e_{j+1,t}), j = 1, 2, \dots, k \tag{11}$$

For multiple model, k loss differential series $\{d_{jt}\}$ in notation of vector is considered as:

$$d_t = (d_{1t}, d_{2t}, \dots, d_{kt})^E, t = 0, \pm 1, \pm 2, \dots \tag{12}$$

In context of Eq. (12), Now the null hypothesis of Eq. (9) can be stated as $H_0 : E[d_t] = 0$. In order to test this null hypothesis, the vector of observed sample means for equal predictive ability test is written as:

$$\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$$

For example d_t is the ccc (AFE). So it can be expressed as:

$$d_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots \tag{13}$$

Finally, according to above equation the multivariate version of the DM test statistic is designed as:

$$S = T \bar{d}^T \hat{\Omega}^{-1} \bar{d} \rightarrow \chi_k^2 \tag{14}$$

where T is total number of years in the predictive set and transpose matrix of \bar{d} is denoted as \bar{d}^T . While $\hat{\Omega}$ is the stable estimator of Ω and k is the degree of freedom that is number of time series in the system. Heteroscedasticity and autocorrelation consistent (HAC) estimate of Ω can provide improvement in the time related structure without distressing test statistic distribution [44, 45]. In addition, a study based on Monto Carlo simulation suggested that a reasonable estimator of Ω can also be attained using the sample variance [46].

Level of significance alpha (α) can be considered for decision of null hypothesis (H_0). Rejection of null hypothesis suggests that not all tested model holds equal predictive ability. Hence, loss dissimilarities in \bar{d} are consider for the choice of the outstanding predictive model.

Results of empirical analysis

Descriptive epidemiology

The annual four Asian countries mortality rates due to breast cancer (BC) from 1990 to 2017 for age range 20 to 80+ year was considered to run the application of three stochastic mortality models with smoothing p-splines approach. The data related to four Asian countries, China, Pakistan, India and Thailand was downloaded from the Institute for Health Metrics and Evaluation (IHME) <http://ghdx.healthdata.org/gbd-results-tool>. The accessibility of

data and sources are mentioned in “availability of data and materials” statement at the end of the paper. We considered women BC mortality rates from 20 to 80+ years of age and for each calendar year 1990 to 2017 for four Asian countries. The mortality rates due to breast cancer given by the ratio between the “number of deaths” and the “exposure to risk,” were arranged in a matrix for specific age x and time t . In order to assess the differences in BC mortality rates corresponding to both age x and time t , we realized that BC mortality has shown a gradual increase with time. To have an idea of this progression, the general trends in the BC mortality rates and log mortality rates during the period 1990–2017 for four Asian countries are depicted in Fig. 1. We can see that mortality improvements are not similar across the ages and the years. As it is clear, there is an increasing difference for higher ages (> 50), particularly around age = 80 for each country except Thailand where slightly declining trends are observed with time.

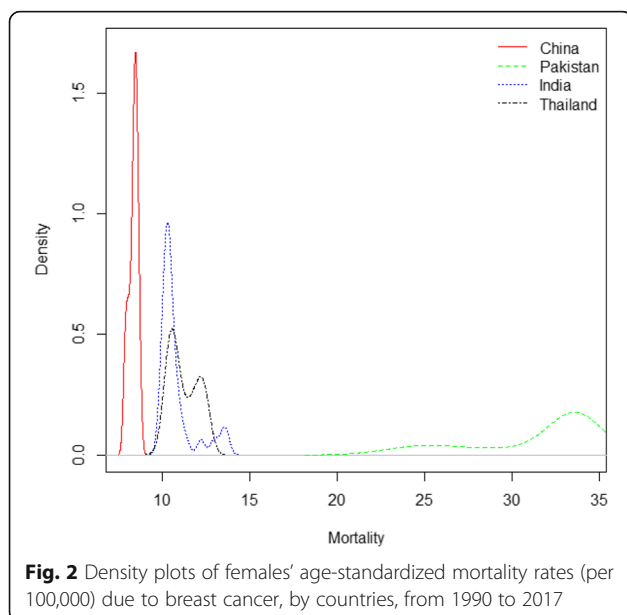
Additionally, Density plot of age standardized BC mortality rates for four countries’ data sets are presented in Fig. 2. Rates are more dispersed in Pakistani women while India and Thailand data showed the monotonic differences with improvement in BC mortality.

Model fits

In order to assess the in-sample and out-sample model performance, we fit the models by excluding the last 7 years of data from each country’s data set. The first step of analytical analysis is fitting the three stochastic mortality models after applying the smoothing p-splines approach for each country’s data sets under consideration. Which later called as, smoothed lee carter (SLC) model, Functional demographic model (FDM) and

Booth-Maindonald-Smith (BMS) model. Supplementary Figs. S1, S2 and S3 show the estimated parameters from these three models. The percentage of variation (PV) explained by the SLC model is 97.3, 89.3, 63.7 and 87.8% for the China, Pakistan, India and Thailand data respectively. The difference among the PV of four countries’ data set is due to their BC mortality behavior and different features of data as shown in Figs. 1 and 2. We can see that the BC mortality rates at older ages in China data are more consistent than other countries; as a result, the SLC model fitted the China data better than the other country’s data set.

Moving from the SLC to the second stochastic mortality model that called as FDM, we notice that the PV explained by the FDM model increases for Pakistan and Thailand while decreases for India and remain almost constant for China. Specifically, in context of order 3 of FDM, the basis functions explain respectively 97.1, 1.9, 0.4% of the variation for China, 90.0, 8.7, 0.8% of the variation for Pakistan, 61.6, 20.5, 7.5% of the variation for India and 88.4, 10.2, 0.7% of the variation for Thailand data. The difference of variations among basis functions is because these basis functions model the different movements in BC mortality across the ages. Specifically, the first basis function usually models the movements in BC mortality for younger ages. Let consider the Figs. S1, S2 and S3 for SLC, BMS and FDM model respectively, to highlight some differences among four countries death rates due to breast cancer. Keeping in view the Fig. S3, we notice that the function fitted on China, Pakistan and Thailand data has a stronger slope between age 20 and 30, 20 and 40 and 30 and 40 respectively than the fitted function on India data. Therefore, the BC mortality reduction at younger ages is greater for China, Pakistan and Thailand comparatively to India. Correspondingly, the first fitted coefficient plot shows the evidence about the mortality improvements at younger age. We can observe that, the BC mortality rates for younger ages have dropped over the period 1990–2005 and this phenomenon is captured by the decreasing trend of the coefficient 1 for China and India. On the other hand, a gradually increasing trend of the coefficient 1 is observed over the whole period for Pakistan data. The second basis function provide us evidence about the BC mortality rates differences between 30 and 60 years old. It can be clearly observe that almost each country have more strained rate differences but this differences is more stressed for Pakistan data as compare to other countries. Finally, the function 3 represents patterns that are more complex and it model the BC mortality differences between all the cohorts. In the current situation, these differences are not much significant for each country under consideration. Therefore, third functions show a lower variability between ages 65 to 80.



