Functional time series forecasting

Rob J Hyndman*, Han Lin Shang

Department of Econometrics & Business Statistics, Monash University, VIC, 3800, Melbourne, Australia

---

Abstract

We propose forecasting functional time series using weighted functional principal component regression and weighted functional partial least squares regression. These approaches allow for smooth functions, assign higher weights to more recent data, and provide a modeling scheme that is easily adapted to allow for constraints and other information. We illustrate our approaches using age-specific French female mortality rates from 1816 to 2006 and age-specific Australian fertility rates from 1921 to 2006, and show that these weighted methods improve forecast accuracy in comparison to their unweighted counterparts. We also propose two new bootstrap methods to construct prediction intervals, and evaluate and compare their empirical coverage probabilities.

Key words: Demographic forecasting, Functional data, Functional partial least squares, Functional principal components, Functional time series.

2000 MSC: primary, 62G08, 62H25; secondary, 62G09, 62J07, 62P05

---

*Corresponding author.

Telephone number: +61 3 9905 5141

Fax number: +61 3 9905 5474

Email address: `Rob.Hyndman@buseco.monash.edu.au` (Rob J Hyndman)

1. Introduction

We are interested in forecasting a series of functional data observed over time. The functions are observed (with error) at times $t = 1, \dots, n$, and we wish to forecast the functions for times $t = n+1, \dots, n+h$. Let $\{y_t(x_i)\}$ denote the observed data, where $i = 1, \dots, p$. We assume that there are underlying L_2 continuous and smooth functions $\{f_t(x)\}$ such that

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \tag{1}$$

where $\{\varepsilon_{t,i}\}$ are independent and identically distributed variables with zero mean and unit variance, and $\sigma_t(x)$ allows for heteroskedasticity.

For example, the curves may be weekly "yield curves" where x denotes the time to maturity of a debt, and $y_t(x)$ is the interest rate applicable for loans taken out in week t. A demographic application occurs when $y_t(x)$ is the mortality rate for people of age x in year t. There are many other applications involving functional time series including those studied in Yao et al. (2005), Erbas et al. (2007), Hyndman and Ullah (2007), and Hyndman and Booth (2008).

Hyndman and Ullah (2007) used nonparametric smoothing on each curve $y_t(x)$ separately to obtain estimates of the smooth functions $\{f_t(x)\}$. Then they proposed a functional principal component approach to decompose the time series of functional data into a number of principal components and their scores. Their model can be written as follows:

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x), \tag{2}$$

where $\phi_k(x)$ is the kth principal component function, $\{\beta_{1,k}, \dots, \beta_{n,k}\}$ are the corresponding scores, $\{e_t(x)\}$ denote independent and identically distributed random functions with zero mean, and $K < n$. Because principal component scores are uncorrelated, Hyndman and Ullah (2007) suggested that each univariate time series $\{\beta_{t,k}\}$, $k = 1, \dots, K$, can be forecasted independently using a univariate time series model. By multiplying the forecasted principal component scores with the principal components, estimated future curves are obtained.

In this paper, we extend the Hyndman-Ullah approach in several directions. First, in Section 2, we introduce geometrically decreasing weights in the principal component decomposition to allow more recent data to affect the results more than data in the distant past. This is particularly important in demography where we can have over 200 years of data, and data from the 18th and 19th centuries may not be so useful in determining the best principal components for forecasting. A second extension we propose in Section 3 is to use functional partial least squares regression instead of functional principal components. The advantage of this proposal is that the decomposition is derived with an eye to forecasting rather than simply modeling the historical data. Finally, in Section 4 we propose some new approaches to constructing prediction intervals for functional time series forecasts based on nonparametric bootstrap methods.

In Section 5, we evaluate the point forecast accuracy, computational time, and empirical coverage probabilities of the prediction intervals using two real data sets. Section 6 concludes with some thoughts on how the methods developed here might be further extended.

2. Weighted functional principal components

The mean function $\mu(x)$ is estimated using a weighted average

$$\hat{\mu}(x) = \sum_{t=1}^{n} w_t \hat{f}_t(x),$$

where $\hat{f}_t(x)$ is the smoothed curve estimated from $y_t(x)$, and $w_t = \kappa(1-\kappa)^{n-t}$ is a geometrically decreasing weight with $0 < \kappa < 1$. If one seeks a robust estimator, then the L_1 median of the estimated smoothed curves can be used instead (see Hyndman and Ullah, 2007). The mean- or median-adjusted functional data are denoted as $\hat{f}_t^*(x) = \hat{f}_t(x) - \hat{\mu}(x)$.

We can then apply functional principal components (FPC) analysis (Ramsay and Dalzell, 1991) to $\{\hat{f}_t^*(x)\}$ to obtain the principal components $\{\phi_k(x)\}$ and their scores $\{\beta_{t,k}\}$. It is well-known that FPC are constructed by maximizing variance within predictors to achieve minimal L_2 loss of information. An account of the statistical properties of FPC, along with applications of the methodology, are given by Ramsay and Silverman (2002, 2005). Papers covering the development of FPC include those of Rice and Silverman (1991), Silverman (1995, 1996), and Reiss and Ogden (2007). Significant treatments of the theory of FPC are given by Dauxois et al. (1982), Cai and Hall (2006), Hall and Hosseini-Nasab (2006), Hall et al. (2006), Hall and Horowitz (2007), Hall and Hosseini-Nasab (2009) and Delaigle et al. (2009).

For computational ease, in practice we discretize $\hat{f}_t^*(x)$ on a dense grid of q equally spaced points $\{x_1^*, \dots, x_q^*\}$ that span the interval $[x_1, x_p]$. Denote the discretized $\hat{f}_t^*(x)$ as an $n \times q$ matrix $\boldsymbol{G}^*$ and let $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{G}^*$, where $\boldsymbol{W} = \text{diagonal}(w_1, \dots, w_n)$. Applying singular value decomposition to $\boldsymbol{G}$ gives $\boldsymbol{G} = \boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{V}'$, where $\phi_k(x_j^*)$ is the (j,k)th element of $\boldsymbol{\Psi}$. If $\boldsymbol{B} = \boldsymbol{G}\boldsymbol{\Psi}$, then $\beta_{t,k}$ is the (t,k)th element of $\boldsymbol{B}$. Other values of $\phi_k(x)$ can be computed using linear interpolation.

Combining (1) and (2), and replacing all terms with their estimates, we obtain

$$y_t(x_i) = \hat{\mu}(x_i) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x_i) + \hat{e}_t(x_i) + \hat{\sigma}_t(x_i)\hat{\varepsilon}_{t,i}. \tag{3}$$

The scores $\{\beta_{t,k}\}$ are assumed to follow independent univariate time series models. Then, conditioning on the observed data $\boldsymbol{\mathcal{I}} = \{y_t(x_i) : t = 1, \dots, n;\, i = 1, \dots, p\}$ and the set of principal components $\boldsymbol{\Phi} = \{\phi_1(x), \dots, \phi_K(x)\}$, we can obtain h-step-ahead forecasts of $y_{n+h}(x)$:

$$\hat{y}_{n+h|n}(x) = \mathrm{E}[y_{n+h}(x)|\boldsymbol{\mathcal{I}}, \boldsymbol{\Phi}] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n+h|n,k}\phi_k(x),$$

where $\hat{\beta}_{n+h|n,k}$ denotes the h-step-ahead forecast of $\beta_{n+h,k}$.

The above method is identical to that given by Hyndman and Ullah (2007) except that they selected the weights to be zero for outlying observations, and to be one for non-outlying observations, based on a simple statistical test for outliers. However, we propose using geometrically decaying weights applied to the principal component to take account of changes in the functional shapes over time. The value of parameter κ can be determined empirically by minimizing the mean integrated forecast error (MISFE) (Hyndman and Ullah, 2007):

$$\text{MISFE}(h) = \int_{x_1}^{x_p} \big(y_{n+h}(x) - \hat{y}_{n+h|n}(x)\big)^2\, dx.$$

Hyndman and Booth (2008) found that the forecasts are insensitive to the choice of K, provided K is large enough. That is, while there is little cost apart from computational speed in choosing a large K, a small K may result in poor forecast accuracy. Greenshtein and Ritov (2004) and Greenshtein (2006) termed this phenomenon "persistence in functional linear predictor selection". Consequently, in this analysis we choose $K = 6$ for all components, which should be larger than any of the components really require (Hyndman and Booth, 2008).

