ORIGINAL RESEARCH PAPER

Coherent mortality forecasting by the weighted multilevel functional principal component approach

Ruhao Wu and Bo Wang

Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK

ARTICLE HISTORY

Compiled December 11, 2018

ABSTRACT

In human mortality modelling, if a population consists of several subpopulations it can be desirable to model their mortality rates simultaneously while taking into account the heterogeneity among them. The mortality forecasting methods tend to result in divergent forecasts for subpopulations when independence is assumed. However, under closely related social, economic and biological backgrounds, mortality patterns of these subpopulations are expected to be non-divergent in the future. In this article, we propose a new method for coherent modelling and forecasting of mortality rates for multiple subpopulations, in the sense of nondivergent life expectancy among subpopulations. The mortality rates of subpopulations are treated as multilevel functional data and a weighted multilevel functional principal component (wmFPCA) approach is proposed to model and forecast them. The proposed model is applied to sex-specific data for nine developed countries, and the results show that, in terms of overall forecasting accuracy, the model outperforms the independent model and the Product-Ratio model as well as the unweighted multilevel functional principal component approach.

KEYWORDS

Human mortality; coherent forecasts; functional principal component analysis; Lee-Carter model; Product-Ratio model

1. Introduction

As governments, insurance companies and pension funds are obliged to pay billions of pensions and annuities, they are heavily exposed to longevity risk due to the increasing life expectancy of pensioners and policy holders. Hence it is of great importance for them to estimate people's future life expectancy as accurately as possible. In the past decades, there have been many statistical models developed for this purpose. In 1992, the pioneering Lee-Carter model was introduced as an extrapolative method for forecasting mortality [21], which has been followed by a series of extensions and modifications, including Lee and Miller [22], Booth et al [3], Renshaw and Haberman [32], among others. Hyndman and Ullah [17] further extends the methods to a functional data framework which incorporates nonparametric smoothing, functional principal component decomposition and times series analysis into their model. However, all of these works deal with forecasting mortality for a single population. When applied to multiple populations independently, they tend to result in divergent forecasts of life

expectancy in the long run, even though those populations may share extensive similarities in socioeconomic, environmental and biological conditions. As a result, it is often desirable that the forecasts do not diverge when modelling the mortality rates simultaneously for multiple correlated subpopulations, such as different sexes or states in a country, or several countries within a larger region (for instance Europe). As such, one of the challenges in human mortality modelling is to model multiple subpopulations simultaneously (while taking into account the heterogeneity among them) which results in nondivergent forecasts when the subpopulations are highly correlated.

Such nondivergent forecasts are termed as coherent forecasts in the literature (see for example, Li and Lee [27]). Li and Lee [27] extends the Lee-Carter model to a group of populations for coherent forecasting by proposing an additional common factor between the multiple populations and projecting the population period effects as stationary time series. Some variants of Li and Lee [27] model have also been developed, such as Li [23]. Oeppen [30] applies the methods of compositional data analysis to transform the Lee-Carter model and then extends it to a multiple-decrement life table. Cairns et al [5] proposes an Age-Period-Cohort model for two related populations which incorporates a mean-reverting stochastic spread that allows for different trends in mortality improvement rates in the short-run but parallel improvements in the long run. Hyndman et al [14] develops the product-ratio method which achieves coherent mortality forecasting by the convergence of the ratios of the forecast age-specific death rates from any two subpopulations to appropriate constants. Their numerical examples for Swedish sex-specific mortality data and Australian state-specific mortality data show that the forecasting accuracy is homogenised across subpopulations while the life expectancy in the long run presents convergence. Enchev et al [8] reviews and compares some multi-population mortality models, including some variations of the Li and Lee [27] model and the common-age-effect (CAE) model of Kleinow [20]. Other developments in this area include Jarner and Kryger [18], Li and Hardy [25], Hatzopoulos and Haberman [12], Antonio et al [1], Li et al [26], Wan and Bertschi [34], Shang [33], and Li et al [24].

In this paper, we propose a new method, namely weighted multilevel functional principal component analysis (wMFPCA), for coherent mortality forecasting, in the sense that the life expectancy in different populations does not diverge in the long run. We treat the mortality rates of subpopulations within a large population as a set of multilevel functional data (e.g. male and female mortality rates within a country), and use wMFPCA to extract core information from the functional data and analyse them at multilevel scale, so that the model incorporates both overall information from the population as a whole and specific information from the subpopulations. Comparing to the Product-Ratio model [14], the proposed model possesses several major differences in nature which will be discussed later in the paper. The usefulness of the proposed model is demonstrated by the application to the sex-specific mortality data for nine developed countries.

The rest of the paper is organised as follows. In Section 2, we introduce the model and the framework of wMFPCA in details. In Section 3, we demonstrate the performance of the model in terms of long-run forecasting convergence in the context of sex-specific mortality data for the UK. The model is also applied to the sex-specific mortality data of nine developed countries and the performance is compared with the independent model [17] as well as the Product-Ratio model [14]. The paper is concluded in Section 4 with some discussions.

2. Methodology

2.1. A review of the functional principal component analysis (FPCA)

Functional principal component analysis (FPCA) is established as a statistical methodology to analyse functional data. It is a natural extension of the multivariate PCA, allowing the data dimension to increase from finite space to infinite space. The core of this technique is based on the Karhunen-Loève (KL) expansion of a stochastic process [19, 28]. Since then, there have been many theoretical developments of FPCA mainly in two streams: the linear operator point of view and the kernel operator point of view [2, 4, 10, 31, 35]. The ultimate goal of FPCA is to reduce the infinite dimension of functional data into finite dimensions in principal directions of variation.

