MONASH University

Department of Econometrics and Business Statistics

A survey of functional principal component analysis

Han Lin Shang

May 2011

A survey of functional principal component analysis

Han Lin Shang

Department of Econometrics & Business Statistics, Monash University,

Caulfield East, VIC 3145, Melbourne, Australia

Email: HanLin.Shang@monash.edu.au

Summary

Advances in data collection and storage have tremendously increased the presence of functional data, whose graphical representations are curves, images or shapes. As a new area of Statistics, functional data analysis extends existing methodologies and theories from the fields of functional analysis, generalized linear models, multivariate data analysis, nonparametric statistics and many others. This paper provides a review into functional data analysis with main emphasis on functional principal component analysis, functional principal component regression, and bootstrap in functional principal component regression. Recent trends as well as open problems in the area are discussed.

Key words: Bootstrap, functional principal component regression, functional time series, Stiefel manifold, Von Mise-Fisher distribution.

1 Introduction

In probability theory, random functions have being studied for quite a long time. Due to recent advances in computing and the opportunity to collect and store high-dimensional data, statisticians can now study models for "infinite-dimensional functional data". Analyzing functional data has had a significant impact on statistical methods and thinking, changing forever the way in which we display, model and

forecast data.

The attention paid to functional data analysis has contributed to a rapidly increasing body of published research. A 2002 joint summer research conference on 'Emerging Issues in Longitudinal Analysis' provided a platform for emerging ideas from longitudinal data analysis and functional data analysis. In 2004, *Statistica Sinica* published a special issue (vol 14, issue 3) based on that conference, which dealt exclusively with the close connection between longitudinal data and functional data, and also contained two review articles by Rice (2004) and Davidian et al. (2004). In 2007, *Computational Statistics & Data Analysis* published a special issue (vol 51, issue 10) on functional data analysis, along with a review article by González-Manteiga & Vieu (2007). *Computational Statistics* also published a special issue (vol 22, issue 3) on modeling functional data, along with a review article by Valderrama (2007). In 2008, a workshop on "Functional and Operatorial Statistics" at Université Paul Sabatier provided a platform for emerging ideas from functional data analysis and operatorial statistics. Based on that conference, *Journal of Multivariate Analysis* published a special issue (vol 101, issue 2), which drew a close connection between functional data analysis and nonparametric function estimation.

The theoretical and practical developments in functional data analysis are mainly from the last four decades, due to the rapid development of computer recording and storing facilities. Some common techniques in functional data analysis are reviewed in this paper.

This paper contains six sections, and reviews the research on functional data analysis undertaken in both the statistics and probabilistic fields. Section 2 provides the background to functional data analysis and the two schools of thought in defining functional data analysis. Section 3 reviews the functional principal component analysis (FPCA), which plays a significant role in the development of functional data analysis. It is also an essential ingredient of functional principal component

regression (FPCR) outlined in Section 4. Section 5 reviews bootstrap techniques in FPCR. A conclusion is presented in Section 6.

2 Functional data analysis

Within the field of functional data analysis, there exist two schools of thought based on how they conceptualize functional data (Ramsay 1982). On the one hand, some authors believe that functional data analysis can be considered as a smoothed version of multivariate data analysis, and functional data analysis expresses the multivariate data analytical tools in the language of functional analysis. On the other hand, the second line of development has been the statistical application of spline functions, especially in the scope of nonparametric function estimation (e.g., Silverman 1985, Wahba 1990, Green & Silverman 1994, Eubank 1999).

Although there is a difference between the two viewpoints, the individual datum in functional data analysis is a whole function defined on a bounded common interval, rather than concentrating on the observed values at particular points in the interval. This importance for understanding functional data analysis has long been emphasized by Dieudonné (1960, p.1), who stated that

> the idea that a function f is a single object, which may itself "vary" and is in general to be thought of as a "point" in a large "functional space"; indeed, it may be said that one of the main differences between the classical and modern concepts of Analysis is that, in classical mathematics, when one writes $f(x)$, f is visualized as "fixed" and x as "variables", whereas nowadays both f and x are considered as "variables" (and sometimes it is x which is fixed, and f which becomes the "varying" object).

Although functional data analysis was in use in the 1960s, it was not popularized until

the work of Ramsay & Dalzell (1991). They not only named functional data analysis, but also highlighted the great advantages of applying functional data analysis in practice. These advantages include the fact that:

1. smoothing and interpolation procedures can yield a functional representation of a finite set of observations;

2. modeling problems are more natural to consider functionally;

3. the objective of an analysis can be functional in nature; and

4. functional pre-processing, such as derivatives, can provide insights into functional data display and functional linear regression models.

These advantages strongly reflect the aims of the functional data paradigm outlined by Ramsay & Silverman (2005, p.9), namely:

- to represent the data in ways that aid further analysis;

- to display the data so as to highlight various characteristics;

- to study important sources of pattern and variation among the data;

- to explain variation in a response variable by using predictor information; and

- to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates.

Due to its practical advantages, functional data analysis has received considerable attention in diverse areas of application, such as: the study of acidification processes (Abraham et al. 2003), the analysis of growth curve (Rao 1958), the analysis of handwritten statistics in Chinese (Ramsay 2000), the analysis of price dynamics in online auctions (Wang, Jank, Shmueli & Smith 2008), agricultural sciences (Ogden et al. 2002), behavioral sciences (Rossi et al. 2002), chemometrics (Burba

et al. 2009), climatic variation forecasting (Besse et al. 2000), climatology (Meiring 2007), criminology (Berk 2008), data mining (Hand 2007), demographic forecasting (Hyndman & Ullah 2007, Hyndman & Booth 2008, Hyndman & Shang 2009), electronic commerce research (Wang, Jank & Shmueli 2008), marketing science (Wang, Jank, Shmueli & Smith 2008), medical research (Erbas et al. 2007), ozone population forecasting (Damon & Guillas 2002), and many more. In another book named *Applied Functional Data Analysis*, Ramsay & Silverman (2002) gave a number of exciting applications with a continuous functional variable. Often, the continuous functional variable is time, even though functional data may be observed over age, space, wavelength, molecular weight or other continuums.

Ferraty & Vieu (2006) incorporated additional nonparametric features into functional data analysis, and provided an excellent resource for revisiting theoretical developments of functional kernel regression — a generalization of Nadaraya-Watson kernel regression. Recent developments in nonparametric functional data analysis have been reviewed in Ferraty et al. (2007).

Functional data analysis research is now being conducted at a time of rapid change in computer technology. In particular, the Internet facilitates the rapid and convenient dissemination of code. Computational code for functional data analysis developed for MATLAB©, S-PLUS©, and R (R Development Core Team 2009) is available from the Functional Data Analysis web site (Ramsay 2011). A recent book by Ramsay, Hooker & Graves (2009) drew strong connections between functional data analysis techniques and their practical applications using R and MATLAB. Furthermore, the *fda* package in R (Ramsay, Wickham, Graves & Hooker 2009) provides a wide range of smoothing and modeling tools, along with a number of classical functional data sets. Computational code for nonparametric functional data analysis has been developed by a research group known as STAPH (Groupe de Travail en Statistique Fonctionnelle et Opeatorielle 2011) based at Université Paul Sabatier.

The rapidity with which functional data analysis is growing as a field means that a concise and informative review of the last forty years is very challenging. Although there are a vast number of studies, it is possible to identify three major streams, namely FPCA, FPCR and bootstrap in FPCR. The latter two are applications of FPCA.

3 Functional principal component analysis

Before reviewing FPCA, it is necessary to revisit multivariate PCA that is used to reduce dimensionality for multivariate data.