3. Weighted functional partial least squares regression

Although the first few FPC explain a large amount of variation in the historical data, they do not necessarily provide good predictors for future data (Escabias et al., 2004; Aguilera et al., 2006). To address this problem, Escabias et al. (2004) considered a stepwise method for selecting which principal components to include. Recently, Kargin and Onatski (2008) proposed a functional predictive factor method, and selected latent components that have maximal correlation with responses. In this section, we introduce another basis decomposition method: weighted functional partial least squares regression (FPLSR).

Partial least squares regression has found numerous applications in multivariate calibration, since it was proposed by Wold (1975) and its subsequent development by Martens and Naes (1989). It has recently been extended to the functional setting by Preda and Saporta (2005a,b), Reiss and Ogden (2007) and Krämer et al. (2008). In contrast to FPC, FPLSR extracts uncorrelated latent component scores by maximizing the covariance function between functional predictors and functional responses. For a functional time series, we consider the situation of a lagged predictor $\hat{f}_{t-1}(x)$ and functional response $\hat{f}_t(x)$, for $t = 2, \dots, n$. We express the so-called outer-relationship as

$$\hat{\boldsymbol{f}}^*(x) = \sum_{k=1}^{\infty} \boldsymbol{\beta}_k \psi_k(x), \quad \text{and} \quad \hat{\boldsymbol{g}}^*(x) = \sum_{k=1}^{\infty} \boldsymbol{\beta}_k \phi_k(x), \tag{4}$$

where $\hat{\boldsymbol{f}}^*(x) = \boldsymbol{W}[f_1^*(x), \dots, f_{n-1}^*(x)]'$ and $\hat{\boldsymbol{g}}^*(x) = \boldsymbol{W}[f_2^*(x), \dots, f_n^*(x)]'$ represent the weighted decentralized functional predictors and responses with zero means, $\boldsymbol{W}$ represents the weight matrix as before, $\boldsymbol{\beta}_k$ represents the common latent component scores, and $\psi_k(x)$ and $\phi_k(x)$ are the kth latent components of predictors and responses respectively.

The common weighted latent component scores are linear combinations of predictors and responses:

$$\boldsymbol{\beta}_k = \int_{x_1}^{x_p} \hat{\boldsymbol{f}}^*(x) w_k(x) dx = \int_{x_1}^{x_p} \hat{\boldsymbol{g}}^*(x) m_k(x)\, dx, \tag{5}$$

To ensure orthogonality, FPLSR uses a deflation procedure to reduce the rank of predictors and responses successively by one at each iteration. This attractive property is known as the Wedderburn rank diminishing operator (Wedderburn, 1974). To be more specific, $\hat{\boldsymbol{f}}^*(x)$ and $\hat{\boldsymbol{g}}^*(x)$ must be deflated at each step by subtracting the current latent component score to give

$$\hat{\boldsymbol{f}}_k^*(x) = (\boldsymbol{I} - \boldsymbol{\beta}_k \boldsymbol{\beta}_k') \hat{\boldsymbol{f}}_{k-1}^*(x) \qquad \text{and} \qquad \hat{\boldsymbol{g}}_k^*(x) = (\boldsymbol{I} - \boldsymbol{\beta}_k \boldsymbol{\beta}_k') \hat{\boldsymbol{g}}_{k-1}^*(x), \tag{6}$$

starting with $\boldsymbol{f}_0^*(x) = \boldsymbol{f}^*(x)$ and $\boldsymbol{g}_0^*(x) = \boldsymbol{g}^*(x)$. Then the functions, $w_k(x)$ and $m_k(x)$, are obtained as follows:

$$w_k(x) = \boldsymbol{\beta}_k' \hat{\boldsymbol{f}}_{k-1}^*(x) = \int_{x_1}^{x_p} m_k(u) [\hat{\boldsymbol{g}}_{k-1}^*(u)]' du\, \hat{\boldsymbol{f}}_{k-1}^*(x), \tag{7}$$

$$\text{and} \qquad m_k(x) = \boldsymbol{\beta}_k' \hat{\boldsymbol{g}}_{k-1}^*(x) = \int_{x_1}^{x_p} w_k(v) [\hat{\boldsymbol{f}}_{k-1}^*(v)]' dv\, \hat{\boldsymbol{g}}_{k-1}^*(x). \tag{8}$$

Plugging (8) into (7) gives the following relationship

$$w_k(x) = \int_{x_1}^{x_p} \int_{x_1}^{x_p} w_k(v) [\hat{\boldsymbol{f}}_{k-1}^*(v)]' \hat{\boldsymbol{g}}_{k-1}^*(u) [\hat{\boldsymbol{g}}_{k-1}^*(u)]' \hat{\boldsymbol{f}}_{k-1}^*(x)\, dv\, du.$$

Thus we can obtain $w_k(x)$ iteratively, starting with the estimate $w_k^{(0)}(x)$, via the updating equation

$$w_k^{(i)}(x) = \int_{x_1}^{x_p}\int_{x_1}^{x_p} w_k^{(i-1)}(v)[\hat{\boldsymbol{f}}_{k-1}^*(v)]'\hat{\boldsymbol{g}}_{k-1}^*(u)[\hat{\boldsymbol{g}}_{k-1}^*(u)]'\hat{\boldsymbol{f}}_{k-1}^*(x)\,dv\,du. \tag{9}$$

Computationally, this is asymptotically equivalent to setting $w_1(x)$ to be the largest eigenvector of $\boldsymbol{F}'\boldsymbol{G}\boldsymbol{G}'\boldsymbol{F}$, where $\boldsymbol{F} = \boldsymbol{W}\boldsymbol{G}^*$ with the last row omitted, and $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{G}^*$ but with the first row omitted. This, in turn, is equivalent to putting $w_1(x)$ equal to the largest eigenvector of $\boldsymbol{F}'\boldsymbol{G}$. Similarly, $m_1(x)$ equals the largest eigenvector of $\boldsymbol{G}'\boldsymbol{F}$, $m_k(x)$ is the largest eigenvector of $\boldsymbol{G}_{k-1}'\boldsymbol{F}_{k-1}$ and $w_k(x)$ is the largest eigenvector of $\boldsymbol{F}_{k-1}'\boldsymbol{G}_{k-1}$, where $\boldsymbol{G}_{k-1}$ and $\boldsymbol{F}_{k-1}$ are constructed analogously to $\boldsymbol{G}$ and $\boldsymbol{F}$.

From (4) and (5), we obtain

$$\hat{\boldsymbol{g}}^*(x) = \sum_{k=1}^{\infty}\int_{x_1}^{x_p}\hat{\boldsymbol{f}}^*(u)w_k(u)du\,\phi_k(x) = \int_{x_1}^{x_p}\hat{\boldsymbol{f}}^*(u)b(x,u)du,$$

where

$$b(x,u) = \sum_{k=1}^{\infty} w_k(u)\phi_k(x)$$

is the functional autoregression coefficient. Then using (4) again, and the orthogonality of $\boldsymbol{\beta}_k$, we find that

$$\phi_k(x) = (\boldsymbol{\beta}_k'\boldsymbol{\beta}_k)^{-1}\boldsymbol{\beta}_k'\hat{\boldsymbol{g}}^*(x).$$

Thus, we approximate $b(x,u)$ by

$$\hat{b}(x,u) = \sum_{k=1}^{K} w_k(u)(\boldsymbol{\beta}_k'\boldsymbol{\beta}_k)^{-1}\boldsymbol{\beta}_k'\hat{\boldsymbol{g}}^*(x), \tag{10}$$

for some finite K. The forecasted curves are then given by

$$\hat{f}_{n+1|n}(x) = \hat{\mu}(x) + \int_{x_1}^{x_p}[\hat{f}_n(u) - \hat{\mu}(u)]\hat{b}(x,u)du. \tag{11}$$

For $h > 1$, (11) can be applied iteratively.