Further, shifting from the SLC and FDM to the BMS model, we observed that the PV explained by the model was consistent with the explained variation by SLC model. On the other hand, fitting from the both SLC and BMS model shows the higher variances in upper ages (> 50) for Pakistan than other Asian countries while with time these mortality differences are increases after 2005 period particularly for Pakistan and India (Figs. S1 and S2).

Goodness of fits of models

Residual analysis

A good fit is achieved when the residuals are independent and identically distributed. In order to verify this condition, residuals mortality rate by age in different years was derived from three fitted models for all four countries' data set (Fig. 3). Residuals mortality rate by age were seem to be more consistent in FDM than in SLC and BMS for all four countries almost everywhere.

Whereas, these errors were lower in SLC and BMS model on older ages for each country's data set. In addition, the error measures were also calculated to conform the error differences among different models as shown in Table 1. By comparing the SLC, FDM and BMS model, we notice that overall the error measures in FDM are smaller as compare to other models, even if the PV described by model is greater in the both SLC and BMS as compare to FDM model. Specifically, the mean square error (MSE) of the SLC model is lower than the MSE of the other two model for China while for other three countries's data set this error is the lower in FDM as compare to other two model.

In-sample and out-sample predictive accuracy by multivariate Diebold-Mariano (DM) test

In order to test and compare the model predictive accuracy we used the multivariate Diebold-Mariano (DM) test, which is advance approach for testing and

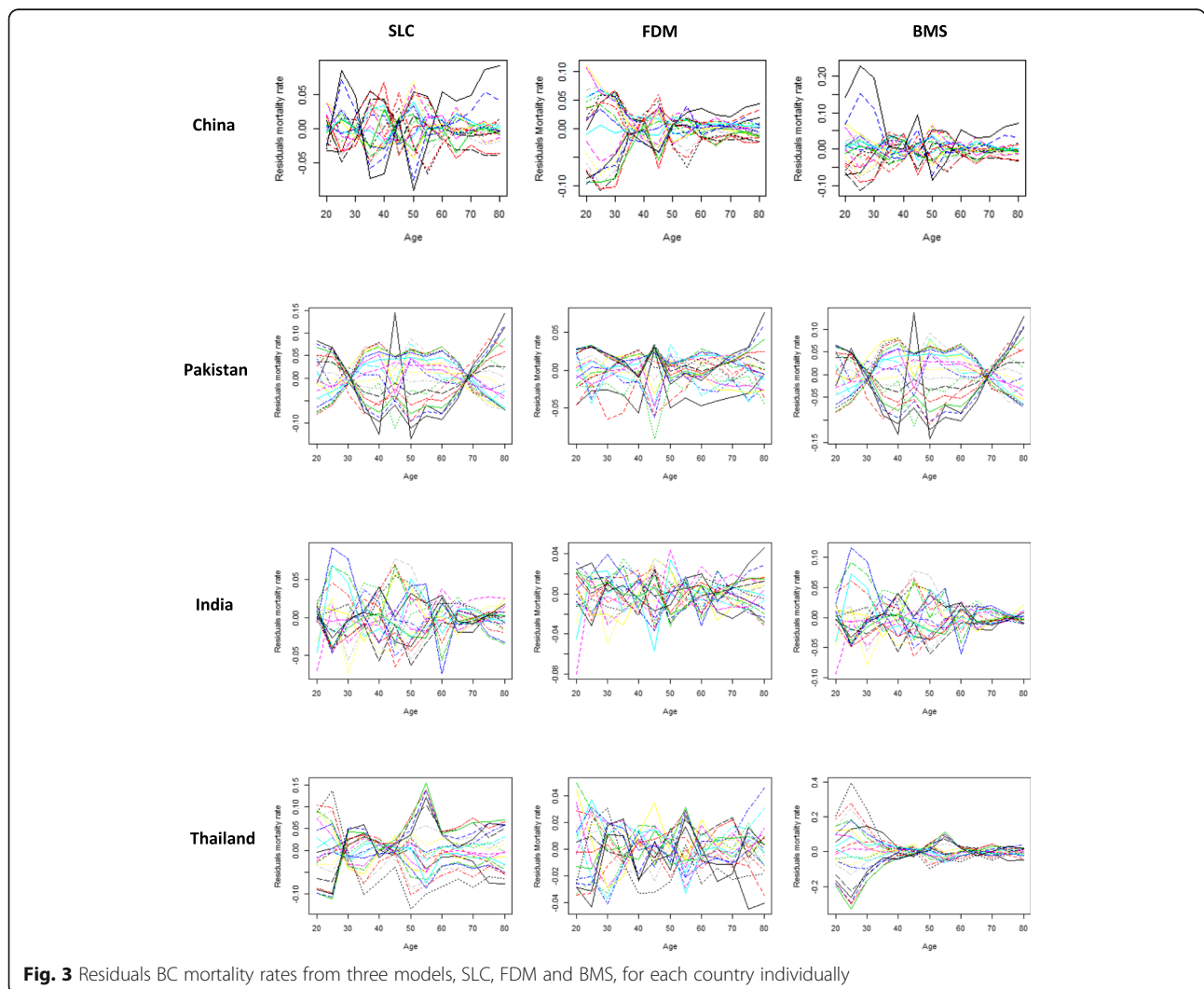


Fig. 3 Residuals BC mortality rates from three models, SLC, FDM and BMS, for each country individually