Let $Y(x)$ be a squared integrable stochastic process defined on a set $\mathcal{T}$ (time interval for instance) with finite variance $\int_{\mathcal{T}} E[Y(x)^2]dx < \infty$. Let $\mu(x)$ be its mean function and $K(x, x')$ be its covariance function. By Karhunen-Loève (KL) expansion, under certain continuity conditions $Y(x)$ can be expressed as:

$$Y(x) = \mu(x) + \sum_{k=1}^{\infty} \beta_k \phi_k(x), \quad x \in \mathcal{T},$$

where, $\{\phi_k\}_{k \geq 1}$ is a set of orthonormal eigenfunctions such that

$$K(x, x') = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(x')$$

with $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ being the corresponding eigenvalues, and $\beta_k = \int_{\mathcal{T}} (Y(x) - \mu(x)) \phi_k(x) dx$ are uncorrelated random variables with mean zero and variance λ_k . **These random variables are the projection of the centred stochastic process $Y(x) - \mu(x)$ in the direction of the k -th eigenfunction ϕ_k and are called principal component scores.** In practice, only the first few eigenfunctions are needed to represent the important features of a sample of random functions. Therefore we can truncate the expansion at the first N terms to obtain an approximation of $Y(x)$:

$$Y(x) \approx \mu(x) + \sum_{k=1}^N \beta_k \phi_k(x), \quad x \in \mathcal{T}.$$

2.2. Multilevel FPCA

In practice it is often the case that a set of functional data comprise a number of subsets with strong correlation so that the functional data have a multilevel structure, such as the mortality rates for males and females in a country. The conventional FPCA is not suitable for this type of functional data since it ignores the heterogeneity among subgroups. To address the challenges, Di et al [7] proposes a multilevel FPCA (MFPCA) which combines FPCA and multilevel mixed effect models, Crainiceanu et al [6] introduces generalized multilevel functional linear regression models, Zipunnikov et al [36] considers multilevel functional principal component analysis for high-dimensional data, Goldsmith et al [9] studies generalized multilevel functional-on-scalar regression, and Meyer et al [29] develops Bayesian function-on-function regression for multilevel

functional data.

Let $Y_{i,j}(x)$ denote a random function on $[0, 1]$ for subject i and group j , $i = 1, 2, \dots, I$, $j = 1, 2, \dots, J$. Consider a two-way functional ANOVA model

$$Y_{i,j}(x) = \mu(x) + \eta_j(x) + Z_i(x) + W_{i,j}(x),$$

where $\mu(x)$ is the overall mean function, $\eta_j(x)$ is the group-specific mean shift from the overall mean, $Z_i(x)$ is the subject-specific deviation from the group-specific mean, and $W_{i,j}(x)$ is the residual subject and group specific deviation from the subject-specific mean. In such a model, $\mu(x)$ and $\eta_j(x)$ are fixed functions while $Z_i(x)$ and $W_{i,j}(x)$ are zero-mean stochastic processes. $Z_i(x)$, termed as the level-one functions, and $W_{i,j}(x)$, the level-two functions, are assumed to be uncorrelated. Both the level-one and level-two functions can then be decomposed using the Karhunen-Loève (KL) expansion as follows:

$$Z_i(x) = \sum_{k=1}^{\infty} \beta_{i,k} \phi_k^{(1)}(x),$$

$$W_{i,j}(x) = \sum_{l=1}^{\infty} \gamma_{i,j,l} \phi_l^{(2)}(x),$$

where $\phi_k^{(1)}(x)$ and $\beta_{i,k}$ are the level-one eigenfunctions and the corresponding principal component scores with mean 0 and variance $\lambda_k^{(1)}$, and $\phi_l^{(2)}(x)$ and $\gamma_{i,j,l}$ are the level-two eigenfunctions and the scores with mean 0 and variance $\lambda_l^{(2)}$. It follows that $\{\beta_{i,k} : k = 1, 2, \dots\}$ are uncorrelated with $\{\gamma_{i,j,l} : l = 1, 2, \dots\}$. By these expansions, the model can then be expressed as

$$Y_{i,j}(x) = \mu(x) + \eta_j(x) + \sum_{k=1}^{\infty} \beta_{i,k} \phi_k^{(1)}(x) + \sum_{l=1}^{\infty} \gamma_{i,j,l} \phi_l^{(2)}(x). \quad (1)$$

To obtain the eigenfunctions (or the principal components), we first estimate the covariance functions. Let $K_T(x_s, x_r) = \text{cov}\{Y_{i,j}(x_s), Y_{i,j}(x_r)\}$ be the overall covariance function, $K_B(x_s, x_r) = \text{cov}\{Y_{i,j}(x_s), Y_{i,k}(x_r)\}$ the covariance function between the level-two units within the same level-one unit. And define $K_W(x_s, x_r) = K_T(x_s, x_r) - K_B(x_s, x_r)$. It follows from (1) that

$$K_T(x_s, x_r) = \sum_{k=1}^{\infty} \lambda_k^{(1)} \phi_k^{(1)}(x_s) \phi_k^{(1)}(x_r) + \sum_{l=1}^{\infty} \lambda_l^{(2)} \phi_l^{(2)}(x_s) \phi_l^{(2)}(x_r),$$

$$K_B(x_s, x_r) = \sum_{k=1}^{\infty} \lambda_k^{(1)} \phi_k^{(1)}(x_s) \phi_k^{(1)}(x_r),$$

$$K_W(x_s, x_r) = \sum_{l=1}^{\infty} \lambda_l^{(2)} \phi_l^{(2)}(x_s) \phi_l^{(2)}(x_r).$$

In practice, each function $Y_{i,j}(x)$ is observed at a set of grid points. We assume that a common grid $\{x_1, \dots, x_p\}$ is used for every subject and group. Then, $\mu(x_s)$, $\eta_j(x_s)$, $K_T(x_s, x_r)$ and $K_B(x_s, x_r)$ can be estimated as follows:

$$\hat{\mu}(x_s) = \frac{1}{IJ} \sum_{i,j} Y_{i,j}(x_s),$$

$$\hat{\eta}_j(x_s) = \frac{1}{I} \sum_i Y_{i,j}(x_s) - \hat{\mu}(x_s),$$

$$\hat{K}_T(x_s, x_r) = \frac{1}{IJ} \sum_{i,j} \{Y_{i,j}(x_s) - \hat{\mu}(x_s) - \hat{\eta}_j(x_s)\} \{Y_{i,j}(x_r) - \hat{\mu}(x_r) - \hat{\eta}_j(x_r)\},$$

$$\hat{K}_B(x_s, x_r) = \frac{2}{IJ(J-1)} \sum_i \sum_{j_1 < j_2} \{Y_{i,j_1}(x_s) - \hat{\mu}(x_s) - \hat{\eta}_{j_1}(x_s)\} \{Y_{i,j_2}(x_r) - \hat{\mu}(x_r) - \hat{\eta}_{j_2}(x_r)\}.$$