3.1 Multivariate principal component analysis (PCA)

Proposed by Pearson (1901), PCA becomes an essential tool for multivariate data analysis and unsupervised dimension reduction. The goal of PCA is to find the sequence of orthogonal components that most efficiently explains the variance of the observations. Depending on the field of application, principal components are also known as the discrete Karhunen-Loève transformation (especially in signal analysis), empirical orthogonal basis functions (especially in meteorology and atmospheric research), latent semantic indexes (especially in information retrieval), the Hotelling transformation, or proper orthogonal decomposition (Izenman 2008). Originally, Pearson intended PCA as the correct solution to some of the problems that were of interest to biometrician at that time, although his study did not consider a practical method for calculating two or more components (Møller et al. 2005). A detailed description of how to compute principal components came much later from Hotelling (1933). However, the calculations were extremely difficult for more than a few variables, since they had to be done by hand. It was not until computers

became generally accessible that PCA achieved its current widespread popularity. Nowadays, calculating hundreds of components takes only few seconds, which has thus popularized this method to a wide spectrum of users.

PCA played an essential part in the development of multivariate data analysis. Notably, PCA is covered in almost all textbooks on multivariate data analysis, and in particular by Jackson (1991), Jolliffe (2002), Izenman (2008) and Hastie et al. (2009). It is also applied widely in the field of social sciences (e.g., Berk 2008).

The advantage of PCA is that it finds a lower-dimensional representation, while preserving the maximum amount of information from the original variables. For a centered data matrix $\boldsymbol{X}_0$ (where the columns represent p variables, and the rows represent n observations), PCA yields an orthogonal decomposition of $\boldsymbol{X}_0$ that is optimal for a given number of principal components. The principal component decomposition provides the minimum mean squared error approximation to $\boldsymbol{X}_0$. Moreover, the explained variation of the excluded principal components converges to zero as K increases, where K denotes the retained number of principal components. The principal component decomposition is given by

$$\underset{1 \times n}{\boldsymbol{\beta}_1} = \underset{1 \times p}{\boldsymbol{\phi}_1} \underset{p \times n}{\boldsymbol{X}_0^{'}},$$

where $\boldsymbol{\beta}_1$ represents a set of the first principal component scores with mean zero; $\boldsymbol{\phi}_1$ is the first principal component; and $'$ symbolizes a vector or matrix transposition. The first principal component $\boldsymbol{\phi}_1$ can be calculated by maximizing the variance of $\boldsymbol{\phi}_1 \boldsymbol{X}_0^{'}$, that is,

$$\begin{aligned} \boldsymbol{\phi}_1 &= \underset{||\boldsymbol{\phi}_1||=1}{\arg\max} \operatorname{Var}(\boldsymbol{\phi}_1 \boldsymbol{X}_0^{'}) \\ &= \underset{||\boldsymbol{\phi}_1||=1}{\arg\max} \boldsymbol{\phi}_1 \boldsymbol{X}_0^{'} \boldsymbol{X}_0 \boldsymbol{\phi}_1^{'}. \end{aligned}$$

Successive principal components can be obtained iteratively by subtracting the first k principal components from $\boldsymbol{X}_0$, for $1 \leq k \leq K < \min(n, p)$. That is,

$$\boldsymbol{X}_k = \boldsymbol{X}_{k-1} - \boldsymbol{X}_{k-1}\boldsymbol{\phi}_k'\boldsymbol{\phi}_k,$$

and then treating $\boldsymbol{X}_k$ as the new data matrix to find $\boldsymbol{\phi}_{k+1}$ by maximizing the variance of $\boldsymbol{\phi}_{k+1}\boldsymbol{X}_k'$, that is,

$$\boldsymbol{\phi}_{k+1} = \underset{\boldsymbol{\phi}_{k+1}}{\arg\max}\,\mathrm{Var}(\boldsymbol{\phi}_{k+1}\boldsymbol{X}_k')$$

$$= \underset{\boldsymbol{\phi}_{k+1}}{\arg\max}\,\boldsymbol{\phi}_{k+1}\boldsymbol{X}_k'\boldsymbol{X}_k\boldsymbol{\phi}_{k+1}',$$

subject to $\|\boldsymbol{\phi}_{k+1}\| = (\sum_{j=1}^{p} \phi_{k+1,j}^2)^{\frac{1}{2}} = 1$ and $\boldsymbol{\phi}_{k+1}\perp\boldsymbol{\phi}_j$, for $j = 1, \ldots, k$.

Alternatively, a simple and effective algorithm known as singular value decomposition (SVD) can be applied. For a centered data matrix $\boldsymbol{X}_0$, the SVD of $\boldsymbol{X}_0$ can be expressed as

$$\underset{n\times p}{\boldsymbol{X}_0} = \underset{n\times K}{\boldsymbol{U}}\,\underset{K\times K}{\boldsymbol{D}}\,\underset{K\times p}{\boldsymbol{V}'}, \tag{1}$$

where $K \leq \min(n, p)$; $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$; and $\boldsymbol{D}$ is a diagonal matrix with $d_1 > d_2 \cdots > d_K$ on the diagonal. The matrix $\boldsymbol{U}\boldsymbol{D}$ again contains principal component scores, which are the coordinates of variables in the space of principal components.

Jolliffe (2002) discussed four advantages of applying the SVD technique in high-dimensional data analysis, which are listed below.

1. It is a computationally efficient method for finding orthogonal principal components, thus achieving a minimal squared loss of information.

2. It provides additional insights into what PCA does.

3. It provides useful graphical and algebraic means of representing the results of

PCA.

4. It computes uncorrelated principal component scores.

Despite the popularity of PCA, its application has often been restricted when the sample size of multivariate data is larger than the number of variables; and the multivariate data are equally spaced on a dense grid. However, these conditions may no longer hold in many fields, such as astronomy, biostatistics, chemometrics, genomics, spectroscopy, and many others.

3.2 Functional principal component analysis (FPCA)

Many authors, such as Croux & Ruiz-Gazen (2005) and Ferraty & Vieu (2006), have realized that the computation of PCA runs into serious difficulties in analyzing functional data because of the "curse of dimensionality" (Bellman 1961). The "curse of dimensionality" originates from data sparsity in high-dimensional space. Even if the geometric properties of PCA remain valid, and even if numerical techniques deliver stable results, the sample covariance matrix is sometimes a poor estimate of the population covariance matrix. To overcome this difficulty, FPCA provides a much more informative way of examining the sample covariance structure than PCA, and it can also complement a direct examination of the variance-covariance structure.

PCA was one of the first multivariate data analysis methods to be adapted to functional data (Dauxois et al. 1982). The main idea of this extension is simply to replace vectors by functions, matrices by compact linear operators, covariance matrices by covariance operators, and scalar products in vector space by scalar products in square-integrable functional space. The differences in notation between PCA and FPCA are summarized in Table 1.

	PCA	FPCA
Variables	$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$, $\boldsymbol{x}_i = [x_{1i}, \ldots, x_{ni}]'$, $i = 1, \ldots, p$	$\boldsymbol{f}(x) = [f_1(x), \ldots, f_n(x)]$, $x \in [x_1, x_p]$
Data	Vectors $\in R^p$	Curves $\in L_2[x_1, x_p]$
Covariance	Matrix $\boldsymbol{V} = \mathrm{Cov}(\boldsymbol{X}) \in R^p$	Operator T bounded between x_1 and x_p, $T : L_2[x_1, x_p] \rightarrow L_2[x_1, x_p]$
Eigen structure	Vector $\xi_k \in R$, $\boldsymbol{V}\xi_k = \lambda_k\xi_k$, for $1 \leq k < \min(n, p)$	Function $\xi_k(x) \in L_2[x_1, x_p]$, $\int_{x_1}^{x_p} T\xi_k(x)dx = \lambda_k\xi_k(x)$, for $1 \leq k < n$
Components	Random variables in R^p	Random variables in $L_2[x_1, x_p]$

Table 1: The differences in notation between PCA and FPCA.

FPCA finds the set of orthogonal principal component functions that maximize the variance along each component. It finds the first functional principal component $\phi_1(x)$, which the variance of principal component scores

$$\boldsymbol{\beta}_1 = \int_{x_1}^{x_p} \phi_1(x)\boldsymbol{f}(x)dx, \tag{2}$$

is maximized subject to $\|\phi_1^2(x)\| = \int_{x_1}^{x_p} \phi_1^2(x)dx = 1$. $\boldsymbol{\beta}_1$ represents a set of the first principal component scores with mean zero, $\boldsymbol{f}(x)$ is a set of decentralized functional curves, and $[x_1, x_p]$ is the function support range.