Table 1 Error measures from three fitted models individually for each country

Smoothed Lee Carter (SLC) Model				
Average across ages	ME	MSE	MPE	MAPE
China	0.000	0.0008	0.020	0.036
Pakistan	0.000	0.0024	0.0004	0.016
India	0.000	0.0007	0.0232	0.073
Thailand	0.000	0.0023	0.0110	0.056
Average across years				
China	0.000	0.046	1.509	2.736
Pakistan	0.000	0.134	0.018	0.794
India	0.000	0.0418	1.7075	5.827
Thailand	0.000	0.133	0.803	4.153
Average across ages Functional Demographic Model (FDM)				
China	0.000	0.0013	-0.029	0.065
Pakistan	-0.000	0.0004	0.001	0.008
India	0.000	0.0003	0.0128	0.046
Thailand	-0.000	0.000	0.005	0.019
Average across years				
China	-0.000	0.0700	-2.199	4.815
Pakistan	-0.000	0.026	0.0437	0.367
India	-0.000	0.016	0.939	3.575
Thailand	-0.000	0.015	0.378	1.359
Average across ages Booth-Maindonald-Smith (BMS) Model				
China	-0.000	0.0014	0.033	0.054
Pakistan	0.001	0.002	0.001	0.016
India	0.000	0.001	0.022	0.075
Thailand	-0.000	0.006	0.032	0.141
Average across years				
China	-0.007	0.084	2.466	4.109
Pakistan	0.077	0.139	0.048	0.788
India	-0.000	0.045	1.675	5.909
Thailand	-0.004	0.362	2.348	10.701

ME mean error, MSE mean square error; MPE mean percent error; MAPE mean absolute percent error

comparing of predictive ability of models. To measure in-sample forecast accuracy, the multivariate DM test is applied to average Absolute Forecast Error (AFE) across ages and years from three models for each country’s data set. According to the test, we have formulated our null hypothesis (H_0) that “predictive accuracy of the three model is the same” that implies zero difference of average absolute forecast errors among competing models. Based on our results in Table 2, the null hypothesis is not accepted at 1% level of significance for all four countries results ($p < 0.01$). Therefore, based on the results we may conclude that three models hold different predictive accuracy across years for each country’s data set.

On the other hand, Table 3 test the predictive accuracy of three models across ages we observe that three models hold equal predictive accuracy across ages.

In addition, we also test the model for outstanding predicting ability based on multivariate DM goodness of fit test. The DM test assigned rank 1 to the model which have minimum mean loss value among the models which indicating best fit of that model. According to the test, null hypothesis indicate the best fit of ranked 1 model. So, Fig. 4 depicts the actual and fitted log BC mortality rates across years from three models. We can see that according to multivariate DM goodness of fit test, the SLC model is ranked 1 which indicating best fit of model for China (Mean loss = 0.0064, p -value = 0.918) and Pakistan (Mean loss = 0.0059, p -value = 0.999). On the other hand, FDM model showed best fit on India (Mean loss = 0.0028, p -value = 0.999) and Thailand (Mean loss = 0.0054, p -value = 0.999) data.

In order to compare the in-sample predictive accuracy among three models, we demonstrated that SLC model provided the outstanding predictive ability for China and Pakistan data set while FDM delivered the outstanding predicting ability on India and Thailand data.

In order to conform the forecast ability of models, it was also tested, the model that provide a good in-sample fit to historical data whether it still produce good out-sample forecasts. A good model should provide accurate in-sample fits to the historical data as well as out-sample plausible forecasts. Therefore, out-sample predictive accuracy was also considered to verify the model predictive accuracy consistency. To evaluate the forecast accuracy, the following steps are required. Initially, it is needed to choose the matrix of interest included the forecasted variable. Possible metrics of forecasted variables may include the mortality rates, life expectancy or future survival rates. Different metrics are relevant for different purposes. As, our study goal is to examine the feasibility of different stochastic mortality models, hence, we emphasis on the BC mortality rates. We project the BC mortality rates from 2011 to 2017 according to the fitted models and derive the life expectancy to compare the forecasts with the actual. In our study forecasts are calculated based on the evolution of time parameter (k_t) and errors in age parameter (a_x) and (b_x) are not considered. Because according to the literature the standard errors of (a_x) and (b_x) to become less significant over forecast time in comparison to the standard error of parameter (k_t) [11]. The forecast errors in the SLC, FDM and BMS model for four countries’ data set are shown in Fig. 5. To compare the out-sample forecast accuracy among three models, average projection error in life expectancy by computing mean over forecast years were considered. Table 4 displays the comparison of out-sample forecast accuracy based on forecast errors in life

Table 2 Comparison of predictive accuracy of three models across years for each country using multivariate Diebold-Mariano (DM) test

Country	Null Hypothesis	Test Statistic Value	P-Value	Decision
China	H_0 : SLC=FDM = BMS	-1796.2	0.000	H_0 is rejected
Pakistan	OR	-20.151	0.000	H_0 is rejected
India	H_0 : All model hold Equal predictive accuracy	-24.314	0.000	H_0 is rejected
Thailand		-3.9457	0.000	H_0 is rejected

Predictive period, 1990–2010

expectancy. Based on multivariate DM test results we determine that three models have significantly different ($p < 0.01$) out-sample predictive accuracy for all countries' data set (Table 4). Moreover, we also find the model with outstanding forecast ability for four countries' data set. According to the test results, SLC model provided the outstanding forecast ability on china and Pakistan data, while FDM has rank 1 for providing outstanding forecast ability for India and Thailand data set. At the same time, SLC model also offered the quite comparable results for India and Thailand data (Fig. 6).

Further, we also calculate the mean and variance of life expectancy forecast errors across forecast years. Table 5 depicts the minimum variance of life expectancy forecast error for SLC model in China and Pakistan data set. On the other hand, the FDM and SLC model showed the almost same variance for India and Thailand data. Moreover, 95% confidence interval plot of BC mortality forecast error were also shown in Fig. 7. The confidence intervals for the mean error are tighter for SLC model than other models in both China and Pakistan data set. However, the confidence intervals for the both model FDM and SLC are narrow for other two countries' data set.

Discussion

Although, the forecasting mortality methods have been broadly studied by a group of researchers like Booth [22] and Booth & Tickle [23]. In the present study, we apply the three stochastic mortality models belong to family of generalized lee-carter with smoothing p-splines approach to four Asian countries' data sets, evaluate the performance of these three methods, compare the similarities of different Asian regions and propose potential improvements. Furthermore, study is also contain the

derivation of future stochastic life expectancy at birth calculated from mortality forecasts and use it for models assessment.