4. Distributional forecasts

Prediction intervals are a valuable tool for assessing the probabilistic uncertainty associated with point forecasts. As emphasized in Chatfield (1993) and Chatfield (2000), it is important to provide interval forecasts as well as point forecasts so as to:

1. assess future uncertainty;
2. enable different strategies to be planned for the range of possible outcomes indicated by the interval forecasts;
3. compare forecasts from different methods more thoroughly; and
4. explore different scenarios based on different assumptions.

In the weighted FPC and FPLSR, there are four sources of errors that need to be taken into account: (1) smoothing error in estimating $f_t(x)$, (2) mean function error in estimating $\mu(x)$, (3) error in forecasting the scores $\beta_{t,k}$, and (4) error in the model residuals $e_t(x)$. While the first two sources of errors depend on the aptness of smoothing, the last two sources of errors depend on the predictiveness of the fitted model.

In Section 4.1, we introduce a nonparametric bootstrap method to construct prediction intervals for the weighted FPC, which overcomes the normality assumption in Hyndman and Ullah (2007) and Hyndman and Booth (2008). In Section 4.2, we propose a nonparametric bootstrap method to construct prediction intervals for the weighted FPLSR.

4.1. Bootstrap curves for FPC

Using univariate time series models, we can obtain multi-step-ahead forecasts for the principal component scores, $\{\hat{\beta}_{1,k}, \dots, \hat{\beta}_{n,k}\}$. Let the h-step-ahead forecast errors be given by $\hat{\xi}_{t,h,k} = \hat{\beta}_{t,k} - \hat{\beta}_{t|t-h,k}$ for $t = h+1, \dots, n$. These can then be sampled with replacement to give a bootstrap sample of $\beta_{n+h,k}$:

$$\hat{\beta}^{(\ell)}_{n+h|n,k} = \hat{\beta}_{n+h|n,k} + \hat{\xi}^{(\ell)}_{*,h,k}, \qquad \ell = 1, \dots, L,$$

where $\hat{\xi}^{(\ell)}_{*,h,k}$ are sampled with replacement from $\{\hat{\xi}_{t,h,k}\}$.

Assuming the first K principal components approximate the data relatively well, the model residual should contribute nothing but random noise. Consequently, we can bootstrap the model fit errors $\{\boldsymbol{e}_t(x)\}$ in (3) by sampling with replacement from the residual term $\{\hat{e}_1(x), \dots, \hat{e}_n(x)\}$. Further, we can obtain bootstrapped smoothing errors $\hat{\varepsilon}^{(\ell)}_{n+h,i}$ by randomly sampling with replacement from $\{\hat{\varepsilon}_{1,i}, \dots, \hat{\varepsilon}_{n,i}\}$.

Adding all possible components of variability and assuming that those components of variability are uncorrelated with each other, we obtain L variants for $\hat{y}^{(\ell)}_{n+h|n}(x)$,

$$\hat{y}^{(\ell)}_{n+h|n}(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}^{(\ell)}_{n+h|n,k}\phi_k(x) + \hat{e}^{(\ell)}_{n+h|n}(x) + \hat{\sigma}_{n+h}(x)\hat{\varepsilon}^{(\ell)}_{n+h,i}.$$

Prediction intervals are produced from the bootstrap variants using percentiles.

4.2. Bootstrap residuals for FPLSR

For the weighted FPLSR model, we can adapt the methods of Faber (2002) and Fernández Pierna et al. (2003). If the model fits well (i.e., K is large enough), the residual function should be independent and identically distributed random noise. Thus it is possible to reconstitute bootstrap residuals, and thence bootstrapped data, such that

$$\hat{\boldsymbol{f}}^{(\ell)}(x) = \hat{\mu}(x) + \hat{\boldsymbol{f}}^{*}(x) + \hat{\boldsymbol{e}}^{(\ell)}(x),$$
$$\hat{\boldsymbol{g}}^{(\ell)}(x) = \hat{\mu}(x) + \hat{\boldsymbol{g}}^{*}(x) + \hat{\boldsymbol{e}}^{(\ell)}(x).$$

With L bootstrapped data and the new functional predictors, we can construct weighted FPLSR models and estimate L pseudo-regression coefficients, from which we obtain L bootstrapped forecast variants:

$$\hat{f}^{(\ell)}_{n+1|n}(x) = \hat{\mu}(x) + \int_{x_1}^{x_p} (f_n(u) - \hat{\mu}(u))\hat{b}^{(\ell)}(x,u)du + \hat{\sigma}_{n+1}(x)\hat{\varepsilon}^{(\ell)}_{n+1,i}.$$

where $\hat{b}^{(\ell)}(x,u)$ denotes the bootstrapped regression coefficients. Prediction intervals are produced from the bootstrap variants using percentiles.

4.3. Coverage probability

To examine the performance of prediction intervals constructed via these two proposed methods, we calculate the (nominally 95%) empirical coverage probability, denoted as

$$\frac{1}{mph}\sum_{t=n-m+1}^{n}\sum_{j=1}^{h}\sum_{i=1}^{p}1\left(\hat{y}_{t+j|t}^{(0.025)}(x_i) < y_{t+j}(x_i) < \hat{y}_{t+j|t}^{(0.975)}(x_i)\right)$$

where $1(\cdot)$ is the indicator function, and $\hat{y}_{t+j|t}^{(\alpha)}(x_i)$ denotes the α-quantile from the bootstrapped samples.

For $h = 1$, we can check the validity of the calculated prediction intervals by computing the difference between the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles based on the in-sample one-step-ahead forecast errors: $\{\hat{f}_{t+1}(x) - \hat{f}_{t+1|t}(x);\ t = m, \dots, n-1\}$, where m is the smallest number of observations used to fit a model. Let the averaged in-sample empirical quantile difference be denoted by $d(x)$, and let the $100(1-\alpha)\%$ h-step-ahead prediction intervals obtained from one of the bootstrap methods be denoted by $[\ell_h(x), u_h(x)]$. We advocate the adjusted intervals

$$\Big[0.5\{\ell_h(x)+u_h(x)\}-\{u_h(x)-\ell_h(x)\}p(x)\ ,\ 0.5\{\ell_h(x)+u_h(x)\}+\{u_h(x)-\ell_h(x)\}p(x)\Big], \quad (12)$$

where $p(x) = d(x)/[u_1(x) - \ell_1(x)]$. For $h = 1$, these will have the same coverage as the in-sample intervals if the forecast distributions are symmetric. We assume the same adjustment is applicable at higher forecast horizons.

5. Applications

5.1. Australian fertility data

Fertility rates in Australia, as in most other developed countries have been falling for a considerable time. Consider annual Australian fertility rates (1921–2006) for ages 15–49, obtained from the Australian Demographic Data Bank (Hyndman, 2007). These are defined as the number of live births during each calendar year, according to the age of the mother, per 1000 female resident population of the same age at 30 June.

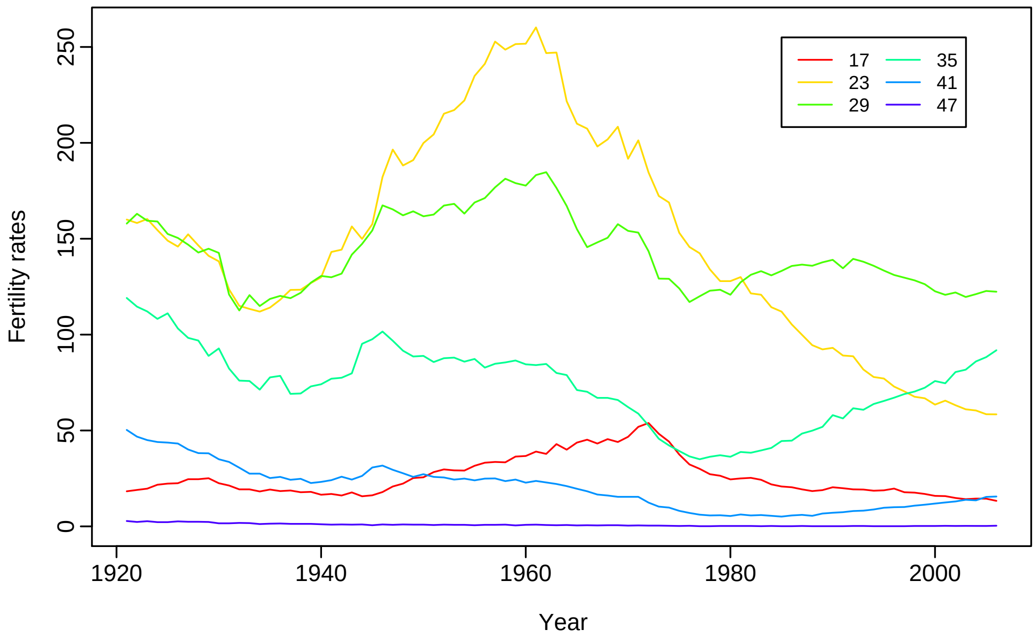


Figure 1: *Australian fertility rates from 1921 to 2006 viewed as univariate time series for ages* $\{17, 23, 29, 35, 41, 47\}$.