Successive principal component functions can be obtained iteratively by subtracting the first k principal component functions from $\boldsymbol{f}^0(x) = \boldsymbol{f}(x)$ for $1 \leq k \leq K < n$, that is,

$$\boldsymbol{f}^k(x) = \boldsymbol{f}^{k-1}(x) - \boldsymbol{\beta}_k\phi_k(x),$$

and then treating $\boldsymbol{f}^k(x)$ as the new collection of functions to find $\phi_{k+1}(x)$, which the variance of principal component scores

$$\boldsymbol{\beta}_{k+1} = \int_{x_1}^{x_p} \phi_{k+1}(x)\boldsymbol{f}^k(x)dx,$$

is maximized subject to

$$\int_{x_1}^{x_p} \phi_{k+1}^2(x)dx = \|\phi_{k+1}^2(x)\| = 1, \quad \text{and} \tag{3}$$

$$\int_{x_1}^{x_p} \phi_{k+1}(x)\phi_j(x)dx = 0, \quad \text{for} \quad j = 1, \ldots, k. \tag{4}$$

The computational difficulty of the integration in (2)-(4) for calculating FPCA can be overcome by any of the following three approaches.

- Discretization: FPCA is carried out in a similar fashion to PCA, except that it is necessary to renormalize the eigenvectors and interpolate them with a suitable smoother (Rao 1958, Ramsay & Silverman 2005). This discretization approach was the earliest method to utilize FPCA.

- Basis function expansion: The second approach involves expressing each function as a linear combination of basis functions $f_t(x) \approx \sum_{k=1}^{K} \beta_{t,k}\phi_k(x)$, and approximating each function by a finite number of basis functions (Rice & Silverman 1991). Some popular basis functions are polynomial basis functions (which are constructed from the monomials $\phi_k(x) = x^{k-1}$), Bernstein polynomial basis functions (which are constructed from $1, 1-x, x, (1-x)^2, 2x(1-x), x^2, \ldots$), Fourier basis functions (which are constructed from $1, \sin(wt), \cos(wt), \sin(2wt), \cos(2wt), \ldots$), radial basis functions, wavelet basis functions, and orthogonal basis functions.

- Numerical approximation: As was observed by Jolliffe (2002, p.411) and Ramsay & Silverman (2005, pp.164-166), the third approach is to use quadrature rules to approximate FPCA. Castro et al. (1986) gave some interesting examples to demonstrate this numerical approach, which produced fairly stable and consistent estimates, in contrast to PCA.

The advance of FPCA dates back to the early forties when Karhunen (1946) and

Loève (1946) independently developed a theory on the optimal series expansion of a continuous stochastic process. Later, Grenander (1950) provided the first application of the Karhunen-Loève expansion to functional data, including the first outlook for functional regression. Motivated by a data set of growth curve, Rao (1958) developed some preliminary ideas on FPCA and proposed statistical tests for the equality of average growth curves over a period of time. Much later, Deville (1974) carried out a PCA of functional data, and Dauxois & Pousse (1976) introduced a functional exposition of PCA with applications to statistical inference, including the statistical structure, point estimation, confidence sets and hypothesis testing. Several other notable developments have arisen out of the systematic research of the functional data analysis group, namely the Toulouse School of Functional Data Analysis (Dauxois et al. 1982). In recent years, Hall et al. (2006) and Hall & Hosseini-Nasab (2009) have investigated the properties of FPCA, and have given some insights into methodology and convergence rates for FPCA.

Several extensions and modifications of FPCA have been put forward. Rice & Silverman (1991), Silverman (1996) and Reiss & Ogden (2007) incorporated the smoothness in the estimation of FPCA through different roughness penalty approaches. Further, Reiss & Ogden (2007) compared the advantages and disadvantages of these three smoothed FPCA approaches from both the theoretical and practical aspects. Jones & Rice (1992) presented an interesting proposal to describe samples of random curves through principal component scores. Locantore et al. (1999) proposed a robust FPCA which deals with the presence of outliers. James et al. (2000) developed an extension of FPCA that allows the estimation of harmonics from fragments of curves. Yao et al. (2005a) proposed a FPCA procedure via a conditional expectation method, which is aimed at estimating functional principal component scores for sparse longitudinal data. Benko et al. (2009) proposed common functional principal component estimation from discrete noisy data, and presented a bootstrap test for

examining the equality of the eigenvalues, eigenfunctions, and mean functions of two functional samples. Di et al. (2009) introduced multilevel FPCA, which is designed to extract the intra- and inter-subject geometric components of multilevel functional data. Based on FPCA, Hyndman & Shang (2010) proposed graphical tools for visualizing functional data and detecting functional outliers.

Due to the theoretical and practical developments, FPCA has been successfully applied to many practical problems, such as the analysis of cornea curvature in the human eye (Locantore et al. 1999), the analysis of electronic commerce (Wang, Jank & Shmueli 2008), the analysis of growth curve (Chiou & Li 2007), the analysis of income density curves (Kneip & Utikal 2001), the analysis of implied volatility surface in finance (Cont & de Fonseca 2002), the analysis of longitudinal primary biliary liver cirrhosis (Yao et al. 2005b), the analysis of spectroscopy data (Yao & Müller 2010), signal discrimination (Hall et al. 2001), and time-course gene expression (Yao et al. 2005a). Furthermore, Hyndman & Ullah (2007) proposed a smoothed and robust FPCA, and used it to forecast age-specific mortality and fertility rates. This approach has been applied by Erbas et al. (2007) to forecast breast cancer mortality rates in Australia.

Overall, FPCA has played an important role in the development of functional data analysis. It is also an essential ingredient of FPCR, which is one of the most popular techniques in functional linear models (e.g., Cardot et al. 2003, Yao et al. 2005b).

4 Functional principal component regression

Before reviewing FPCR, it is necessary to revisit multivariate PCR that is used to solve multicollinearity in multivariate linear regression.

4.1 Multicollinearity in multivariate linear regression

Consider a multivariate linear regression model,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is a $(n \times 1)$ vector of "centered" responses, $\boldsymbol{X}$ is a $(n \times p)$ matrix of predictors, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown regression coefficients, and $\boldsymbol{e}$ is a $(n \times 1)$ vector of random errors with $\mathrm{E}(\boldsymbol{e}) = 0$ and $\mathrm{Var}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is an $(n \times n)$ identity matrix.

By using ordinary least squares (OLS) regression, a closed form of the unbiased regression coefficient estimates can be obtained as

$$\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

Using the SVD of $\boldsymbol{X}$ expressed in (1), the OLS regression coefficients can also be written as

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} &= \left[(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}')'\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'\right]^{-1}(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}')'\boldsymbol{y}, \\
&= \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}'\boldsymbol{y}, \\
&= \sum_{k=1}^{p} \frac{v_k u_k'}{d_k}\boldsymbol{y},
\end{aligned}
\tag{5}
$$

where p is the number of variables.

However, in the presence of multicollinearity (when the smallest eigenvalue of the predictors is close to zero), sample covariance matrix can be near singular, and thus statistical inference drawn from the singular covariance matrix could be erroneous. For instance, the OLS estimates of the regression coefficients are likely to be too large in absolute values, and possibly of the wrong sign (Wichern & Churchill 1978).