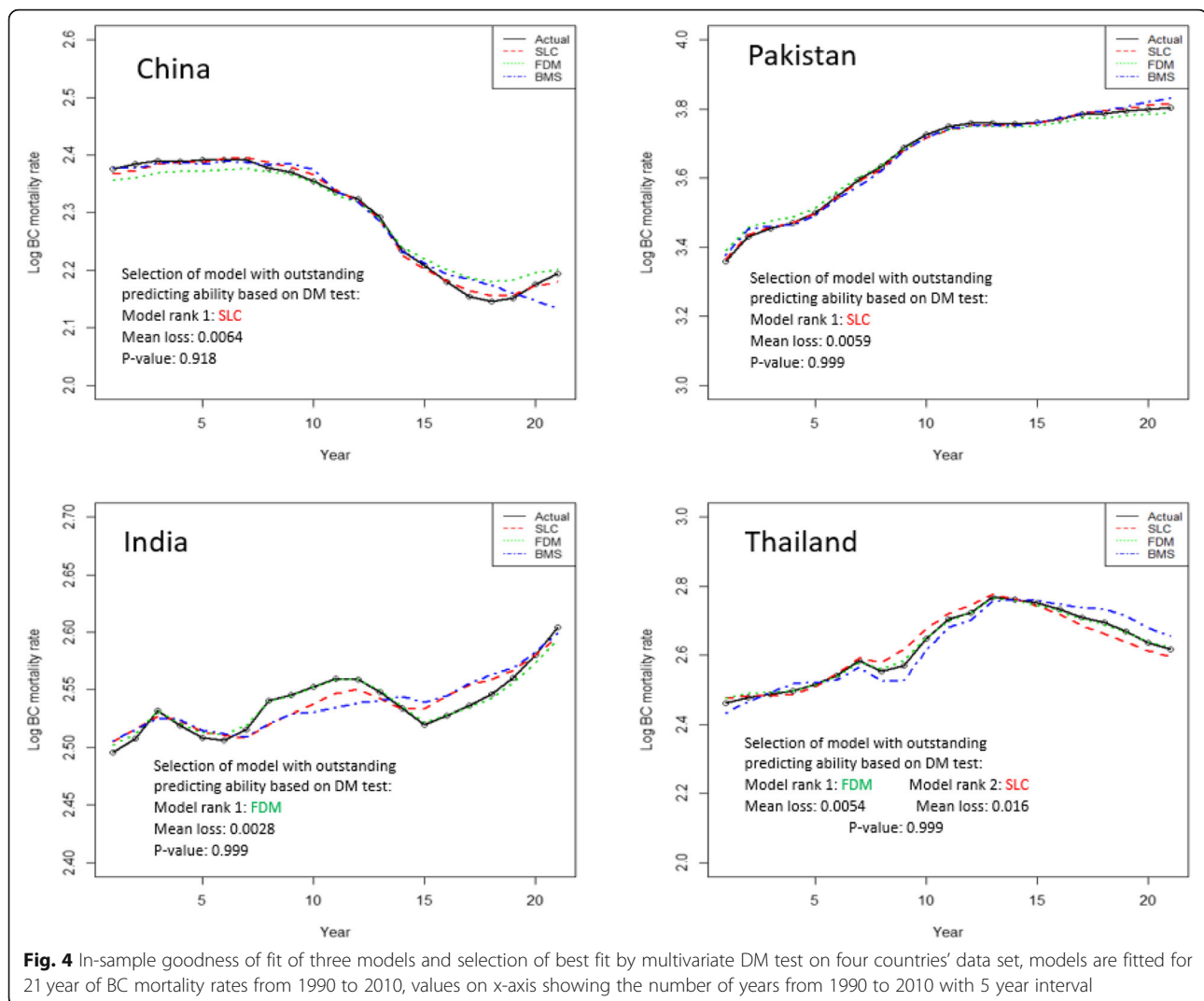
The age-specific BC mortality change by different years for four Asian countries' data set was depicted in Fig. 1. We can observe that, the mortality rates has followed a smooth function with some observational error. The observational error has higher variance at very old ages where the population are small and also at young ages where the mortality rates are small. This phenomenon can also be clearly shown from Fig. 2 that have higher intensity of mortality for small population of Pakistan while remaining countries have relatively smaller mortality rates for large number of population. A study also reported the similar pattern in USA mortality data, where researchers observed that age-specific mortality rates are higher than 1/100,000 for very small populations [38]. Mortality forecasting based on such a data pattern may be well representative by stochastic mortality models with P-splines smoothing approach. This smoothing approach have been extensively used to smooth mortality data for application of these models in various studies [10, 26, 27, 38].

Application of stochastic smoothed mortality models namely SLC, FDM and BMS model on BC mortality data provided the different PV explained by the models. Consistent PV among three models were observed for China data set. The reason may be the method assume the homoscedastic variances as this data is follow to this assumption to some extent [38]. The other reason might be the China's data sets contain less wiggly data in a shorter and older age range. While for other countries these PV were little bit different among the three models, which shows the capability of the models to deal with the heteroscedastic data. It observed that for these

Table 3 Comparison of predictive accuracy of three models across ages for each country using multivariate Diebold-Mariano (DM) test

Country	Null Hypothesis	Test Statistic Value	P-Value	Decision
China	H_0 : SLC=FDM = BMS	14.165	0.999	H_0 is not rejected
Pakistan	OR	2617.9	1	H_0 is not rejected
India	H_0 : All model hold Equal predictive accuracy	1034.7	1	H_0 is not rejected
Thailand		89.218	1	H_0 is not rejected

Predictive period, 1990–2010



type of data FDM provided the larger PV and smaller measure error than the other models. In line with previous literature, the basis functions used in FDM model have the ability to account different mortality movements and model them across the ages [15]. Therefore, larger PV explained by FDM in irregular data set might be the reason of presence of different basis function in the model.

To highlight some differences among four countries' death rates due to breast cancer we fitted the three models on each country's data set separately (supplementary Figs. S1, S2 and S3). Keeping in view the FDM results (Fig. S3), we observed that the fitted mortality functions, from age 20 to 30 for China, 20 to 40 for Pakistan and 30 to 40 for Thailand has a stronger negative slope. Therefore, the BC mortality reduction at younger ages was greater for these countries as compare to India. The decreasing trend in BC mortality in these countries has been reported previously and reason of this decline may be due to higher pregnancy

rates and related reproductive factors in that countries [47, 48]. In addition, low or never use of oral contraceptives may also be the possible reason for the decreasing BC mortality rates in younger women [4, 49]. Both SLC and BMS model showed the higher mortality variances in upper ages (> 50) for Pakistan than other Asian countries during the study period. Mortality differences gradually improved after 2005 period in Pakistan and India and remained consistent or move to decline in China and Thailand during the specified time period. The actual cause of the reduction in breast cancer mortality is still unknown yet and required more research to explore the possibility. Proper Access to medical facilities and advanced treatments are likely reasons for the decline in mortality in China: such as improved early detection combined with effective treatment. Most women under 50 years of age working in urban areas have employer-sponsored benefits like medical examinations and free breast ultrasounds once or twice a year. Previous studies