Since $K_W(x_s, x_r)$ is estimated by the difference between $\hat{K}_T(x_s, x_r)$ and $\hat{K}_B(x_s, x_r)$, it may not always be positive definite. This issue can be solved by trimming the pairs of eigenvalue and eigenfunction where the eigenvalue is negative [11]. Therefore the eigenfunctions can be estimated by the following steps:

Step 1. Estimate the mean and covariance functions $\hat{\mu}(x_s)$, $\hat{\eta}_j(x_s)$, $\hat{K}_T(x_s, x_r)$ and $\hat{K}_B(x_s, x_r)$; set $\hat{K}_W(x_s, x_r) = \hat{K}_T(x_s, x_r) - \hat{K}_B(x_s, x_r)$.

Step 2. Decompose $\hat{K}_B(x_s, x_r)$ to obtain $\hat{\lambda}_k^{(1)}$, $\hat{\phi}_k^{(1)}(x)$.

Step 3. Decompose $\hat{K}_W(x_s, x_r)$ to obtain $\hat{\lambda}_l^{(2)}$, $\hat{\phi}_l^{(2)}(x)$; then trim those pairs of eigenvalues and eigenfunctions where the eigenvalue is negative.

Step 4. Estimate the principal component scores for both levels.

The technical details for estimating the principal component scores for both levels are provided in Appendix A. In practice determining the number of eigenfunctions in the approximation is always important in the context of principal component analysis. One method is to use a cross validation procedure. But a relatively simple solution is to estimate the number of components based on the proportion of variance explained. In the MFPCA model, the cumulative percentage of the total variation explained by the first N_1 components at level-one is given by the ratio $\sum_{k=1}^{N_1} \lambda_k^{(1)} / \sum_{k=1}^{\infty} \lambda_k^{(1)}$. Similarly, the cumulative percentage of the total variation explained by the first N_2 components at level-two is given by the ratio $\sum_{l=1}^{N_2} \lambda_l^{(2)} / \sum_{l=1}^{\infty} \lambda_l^{(2)}$. Therefore the number of eigenfunctions can be determined given a threshold for the cumulative percentage of the total variation explained by the first N_1 (N_2) components at level-one (level-two). And the proportion of the variation explained by the level-one and level-two against the overall variation of the entire data set can be calculated as $\frac{\sum_{k=1}^{\infty} \lambda_k^{(1)}}{\sum_{k=1}^{\infty} \lambda_k^{(1)} + \sum_{l=1}^{\infty} \lambda_l^{(2)}}$ and

$\frac{\sum_{l=1}^{\infty} \lambda_l^{(2)}}{\sum_{k=1}^{\infty} \lambda_k^{(1)} + \sum_{l=1}^{\infty} \lambda_l^{(2)}}$ respectively.

In reality, functional data are usually observed with error. Smoothing is therefore needed before the analysis is carried out. In the literature on FPCA, there exist three methods commonly used for smoothing: smoothing the data before applying FPCA [31], introducing a penalty term [31], and smoothing the covariance function [35]. In our later application in mortality modelling, we will adopt the first method to smooth the raw data first before the MFPCA is conducted. This will be discussed again in next section.

2.3. Weighted MFPCA for coherent mortality forecasting

In this section we propose a weighted MFPCA (wMFPCA) approach for coherent forecast of the future mortality for several subpopulations within a large population. Let $y_{t,j}(x)$ denote the log of the death rate in the j th subpopulation in year t for age x ($t = 1, \dots, n, j = 1, \dots, m$). And we assume that, for each t and j , there exists an underlying smooth function $f_{t,j}(x)$ that we are observing with error at discrete points of x . Suppose we have observed $\{x_i, y_{t,j}(x_i)\}$, $i = 1, \dots, p$, $t = 1, \dots, n$, $j = 1, \dots, m$. Then,

$$y_{t,j}(x_i) = f_{t,j}(x_i) + \sigma_{t,j}(x_i)e_{t,j,i},$$

where $e_{t,j,i}$ is i.i.d standard normal random variables and $\sigma_{t,j}(x_i)$ allows the amount of noise to vary with age x .

In demographic modelling, it is often the case that more recent data tend to have more impact on the results than those in the distant past. Hence, we propose to incorporate the concept of weights into the MFPCA model to allow recent mortality data to have more significant effect on the forecasting results, as discussed below.

Let $\hat{f}_{t,j}(x)$ be the smoothed function from $y_{t,j}(x)$, and $w_t = \kappa(1 - \kappa)^{n-t}$ be a geometrically decaying weight with $0 < \kappa < 1$. Then we estimate the overall mean function $\mu(x)$ using a weighted average:

$$\hat{\mu}(x) = \sum_{t,j} \frac{w_t}{m} \hat{f}_{t,j}(x),$$

and estimate the subpopulation-specific shifts from the overall mean, $\eta_j(x)$, by