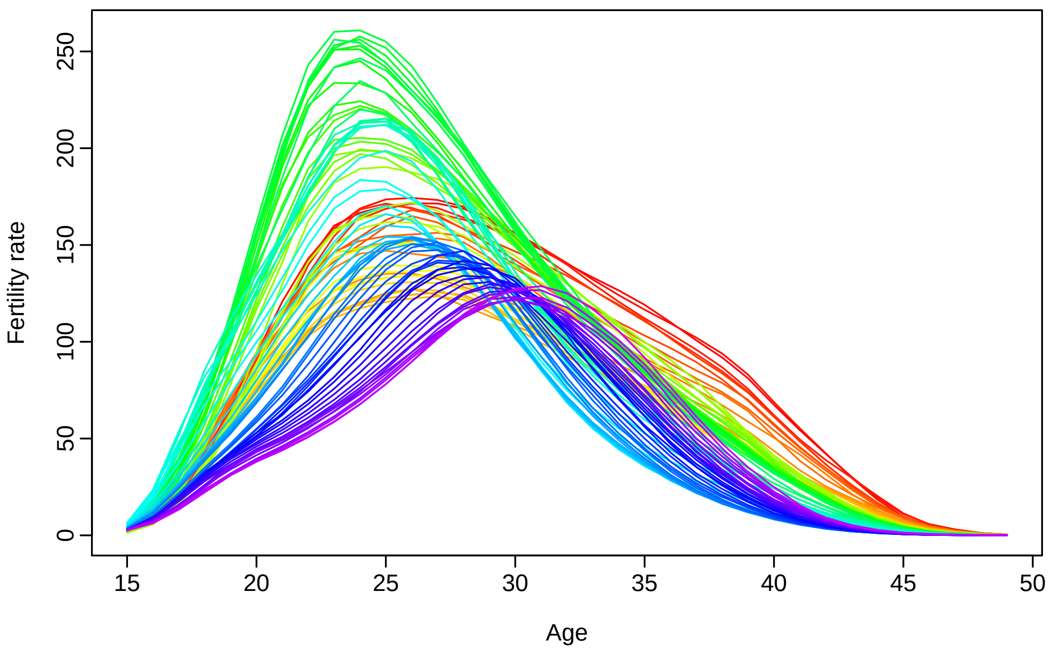


Figure 2: *Smoothed age-specific Australian fertility rates viewed as functional time series for ages 15–49, observed from 1921 to 2006.*

Figure 1 shows the fertility rates viewed as univariate time series for ages 17, 23, 29, 35, 41 and 47, while a rainbow plot (Hyndman and Shang, 2008b) is presented in Figure 2, where the colors indicate the time ordering of the curves in the same order as the colors in a rainbow (the oldest curves are red and the most recent curves are purple). In Figure 2, the fertility curves have been estimated using a weighted median smoothing B-splines with concave constraint (see Hyndman and Ullah, 2007; Ng and Maechler, 2008).

Figures 1 and 2 reflect the changing social conditions affecting fertility rates. For instance, there was an increase in fertility rates in all age groups around the end of World War II, achieving a peak in 1961, followed by a rapid decrease during the 1970s due to the increasing use of contraceptive pills, and then an increase in fertility rates at higher ages in more recent years caused by a tendency to postpone child-bearing while pursuing careers.

As the presence of outliers can seriously affect forecast accuracy, we applied the functional bagplot and functional HDR boxplot of Hyndman and Shang (2008b) to identify outliers. In these data, there were no outliers detected.

5.2. French female mortality data

Annual French female mortality rates (1816–2006) for single year-of-age from 0 to 100 were obtained from the Human Mortality Database (2008). These are the ratio of death counts per calendar year to population exposure (i.e., mid-year populations) in the relevant intervals of age and time (Wilmoth et al., 2007).

Figure 3 shows the female mortality rates of ages 0, 15, 30, 45, 60 and 75, viewed as univariate time series. A rainbow plot is presented in Figure 4, where the mortality rates are smoothed by a weighted penalized regression with monotonic constraint (see Hyndman and Ullah, 2007; Wood, 1994).

Using the method of Hyndman and Shang (2008b), we detected three functional outliers corresponding to the years 1871 (due to the Commune crisis; see Mesle and Vallin, 1991, p.53), 1918 (due to World War I and the Spanish flu pandemic) and 1944 (due to World War II). These three outliers are consequently removed from further analysis.

5.3. Point forecasts for fertility data

For simplicity of presentation, we present and interpret only the first three weighted principal components and their associated scores, although we use $K = 6$ in modeling and forecasting. Figure 5 shows the forecasted principal component scores, and their 80% and 95% prediction intervals based on exponential smoothing state space models (Hyndman et al., 2008). It is clear that the weighted principal components are modeling the fertility rates of women at different ages. For example, $\phi_1(x)$ models women in their early 20s, $\phi_2(x)$ models women at around the age of 30, and $\phi_3(x)$ models women in their late 30s. Similar patterns are also shown in Figure 6, which presents the first three predictor weights and response loadings using weighted FPLSR.

Figure 7 presents the point forecasts of Australian fertility rates from 1987 to 2006 highlighted in rainbow colors (with the longest forecast horizons shown in purple), while the data used for estimation are greyed out. The forecasts show a continuing shift to older ages for peak fertility. There is little visible difference between the forecasts of the two methods.

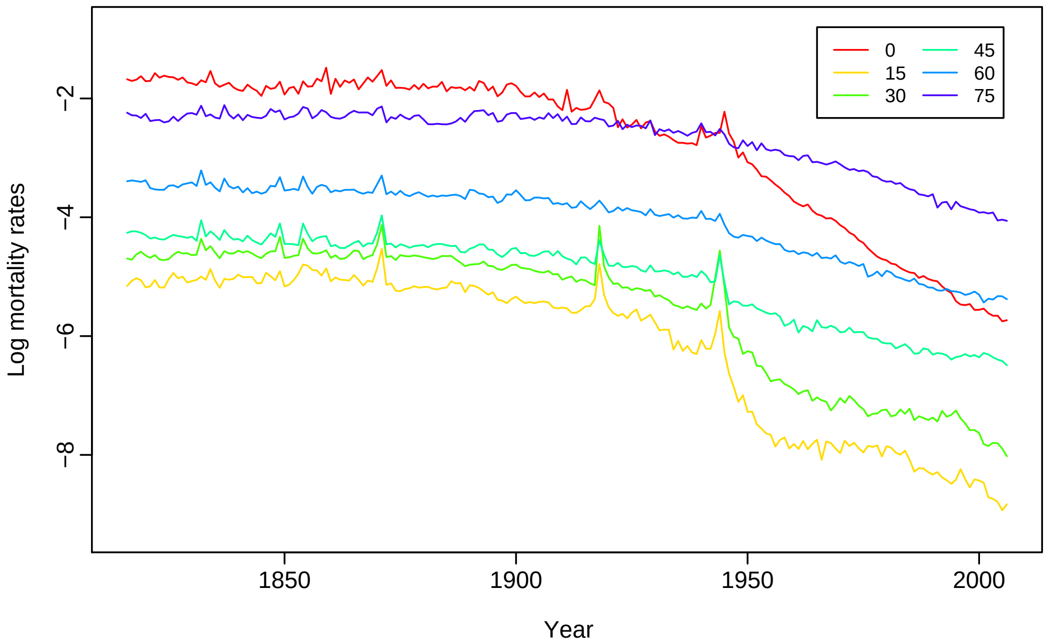


Figure 3: *French female log mortality rates from 1816 to 2006 viewed as univariate time series for ages* $\{0, 15, 30, 45, 60, 75\}$.