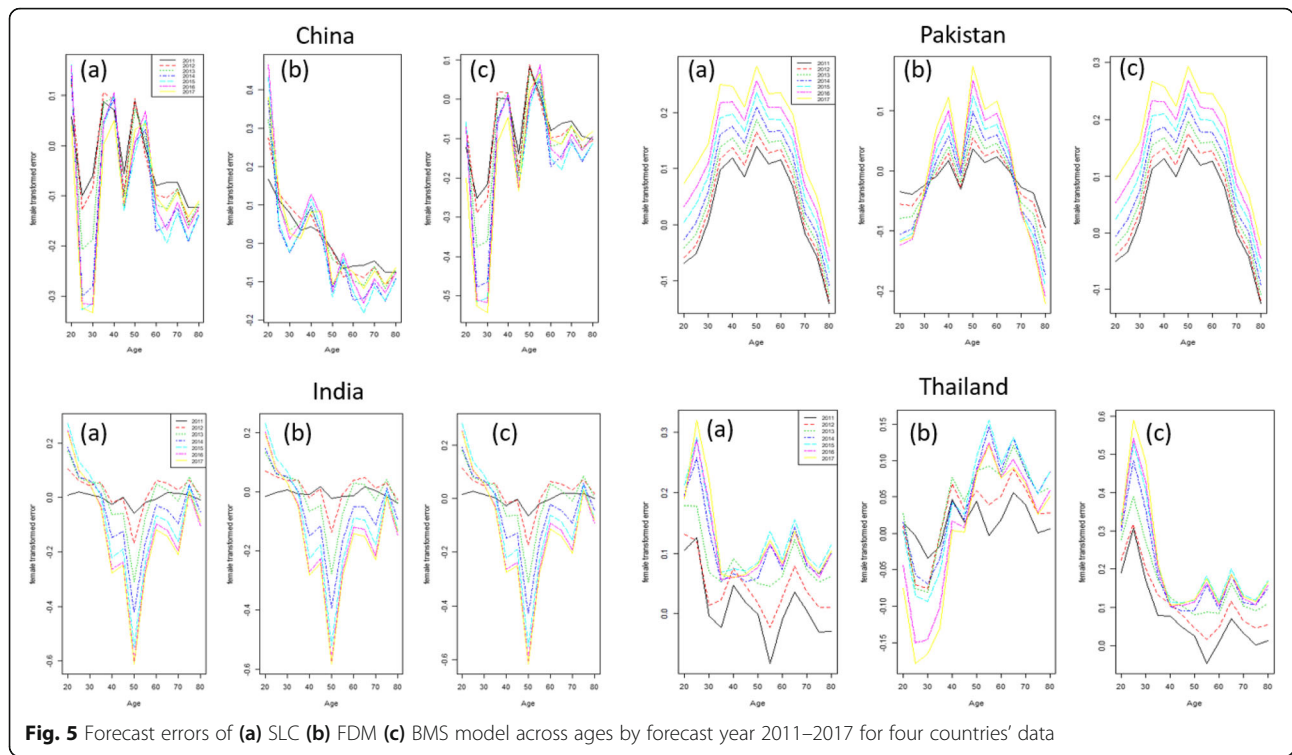


Fig. 5 Forecast errors of (a) SLC (b) FDM (c) BMS model across ages by forecast year 2011–2017 for four countries' data

have shown that ultrasound is superior to Chinese women’s mammography for the prevention and control of BC [50]. Mubarik et al. [4, 7] conducted a study on trends and projections in BC mortality, and they reported the higher BC mortality rates during the period 1990–2017 in some of Asian countries including Pakistan and India. Our study results has also reported the similar findings in these countries. The rising trends in BC mortality in these countries might be due to that they did not have much success in disease diagnosis and treatment programs as compare to developed Asian and European countries during the developmental period.

In order to evaluate the models performance, both in their goodness of fit on historical data and their forecasting ability (in-sample and out-sample), all models were tested and ranked by using residual analysis and an advance statistical multivariate DM test. This test is widely used for measuring predictive ability across multiple

forecasting methods [51]. According to goodness of fit analysis, FDM errors were generally lower and residual trends were more consistent. In addition, the error measures suggested by Cairns et al. [18] were also computed and compared for the three models. We noticed that overall the error measures in FDM were smaller as compare to other models, even if the PV explained by model was higher in the SLC and BMS than in FDM. Specifically, the MSE of the FDM model was lower than the MSE of the other two model for each country’s data set except China. Furthermore, multivariate DM test had showed the significant difference among the models predictive performance for each country’s data set. When outstanding forecasting ability of models was tested for both in-sample and out-sample forecasts, SLC model showed the best predicting ability in both cases for China (in-sample mean loss: 0.006, $p = 0.918$; out-sample mean loss: 0.060, $p = 0.999$) and Pakistan (in-sample

Table 4 Comparison of forecast accuracy of three models across years for each country using multivariate Diebold-Mariano (DM) test

Country	Null Hypothesis	Test Statistic Value	P-Value	Decision
China	$H_0: SLC=FDM = BMS$	64.354	0.000	H_0 is rejected
Pakistan	OR $H_0: All model hold Equal predictive accuracy$	7.636	0.000	H_0 is rejected
India		3624.1	0.000	H_0 is rejected
Thailand		286.68	0.000	H_0 is rejected