$$\hat{\eta}_j(x) = \sum_t w_t \hat{f}_{t,j}(x) - \hat{\mu}(x).$$

The mean-adjusted functional data are denoted as $\hat{f}_{t,j}^*(x) = \hat{f}_{t,j}(x) - \hat{\mu}(x) - \hat{\eta}_j(x)$. Then the weights are incorporated into the mean-adjusted data $\hat{f}_{t,j}^*(x)$ in order to calculate the weighted functional principal components. Similar to Hyndman and Ullah [17], the weighted multilevel functional principal components $\hat{\phi}_k^{(1)}(x)$ and $\hat{\phi}_l^{(2)}(x)$ can be obtained as follows. We first denote $\{\hat{f}_{t,j}^*(x_i) : i = 1, \dots, p, t = 1, \dots, n, j = 1, \dots, m\}$ as an $nm \times p$ matrix F . Then the weights are incorporated by multiplying a weight matrix W to F , where $W = \text{diag}(\underbrace{\frac{w_1}{m}, \frac{w_1}{m}, \dots, \frac{w_1}{m}}_m, \dots, \underbrace{\frac{w_n}{m}, \frac{w_n}{m}, \dots, \frac{w_n}{m}}_m)$. Thus F is

transformed into $F^* = WF$. Based on the weighted mean-adjusted matrix F^* , the covariance functions $\hat{K}_T(x_s, x_r)$ and $\hat{K}_B(x_s, x_r)$ can be estimated using the methods

discussed in the previous subsection and $\widehat{K}_W(x_s, x_r) = \widehat{K}_T(x_s, x_r) - \widehat{K}_B(x_s, x_r)$ is calculated accordingly. The weighted multilevel functional principal components $\widehat{\phi}_k^{(1)}(x)$ and $\widehat{\phi}_l^{(2)}(x)$ can then be obtained by decomposing $\widehat{K}_B(x_s, x_r)$ and $\widehat{K}_W(x_s, x_r)$. Finally, the principal component scores at the two levels can be estimated using the methods described in Appendix A. Therefore, the entire weighted MFPCA model for coherent mortality forecasting of the subpopulations is given as:

$$y_{t,j}(x_i) = \mu_j(x_i) + \sum_{k=1}^{N_1} \beta_{t,k} \phi_k^{(1)}(x_i) + \sum_{l=1}^{N_2} \gamma_{t,j,l} \phi_l^{(2)}(x_i) + \sigma_{t,j}(x_i) e_{t,j,i},$$

where $\mu_j(x_i) = \mu(x_i) + \eta_j(x_i)$ is the mean of the subpopulation j . Forecasts can then be achieved by extrapolating the level-one scores $\{\beta_{t,1}, \dots, \beta_{t,N_1}\}$ and the level-two scores $\{\gamma_{t,j,1}, \dots, \gamma_{t,j,N_2}\}$, using time series models. Since the level-one scores are independent we assume independent and possibly non-stationary autoregressive integrated moving average (ARIMA) models for $\beta_{t,1}, \dots, \beta_{t,N_1}$. As for the level-two scores $\gamma_{t,j,l}$, it is noted that, for the same order l ($l = 1, \dots, N_2$), the set of scores for different subpopulations $\{\gamma_{t,1,l}, \dots, \gamma_{t,m,l}\}$ share the same basis $\phi_l^{(2)}(x)$, which implies that the scores $\{\gamma_{t,1,l}, \dots, \gamma_{t,m,l}\}$ are not independent and hence multivariate autoregressive moving average (ARMA) models with stationary restriction are used for the level-two scores. The stationary constraint is to ensure the coherence in the mortality forecasting.

Let $\widehat{\beta}_{(n+h),k}$ denote the h -step ahead forecast of $\beta_{(n+h),k}$ and $\widehat{\gamma}_{(n+h),j,l}$ denote the h -step ahead forecast of $\gamma_{(n+h),j,l}$. Then the h -step ahead forecast of $y_{(n+h),j}(x)$ is obtained as:

$$\widehat{y}_{(n+h),j}(x) = \widehat{\mu}_j(x) + \sum_{k=1}^{N_1} \widehat{\beta}_{(n+h),k} \widehat{\phi}_k^{(1)}(x) + \sum_{l=1}^{N_2} \widehat{\gamma}_{(n+h),j,l} \widehat{\phi}_l^{(2)}(x).$$

Due to the fact that each component in the model is uncorrelated with each other, the forecasting variance can be obtained by adding up the variance of each component:

$$\begin{aligned} \text{var}\{y_{(n+h),j}(x)\} &= \widehat{\sigma}_{\mu_j}^2(x) + \sum_{k=1}^{N_1} u_{(n+h),k} \{\widehat{\phi}_k^{(1)}(x)\}^2 \\ &\quad + \sum_{l=1}^{N_2} v_{(n+h),j,l} \{\widehat{\phi}_l^{(2)}(x)\}^2 + \{\widehat{\sigma}_{(n+h),j}(x)\}^2, \end{aligned}$$

where, $\widehat{\sigma}_{\mu_j}^2(x)$ denotes the variance of the smoothed estimates of the means and can be estimated from the smoothing method used; $u_{(n+h),k} = \text{var}\{\widehat{\beta}_{(n+h),k}\}$ and $v_{(n+h),j,l} = \text{var}\{\widehat{\gamma}_{(n+h),j,l}\}$ can be obtained from the time series models; and the estimated variance of the observational errors $\{\widehat{\sigma}_{(n+h),j}(x)\}^2$ can be obtained from the historical data [17]. Thus a $100(1-\alpha)\%$ approximate prediction interval can be constructed as $\widehat{y}_{(n+h),j}(x) \pm z_\alpha \sqrt{\text{var}\{y_{(n+h),j}(x)\}}$, where z_α is the $(1-\alpha/2)$ quantile of the standard normal distribution.

Note that the weights $\{w_t\}$ are controlled by the tuning parameter κ . The larger κ is, the faster the weights for the past years are decaying over time. Therefore, κ represents people's perception on how past data should be weighted. The value of the parameter κ can be specified *a priori* or determined empirically by minimising the average root mean square forecast error (RMSFE):

$$\frac{1}{m} \sum_{j=1}^m \left(\frac{1}{p} \sum_{i=1}^p \{y_{(n+h),j}(x_i) - \hat{y}_{(n+h),j}(x_i)\}^2 \right)^{1/2}.$$

3. Application

In this section we evaluate our weighted MFPCA model by applying it to some sex-specific mortality data. We first consider the coherent forecasting of the sex-specific mortality rates in the UK to test if the forecasting result appears to be non-divergent for the male and female life expectancy. We then compare the forecasting accuracy of our model with those of the unweighted MFPCA and the Product-Ratio model [14] as well as the independent model [17] using the sex-specific mortality rates of nine different countries.