One possibility for evaluating the quantity of an estimator is to determine its mean square error (MSE). The MSE of an estimator $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ for a parameter $\boldsymbol{\beta}$ is defined as

$$
\begin{aligned}
\text{MSE}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) &= \left\| (\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})' \right\| \\
&= \text{E}\left[(\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \right] \\
&= \text{E}\left\{ \left[\hat{\boldsymbol{\beta}}_{\text{OLS}} - \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) + \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) - \boldsymbol{\beta} \right]' \left[\hat{\boldsymbol{\beta}}_{\text{OLS}} - \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) + \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) - \boldsymbol{\beta} \right] \right\} \\
&= \underbrace{\text{E}\left\{ \left[\hat{\boldsymbol{\beta}}_{\text{OLS}} - \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) \right]' \left[\hat{\boldsymbol{\beta}}_{\text{OLS}} - \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) \right] \right\}}_{\text{variance}} + \underbrace{\left[\text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) - \boldsymbol{\beta} \right]' \left[\text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) - \boldsymbol{\beta} \right]}_{\text{bias}^2} \\
&= \underbrace{\text{E}\left\{ \left[\hat{\boldsymbol{\beta}}_{\text{OLS}} - \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) \right]' \left[\hat{\boldsymbol{\beta}}_{\text{OLS}} - \text{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) \right] \right\}}_{\text{variance}}.
\end{aligned}
$$

This is the well-known bias-variance tradeoff decomposition of the MSE. The first part is the variance of the estimator, while the second part is the squared bias of the estimator. $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is an unbiased estimator and has minimum variance among all linear unbiased estimators. However, it inflates the variance of the estimator in the case of multicollinearity. Consequently, the overall MSE tends to be large. This motivates the development of the shrinkage estimates of the regression coefficients, such as $\hat{\boldsymbol{\beta}}_{\text{PCR}}$.

4.2 Multivariate PCR

PCR starts by using the principal components of the predictor variables in place of the predictors (Jolliffe 2002). Since the retained principal components are uncorrelated, it solves the multicollinearity problem by deleting those principal components that have low variances. As a result, a much more stable estimate of $\boldsymbol{\beta}$ can be obtained and regression calculations are also simplified.

The use of PCA in regression dates back to the work of Kendall (1957) and Hotelling (1957). While Massy (1965) and Jeffers (1967) presented two well-known examples

(the studies of pitprops and alate adelges) of the use of PCR, Mosteller & Tukey (1977, pp.397-406), Mardia et al. (1979), Gunst & Mason (1980) and Naes & Martens (1988) emphasized the key points in choosing relevant principal components. The idea of PCR as envisaged by these authors is to replace the original regressors by their first few principle components, thus orthogonalizing the regressors and making computations easier and more stable (Jolliffe 1982).

PCR can be defined mathematically as

$$\boldsymbol{y} = \boldsymbol{\phi}\boldsymbol{\beta}_{\mathrm{PCR}} + \boldsymbol{e},$$

where $\boldsymbol{\phi}$ is an $(n \times K)$ matrix of principal components, $\boldsymbol{\beta}_{\mathrm{PCR}}$ represents the first K number of principal component scores, and $\boldsymbol{e}$ is a $(n \times 1)$ vector of random errors. The principal component scores are calculated via the OLS method, and they are given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = (\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\boldsymbol{y},$$
$$= (\boldsymbol{L}^2)^{-1}\boldsymbol{\phi}'\boldsymbol{y},$$

where $\boldsymbol{L}^2$ represents the diagonal matrix whose k^{th} diagonal element is λ_k (the kth largest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$).

The shrinkage estimates of the regression coefficients using the PCR can be expressed, similarly to OLS, as linear combinations of the eigenvector of $\boldsymbol{X}$ (Van Huffel & Vandewalle 1991). The PCR coefficient estimates truncate the expansion (5) after a certain term. Thus,

$$\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = \sum_{k=1}^{K} \frac{v_k u_k'}{d_k}\boldsymbol{y}, \quad K < \min(n, p).$$

Jeffers (1967) and Krzanowski & Marriott (1994) point out some practical advantages

of PCR. These include

- orthogonalization of regression variables;

- examination of the grouping of individuals in high-dimensional space;

- determination of the objective weighting of measured variables in the construction of meaningful indexes;

- elimination of variables which contribute relatively little information, thus easing interpretation;

- construction of principal components depending on a complete data set rather than on a single variable; and

- reduction of variability in the measured set to the smallest number of meaningful dimensions.

Despite the popularity of PCR, its application has been limited when the number of variables is less than the sample size of multivariate data; and the multivariate data are equally spaced on a dense grid. However, these conditions may no longer hold in many fields, such as astronomy, chemometrics, biostatistics, genomics, electronic commerce and many others.

4.3 Functional principal component regression (FPCR)

FPCR describes the relationship between the functional predictors and responses, where the response variable can be scalar or function. The first functional formulation of a PCR dates back to Ramsay & Dalzell (1991). Since then, it has gained an increasing popularity in high-dimensional prediction problems.

In this paper, I shall demonstrate the applicability of FPCR for modeling and forecasting a time series of functions (also known as functional time series (Hyndman

& Shang 2009)). Functional time series is a type of functional data, where the same functional object is observed over a period of time. In modeling and forecasting functional time series, numerous examples of using FPCR can be found in different fields of application, such as breast cancer mortality rate modeling and forecasting (Erbas et al. 2007), climate variation forecasting (Shang & Hyndman 2011), call volume forecasting (Shen & Huang 2008, Shen 2009), demographic modeling and forecasting (Hyndman & Ullah 2007, Hyndman & Booth 2008, Hyndman & Shang 2009), and electricity demand forecasting (Antoch et al. 2008).

To motivate the discussion, Figure 1 shows annual age-specific Australian fertility curves from 1921 to 2006 for ages 15-49. The data were taken from the Australian Demographic Data Bank (Hyndman 2007), and were smoothed using a weighted median smoothing B-splines with concave constraint (see Hyndman & Ullah 2007, for details).
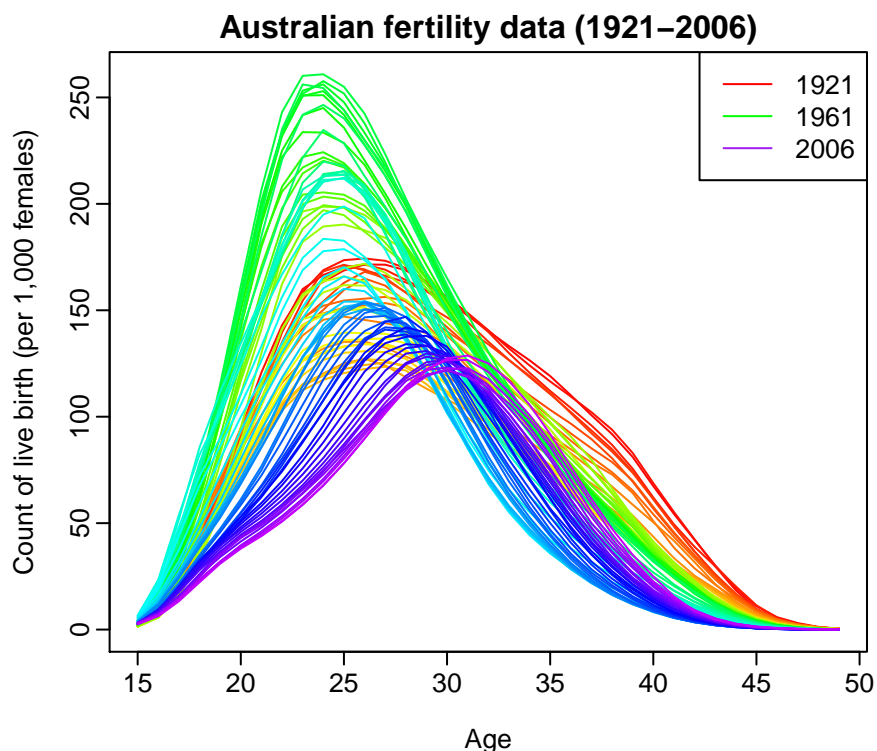

Figure 1: Smoothed age-specific Australian fertility rates for ages 15-49, observed from 1921 to 2006. The oldest years are shown in red, while the most recent years in violet. Curves are order chronologically according to the colors of the rainbow.

In this example, the age-specific fertility rates in each year is considered as an observation of a stochastic process in $\mathcal{L}^2_{[15,49]}$. $\mathcal{L}^2_{[15,49]}$ is the Hilbert space of square integrable functions on the interval $[15, 49]$, with the inner product $< f, g >= \int_{15}^{49} f(x)g(x)dx$ for any two functions $f, g \in \mathcal{L}^2_{[15,49]}$ and induced squared norm $\| \cdot \|^2 =< \cdot, \cdot >$. Within the framework of functional time series, FPCR can be expressed as follows

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \epsilon_t(x), \quad t = 1, 2, \ldots, n, \tag{6}$$

where

1. $\mu(x) = \mathrm{E}[\boldsymbol{f}(x)]$ is the mean function, and $\boldsymbol{f}(x) = [f_1(x), \ldots, f_n(x)]'$ is a vector of n realizations of a stochastic process.