Forecast period, 2011–2017

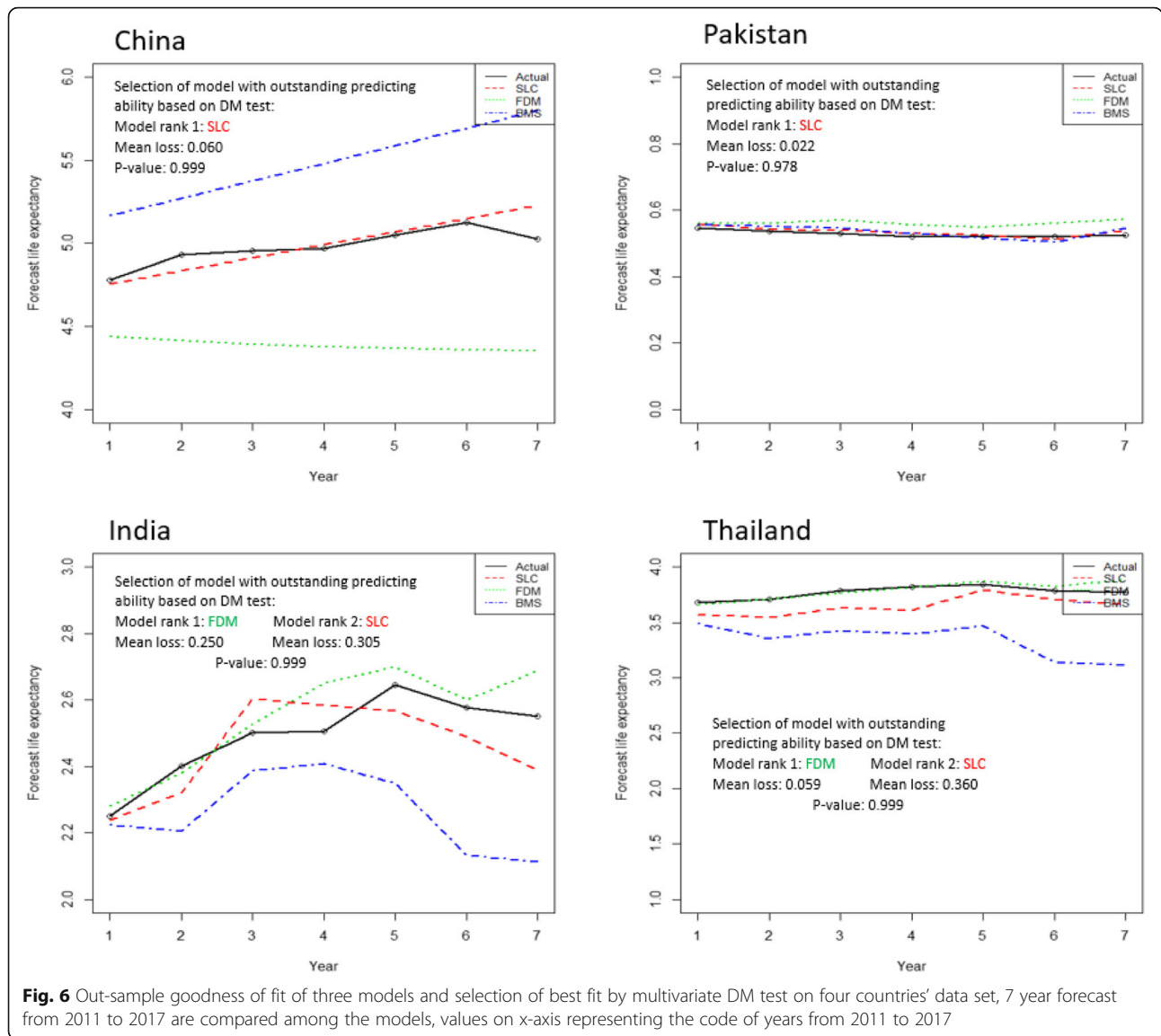


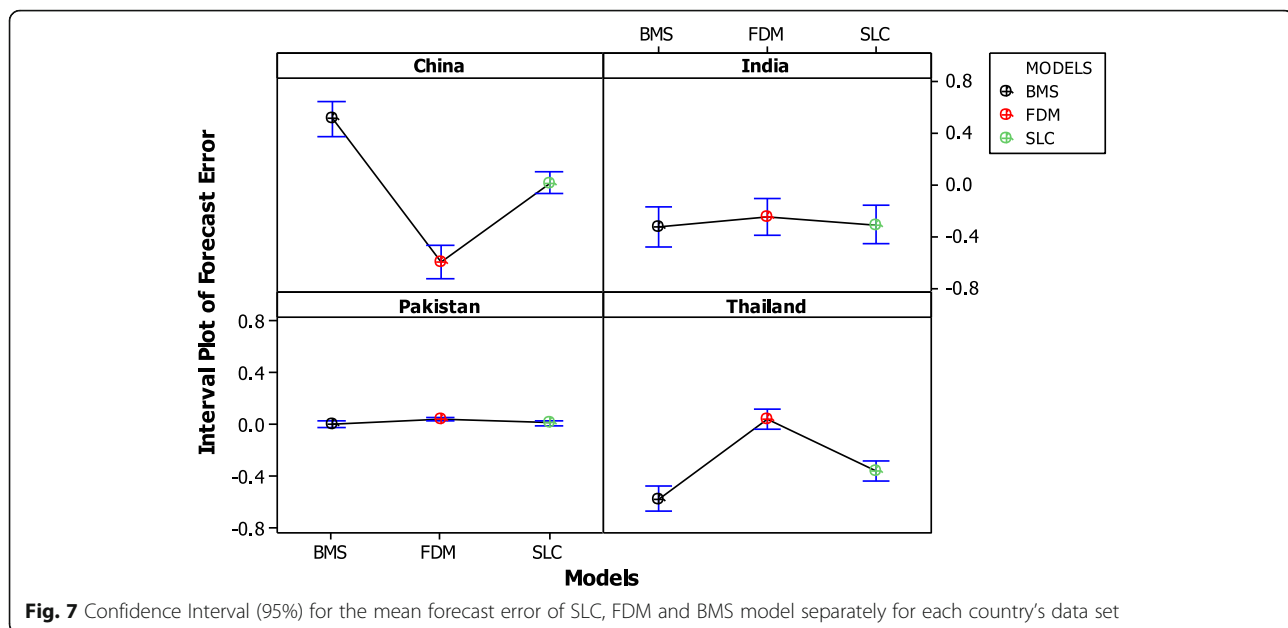
Fig. 6 Out-sample goodness of fit of three models and selection of best fit by multivariate DM test on four countries' data set, 7 year forecast from 2011 to 2017 are compared among the models, values on x-axis representing the code of years from 2011 to 2017

mean loss: 0.005, $p = 0.999$; out-sample mean loss: 0.020, $p = 0.978$). While FDM model indicated the best predictive performance within and outside forecasts for data set of other two countries. In contrast, model choice based on goodness of fit testing was ambiguous. Overall, study results indicated that the model/models, which had

similar predictive ability within data, they also provided the almost same predictive ability of outside of fitted data. Nevertheless, there was no single model that we called the overall best model for all countries' data set. Overall, based on model predictive accuracy tests, SLC model generally outperform the FDM for China and

Table 5 Mean and variance of forecast error in life expectancy

Country	SLC		FDM		BMS	
	Mean	Variance	Mean	Variance	Mean	Variance
China	0.015969	0.007437	-0.58973	0.01647	0.507574	0.017841
Pakistan	0.012938	0.000478	0.044085	0.000288	0.003414	0.000489
India	-0.30533	0.023026	-0.24113	0.020099	-0.32129	0.023243
Thailand	-0.36042	0.006641	0.044443	0.006649	-0.57447	0.010475



Pakistan data set, while FDM was ranked 1 on India and Thailand data both on goodness of fit and on forecast accuracy. These findings are consistent with the claim by other studies that have reported their findings in favor of the FDM model than the Lee-Carter model in term of better performance [10, 52].