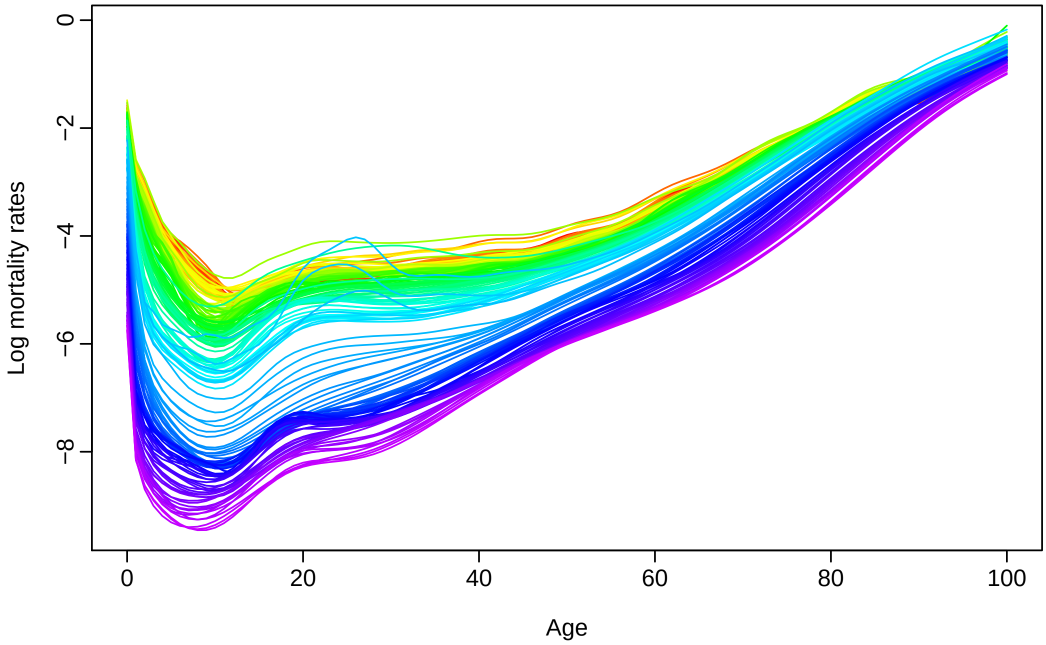


Figure 4: *Smoothed age-specific French female log mortality rates viewed as functional time series for ages 0–100, observed from 1816 to 2006.*

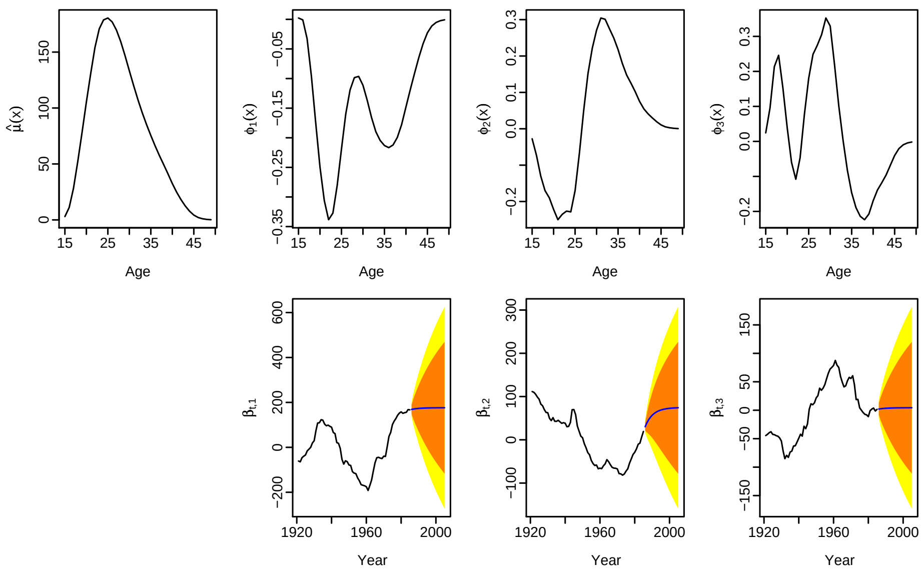


Figure 5: *The first three weighted functional principal components and associated scores for Australian fertility data. Forecasts of the principal component scores from 1987 to 2006 are shown with 80% and 95% prediction intervals, using exponential smoothing state space models.*

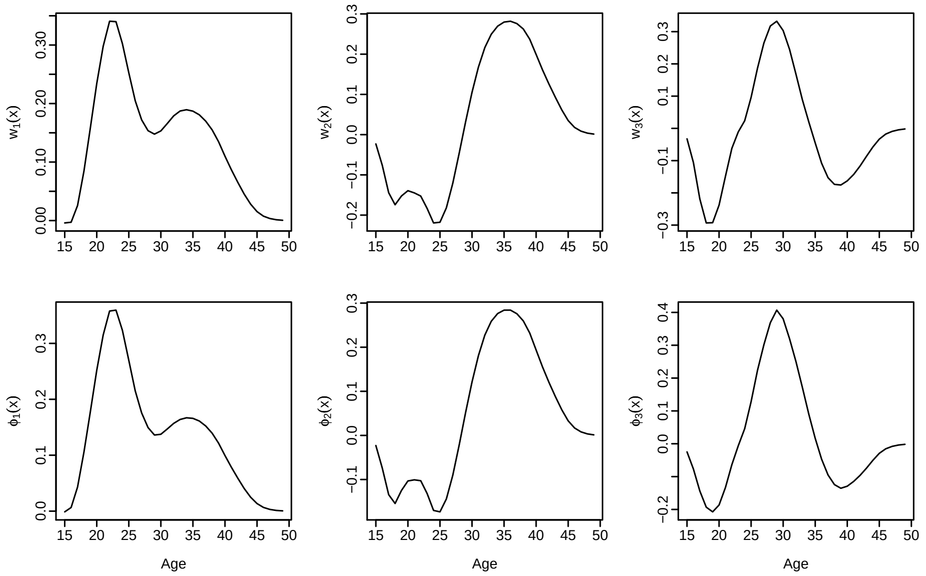
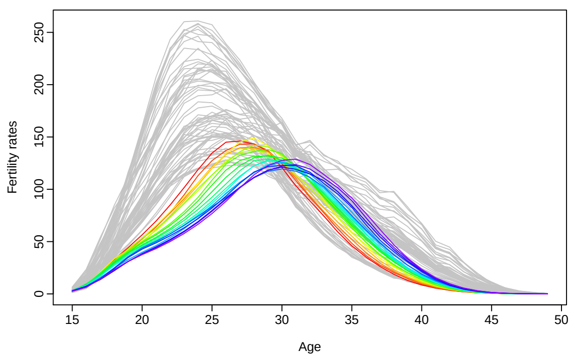
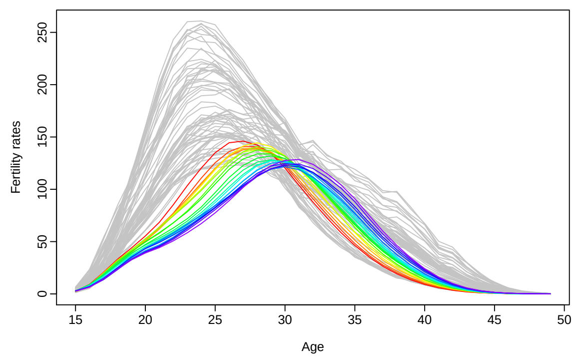


Figure 6: *The first three predictor weights and response loadings for Australian fertility data.*