3.1. Coherent forecasting for the male and female mortality in the UK

The sex-specific mortality data in the UK are obtained from the Human Mortality Database [13]. The data provide the observed mortality rates for every one year at each age. We select the years from 1950 to 2010 and aggregate the data for the ages 100 and above to avoid the anomalous mortality rates during the first and second world wars and the erratic rates above 100. The mortality curves are first smoothed using the weighted penalized regression splines with a monotonicity constraint [17]. The smoothed curves are shown in Figure 1.

The weighted MFPCA model is then fitted to the data. The weight parameter κ is determined by minimizing the average root mean square forecast error (RMSFE) based on a rolling window approach; see Section 3.2 for the details. The mean functions and the level-one and level-two functional principal components are estimated as discussed in the previous section. The level-one scores are modelled by univariate ARIMA time series using the R package ‘forecast’ [15, 16]. The level-two scores are modelled by multivariate stationary ARMA time series using the R package ‘MTS’. **The orders of the ARIMA and ARMA models are selected based on the Akaike information criterion (AIC).**

The results show that the level-one principal components explain 94% of the total variation whilst the level-two principal components do less than 6%. The first three level-one principal components account for around 97.7%, 1.6% and 0.3% of the total level-one variation respectively, and the first three level-two principal components account for around 61.8%, 17.2% and 7.0% of the total level-two variation respectively. Together, the first three principal components in both levels explain about 98.8% of the total variation. Hence three principal components in both levels are used in the approximation ($N_1 = N_2 = 3$); more principal components are tested in the experiments but no significant differences are found in the results. The estimates and the

Figure 1. The smoothed log death rates for the males and females in the UK from 1950 to 2010, viewed as functional data series. Different colours represent different years.

forecasts with 30-year horizon are shown in Figures 2 and 3 respectively.

Figure 2 displays the level-one components which are shared by both male and female subpopulations. As can be seen in the figure, the first level-one principal component models the degree of variation in mortality among different age groups, which is very large for children age groups and relatively small for twenties and old age groups. The second and third principal components present more complicated patterns and are difficult to interpret.

Figure 3 displays the sex-specific deviations from the overall mean function, the level-two principal component scores which are specific for male and female subpopulations, and the common level-two principal components. It can be observed that the male and female deviations from the overall mean are complementary to each other, because of the method used for estimating the mean and deviations.

As demonstration, Figure 4 shows the forecasted log mortality rates for males and females in the UK by the weighted MFPCA model for the years 2025 (15-year forecast horizon) and 2040 (30-year forecast horizon). Figure 5 shows the 30-year forecasts of the male and female cohort life expectancy at birth by our proposed model as well as the independent model [17]. It can be observed that the independent model produces a more divergent forecasting result than our coherent model does. **From the observed data, the actual life expectancy of the males and females in 2010 are 78.4 years and 82.3 years respectively, indicating a 3.9 years sex gap. The life expectancy of the males and females in 2040 forecasted by the independent model are 82.5 years and 86.8 years respectively, in contrast to 84.1 years and 87.6 years by our weighted MFPCA model. Thus the sex gap is increased by the independent model to 4.3 years in 2040 while it is narrowed down by our coherent model to 3.5 years.** Hence, our coherent model demonstrates slow convergence in the forecasted life expectancy in the UK.

3.2. Comparison with other models

We now compare the forecasting accuracy of our proposed weighted MFPCA model (denoted by wMFPCA) with the unweighted (or equally weighted) MFPCA model

Figure 2. Level-one decomposition: the overall mean function, the first three level-one functional principal components and their corresponding scores with 30-year forecast horizon and 80% prediction intervals.

Figure 3. Level-two decomposition: the sex-specific deviations from the overall mean function, the first three level-two functional principal components and their corresponding scores with 30-year forecast horizon and 80% prediction intervals using stationary multivariate ARMA models.

Figure 4. Forecasted log mortality rates for males and females in the UK by the weighted MFPCA model. Top panels: 15-year forecast horizon; bottom panels: 30-year forecast horizon. Left panels: UK males; right panel: UK females. The shaded areas represent the 80% prediction intervals.

Forecasted life expectancy in UK

Figure 5. 30-year forecasted life expectancy for males and females in the UK by the weighted MFPCA model (solid lines) and the independent model (dotted lines). The top (red) lines represent female group and the bottom (blue) lines represent male group. The shaded areas represent the 80% prediction intervals by the weighted MFPCA model.

(uMFPCA), the Product-Ratio model (Product-Ratio) and the independent model (Independent). We first consider the UK male and female mortality data from 1950 to 2010. We adopt a rolling window approach, that is, to use the data from 1950 to $1981+t$ as the observations and forecast the mortality rates for years $1981+t+1, \dots, 1981+t+20$, for $t = 0, \dots, 9$. Thus, ten sets of forecasts for each of 1 to 20 year horizons are obtained and compared with the actual values. For a given forecast horizon h ($h = 1, \dots, 20$), the out-of-sample root mean square forecast error (RMSFE) for the j th subpopulation is defined as:

$$RMSFE_j(h) = \sqrt{\frac{1}{10p} \sum_{t=0}^9 \sum_{i=1}^p \{y_{(1981+t+h),j}(x_i) - \hat{y}_{(1981+t+h),j}(x_i)\}^2}$$

where p is the number of age groups ($p=101$ in this example). For the UK male and female mortality data, the values of the out-of-sample RMSFE are obtained using the weighted MFPCA model, the unweighted MFPCA model, the Product-Ratio model and the independent model, and are illustrated in Figure 6.