In order to confirm the predictive accuracy results, we also calculated the mean and variance of forecasting errors in life expectancy over forecast years. The minimum mean error and variance of forecast error in life expectancy for SLC model for China and Pakistan data was confirming the predicted accuracy of that model. On the other hand, the FDM and SLC model showed the almost same variance for India and Thailand data. Moreover, 95% confidence interval plot of BC mortality forecast error were also shown. The confidence intervals for the mean error was tighter for both SLC model and FDM for Pakistan and Thailand. A quantitative comparison study of different stochastic mortality models has also reported the consistency in performance of lee-carter and functional demographic models on Italian mortality data [19].

Conclusion

According to the study findings, we conclude that in selecting and ranking the mortality models, they should carried out for both their goodness of fit and testing of within and out data forecasting ability. In particular, application on BC mortality data under study, stochastic smoothed mortality models based on functional data analysis, generally had better perform on quadratic structure of data, both in term of goodness of fit and on

forecast accuracy. Commonly, their errors were lower and their error distributions although were not ideally satisfying but showed less inconsistent effects. In some of previous studies, application of these models on mortality data also claim in favor of FDM that it outperform the lee-carter model [10, 18]. On the other hand, SLC model outperform the FDM in case of symmetric data, and also shows almost comparable results to FDM for both within and outside data forecast accuracy in case of asymmetric or quadratic structure of data.

Therefore, our study recommends considering the SLC model in comparison to the other, since it provides even better results in some cases. Moreover, each model considered has certain limitations and shows certain failures to present all data. Therefore, no one can eventually be convinced of "the finest" in fitting and predicting BC mortalities. Although the results of both models, FDM and SLC included, were generally fair. However, they still display some nonrandom errors because of the cohort effect. Therefore, we plan to consider and compare other stochastic mortality models with cohort effect in future work and test them on smoothed BC mortality data. In conclusion, as already renowned by other studies [19, 52] we also suggest generating mortality forecasts using multiple models rather than relying upon any single model. In the current case, we can assume that there is no single model, which can truly outperform all the others on every population. Additionally, we perceived that these approaches would be most convenient for modeling and projecting the future trends of rare diseases for which there has been very little improvement in treatment and involve least birth cohort effects.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-020-01187-5>.

Additional file 1: Figure S1. The parameter estimates of SLC model for four countries' BC mortality rates, α_x is the derived age pattern averaged across years; β_x stands for the sensitivity of the mortality rates to the change of k_t , reflecting how fast the mortality rate changes over ages; k_t represents the only time-varying index of mortality level. **Figure S2.** The parameter estimates of BMS model for four countries' BC mortality rates, α_x is the derived age pattern averaged across years; β_x stands for the sensitivity of the mortality rates to the change of k_t , reflecting how fast the mortality rate changes over ages; k_t represents the only time-varying index of mortality level. **Figure S3.** The parameter estimates of FDM model on four countries' BC mortality rates, (a) China (b) Pakistan (c) India and (d) Thailand.

Abbreviations

BC: Breast cancer; LC: Lee-Carter; SLC: Smoothed Lee-Carter; FDM: Functional Demographic model; SVD: Singular value decomposition; BMS: Booth-Maindonald-Smith; DM: Diebold-Mariano; ARIMA: Autoregressive integrated moving average; MSE: Mean square error; AFE: Absolute forecast error

Acknowledgements

Not applicable

Authors' contributions

Conceptualization, SM and CY Methodology-Analysis-Writeup, SM Manuscript revision, SM and YH. All authors have read and approved the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant NO.81773552) and the National Key R&D Program of China (Grant No. 2018YFC1315302 and 2017YFC1200502). The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The dataset analyzed during the current study are available in the Institute for Health Metrics and Evaluation (IHME): <http://ghdx.healthdata.org/gbd-results-tool>. Statistical analysis was performed using the R demography and forecast evaluation package, which codes are available on GitHub (<https://github.com/robjhyndman/demography>); (<https://github.com/johntwk/Diebold-Mariano-Test>).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Received: 22 July 2020 Accepted: 3 December 2020