(a) *Weighted FPC.*



(b) *Weighted FPLSR.*

Figure 7: *Forecasts of Australian fertility rates from 1987 to 2006 using the weighted FPC and weighted FPLSR.*

5.4. Point forecasts for mortality data

Figure 8 shows the first three weighted principal components and associated scores for the French female log mortality data (although again we use $K = 6$ in modeling and forecasting). Also shown are forecasted principal component scores, and their 80% and 95% prediction intervals based on exponential smoothing state-space models (Hyndman et al., 2008). It is apparent that the weighted principal components are modeling different movements in

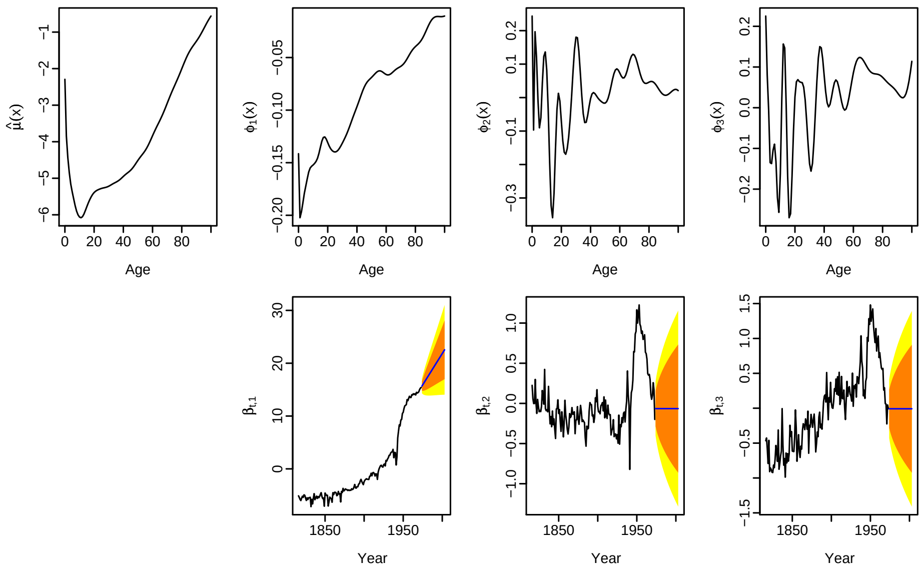


Figure 8: *The first three weighted functional principal components and associated scores for the French female log mortality data. Forecasts of the principal component scores from 1977 to 2006 are shown with 80% and 95% prediction intervals, using exponential smoothing state space models.*

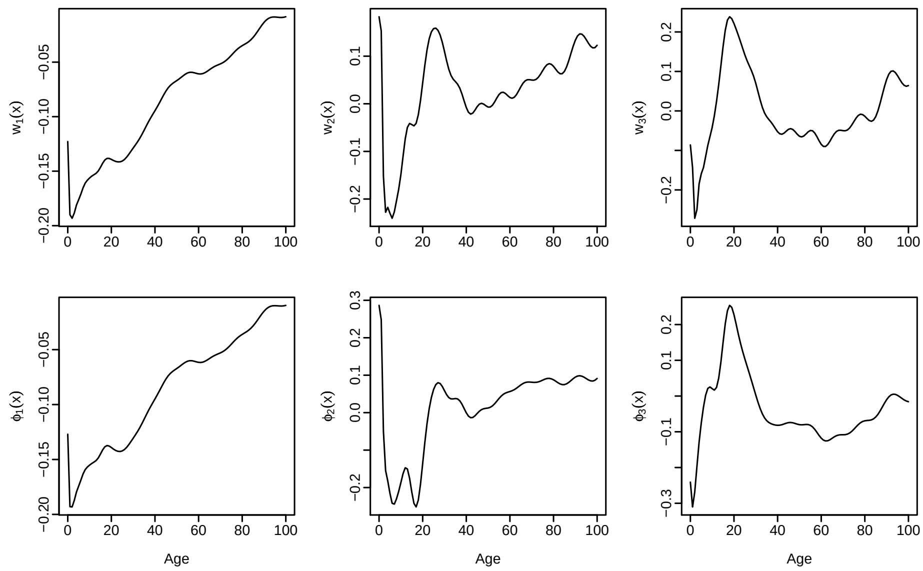
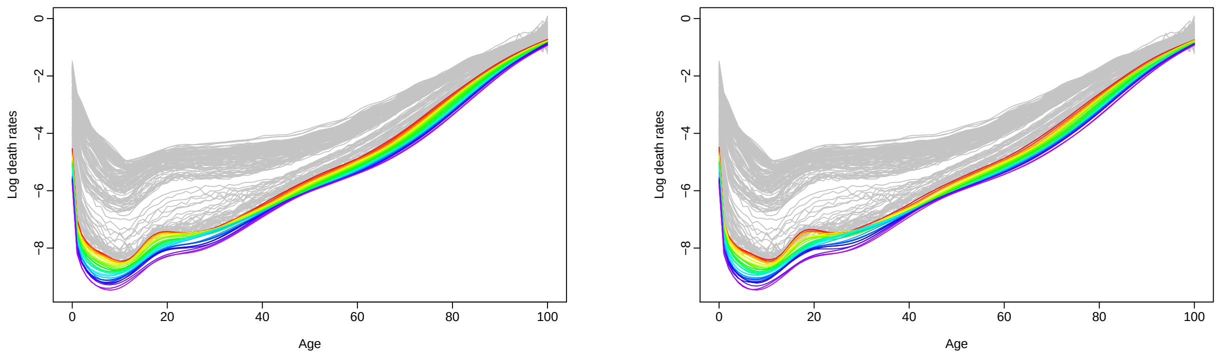


Figure 9: *The first three predictor weights and response loadings for French female log mortality data.*



(a) Weighted FPC. *(b) Weighted FPLSR.*

Figure 10: *Forecasts of French female log mortality rates from 1977 to 2006 using the weighted FPC and weighted FPLSR.*

mortality rates: $\phi_1(x)$ indicates that the major improvements in mortality have been at the younger ages, while $\phi_2(x)$ and $\phi_3(x)$ model some complex differences in mortality patterns between age groups. Similar patterns are also shown in Figure 9, which presents the first three predictor weights and response loadings using the weighted FPLSR.

Figure 10 shows the forecasts of French female log mortality rates from 1977 to 2006 highlighted in rainbow colors, while the data used for estimation are greyed out. The forecasts show a continuing decreasing trend in log mortality rates, but with different rates of decrease for different ages.

5.5. Comparisons of point forecast accuracy and computational time

We use the mean squared error (MSE) to compare the forecast accuracy of the various methods over the observed data:

$$\text{MSE}_t = \frac{1}{p}\sum_{i=1}^{p}\left[y_t(x_i) - \hat{y}_{t|t-1}(x_i)\right]^2.$$

These are then averaged over the last m years of observed data, to give an MSE value for each method and each data set. For the French female mortality data we set $m = 30$, and for the Australian fertility data we set $m = 20$. As a benchmark, we also investigate the forecast performance of the random walk model, which predicts $y_{t+1}(x)$ by $y_t(x)$.

Figure 11 presents the boxplots of MSE_t values using the various forecast methods. The weighted methods provide much smaller forecast errors than the unweighted methods, since they are able to capture the recent patterns more accurately. In addition, the weighted FPLSR shows a slight improvement over the weighted FPC for both data sets, provided K is large enough, as shown in Tables 1 and 2.

Although the difference in forecast accuracy between the weighted FPC and weighted FPLSR is small, the weighted FPLSR is computationally more efficient than the weighted FPC. This is because the forecasts obtained from the latter require separate univariate time series models for each set of principal component scores in each forecasting experiment. Times to fit 100 replicated models on an Intel Xeon 2.33GHz processor with a Windows XP operating system are summarized in Table 3. Clearly, the weighted FPLSR requires lower computational cost than the weighted FPC without sacrificing forecast accuracy.

5.6. Distributional forecasts

Figure 12 exhibits the unadjusted and adjusted prediction intervals for each data set in 2006, constructed by the two nonparametric bootstrap methods based on $L = 1000$ replications. In these two examples, the two methods are giving very similar prediction intervals.