It can be observed that the wMFPCA model performs similarly to the independent model for the UK male mortality and both outperform the unweighted MFPCA model and the Product-Ratio model. On the other hand, for the UK females, the performance of wMFPCA is very close to those of the unweighted MFPCA model and the Product-Ratio model, and they all present significantly higher forecasting accuracy than the independent model. Overall, the average RMSFE (over the forecast horizons and the sex) by the weighted MFPCA model is 0.1491, compared with 0.1766 by the unweighted MFPCA, 0.1705 by the Product-Ratio model and 0.1770 by the independent model. These results show that, in overall terms, the proposed weighted MFPCA

Figure 6. Out-of-sample RMSFEs for males (left panel) and females (right panel) of the UK using the weighted MFPCA model (solid lines), unweighted MFPCA model (dotted lines), the Product-Ratio model (dotted lines) and the independent model (dashed lines).

Table 1. The RMSFEs by the weighted MFPCA model (wMFPCA), the unweighted MFPCA model (uMFPCA), the Product-Ratio model and the independent model for nine developed countries. The smallest forecasting error for each country is highlighted in bold.

	wMFPCA	uMFPCA	Product-Ratio	Independent
AUS	0.1922	0.2077	0.1988	0.2010
USA	0.1065	0.1151	0.1133	0.1285
UK	0.1491	0.1766	0.1705	0.1770
FRA	0.1641	0.2030	0.2020	0.1969
BEL	0.2098	0.2346	0.2187	0.2227
ESP	0.2866	0.2710	0.2886	0.3056
CAN	0.1433	0.1654	0.1557	0.1705
NLD	0.1665	0.1903	0.1812	0.1962
ITA	0.2010	0.2439	0.2371	0.2413
Average	0.1799	0.2009	0.1962	0.2044

model performs considerably better than the unweighted MFPCA, the Product-Ratio model and the independent model for the UK mortality forecast.

The same experiment is further conducted to the male and female mortality data of eight other developed countries, including Australia, USA, France, Belgium, Spain, Canada, Netherlands and Italy. The values of the average RMSFE by the four models for all nine countries are presented in Table 1. It can be seen that the weighted MFPCA model provides the smallest forecasting errors for all the countries except Spain, and it is confirmed that in overall terms the weighted MFPCA model outperforms the other three models.

4. Conclusion and Discussion

We have proposed a new model for coherent forecasting of mortality rates for several correlated subpopulations in a large population. The model is based on the multilevel functional principal component analysis (MFPCA) framework and is incorporated with weights which allow more recent data to have more significant impact on forecasting results. The mortality curves of different subpopulations are treated as a set of mul-

tilevel functional data, and the principal components and their corresponding scores are obtained at two levels and forecasts are made by extrapolating the scores using time series models. The numerical example shows that the proposed model can achieve nondivergent forecasts for the subpopulations. Using the sex-specific mortality data of nine developed countries, we have demonstrated the usefulness of the proposed model and compared the results with the unweighted MFPCA, the Product-Ratio model and the independent model.

The wMFPCA model consists of a group mean, a decomposition of the level-one functions and a decomposition of the level-two functions. The level-one functions govern the common properties of the entire population while the level-two functions involve the properties specific to the subpopulations. Therefore the model possesses a simple and explicit form which makes the modelling procedure easy and interpretable. Since the level-one functions are largely dominated by their first principal component, the age pattern of change at this level is relatively fixed. However, the governance of the level-two functions are separated into several principal components, which gives a flexible age pattern of change at level-two. In such way, the wMFPCA model provides more flexibility in the age pattern of change compared with the independent model which contains only a single level of principal components.

It is noted that, although the wMFPCA model has a similar structure to the Product-Ratio model the former has several advantages over the latter in its nature. The core of the wMFPCA method is to find the variations of multilevel functional data at different levels and then decompose them using the Karhunen-Loève expansion. There is no need to transform the data as done in the Product-Ratio method (calculating the product functions and the ratio functions) so the functional data are allowed to speak more for themselves within the wMFPCA framework. The Product-Ratio model assumes the mortality rates in the subpopulations having equal variances, which is not needed in the wMFPCA model. Moreover, within the wMFPCA framework, the percentage of variance explained by each principal component at each level can be calculated explicitly and easily. Knowing these percentages of variance explained, we are able to explicitly evaluate the importance of the principal components at different levels.

Coherence is another important issue to be discussed. In our model, coherence is ensured by applying stationary multivariate ARMA models to the level-two principal component scores. The convergence of the level-two principal component scores to certain constants under stationary ARMA models guarantees that the long-term forecasts of the level-two functions also converge to their age-specific constants. As the level-two functions converge to constants, they gradually lose ability to affect the change of mortality. The change of mortality is then entirely dominated by level-one functions. Since the level-one functions are commonly shared by different subpopulations, their impact on the changes of mortality in those subpopulations is equal. The forecasted mortality differences among subpopulations are therefore constrained, leading to a similar constraint on the forecasted life expectancy as well.

Acknowledgements

This work is based on part of the PhD thesis entitled “Gaussian Process and Functional Data Methods for Mortality Modelling” by the first author from the University of Leicester. The work was supported by the Institute and Faculty of Actuaries (IFoA) and the College of Science and Engineering of the University of Leicester (UoL). The

authors thank the Associate Editor and the reviewers for their constructive suggestions and very helpful comments.

