SFB 649 Discussion Paper 2005-016

Common Functional Component Modelling

Alois Kneip*
Michal Benko**

* Department Statistics and Mathematics, Johannes Gutenberg-Universität Mainz, Germany
** CASE - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

1 Common functional component modelling

Alois Kneip and *Michal Benko*

ACKNOWLEDGEMENT

Financial support of *Deutsche Forschungsgemeinschaft* via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin, is gratefully acknowledged.

1.1 Introduction

Functional data analysis (FDA) has become a popular technique in applied statistics. In particular, this methodology has received considerable attention in recent studies in empirical finance. In this talk we discuss selected topics of functional principal components analysis that are motivated by financial data.

By definition, FDA deals with the analysis of samples of functions. However, in practice the functions of interest are often not directly observed but are regression curves which have to be reconstructed from discrete, noisy data. In Section 2 we present a new method for efficient estimation of functional principal components in such situations. It consists in an adaptation of a technique introduced by Kneip and Utikal (2001) for the case of density functions.

Inference for two independent functional samples is considered. Bootstrap tests are developed to test whether principal components coincide and the two samples thus possess "common" functional principal components. The procedure possesses an important application in modelling implied volatilites as described in Benko and Härdle (2004).

In this section we will focus on one sample of i.i.d. smooth random functions $x_1(t), \ldots, x_N(t) \in L^2[0,1]$, $t \in [0,1]$. For $v, w \in L^2[0,1]$ let $\langle v, w \rangle = \int_a^b v(t)w(t)dt$, and let $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$ denote the usual L^2-norm. The Karhunen-Loève decomposition then

provides a basic tool to describe the distribution of the random functions x_i. With $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\gamma_1, \gamma_2, \ldots$ denoting eigenvalues and corresponding orthonormal eigenfunctions of the covariance operator $\mathcal{C}$ of x_i we obtain

$$x_i = \mu + \sum_{j=1} \beta_{ij}\gamma_j, \quad i = 1, \ldots, N, \tag{1.1}$$

where $\mu = \mathbf{E}(x_i)$ is the mean function and $\beta_{ij} = \langle x_i - \mu, \gamma_j \rangle$ are (scalar) factor loadings with $\mathbf{E}(\beta_{ij}^2) = \lambda_j$. Structure and dynamics of the random functions can be assessed by analyzing the "functional principal components" γ_j as well as the distribution of the factor loadings.

An important property of (1.1) consists in the known fact that the first L principal components provide a "best basis" for approximating the sample functions in terms of the integrated square error.

For any choice of L orthonormal basis functions $v_1, \ldots, v_L$

$$\rho(v_1, \ldots, v_L) = \mathbf{E}(\| x_i - \mu - \sum_{j=1}^{L} \langle x_i - \mu, v_j \rangle v_j \|^2)$$

is minimized by $v_j = \gamma_j$. In many important applications a small number of functional principal components will suffice to approximate the functions x_i with a high degree of accuracy.

For a given sample an empirical analog of (1.1) can be constructed by using eigenvalues $\lambda_{1,n} \geq \lambda_{2,N} \geq \ldots$ and orthonormal eigenfunctions $\gamma_{1,N}, \gamma_{2,N}, \ldots$ of the empirical covariance operator $\mathcal{C}_N$, where $\mathcal{C}_N = \frac{1}{N} \sum_{i=1}^{N} \langle x_i - \bar{x}, \xi \rangle (x_i - \bar{x})$. If K denotes the number of nonzero eigenvalues of $\mathcal{C}_N$ then

$$x_i = \bar{x} + \sum_{j=1}^{K} \beta_{ij,N}\gamma_{j,N}, \quad i = 1, \ldots, N, \tag{1.2}$$

where $\bar{x}$ is the sample mean, and $\frac{1}{N}\sum_i \beta_{ij,N}^2 = \lambda_{j,N}$. Obviously, $\lambda_{j,N}$ and $\gamma_{j,N}$ estimate λ_j and γ_j for $j = 1, 2, \ldots$.