Published online: 09 December 2020

References

- Hakulinen T. The future cancer burden as a study subject. *Acta Oncol.* 1996; 35(6):665–70.
- Fan L, Goss PE, Strasser-Weippl K. Current status and future projections of breast cancer in Asia. *Breast Care.* 2015;10(6):372–8.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
- Mubarik S, Wang F, Fawad M, Wang Y, Ahmad I, Yu C. Trends and projections in breast cancer mortality among four Asian countries (1990–2017): evidence from five stochastic mortality models. *Sci Rep.* 2020;10(1):1–12.
- Kim Y, Yoo K-Y, Goodman MT. Differences in incidence, mortality and survival of breast cancer by regions and countries in Asia and contributing factors. *Asian Pac J Cancer Prev.* 2015;16(7):2857–70.
- Mubarik S, Malik SS, Wang Z, Li C, Fawad M, Yu C. Recent insights into breast cancer incidence trends among four Asian countries using age-period-cohort model. *Cancer Manag Res.* 2019;11:8145.
- Mubarik S, Wang F, Malik SS, Shi F, Wang Y, Yu C. A hierarchical age-period-cohort analysis of breast Cancer mortality and disability adjusted life years (1990–2015) attributable to modified risk factors among Chinese women. *Int J Environ Res Public Health.* 2020;17(4):1367.
- Li N, Lee R, Tuljapurkar S. Using the Lee–Carter method to forecast mortality for populations with limited data. *Int Stat Rev.* 2004;72(1):19–36.
- Camarda CG. Smooth constrained mortality forecasting. *Demogr Res.* 2019; 41:1091–130.
- D'Amato V, Piscopo G, Russolillo M. The mortality of the Italian population: smoothing techniques on the Lee–Carter model. *Ann Appl Stat.* 2011;705–24.
- Lee RD, Carter LR. Modeling and forecasting US mortality. *J Am Stat Assoc.* 1992;87(419):659–71.
- Booth H, Maindonald J, Smith L. Applying Lee–Carter under conditions of variable mortality decline. *Popul Stud.* 2002;56(3):325–36.
- Renshaw AE, Haberman S. On the forecasting of mortality reduction factors. *Insurance Math Econ.* 2003;32(3):379–401.
- Renshaw AE, Haberman S. On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling. *Insurance Math Econ.* 2008;42(2):797–816.
- Hyndman RJ, Ullah MS. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput Stat Data Anal.* 2007;51(10):4942–56.
- Renshaw A, Haberman S. Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. *J R Stat Soc: Ser C: Appl Stat.* 2003;52(1):119–37.
- Renshaw AE, Haberman S. Lee–Carter mortality forecasting with age-specific enhancement. *Insurance Math Econ.* 2003;33(2):255–72.
- Cairns AJ, Blake D, Dowd K, Coughlan GD, Epstein D, Ong A, Balevich I. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North Am Actuarial J.* 2009;13(1):1–35.
- Carfora MF, Cutillo L, Orlando A. A quantitative comparison of stochastic mortality models on Italian population data. *Comput Stat Data Anal.* 2017; 112:198–214.
- Dowd K, Cairns AJ, Blake D, Coughlan GD, Epstein D, Khalaf-Allah M. Evaluating the goodness of fit of stochastic mortality models. *Insurance Math Econ.* 2010;47(3):255–65.
- Erbas B, Akram M, Gertig DM, English D, Hopper JL, Kavanagh AM, Hyndman R. Using functional data analysis models to estimate future time trends in age-specific breast cancer mortality for the United States and England–Wales. *J Epidemiol.* 2010;20(2):159–65.
- Booth H. Demographic forecasting: 1980 to 2005 in review. *Int J Forecast.* 2006;22(3):547–81.
- Booth H, Tickle L. Mortality modelling and forecasting: a review of methods. *Ann Actuarial Sci.* 2008;3(1–2):3–43.
- Delwarde A, Denuit M, Eilers P. Smoothing the Lee–Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Stat Model.* 2007;7(1):29–48.
- Eilers PH, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci.* 1996;89–102.
- Currie ID, Durban M, Eilers PH. Smoothing and forecasting mortality rates. *Stat Model.* 2004;4(4):279–98.
- Camarda CG. MortalitySmooth: an R package for smoothing Poisson counts with P-splines. *J Stat Softw.* 2012;50(1):1–24.
- Currie ID, Durban M, Eilers PH. Generalized linear array models with applications to multidimensional smoothing. *J Royal Stat Soc Ser B.* 2006;68(2):259–80.
- Lee R, Miller T. Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography.* 2001;38(4):537–49.
- Hyndman RJ, Shang HL. Forecasting functional time series. *J Korean Statist Soc.* 2009;38(3):199–211.
- Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. United Kingdom: Wiley; 2015. ISBN: 978-1-118-67502-1.
- Hamilton JD. Time series analysis, vol. 2. Princeton New Jersey: Princeton University Press; 1994. ISBN: 9780691042893.
- Booth H, Hyndman RJ, Tickle L, De Jong P. Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demogr Res.* 2006; 15:289–310.

34. Booth H, Tickle L, Smith L. Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison. *N Z Popul Rev.* 2005;31(1):13–34.
35. Erbas B, Hyndman RJ, Gertig DM. Forecasting age-specific breast cancer mortality using functional data models. *Stat Med.* 2007;26(2):458–70.
36. Hyndman RJ, Booth H. Stochastic population forecasts using functional data models for mortality, fertility and migration. *Int J Forecast.* 2008;24(3):323–42.
37. Koissi M-C, Shapiro AF, Högnäs G. Evaluating and extending the Lee-Carter model for mortality forecasting: bootstrap confidence interval. *Insurance Math Econ.* 2006;38(1):1–20.
38. Booth H, Hyndman RJ, Tickle L. Prospective life tables. *Computational Actuarial Science with R*; 2014. p. 319–44.
39. Shang HL, Haberman S. Model confidence sets and forecast combination: an application to age-specific mortality. *Genus.* 2018;74(1):19.
40. Hering AS, Genton MG. Comparing spatial predictions. *Technometrics.* 2011; 53(4):414–25.
41. Diebold FX. Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano tests. *J Bus Econ Stat.* 2015;33(1):1–1.
42. Gilleland E, Roux G. A new approach to testing forecast predictive accuracy. *Meteorol Appl.* 2015;22(3):534–43.
43. Fawad M, Yan T, Chen L, Huang K, Singh VP. Multiparameter probability distributions for at-site frequency analysis of annual maximum wind speed with L-moments for parameter estimation. *Energy.* 2019;181:724–37.
44. Mariano RS, Preve D. Statistical tests for multiple forecast comparison. *J Econ.* 2012;169(1):123–30.
45. Diks C, Panchenko V, Van Dijk D. Out-of-sample comparison of copula specifications in multivariate density forecasts. *J Econ Dyn Control.* 2010; 34(9):1596–609.
46. Giacomini R, White H. Tests of conditional predictive ability. *Econometrica.* 2006;74(6):1545–78.
47. Khan M, Hossain M, Khan A, Nojima M, Suzuki S, Fujino Y, Tokudome S, Tamakoshi K, Mori M, Tamakoshi A. Ovarian cancer mortality among women aged 40–79 years in relation to reproductive factors and body mass index: latest evidence from the Japan collaborative cohort study. *J Gynecol Oncol.* 2013;24(3):249–57.
48. Mubarik S, Shakil Malik S, Wang Z, Li C, Fawad M, Yu C. <p>recent insights into breast cancer incidence trends among four Asian countries using age-period-cohort model</p>. *Cancer Manag Res.* 2019;11:8145–55.
49. Wang J, Lv H, Xue Z, Wang L, Bai Z. Temporal trends of common female malignancies on breast, cervical, and ovarian cancer mortality in Japan, Republic of Korea, and Singapore: application of the age-period-cohort model. *Biomed Res Int.* 2018;2018.
50. Shen S, Zhou Y, Xu Y, Zhang B, Duan X, Huang R, Li B, Shi Y, Shao Z, Liao H. A multi-Centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *Br J Cancer.* 2015; 112(6):998–1004.
51. Borup D, Thyrgaard M. Statistical tests for equal predictive ability across multiple forecasting methods. Denmark: Department of Economics and Business Economics, Aarhus University; 2017.
52. Cairns AJ, Blake D, Dowd K, Coughlan GD, Epstein D, Khalaf-Allah M. Mortality density forecasts: an analysis of six stochastic mortality models. *Insurance Math Econ.* 2011;48(3):355–67.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