2. $\phi_k(x)$ is the kth orthonormal eigenfunction of the variance kernel $\Gamma(x) = \mathrm{Var}[\boldsymbol{f}(x)]$. The eigenvalues corresponding to $\phi_k(x)$ are listed in a decreasing order, such that $\lambda_1 > \lambda_2, \cdots > \lambda_K > \lambda_{K+1}, \ldots$, where

$$\int_{15}^{49} \Gamma(x)\phi_k(x)dx = \lambda_k\phi_k(x), \quad \Gamma(x) = \sum_{k=1}^{K} \lambda_k\phi_k(x)\phi_k(x).$$

3. The coefficient $\boldsymbol{\beta}_k$ is the kth functional principal component scores, they are given by the projection of $\boldsymbol{f}(x) - \mu(x)$ in the direction of kth eigenfunction $\phi_k(x)$, that is $\boldsymbol{\beta}_k =< \boldsymbol{f}(x) - \mu(x), \phi_k(x) >$.

4. $\epsilon_t(x)$ is the error function for the tth observation, and it contains the excluded functional principal component expansion.

5. K is the number of retained functional principal components.

Using FPCA, Figure 2 presents the first four functional principal components and their associated principal component scores. The bottom panel of Figure 2 also

shows the forecasted principal component scores, and their 80% and 95% prediction intervals using exponential smoothing state space models (Hyndman et al. 2008).
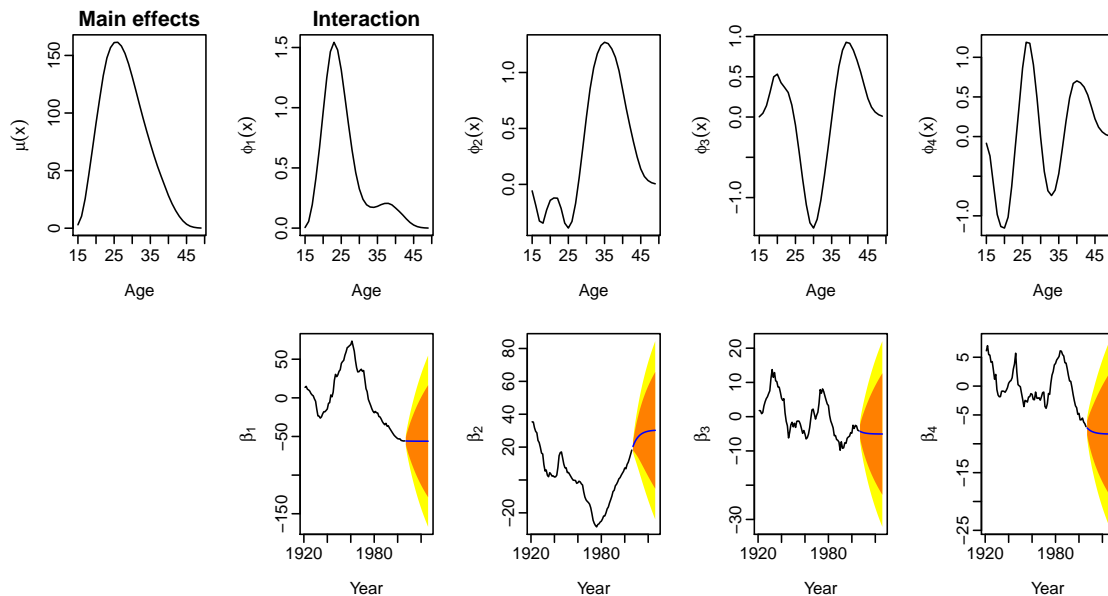

Figure 2: The first four functional principal components and their associated principal component scores for Australian fertility data. Forecasts of the principal component scores from 2007 to 2026 are shown with 80% and 95% prediction intervals, using the exponential smoothing state space models.

By conditioning on the observed data $\mathcal{I} = \{f_1(x), \ldots, f_n(x)\}$ and the fixed functional principal components $\mathcal{B} = \{\phi_1(x), \ldots, \phi_K(x)\}$, the h-step-ahead forecast of $y_{n+h}(x)$ can be obtained by

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

where $\hat{\mu}(x) = \frac{1}{n}\sum_{t=1}^{n} f_t(x)$ is the estimated mean function, and $\hat{\beta}_{n+h|n,k}$ denotes the h-step-ahead forecast of $\beta_{n+h,k}$ using a univariate time series, such as an exponential smoothing state space model (Hyndman et al. 2008).

Figure 3 shows the forecasts of Australian fertility rates from 2007 to 2026 highlighted in rainbow color, while the data used for estimation are grayed out. The forecasts exhibit a continuing shift to older ages for peak fertility rates.
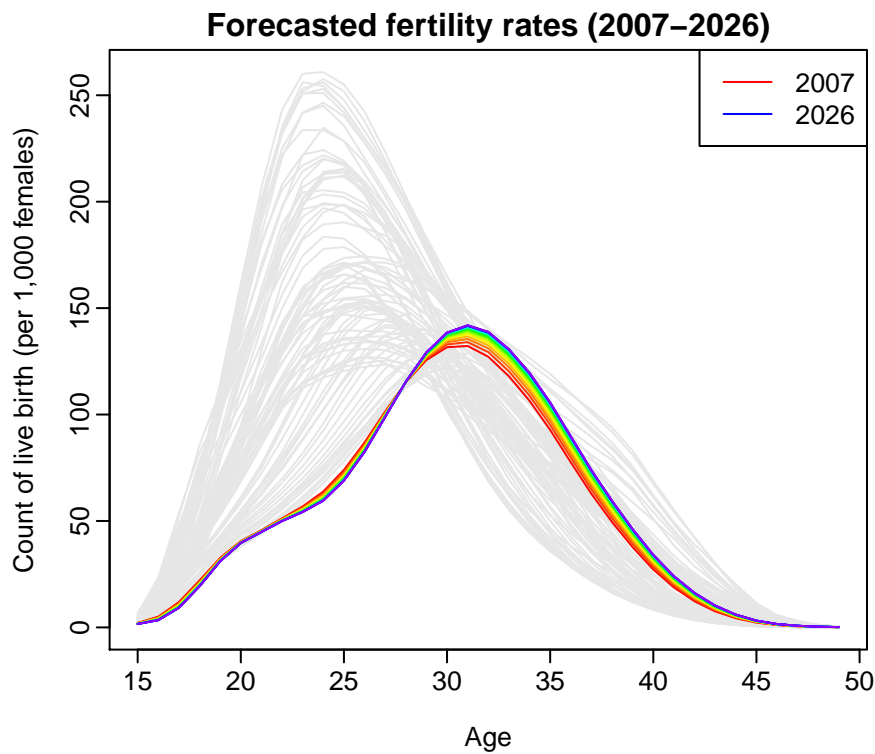
Figure 3: Forecasts of Australian fertility rates from 2007 to 2026, based on the first four functional principal components for illustration.

Illustrated by the Australian fertility data, I demonstrated the use of FPCR in modeling and forecasting functional time series. In a similar vein, Hyndman & Booth (2008) used this technique to produce point and interval forecasts of mortality rates and immigration rates, while Erbas et al. (2007) applied this technique to model and forecast breast cancer mortality rates. Shen & Huang (2008) and Shen (2009) applied FPCR to model and forecast the volume of calls in a call center. Antoch et al. (2008) implemented FPCR to model and forecast electricity consumption.