However, in practice, the sample functions x_i are often not directly observed, but have to be reconstructed from noisy observations Y_{ij} at discrete design points t_{ik}:

$$y_{ik} = x_i(t_{ik}) + \varepsilon_{ik}, \quad k = 1, \ldots, T_i, \tag{1.3}$$

where ε_{ik} are independent noise terms with $\mathrm{E}(\varepsilon_{ik}) = 0$, $\mathrm{Var}(\varepsilon_{ik}) = \sigma_i^2$.

In this context the standard approach to estimate functional principal components is to first estimate individual functions nonparametrically and then to determine eigenfunctions of the resulting estimated empirical covariance operator (compare, e.g., Ramsay and Silverman (1997)).

We propose an alternative approach which in a first step relies on estimating the elements of the matrix

$$M_{lk} = \langle x_l - \bar{x}, x_k - \bar{x} \rangle, \quad l, k = 1, \ldots, N. \tag{1.4}$$

Some simple linear algebra shows that all nonzero eigenvalues $\lambda_{1,N} \geq \lambda_{2,N} \ldots$ of $\mathcal{C}_N$ and $l_1 \geq l_2 \ldots$ of M are related by $\lambda_{j,N} = l_j/N$. When using additionally the corresponding orthonormal eigenvectors $p_1, p_2, \ldots$ of M, the empirical scores $\beta_{jr,N} = \langle x_j - \bar{x}, \gamma_{r,N} \rangle$ as well as the empirical eigenfunctions $\gamma_{r,N}$ are obtained by $\beta_{jr,N} = \sqrt{l_r} p_{jr}$ and

$$\gamma_{r,N} = \left(\sqrt{l_r}\right)^{-1} \sum_{i=1}^{N} p_{ir} \left(x_i - \bar{x}\right) = \left(\sqrt{l_r}\right)^{-1} \sum_{i=1}^{N} p_{ir} x_i. \tag{1.5}$$

The elements of M are functionals which can be estimated with asympotically negligible bias and a parametric rate of convergence $T_i^{-1/2}$. If the data in (1.3) is generated from a balanced, equidistant design, then it is easily seen that for $i \neq j$ this rate of convergence is achieved by the estimator $\widehat{M}_{ij} = \frac{1}{T} \sum_{k=1}^{T} y_{ik} y_{jk}$.

In the case of a random design some adjustment is necessary: Set $T \stackrel{\text{def}}{=} \min\{T_1, T_2, \ldots, T_N\}$ and define an equidistant grid $\{t_k \stackrel{\text{def}}{=} k/(T-1), k = 0, 1, \ldots, T-1\}$ on $[0, 1]$. Then, for each $i = 1, \ldots, N$ and $k = 1, \ldots, T$ find the index of the first and second nearest neighbor of t_k:

$$k(i) = \underset{j=1,\ldots,T_i}{\arg\min} |t_{ij} - t_k| \text{ and } k'(i) = \underset{j \neq k'(i)}{\arg\min} |t_{ij} - t_k|.$$

Finally, construct the estimators

$$\widehat{M}_{ij} = \frac{1}{T} \sum_{k=1}^{T} y_{ik(i)} y_{jk(j)} \text{ for } i \neq j \text{ and } \widehat{M}_{ii} = \frac{1}{T} \sum_{k=1}^{T} y_{ik(i)} y_{ik'(i)}. \tag{1.6}$$

The aim of using special estimator (1.6) for the diagonal terms is to avoid the additional bias. Alternatively we can construct a bias corrected estimator using some nonparametric estimation of variance σ_i^2.

The eigenvalues $\hat{l}_1 \geq \hat{l}_2 \ldots$ and eigenvectors $p_1, p_2, \ldots$ of the resulting matrix $\widehat{M}$ then provide estimates $\hat{\lambda}_{r,N} = \hat{l}_r/N$ and $\hat{\beta}_{jr} = \sqrt{\hat{l}_r} \hat{p}_{jr}$. Estimates $\hat{\gamma}_{r,N}$ of the empirical

functional principal component $\gamma_{r,N}$ can be determined from (1.5) when replacing the unknown true functions x_i by nonparametric estimates $\hat{x}_i$ (as, for example, local polynomial estimates with bandwidths h_i):

$$\hat{\gamma}_{r,N} = \left(\sqrt{\hat{l}_r}\right)^{-1} \sum_{i=1}^{N} \hat{p}_{ir}\hat{x}_i. \tag{1.7}$$

When considering (1.7), it is important to note that $\hat{\gamma}_{r,N}$ is defined as *a weighted average* of all estimated sample functions. Averaging reduces variance, and efficient estimation of $\gamma_{r,N}$ therefore requires *undersmoothing* of individual function estimates $\hat{x}_i$. Indeed, under suitable additional regularity conditions it can be shown that for an optimal choice of smoothing parameters and twice continuously differentiable x_i, we obtain the rate of convergence $\| \gamma_{r,N} - \hat{\gamma}_{r,N} \| = O_P((NT)^{-2/5})$. Here, $T = \min_i T_i$.