To evaluate the empirical coverage probabilities of prediction intervals, we compare the intervals with the originally observed data. The calculation process was performed as follows: for each testing sample, prediction intervals were generated by the two bootstrap methods at a nominal coverage probability of 95%, and were tested to check if the known values fall within the specific prediction intervals. The empirical coverage probability was calculated as the ratio of the number of samples falling into their calculated prediction intervals and the number of total testing samples. As displayed in Table 4, the empirical coverage probabilities are considerably higher than the nominal coverage probability.

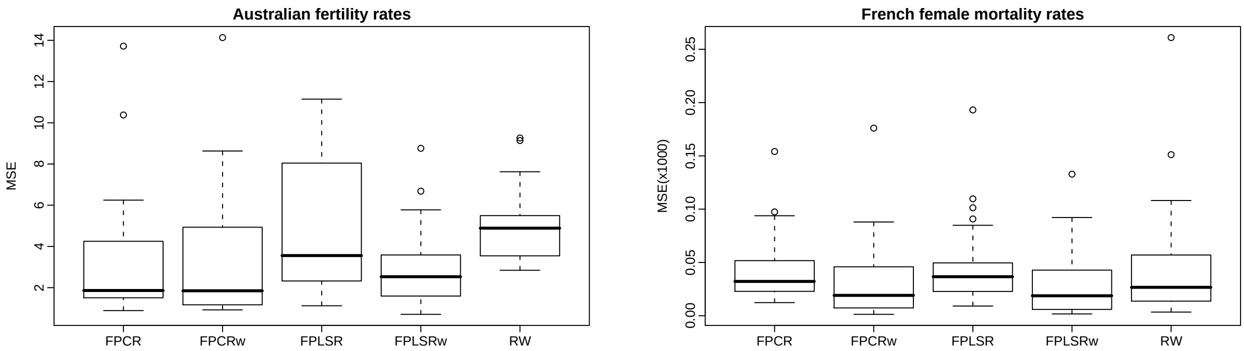


Figure 11: *Boxplots of MSE_t using the FPC, weighted FPC, FPLSR, weighted FPLSR, and random walk methods for 20 successive one-step-ahead forecasts in Australian fertility data and 30 successive one-step-ahead forecasts in French female log mortality data.*

K	FPC	FPC$_w$	FPLSR	FPLSR$_w$	RW
1	99.0611	16.7304	94.0311	53.8186	
2	56.3095	3.3019	54.3410	17.5883	
3	24.9330	3.2580	26.0428	10.2599	
4	15.6845	3.1995	19.7227	4.4818	
5	4.4495	3.2132	5.9299	4.0573	
6	3.4310	3.2123	4.9205	2.9046	4.9800

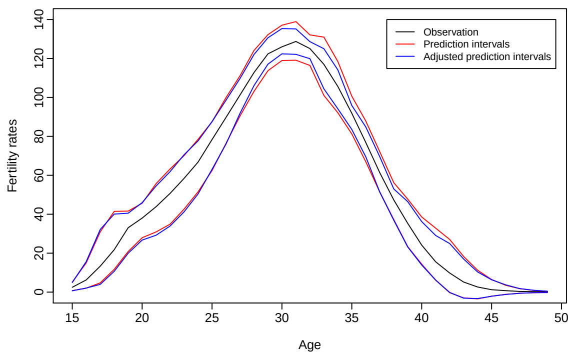
Table 1: *MSE: Australian fertility rates. Forecasts based on FPC, weighted FPC, FPLSR, weighted FPLSR, and a random walk.*

K	FPC	FPC$_w$	FPLSR	FPLSR$_w$	RW
1	0.5956	0.0293	0.5994	0.0607	
2	0.0537	0.0310	0.0738	0.0288	
3	0.0316	0.0310	0.0445	0.0288	
4	0.0296	0.0311	0.0428	0.0288	
5	0.0287	0.0311	0.0472	0.0297	
6	0.0425	0.0311	0.0474	0.0291	0.0437

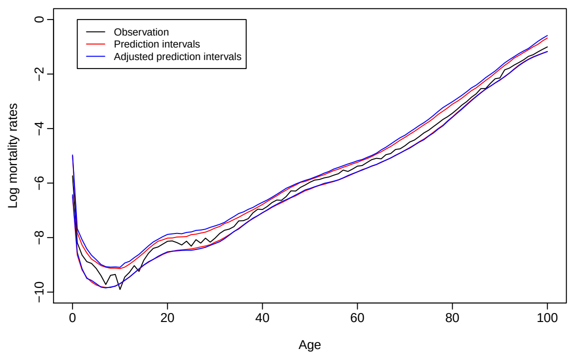
Table 2: *MSE ($\times$ 1000): French female log mortality rates. Forecasts based on FPC, weighted FPC, FPLSR, weighted FPLSR, and a random walk.*

Method	Fertility data	Mortality data
FPC	34.1072	62.2797
FPC_w	33.1424	60.8426
FPLSR	0.4287	2.9184
FPLSR_w	0.4537	3.1602
RW	0.0000	0.0002

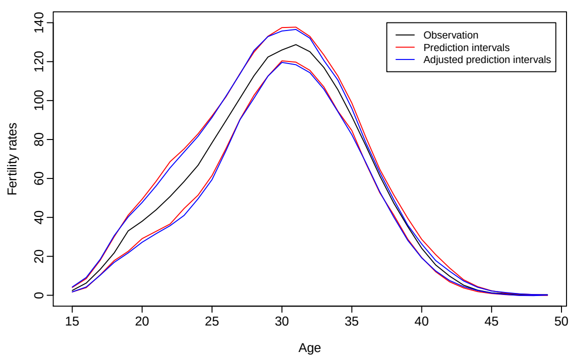
Table 3: *Averaged computation time (in central processing unit seconds) using FPC, weighted FPC, FPLSR, weighted FPLSR, and a random walk based on* 100 *replications.*



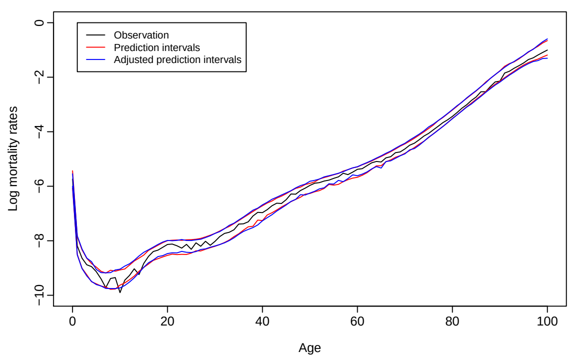
(a) *Bootstrap curve method.*



(b) *Bootstrap curve method.*



(c) *Bootstrap residual method.*



(d) *Bootstrap residual method.*

Figure 12: *Prediction intervals constructed by two proposed methods for Australian fertility rates in 2006 and French female log mortality rates in 2006.*

Method	Fertility data		Mortality data	
	95%	adjusted 95%	95%	adjusted 95%
Bootstrap curves	98.00%	95.86%	97.19%	95.91%
Bootstrap residuals	96.86%	94.89%	97.23%	94.95%

Table 4: *Unadjusted and adjusted empirical coverage probabilities in comparison to the 95% nominal coverage probability.*

Such a deviation of coverage probabilities is a well-known phenomenon in demographic forecasting; see in particular Alho (2005), Alho and Spencer (2005), and Hyndman and Booth (2008). To rectify this problem, we use the adjustment factors given in (12) to correct the lower and upper bounds; the resulting empirical coverage probabilities are closer to the nominal coverage probability, as shown in Table 4.

6. Conclusions

Since its introduction by Ramsay and Dalzell (1991), functional principal components analysis has been widely used in various forms of functional data analysis. In a forecasting context, Hyndman and Ullah (2007) proposed a FPC approach and applied it to forecast demographic data. The virtue of this FPC approach can be summarized as follows: (1) it allows more complex dynamics than other methods, allowing higher order components to be included; (2) it utilizes a nonparametric smoothing technique to reduce the observational error; (3) it eases the interpretability of dynamic changes by separating out the effects of a few orthogonal components; and (4) it solves the inverse and multicollinearity problems caused by the "curse of dimensionality" (Bellman, 1961).