Overall, FPCR has played an important role in the development of functional linear model. It produces a parsimonious model with orthogonal regressors and uncorrelated regression coefficients. However, there is an open question on how to select the optimal number of functional principal components. In this direction, Yao et al. (2005*b*) proposed using a functional version of Akiake's information criterion to select the optimal number of components, justified via an appeal to a pseudo-Gaussian

likelihood argument and results of Shibata (1981). Peres-Neto et al. (2005) provided a comprehensive comparison of 20 stopping rules and proposed a two-step procedure that appears to be highly effective. Hall & Vial (2006) proposed a bootstrap method to determine the optimal number of components, and compared favorably with the two-step procedure of Peres-Neto et al. (2005). Poskitt & Sengarapillai (2009) considered optimal component selection criteria using optimal encoding, description length principles. As elucidated by Hall & Vial (2006) and Poskitt & Sengarapillai (2009), the use of bootstrap techniques allow statisticians to draw inference for functional data. This leads to the next section on bootstrapping functional time series in FPCR.

5 Bootstrapping functional time series in functional principal component regression

Since the seminar work of Efron & Tibshirani (1993), bootstrap techniques have receiving increasing popularity in Statistics. Bootstrap techniques have been widely used for determining the critical value for testing hypothesis, in assessing the "goodness of fit" of error assumption in a linear model, and in constructing confidence intervals. Despite its general applicability, bootstrap techniques have been mainly applied to multivariate or univariate data. There exists comparably less work on bootstrapping functional data, with some exceptions of Cuevas et al. (2006), Hall & Vial (2006), Poskitt & Sengarapillai (2009) and González-Manteiga & Martínez-Calvo (2011).

Cuevas et al. (2006) proposed a bootstrap technique to re-sample the functional data and construct confidence bands for the mean function. They then applied the bootstrap technique to classify functional data in the experimental cardiology. Hall &

Vial (2006) also proposed a bootstrap technique to address the issue of selecting the optimal number of functional principal components. However, the limitation of the bootstrap techniques proposed by Cuevas et al. (2006) and Hall & Vial (2006) is that the original functional data are all independent and identically distributed (i.i.d.). In this section, I revisit two different bootstrap techniques given by González-Manteiga & Martínez-Calvo (2011) and Poskitt & Sengarapillai (2009) that are applicable to re-sample functional time series.

Given the raw time series of functions $\boldsymbol{f}(x) = [f_1(x), \ldots, f_n(x)]'$ of n observations of a stochastic process, an obvious way to get some idea of the sampling variability of a statistics of interest is to re-sample from $\boldsymbol{f}(x)$, and construct a bootstrap replication $\boldsymbol{f}(x) = [f_1^*(x), \ldots, f_n^*(x)]'$. By repeatedly generating different bootstrap replications, an approximation to the statistical distribution can be constructed.

In González-Manteiga & Martínez-Calvo (2011), the random variation observed in $\boldsymbol{f}(x) = [f_1(x), \ldots, f_n(x)]'$ stems from fluctuation in the residual function, denoted by $\epsilon_t(x)$ in (6). Because $\epsilon_1(x), \ldots, \epsilon_n(x)$ are uncorrelated, they can be randomly sampled. From which, the bootstrap replications of $\boldsymbol{f}^*(x)$ are obtained. The algorithm for re-sampling proceeds as follows:

Step 1. Obtain the residuals $\epsilon_t(x) = f_t(x) - \mu(x) - \sum_{k=1}^{K} \beta_{t,k} \phi_k(x)$, for all $t = 1, 2, \ldots, n$.

Step 2. This step alters depending on which bootstrap procedure is applied: naive bootstrap or wild bootstrap.

Naive Draw $\epsilon_1^*(x), \ldots, \epsilon_n^*(x)$ i.i.d. random variables from the cumulative distribution of $[\epsilon_t(x) - \bar{\epsilon}(x)]_{t=1}^{n}$, where $\bar{\epsilon}(x) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(x)$.

Wild For $t = 1, 2, \ldots, n$, define $\epsilon_t^*(x) = \epsilon_t(x) V_t$, and $V_1, \ldots, V_n$ are i.i.d. standard random variables with mean 0 and variance 1.

Step 3. Definite $f_t^*(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \epsilon_t^*(x)$, for all $t = 1, 2, \ldots, n$.

In Poskitt & Sengarapillai (2009), the random variation observed in $\boldsymbol{f}(x) = [f_1(x), \ldots, f_n(x)]'$ emanates from fluctuations in the principal component scores. These coefficients are uncorrelated random variables, with mean zero and unit variance, thus can be sampled without replacement to produce $\beta_{t,k}^*$. The algorithm for re-sampling proceeds as follows:

Step 1. Hold the mean $\mu(x)$, the eigenvalues $\lambda_k, k = 1, \ldots, K$, and the functional principal components $\phi_k(x)$ fixed at their realized values.

Step 2. This step alters depending on which bootstrap procedure is applied: direct bootstrap or Gaussian approximation bootstrap.

Direct For $t = 1, \ldots, n$, generating bootstrap replications $\beta_{t,k}^*, k = 1, \ldots, K$ by taking i.i.d. random draws from $\beta_{t,k}$.

Gaussian approximation For $t = 1, \ldots, n$, generating bootstrap replication $\beta_{t,k}^*, k = 1, \ldots, K$ by taking i.i.d. random draws from a standard normal distribution.

Step 3. Construct the bootstrap sample $\boldsymbol{f}^*(x) = [f_1^*(x), \ldots, f_n^*(x)]'$. The bootstrap realization is constructed as in (6) by simply replacing $\beta_{t,k}$ by $\beta_{t,k}^*$.

In the Gaussian approximation bootstrap, the rational behind generating the $\beta_{t,k}^*$ as independent standard normal variables comes from noticing that the $\{\boldsymbol{\beta}_k, k = 1, \ldots, K\}$ lies in the Stiefel manifold and a natural distribution to take in this manifold is the Von Mises-Fisher distribution (Hoff 2009). As the concentration parameter increases, the Von Mise-Fisher distribution can be well approximated by a standard normal distribution (Poskitt & Sengarapillai 2009).

To demonstrate the usefulness of bootstrap in FPCR, I applied the direct bootstrap technique of Poskitt & Sengarapillai (2009) to Australian fertility data displayed in

Figure 1. By bootstrapping functional principal component scores, I obtain 1000 bootstrap replications. The median of these bootstrap replications for each age and each year is plotted in Figure 4, along with the original data.
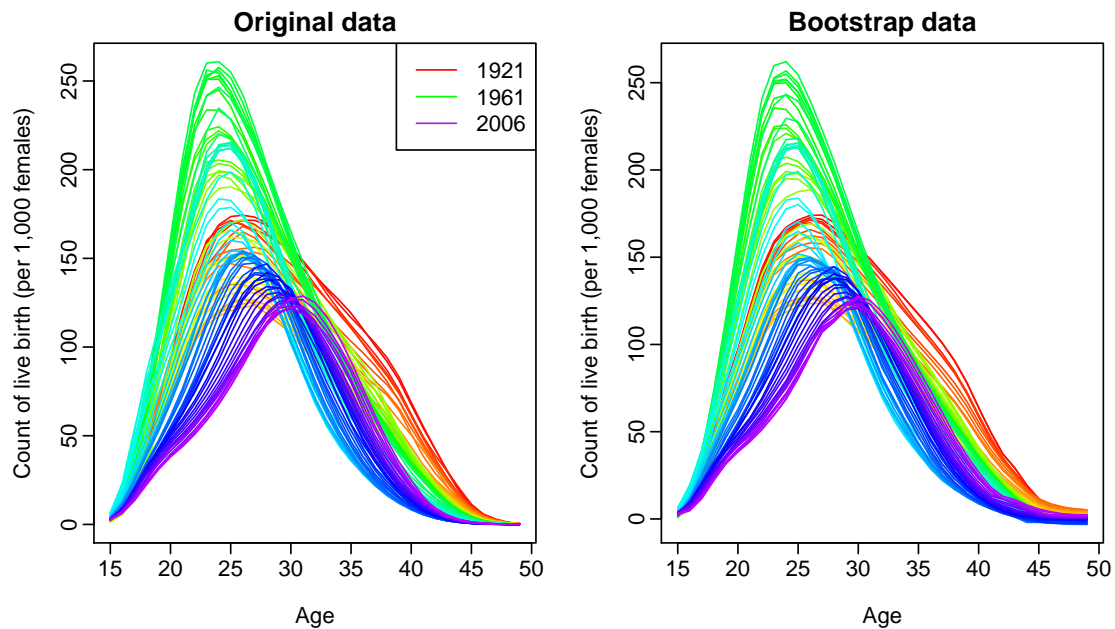

Figure 4: Bootstrapped Australian fertility rates from 1921 to 2006. The median of the bootstrap samples for each age and each year is plotted.