A natural question is how many functions K should be used when approximating the sample functions by a factor model of the form (1.2), when relying on *estimated* principal components and scores. This question is crucial from statistical and practical point of view, although it is not discussed often in the literature. We propose a bootstrap test which successively tests the hypothesis $H_0 : \lambda_{K_0+1,N} = \cdots = \lambda_{N,N} = 0$ for $K_0 = 1, 2, 3, \ldots$ based on the respective residual average integrated square error: $\hat{\rho}_N(\gamma_1, \ldots, \gamma_{K_0}) = \frac{1}{N} \sum_{r=K_0+1}^{N} \hat{l}_r$.

The idea of this procedure is based on the fact that the estimated eigenfunctions that correspond to small estimated eigenvalues (with relative small importance) are rather driven by the sampling noise than can be interpreted as a component of the variation of the underlying population. In other words we do not want to use eigenfunctions whose relative importance cannot be significantly distinguished from random components generated by noise.

1.2 Two sample inference

Clearly, in the framework described by (1.1) - (1.3) we are faced with two sources of variability of estimated functional principal components. Due to sampling variation $\gamma_{r,N}$ will differ from the true component γ_r, and due to (1.3) there will exist an additional estimation error when approximating $\gamma_{r,N}$ by $\hat{\gamma}_{r,N}$. The results of Dauxois, Pousse and Romain (1982) imply that $\| \gamma_r - \gamma_{r,N} \| = O_P(N^{-1/2})$, and the results of the proceeding section therefore imply that the difference between $\hat{\gamma}_{r,N}$ and $\gamma_{r,N}$ is of smaller order of magnitude if T is sufficiently large compared to N. Inference about functional principal components under (1) - (3) will then be first order equivalent to an inference based on known functions x_i.

We are mainly interest in two sample problems. Thus let

$$x_1^1(t), x_2^1(t), \ldots, x_{N_1}^1(t) \text{ and } x_1^2(t), x_2^2(t), \ldots, x_{N_2}^2(t) \tag{1.8}$$

denote two independent samples of smooth functions. The problem of interest is to test whether the functional principal components γ_r in the respective decompositions (1.1) are common (identical) for both groups. In this case only the factor loadings β_{ir} may vary across samples. Then $\gamma_r = \gamma_r^1 = \gamma_r^2$ and

$$x_i^p = \mu^p + \sum_{r=1} \beta_{ir}^p \gamma_r, \quad p = 1, 2. \tag{1.9}$$

This hypothesis has been used in the work of Fengler, Härdle and Villa (2003) and Benko and Härdle (2004) in modelling implied volatilites. It can be seen as a functional generalization of the concept of "common principal components" as introduced by Flury (1988) in multivariate analysis.

If the functions x_i^p were directly observed, then in order to test the hypothesis

$$H_0 : \quad \gamma_r^1 = \gamma_r^2$$

for $r = 1, 2, \ldots$ one could rely on the test statistics

$$D_r = \| \gamma_{r,N}^1 - \gamma_{r,N}^2 \|^2 .$$

It can be shown that critical values of D_r can be determined by a bootstrap procedure: Under H_0 we have $D_r = \| \gamma_{r,N}^1 - \gamma_r^1 - (\gamma_{r,N}^2 - \gamma_r^2) \|^2$. The distribution of $\| \gamma_{r,N}^1 - \gamma_r^1 - (\gamma_{r,N}^2 - \gamma_r^2) \|^2$ can then be approximated by the bootstrap distribution of $\| \gamma_{r,N}^{1*} - \gamma_{r,N}^1 - (\gamma_{r,N}^{2*} - \gamma_{r,N}^2) \|$, where $\gamma_{r,N}^{1*}$ and $\hat{\gamma}_{r,N}^{2*}$ are estimates to be obtained from independent bootstrap samples

$$x_1^{1*}(t), x_2^{1*}(t), \ldots, x_{N_1}^{1*}(t) \text{ and } x_1^{2*}(t), x_2^{2*}(t), \ldots, x_{N_2}^{2*}(t). \tag{1.10}$$