References

- [1] K. Antonio, A. Bardoutsos, and W. Ouburg, *Bayesian Poisson log-bilinear models for mortality projections with multiple populations*, European Actuarial Journal 5(2) (2015), pp. 245-281.
- [2] P. Besse, *PCA stability and choice of dimensionality*, Stat. Probab. Lett. 13(5) (1992), pp. 405-410.
- [3] H. Booth, J. Maindonald, and L. Smith, *Applying Lee-Carter under conditions of variable mortality decline*, Population Studies 56 (2002), pp. 325-336.
- [4] D. Bosq, *Linear Processes in Function Spaces: Theory and Applications*, Springer, New York, 2000.
- [5] A. Cairns, D. Blake, K. Dowd, G. Coughlan, and M. Khalaf-Allah, *Bayesian stochastic mortality modelling for two populations*, ASTIN Bulletin 41 (2011), pp. 29-59.
- [6] C.M. Crainiceanu, A.M. Staicu, and C.Z. Di, *Generalized multilevel functional regression*, Journal of the American Statistical Association 104(488) (2009), pp. 1550-1561.
- [7] C.Z. Di, C.M. Crainiceanu, B.S. Caffo, and N.M. Punjabi, *Multilevel functional principal component analysis*, The Annals of Applied Statistics 3(1) (2009), pp. 458-488.
- [8] V. Enchev, T. Kleinow, and A.J.G. Cairns, *Multi-population mortality models: fitting, forecasting and comparisons*, Scandinavian Actuarial Journal 2017 (2017), pp. 319-342.
- [9] J. Goldsmith, V. Zipunnikov, and J. Schrack, *Generalized multilevel functional-on-scalar regression and principal components analysis*, Biometrics 71 (2015), pp. 344-353.
- [10] P. Hall, H.G. Müller, and J.L. Wang, *Properties of principal component methods for functional and longitudinal data analysis*, Ann. Stat. 34(3) (2006), pp. 1493-1517.
- [11] P. Hall, H.G. Müller, and F. Yao, *Modeling sparse generalized longitudinal observations with latent Gaussian processes*, J. R. Stat. Soc. Ser. B Statist. Methodol. 70 (2008), pp. 703-723.
- [12] P. Hatzopoulos and S. Haberman, *Common mortality modeling and coherent forecasts: An empirical analysis of worldwide mortality data*, Insurance: Mathematics & Economics 52 (2013), pp. 320-337.
- [13] Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2010. Available at <http://www.mortality.org/>.
- [14] R.J. Hyndman, H. Booth, and F. Yasmeen, *Coherent mortality forecasting: The product-ratio method with functional time series models*, Demography 50 (2013), pp. 261-283.
- [15] R.J. Hyndman and Y. Khandakar, *Automatic time series forecasting: The forecast package for R*, Journal of Statistical Software, 27(3) (2008), pp. 1-22.
- [16] R.J. Hyndman, and H.L. Shang, *Forecasting functional time series*, Journal of the Korean Statistical Society 38 (2009), pp. 199-211.
- [17] R.J. Hyndman, and M.S. Ullah, *Robust forecasting of mortality and fertility rates: A functional data approach*, Computational Statistics & Data Analysis 51 (2007), pp. 4942-4956.
- [18] S. Jarner and E. Kryger, *Modelling adult mortality in small populations: The SAINT model*, ASTIN Bulletin 41(2) (2011), pp. 377-418.
- [19] K. Karhunen, *Zur spektraltheorie stochastischer prozesse*, Annales Academiae Scientiarum Fennicae 37 (1946), pp. 1-37.
- [20] T. Kleinow, *A common age effect model for the mortality of multiple populations*, Insurance: Mathematics & Economics 63 (2014), pp. 147-152.
- [21] R. Lee, and L. Carter, *Modeling and forecasting the time series of U.S. mortality*, Journal of the American Statistical Association 87 (1992), pp. 659-671.

- [22] R.D. Lee, and T. Miller, *Evaluating the performance of the Lee-Carter method for forecasting mortality*, Demography 38 (2001), pp. 537-549.
- [23] J. Li, *A Poisson common factor model for projecting mortality and life expectancy jointly for females and males*, Population Studies: A Journal of Demography 67(1) (2013), pp. 111-126.
- [24] J. Li, L. Tickle, and N. Parr, *A multi-population evaluation of the Poisson common factor model for projecting mortality joint for both sexes*, Journal of Population Research 33(4) (2016), pp. 333-360.
- [25] J.S.-H. Li and M. Hardy, *Measuring basis risk in longevity hedges*, North American Actuarial Journal 15(2) (2011), pp. 177-200.
- [26] J.S.-H., Li, R. Zhou, and M. Hardy, *A step-by-step guide to building two-population stochastic mortality models*, Insurance: Mathematics & Economics 63 (2015), pp. 121-134.
- [27] N. Li and R. Lee, *Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method*, Demography 42 (2005), pp. 575-594.
- [28] M. Loève, *Fonctions alatoires a decomposition orthogonale exponentielle*, La Revue Scientifique 84 (1946), pp. 159-162.
- [29] M.J. Meyer, B.A. Coull, F. Versace, P. Cinciripini, and S.M. Jeffrey, *Bayesian function-on-function regression for multilevel functional data*, Biometrics 71(3) (2015), pp. 563-574.
- [30] J. Oeppen, *Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data*, Technical Report, Rostock, Germany: Max Planck Institute for Demographic Research, 2008.
- [31] J.O. Ramsay and B.W. Silverman, *Functional Data Analysis*, Springer, 2005.
- [32] A. Renshaw and S. Haberman, *Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections*, Applied Statistics 52 (2003), pp. 119-137.
- [33] H.L. Shang, *Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method*, The Annals of Applied Statistics 10(3) (2016), pp. 1639-1672.
- [34] C. Wan and L. Bertschi, *Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach*, Insurance: Mathematics & Economics 63 (2015), pp. 66-75.
- [35] F. Yao, H.G. Müller, and J.L. Wang, *Functional data analysis for sparse longitudinal data*, J. Am. Stat. Assoc. 100 (470) (2005), pp. 577-590.
- [36] V. Zupnik, B. Caffo, D.M. Yousem, C. Davatzikos, B.S. Schwartz, and C. Crainiceanu, *Multilevel functional principal component analysis for high-dimensional data*, Journal of Computational and Graphical Statistics 20(4) (2011), pp. 852-873.