There are some differences between the original functional time series and boot-strapped functional time series, especially for the older ages. However, the bootstrap samples are able to capture the main pattern of the underlying stochastic process, especially for the peak fertility rates. Although it is out of the scope of this review paper, it would be interesting to compare the performance of these four different bootstrap algorithms and their combinations.

6 Conclusion

Modern data analysis has had and will continue to benefit greatly from the recent development of functional data analysis. Illustrated by the Australian fertility data, this paper has broadly revisited some functional principal component techniques

for analyzing increasingly high-dimensional data, with the main emphasis being on three popular areas, namely FPCA, FPCR, and bootstrap in FPCR.

This paper is concluded by pointing out a future direction in FPCR. In the literature of FPCR, the main attention has been given to the estimation of mean function and functional principal components. The density estimation of error function in FPCR has played a minor role at most, although it is important to understand the residual behavior and assess the fit of FPCR. To name a few, the density estimation of error function, denoted by $f(\epsilon)$, is useful for testing the adequacy of an assumed error distribution. The estimation of $f(\epsilon)$ allows us to visualize the density shape of residuals, such as heavy tailed residual density exhibited in many financial applications. The estimation of $f(\epsilon)$ is also useful to construct nonparametric prediction intervals of the error and response variables. It is hoped that the error density estimation in FPCR will receive its deserved attention. A recent paper by Gabrys et al. (2010) is an example in this direction.

Acknowledgement

The author would like to thank Professors Rob Hyndman and Donald Poskitt for introducing him to the field of functional data analysis. The author also would like to thank Professor Marc Hallin for providing many suggestions, which lead to a much improved manuscript.

References

Abraham, C., Cornillon, P. A., Matzner-Løber, E. & Molinari, N. (2003), 'Unsupervised curve clustering using B-splines', *Scandinavian Journal of Statistics* **30**(3), 581–595.

Antoch, J., Prchal, L., De Rosa, M. R. & Sarda, P. (2008), Functional linear regression with functional response: application to prediction of electricity consumption, *in* S. Dabo-Niang & F. Ferraty, eds, 'Functional and Operatorial Statistics', Physica-Verlag, Heidelberg, pp. 23–29.

Bellman, R. E. (1961), *Adaptive Control Processes: a Guided Tour*, Princeton University Press, Princeton, New Jersey.

Benko, M., Härdle, W. & Kneip, A. (2009), 'Common functional principal components', *The Annals of Statistics* **37**(1), 1–34.

Berk, R. A. (2008), *Statistical Learning from a Regression Perspective*, Springer, New York.

Besse, P. C., Cardot, H. & Stephenson, D. B. (2000), 'Autoregressive forecasting of some functional climatic variations', *Scandinavian Journal of Statistics* **27**(4), 673–687.

Burba, F., Ferraty, F. & Vieu, P. (2009), 'k-nearest neighbour method in functional nonparametric regression', *Journal of Nonparametric Statistics* **21**(4), 453–469.

Cardot, H., Ferraty, F. & Sarda, P. (2003), 'Spline estimators for the functional linear model', *Statistica Sinica* **13**(3), 571–591.

Castro, P. E., Lawton, W. H. & Sylvestre, E. A. (1986), 'Principal modes of variation for processes with continuous sample curves', *Technometrics* **28**(4), 329–337.

Chiou, J.-M. & Li, P.-L. (2007), 'Functional clustering and identifying substructures

of longitudinal data', *Journal of the Royal Statistical Society: Series B* **69**(4), 679–699.

Cont, R. & de Fonseca, J. (2002), 'The dynamics of implied volatility surfaces', *Quantitative Finance* **2**(1), 45–60.

Croux, C. & Ruiz-Gazen, A. (2005), 'High breakdown estimators for principal components: the projection-pursuit approach revisited', *Journal of Multivariate Analysis* **95**, 206–226.

Cuevas, A., Febrero, M. & Fraiman, R. (2006), 'On the use of the bootstrap for estimating functions with functional data', *Computational Statistics & Data Analysis* **51**(2), 1063–1074.

Damon, J. & Guillas, S. (2002), 'The inclusion of exogenous variables in functional autoregressive ozone forecasting', *Environmetrics* **13**(7), 759–774.

Dauxois, J. & Pousse, A. (1976), Les analyses factorielles en calcul des probabilités et en statistique: essai d'étude synthétique, PhD thesis, l'Université Paul-Sabatier de Toulouse, France.

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.

Davidian, M., Lin, X. & Wang, J.-L. (2004), 'Introduction: emerging issues in longitudinal and functional data analysis', *Statistica Sinica* **14**(3), 613–614.

Deville, J. C. (1974), 'Méthodes statistiques et numériques de l'analyse harmonique', *Annales de l'INSEE* **15**, 3–101.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. & Punjabi, N. M. (2009), 'Multilevel functional principal component analysis', *The Annals of Applied Statistics* **3**(1), 458–

488.

Dieudonné, J. A. E. (1960), *Foundations of Modern Analysis*, Academic Press, New York.

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York.

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.

Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing*, 2nd edn, Marcel Dekker, New York.

Ferraty, F., Mas, A. & Vieu, P. (2007), 'Nonparametric regression on functional data: inference and practical aspects', *Australian & New Zealand Journal of Statistics* **49**(3), 267–286.

Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York.

Gabrys, R., Horvth, L. & Kokoszka, P. (2010), 'Tests for error correlation in the functional linear model', *Journal of the American Statistical Association* **105**(491), 1113–1125.

González-Manteiga, W. & Martínez-Calvo, A. (2011), 'Bootstrap in functional linear regression', *Journal of Statistical Planning and Inference* **141**(1), 453–461.

González-Manteiga, W. & Vieu, P. (2007), 'Statistics for functional data (editorial)', *Computational Statistics & Data Analysis* **51**(10), 4788–4792.

Green, P. J. & Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*, Chapman & Hall, London.

Grenander, U. (1950), 'Stochastic processes and statistical inference', *Arkiv för Matematik* **1**(3), 195–277.

Groupe de Travail en Statistique Fonctionnelle et Opeatorielle (2011), 'Nonparametric functional data analysis', University of Toulouse, France. Accessed 8th March, 2011.
URL: *http: // www. math. univ-toulouse. fr/ staph/ npfda/*

Gunst, R. F. & Mason, R. L. (1980), *Regression Analysis and its Application: a Data-Oriented Approach*, Marcel Dekker, New York.

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.-G. & Wang, J.-L. (2006), 'Properties of principal component methods for functional and longitudinal data analysis', *The Annals of Statistics* **34**(3), 1493–1517.

Hall, P., Poskitt, D. S. & Presnell, B. (2001), 'A functional data-analytic approach to signal discrimination', *Technometrics* **43**(1), 1–9.

Hall, P. & Vial, C. (2006), 'Assessing the finite dimensionality of functional data', *Journal of the Royal Statistical Society (Series B)* **68**(4), 689–705.

Hand, D. J. (2007), *Information Generation: How Data Rule Our World*, Oneworld, Oxford.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd edn, Springer, New York.

Hoff, P. D. (2009), 'Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data', *Journal of Computational*

and Graphical Statistics **18**(2), 438–456.

Hotelling, H. (1933), 'Analysis of a complex of statistical variables into principal components', *Journal of Educational Psychology* **24**(6-7), 417–441; 498–520.

Hotelling, H. (1957), 'The relations of the newer multivariate statistical methods to factor analysis', *British Journal of Statistical Psychology* **10**(2), 69–79.