In the practically more relevant situation that all curves in both samples have to be reconstructed from noisy observations according to (1.3), D_r must be replaced by $\hat{D}_r = \| \hat{\gamma}_{r,N}^1 - \hat{\gamma}_{r,N}^2 \|^2$, where estimates are determined by the procedure described above. Bootstrap estimates are then obtained by resampling the observations corresponding to the unknown curves x_i. The procedure is asymptotically valid if T is sufficiently large such that the additional estimation error is asymptotically negligible. Of course, the test should only be performed for components which can be savely distinguished from noise, and $1 \leq r \leq \min K_0^1, K_0^2$.

Bibliography

Benko, M.and Härdle, W. (2004). Common Functional IV Surface Analysis *Statistical Tools for Finance and Insurance*, edited by Čížek, P., Härdle, W., Weron, R. to appear by Springer.

Dauxois, J., Pousse, A. and 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*: 136-154.

Fan, J. and Huang, L. (1999). Nonparametric Estimation of Quadratic Regression Functionals, *Bernoulli 5,*: 927-949.

Hall, P., Kay, J.W. and Titterington, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika* 77: 520:528.

Flury, B. (1988). *Common Principal Components and Related Models*, Wiley, New York.

Fengler, M., Härdle, W. and Villa, P. (2003). The Dynamics of Implied Volatilities: A common principle components approach, *Review of Derivative Research 6*: 179-202.

Fengler, M., Härdle, W. and Mammen, E. (2003). Implied Volatitlity String Dynamics, CASE Discussion Paper, http://www.case.hu-berlin.de.

Kneip, A. and Utikal, K. (2001). Inference for Density Families Using Functional Principal Components Analysis, *Journal of the American Statistical Association 96*: 519-531.

Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*, Springer, New York.

SFB 649 Discussion Paper Series

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Nonparametric Risk Management with Generalized Hyperbolic Distributions" by Ying Chen, Wolfgang Härdle and Seok-Oh Jeong, January 2005.
002 "Selecting Comparables for the Valuation of the European Firms" by Ingolf Dittmann and Christian Weiner, February 2005.
003 "Competitive Risk Sharing Contracts with One-sided Commitment" by Dirk Krueger and Harald Uhlig, February 2005.
004 "Value-at-Risk Calculations with Time Varying Copulae" by Enzo Giacomini and Wolfgang Härdle, February 2005.
005 "An Optimal Stopping Problem in a Diffusion-type Model with Delay" by Pavel V. Gapeev and Markus Reiß, February 2005.
006 "Conditional and Dynamic Convex Risk Measures" by Kai Detlefsen and Giacomo Scandolo, February 2005.
007 "Implied Trinomial Trees" by Pavel Čížek and Karel Komorád, February 2005.
008 "Stable Distributions" by Szymon Borak, Wolfgang Härdle and Rafal Weron, February 2005.
009 "Predicting Bankruptcy with Support Vector Machines" by Wolfgang Härdle, Rouslan A. Moro and Dorothea Schäfer, February 2005.
010 "Working with the XQC" by Wolfgang Härdle and Heiko Lehmann, February 2005.
011 "FFT Based Option Pricing" by Szymon Borak, Kai Detlefsen and Wolfgang Härdle, February 2005.
012 "Common Functional Implied Volatility Analysis" by Michal Benko and Wolfgang Härdle, February 2005.
013 "Nonparametric Productivity Analysis" by Wolfgang Härdle and Seok-Oh Jeong, March 2005.
014 "Are Eastern European Countries Catching Up? Time Series Evidence for Czech Republic, Hungary, and Poland" by Ralf Brüggemann and Carsten Trenkler, March 2005.
015 "Robust Estimation of Dimension Reduction Space" by Pavel Čížek and Wolfgang Härdle, March 2005.
016 "Common Functional Component Modelling" by Alois Kneip and Michal Benko, March 2005.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".