Appendix A. Estimation of the principal component scores

In standard FPCA the principal component scores (PC scores) can be estimated straightforwardly by numerical integration, using $\beta_k = \int_{\mathcal{T}} (Y(x) - \mu(x)) \phi_k(x) dx$. However, estimating the PC scores for multilevel functional data involves extra complication because the two levels of eigenfunctions, namely $\phi_k^{(1)}(x)$ and $\phi_l^{(2)}(x)$, are not necessarily mutually orthogonal. Di et al [7] assumes that $\beta_{i,k}$ and $\gamma_{i,j,l}$ both follow Gaussian distributions and proposes two methods for estimating the scores. In the first method, the estimated $\mu(x)$, $\eta_j(x)$, $\lambda_k^{(1)}$, $\lambda_l^{(2)}$, $\phi_k^{(1)}(x)$ and $\phi_l^{(2)}(x)$ are treated as fixed while the PC scores $\beta_{i,k}$ and $\gamma_{i,j,l}$ are random effects to be estimated. The model

can then be written as:

$$\begin{cases} Y_{i,j}(x) = \mu(x) + \eta_j(x) + \sum_{k=1}^{N_1} \beta_{i,k} \phi_k^{(1)}(x) + \sum_{l=1}^{N_2} \gamma_{i,j,l} \phi_l^{(2)}(x) + \varepsilon_{i,j}(x); \\ \beta_{i,k} \sim N(0, \lambda_k^{(1)}); \quad \gamma_{i,j,l} \sim N(0, \lambda_l^{(2)}); \quad \varepsilon_{i,j}(x) \sim N(0, \sigma^2), \end{cases}$$

where $\varepsilon_{i,j}(x)$ is i.i.d and appears only when the functional data are observed with error (if the functional data are already smoothed or perfectly observed, simply remove this term from the model). This is in fact a linear mixed effect model in nature, therefore the random effects can be estimated by either best linear unbiased prediction (BLUP) or Markov Chain Monte Carlo (MCMC). This model is appropriate for both dense and sparse functional data. However, it faces computational challenges, especially when dealing with large volume of data. Therefore the second method is adopted in our proposed model.

The intuition behind the second method is to project each mean-adjusted function into the space spanned by each eigenfunction at different levels. We denote

$$A_{i,j,k} = \int_0^1 \{Y_{i,j}(x) - \mu(x) - \eta_j(x)\} \phi_k^{(1)}(x) dx = \beta_{i,k} + \sum_{l=1}^{N_2} \gamma_{i,j,l} c_{k,l} + \epsilon_{i,j,k}^{(1)}, \quad (A.1)$$

$$B_{i,j,l} = \int_0^1 \{Y_{i,j}(x) - \mu(x) - \eta_j(x)\} \phi_l^{(2)}(x) dx = \gamma_{i,j,l} + \sum_{k=1}^{N_1} \beta_{i,k} c_{k,l} + \epsilon_{i,j,l}^{(2)}, \quad (A.2)$$

where $c_{k,l} = \int_0^1 \phi_k^{(1)}(x) \phi_l^{(2)}(x) dx$ is the inner product of two eigenfunctions at different levels, and $\epsilon_{i,j,k}^{(1)}$ and $\epsilon_{i,j,l}^{(2)}$ are the residuals which can be expressed as

$$\epsilon_{i,j,k}^{(1)} = \int_0^1 \left\{ \sum_{l=N_2+1}^{\infty} \gamma_{i,j,l} \phi_l^{(2)}(x) + \varepsilon_{i,j}(x) \right\} \phi_k^{(1)}(x) dx,$$

$$\epsilon_{i,j,l}^{(2)} = \int_0^1 \left\{ \sum_{k=N_1+1}^{\infty} \beta_{i,k} \phi_k^{(1)}(x) + \varepsilon_{i,j}(x) \right\} \phi_l^{(2)}(x) dx$$

respectively. Equations (A.1) and (A.2) can be rewritten in matrix format. Letting

$$\mathbf{A}_{i,j} = (A_{i,j,1}, A_{i,j,2}, \dots, A_{i,j,N_1})^T, \quad \mathbf{B}_{i,j} = (B_{i,j,1}, B_{i,j,2}, \dots, B_{i,j,N_2})^T,$$

$$\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,N_1})^T, \quad \boldsymbol{\gamma}_{i,j} = (\gamma_{i,j,1}, \gamma_{i,j,2}, \dots, \gamma_{i,j,N_2})^T,$$

$$\boldsymbol{\epsilon}_{i,j}^{(1)} = (\epsilon_{i,j,1}^{(1)}, \epsilon_{i,j,2}^{(1)}, \dots, \epsilon_{i,j,N_1}^{(1)})^T, \quad \boldsymbol{\epsilon}_{i,j}^{(2)} = (\epsilon_{i,j,1}^{(2)}, \epsilon_{i,j,2}^{(2)}, \dots, \epsilon_{i,j,N_2}^{(2)})^T,$$

then (A.1) and (A.2) are transformed as follows:

$$\left\{ \begin{array}{ll} \mathbf{A}_{i,j} = \beta_i + C\gamma_{i,j} + \epsilon_{i,j}^{(1)}, & \mathbf{B}_{i,j} = \gamma_{i,j} + C^T\beta_i + \epsilon_{i,j}^{(2)}, \\ \beta_i \sim N(0, \Lambda^{(1)}), & \gamma_{i,j} \sim N(0, \Lambda^{(2)}), \\ \epsilon_{i,j}^{(1)} \sim N(0, \sigma_1^2 \mathbf{I}_{N_1}), & \epsilon_{i,j}^{(2)} \sim N(0, \sigma_2^2 \mathbf{I}_{N_2}), \end{array} \right.$$

where $C = (c_{k,l})$ is an $N_1 \times N_2$ matrix, $\Lambda^{(1)} = \text{diag}(\lambda_1^{(1)}, \lambda_2^{(1)}, \dots, \lambda_{N_1}^{(1)})$, and $\Lambda^{(2)} = \text{diag}(\lambda_1^{(2)}, \lambda_2^{(2)}, \dots, \lambda_{N_2}^{(2)})$. The residual variance σ_1^2 and σ_2^2 can be estimated from the data. This is also a linear mixed effects model so the PC scores can be estimated by BLUP or MCMC.