Hyndman, R. J. (2007), *addb: Australian demographic data bank.* R package version 3.222.
URL: *http: // robjhyndman. com/ software/ 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. (2009), 'Forecasting functional time series (with discussion)', *Journal of the Korean Statistical Society* **38**(3), 199–221.

Hyndman, R. J. & Shang, H. L. (2010), 'Rainbow plots, bagplots, and boxplots for functional data', *Journal of Computational and Graphical Statistics* **19**(1), 29–45.

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.

Izenman, A. J. (2008), *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, New York.

Jackson, J. E. (1991), *A User's Guide to Principal Components*, Wiley, New York.

James, G. M., Hastie, T. J. & Sugar, C. A. (2000), 'Principal component models for

sparse functional data', *Biometrika* **87**(3), 587–602.

Jeffers, J. N. R. (1967), 'Two case studies in the application of principal component analysis', *Journal of the Royal Statistical Society: Series C* **16**(3), 225–236.

Jolliffe, I. T. (1982), 'A note on the use of principal components in regression', *Journal of the Royal Statistical Society: Series C* **31**(3), 300–303.

Jolliffe, I. T. (2002), *Principal Component Analysis*, 2nd edn, Springer, New York.

Jones, M. C. & Rice, J. A. (1992), 'Displaying the important features of large collections of similar curves', *The American Statistician* **46**(2), 140–145.

Karhunen, K. (1946), 'Zur spektraltheorie stochastischer prozesse', *Annales Academiae Scientiarum Fennicae* **37**, 1–37.

Kendall, M. G. (1957), *A Course in Multivariate Analysis*, Hafner Pub. Co., New York.

Kneip, A. & Utikal, K. J. (2001), 'Inference for density families using functional principal component analysis', *Journal of the American Statistical Association* **96**(454), 519–542.

Krzanowski, W. J. & Marriott, F. H. C. (1994), *Multivariate Analysis*, E. Arnold, London.

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. & Cohen, K. L. (1999), 'Robust principal component analysis for functional data', *Test* **8**(1), 1–73.

Loève, M. (1946), 'Fonctions aléatoires a decomposition orthogonale exponentielle', *La Revue Scientifique* **84**, 159–162.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979), *Multivariate Analysis*, Academic Press, London.

Massy, W. F. (1965), 'Principal components regression in exploratory statistical research', *Journal of the American Statistical Association* **60**(309), 234–256.

Meiring, W. (2007), 'Oscillations and time trends in stratospheric ozone levels: a functional data analysis approach', *Journal of the American Statistical Association* **102**(479), 788–802.

Møller, S. F., von Frese, J. & Bro, R. (2005), 'Robust methods for multivariate data analysis', *Journal of Chemometrics* **19**(10), 549–563.

Mosteller, F. & Tukey, J. W. (1977), *Data Analysis and Regression: a Second Course in Statistics*, Addison-Wesley, Reading, Mass, USA.

Naes, T. & Martens, H. (1988), 'Principal component regression in NIR analysis: viewpoints, background details and selection of components', *Journal of Chemometrics* **2**(2), 155–167.

Ogden, R. T., Miller, C. E., Takezawa, K. & Ninomiya, S. (2002), 'Functional regression in crop lodging assessment with digital images', *Journal of Agricultural, Biological, and Environmental Statistics* **7**(3), 389–402.

Pearson, K. (1901), 'On lines and planes of closest fit to systems of points in space', *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572.

Peres-Neto, P., Jackson, D. & Somers, K. (2005), 'How many principal components?stopping rules for determining the number of non-trivial axes revisited', *Computational Statistics & Data Analysis* **49**(4), 974–997.

Poskitt, D. S. & Sengarapillai, A. (2009), Description length and dimensionality reduction in functional data analysis, Working paper 13/09, Department of Econometrics & Business Statistics, Monash University.
URL: *http: // www. buseco. monash. edu. au/ ebs/ pubs/ wpapers/ 2009/*

wp13–09. pdf

R Development Core Team (2009), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http:// www. R-project. org*

Ramsay, J. O. (1982), 'When the data are functions', *Psychometrika* **47**(4), 379–396.

Ramsay, J. O. (2000), 'Functional components of variation in handwriting', *Journal of the American Statistical Association* **95**(449), 9–15.

Ramsay, J. O. (2011), 'Functional data analysis', McGill University, Canada. Accessed 16th March, 2011.
URL: *http:// www. psych. mcgill. ca/ misc/ fda/ software. html*

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., Hooker, G. & Graves, S. (2009), *Functional Data Analysis with R and MATLAB*, Springer, New York.

Ramsay, J. O. & Silverman, B. W. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.

Ramsay, J. O. & Silverman, B. W. (2005), *Functional Data Analysis*, 2nd edn, Springer, New York.

Ramsay, J. O., Wickham, H., Graves, S. & Hooker, G. (2009), *fda: functional data analysis*. R package version 2.2.0.
URL: *http:// cran. r-project. org/ package= fda*

Rao, C. R. (1958), 'Some statistical methods for comparison of growth curves', *Biometrics* **14**(1), 1–17.

Reiss, P. T. & Ogden, R. T. (2007), 'Functional principal component regression and functional partial least squares', *Journal of the American Statistical Association* **102**(479), 984–996.

Rice, J. A. (2004), 'Functional and longitudinal data analysis: perspectives on smoothing', *Statistica Sinica* **14**(3), 631–647.

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.

Rossi, N., Wang, X. & Ramsay, J. O. (2002), 'Nonparametric item response function estimates with the EM algorithm', *Journal of Educational and Behavioral Statistics* **27**(3), 291–317.

Shang, H. L. & Hyndman, R. J. (2011), 'Nonparametric time series forecasting with dynamic updating', *Mathematics and Computers in Simulation* **81**(7), 1310–1324.

Shen, H. (2009), 'On modeling and forecasting time series of smooth curves', *Technometrics* **51**(3), 227–238.

Shen, H. & Huang, J. Z. (2008), 'Interday forecasting and intraday updating of call center arrivals', *Manufacturing & Service Operations Management* **10**(3), 391–410.

Shibata, R. (1981), 'An optimal selection of regression variables', *Biometrika* **68**(1), 45–54.

Silverman, B. W. (1985), 'Two books on density estimation (book review)', *The Annals of Statistics* **13**(4), 1630–1638.

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

Valderrama, M. J. (2007), 'An overview to modelling functional data (editorial)',

Computational Statistics **22**(3), 331–334.

Van Huffel, S. & Vandewalle, J. (1991), *The Total Least Squares Problem: Computational Aspects and Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.

Wahba, G. (1990), *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia.

Wang, S., Jank, W. & Shmueli, G. (2008), 'Explaining and forecasting online auction prices and their dynamics using functional data analysis', *Journal of Business & Economic Statistics* **26**(2), 144–160.

Wang, S., Jank, W., Shmueli, G. & Smith, P. (2008), 'Modeling price dynamics in eBay auctions using differential equations', *Journal of the American Statistical Association* **103**(483), 1100–1118.

Wichern, D. W. & Churchill, G. A. (1978), 'A comparison of ridge estimators', *Technometrics* **20**(3), 301–311.

Yao, F. & Müller, H.-G. (2010), 'Functional quadratic regression', *Biometrika* **97**(1), 49–64.

Yao, F., Müller, H.-G. & Wang, J.-L. (2005*a*), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association* **100**(470), 577–590.

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

Résumé

Les progrès de la collecte de données et de stockage ont considérablement accru la présence de données fonctionnelles, dont les représentations graphiques sont des courbes, des images ou des formes. Comme un nouveau domaine de la statistique, analyse de données fonctionnelles prolonge les méthodologies et théories dans les domaines de l'analyse fonctionnelle, les modèles linéaires généralisés, analyse multivariée des données, la statistique non paramétrique et bien d'autres. Ce document fournit un examen en analyse de données fonctionnelles en mettant l'accent principal sur l'analyse en composantes principales fonctionnelles, fonctionnelles régression en composantes principales, et bootstrap en régression élément fonctionnel principal. Les tendances récentes ainsi que les problèmes ouverts dans la région sont discutées.