In this paper, we have improved on the FPC approach for forecasting, by introducing a FPLSR method. The advantage of this method is that the latent components are more suitable for prediction rather than variance decomposition. In addition, we do not need to fit univariate time series models, thus resulting in faster computation.

By incorporating a geometric decay weighting scheme into the estimation and forecasting procedures, we also propose the weighted FPC and weighted FPLSR approaches, and show an improvement in forecast accuracy over the unweighted approaches measured by MSE.

Both approaches assign more weight to the most recent data than data from the distant past, thus providing a solution to overcome structural changes in demographic data.

In addition, our weighted methods are easy to program and implement. By discretizing the smoothed curves on a dense grid, we can utilize standard principal component and partial least squares methods with functional predictors and functional responses. The fast computational speed of our methods makes them feasible to be applied in empirical studies with a large number of observations.

Further, we have proposed two nonparametric bootstrap methods to construct prediction intervals. The main advantage of our bootstrap approaches is to avoid the normality assumption, but at the expense of higher computational cost. Our prediction interval adjustment method allows the empirical coverage probability to be closer to the nominal coverage probability.

The implementation of our proposed methods is straightforward using the `ftsa` package (Hyndman and Shang, 2008a) in R.

References

Aguilera, A. M., Escabias, M., Valderrama, M. J., 2006. Using principal components for estimating logistic regression with high-dimensional multicollinear data. Computational Statistics & Data Analysis 50 (8), 1905–1924.

Alho, J. M., 2005. Remarks on the use of probabilities in demography and forecasting. In: Keilman, N. (Ed.), Perspectives on mortality forecasting. Swedish Social Insurance Agency, Oslo, pp. 27–38.

Alho, J. M., Spencer, B. D., 2005. Statistical demography and forecasting. Springer, New York.

Bellman, R., 1961. Adaptive control processes: A guided tour. Princeton University Press.

Cai, T., Hall, P., 2006. Prediction in functional linear regression. Annals of Statistics 34 (5), 2159–2179.

Chatfield, C., 1993. Calculating interval forecasts. Journal of Business & Economic Statistics 11 (2), 121–135.

Chatfield, C., 2000. Time series forecasting. Chapman & Hall/CRC, Boca Raton.

Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. Journal of Multivariate Analysis 12 (1), 136–154.

Delaigle, A., Hall, P., Apanasovich, T. V., 2009. Weighted least squares methods for prediction in the functional data linear model, arXiv:0902.3319v1 [stat.ME].
URL `http://arxiv.org/abs/0902.3319v1`

Erbas, B., Hyndman, R. J., Gertig, D. M., 2007. Forecasting age-specific breast cancer mortality using functional data models. Statistics in Medicine 26 (2), 458–470.

Escabias, M., Aguilera, A. M., Valderrama, M. J., 2004. Principal component estimation of function logistic regression: Discussion of two different approaches. Journal of Nonparametric Statistics 16 (3-4), 365–384.

Faber, N., 2002. Uncertainty estimation for multivariate regression coefficients. Chemometrics and Intelligent Laboratory Systems 64 (2), 169–179.

Fernández Pierna, J. A., Jin, L., Wahl, F., Faber, N. M., Massart, D. L., 2003. Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. Chemometrics and Intelligent Laboratory Systems 65 (2), 281–291.

Greenshtein, E., 2006. Best subset selection, persistence in high-dimensional statistical learning and optimization under l_1 constraint. Annals of Statistics 34 (5), 2367–2386.

Greenshtein, E., Ritov, Y., 2004. Persistence in high-dimensional linear predictor selection and the virtue of overparameterization. Bernoulli 10 (6), 971–988.

Hall, P., Horowitz, J., 2007. Methodology and convergence rates for functional linear regression. Annals of Statistics 35 (1), 70–91.

Hall, P., Hosseini-Nasab, M., 2006. On properties of functional principal components analysis. Journal of the Royal Statistical Society: Series B 68 (1), 109–126.

Hall, P., Hosseini-Nasab, M., 2009. Theory for high-order bounds in functional principal components analysis. Mathematical Proceedings of the Cambridge Philosophical Society 146 (1), 225–256.

Hall, P., Müller, H., Wang, J., 2006. Properties of principal component methods for functional and longitudinal data analyss. Annals of Statistics 34 (3), 1493–1517.

Human Mortality Database, 2008. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Data downloaded on 10 Sep 2008.
URL `www.mortality.org`

Hyndman, R. J., 2007. addb: Australian Demographic Data Bank. R package version 3.222.
URL `http://www.robhyndman.info/Rlibrary/addb`

Hyndman, R. J., Booth, H., 2008. Stochastic population forecasts using functional data models for mortality, fertility and migration. International Journal of Forecasting 24 (3), 323–342.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with exponential smoothing: the state space approach. Springer, Berlin.

Hyndman, R. J., Shang, H. L., 2008a. ftsa: Functional time series analysis. R package version 1.0.
URL `http://monashforecasting.com/index.php?title=R_packages`

Hyndman, R. J., Shang, H. L., 2008b. Rainbow plots, bagplots, and boxplots for functional data. Working paper 9/08, Department of Econometrics & Business Statistics, Monash University.
URL `http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2008/9-08.php`

Hyndman, R. J., Ullah, M. S., 2007. Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis 51 (10), 4942–4956.

Kargin, V., Onatski, A., 2008. Curve forecasting by functional autoregression. Journal of Multivariate Analysis 99 (10), 2508–2526.

Krämer, N., Boulesteix, A., Tutz, G., 2008. Penalized partial least squares with applications to B-spline transformations and functional data. Chemometrics and Intelligent Laboratory Systems 94 (1), 60–69.

Martens, H., Naes, T., 1989. Multivariate calibration. Wiley, Chichester.

Mesle, F., Vallin, J., 1991. Reconstitution of annual life tables for nineteenth-century France. Population: An English Selection 3, 33–62.

Ng, P. T., Maechler, M., 2008. COBs: COBS – Constrained B-splines (Sparse matrix based). R package version 1.1-5.
URL `http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:cobs`

Preda, C., Saporta, G., 2005a. Clusterwise PLS regression on a stochastic process. Computational Statistics & Data Analysis 49 (1), 99–108.

Preda, C., Saporta, G., 2005b. PLS regression on a stochastic process. Computational Statistics & Data Analysis 48 (1), 149–158.

Ramsay, J. O., Dalzell, C. J., 1991. Some tools for functional data analysis (with discussion). Journal of the Royal Statistical Society: Series B 53 (3), 539–572.

Ramsay, J. O., Silverman, B. W., 2002. Applied Functional Data Analysis. Springer, New York.

Ramsay, J. O., Silverman, B. W., 2005. Functional Data Analysis, 2nd Edition. Springer, New York.

Reiss, P. T., Ogden, T. R., 2007. Functional principal component regression and functional partial least squares. Journal of the American Statistical Association 102 (479), 984–996.

Rice, J. A., Silverman, B. W., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society: Series B 53 (1), 233–243.

Silverman, B. W., 1995. Incorporating parametric effects into functional principal components analysis. Journal of the Royal Statistical Society: Series B 57 (4), 673–689.

Silverman, B. W., 1996. Smoothed functional principal components analysis by choice of norm. Annals of Statistics 24 (1), 1–24.

Wedderburn, R. W. M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61 (3), 439–447.

Wilmoth, J., Andreev, K. F., Jdanov, D. A., Glei, D. A., 2007. Methods protocol for the Human Mortality Database. 5th Edition.
URL http://www.mortality.org/Public/Docs/MethodsProtocol.pdf

Wold, H., 1975. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. In: Gani, J. (Ed.), Perspectives in Probability and Statistics. Applied Probability Trust, Sheffield, pp. 117–145.

Wood, S. N., 1994. Monotonic smoothing splines fitted by cross validation. SIAM Journal on Scientific Computing 15 (5), 1126–1133.

Yao, F., Müller, H. G., Wang, J. L., 2005. Functional linear regression analysis for longitudinal data. Annals of Statistics 33 (6), 2873–2903.