

MODELING REPEATED FUNCTIONAL OBSERVATIONS

Kehui Chen and Hans-Georg Müller
Department of Statistics, University of California, Davis
Davis, CA 95616 USA

August 2012

ABSTRACT

We introduce a new methodological framework for repeatedly observed and thus dependent functional data, aiming at situations where curves are recorded repeatedly for each subject in a sample. Our methodology covers the case where the recordings of the curves are scheduled on a regular and dense grid and also situations more typical for longitudinal studies, where the timing of recordings is often sparse and random. The proposed models lead to an interpretable and straightforward decomposition of the inherent variation in repeatedly observed functional data and are implemented through a straightforward two-step Functional Principal Component Analysis. We provide consistency results and asymptotic convergence rates for the estimated model components. We compare the proposed model with an alternative approach via a two-dimensional Karhunen-Loève expansion and illustrate it through the analysis of longitudinal mortality data from period lifetables that are repeatedly observed for a sample of countries over many years, and also through simulation studies.

KEY WORDS: Asymptotics, functional data analysis, functional principal components, hierarchical model, lifetable, longitudinal data, mortality, rate of convergence, repeated measures, uniform convergence.

1. INTRODUCTION

Rapid advances in computational and analytical techniques have made it possible to continuously monitor and record many time-dynamic processes. As a result, functional data analysis (FDA) methodology aiming to study the resulting random functions encounters increasing interest. The statistical techniques of this field are still by and large focused on the case of independently sampled functions. A general introduction of the available methods for the i.i.d. case can be found in Ramsay and Silverman (2005).

An important extension of the i.i.d. case that we study here concerns situations where curves are repeatedly observed for each subject at each of a series of measurement times or along a series of locations. We illustrate our methods with the analysis of samples of mortality profiles, which have been collected over many years for various countries and where one wishes to characterize the country-specific mortality curves and their secular evolution over calendar time. Data of this type are common in longitudinal studies where functional measurements are made at each visit, for example daily behavior profiles in animal studies, obtained for many days per subject.

The importance of models for dependent functional data has been recognized previously (Morris et al. 2003; Morris and Carroll 2006; Crainiceanu et al. 2009; Di et al. 2009), where the emphasis has been on functional data with a general hierarchical structure. In these multilevel functional models and functional mixed effects models, the multiple functions observed for each subject are modeled additively for the different levels, similar to a multilevel ANOVA design. Greven et al. (2010) introduced models for longitudinally observed functional data that are more similar to the data structure we are considering here; these authors incorporated the time spacing of the measurements into the coefficients through a linear structure, relying on additive assumptions.

In contrast, we target repeatedly observed functions along a longitudinal time axis for situations where these functions are smoothly changing with the longitudinal times at

which the repeated functions are recorded, leading to a fully nonparametric model. The covariance and mean functions of the repeated functions are assumed to evolve smoothly in longitudinal time. The resulting flexible longitudinal functional model leads to sensible results under minimal assumptions, covers sparse random designs for the longitudinal times where the functions are observed, and also facilitates the study of the asymptotic behavior of methods for hierarchical or repeated functional data.

The proposed double FPCA method emerges as a natural extension of the well-established functional principal component analysis (FPCA) and has several key advantages. First, it relies on mild assumptions and is generally applicable for repeatedly observed functions when the sampling of these functions follows either dense regular or sparse and irregular random designs, where the latter are common in real-life longitudinal studies. Second, it is a straightforward extension of the tried and proven FPCA approach that has emerged as a reliable workhorse for FDA in the i.i.d. situation (Castro et al. 1986; Rice and Silverman 1991; Di et al. 2009). Third, it provides a decomposition of the total variation observed into the variation within the repeatedly observed functions as one component; and the variation between these random functions as a second component. Fourth, it is easy to implement and leads to interpretable results. Fifth, since the proposed double FPCA is derived from the nearly always applicable Karhunen-Loève expansion (Ash and Gardner 1975) of squared integrable random processes, which has been well studied in the i.i.d. case, it is amenable to asymptotic analysis under both longitudinal and dense designs and one obtains reasonably fast rates of convergence.

Referring to the times where the repeated functional measurements are taken from now on as longitudinal time, we consider observations $\{X_i(t_{ijl}|s_{ij}), 1 \leq i \leq n, 1 \leq j \leq m_i, 1 \leq l \leq L_{ij}\}$, where i is a subject index. Functions $X_i(\cdot|s_{ij})$ are repeatedly recorded for subject i at a series of longitudinal times s_{ij} (which might also correspond to spatial distances or other continuously scaled variables, depending on the application), j is the

repeated observation index for the repeated functional data of the i -th subject, and t_{ijl} are the time coordinates where the measurements of the j -th repeated function $X_i(\cdot|s_{ij})$ for the i -th subject are taken. The idea underlying the double FPCA approach is to model the repeated functions through an initial FPCA at a grid of suitable longitudinal times, across all subjects. The resulting principal components then become functions of longitudinal time, and these functions are subjected to a second FPCA.

These ideas motivate a two-step Karhunen-Loève expansion, leading to the following model for the repeated functions, evaluated at argument $t \in \mathcal{T}$ and recorded at longitudinal time $s \in \mathcal{S}$ for intervals \mathcal{S}, \mathcal{T} :

$$X_i(t|s) = \mu(t|s) + \sum_{k=1}^{\infty} \xi_k(s) \phi_k(t|s) = \mu(t|s) + \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \zeta_{ikp} \psi_{kp}(s) \phi_k(t|s). \quad (1)$$

Here $\phi_k(\cdot|s)$ is the k -th eigenfunction of the repeated functions at longitudinal time s , belonging to the first series of FPCAs for the repeated functions, carried out at fixed longitudinal times s , ζ_{ikp} are zero mean random variables and ψ_{kp} is the p -th eigenfunction of the functional principal component function $\xi_k(s)$, where the latter is obtained in the first FPCA series. Eigenfunctions ψ_{kp} belong to the second FPCA step of the proposed double FPCA approach.

In model (1), the variation inherent in the repeated functions $X_i(t|s)$ is decomposed into a series of two-dimensional principal surface components $\varphi_k(t|s)$, as per

$$X_i(t|s) = \mu(t|s) + \sum_{k,p=1}^{\infty} \zeta_{ikp} \varphi_{kp}(t|s), \quad \varphi_{kp}(t|s) = \psi_{kp}(s) \phi_k(t|s). \quad (2)$$

The random scores ζ_{ikp} can be employed for additional modeling tasks such as regression and classification. While decomposition (2) might look similar to the Karhunen-Loève expansion for a random field with two-dimensional argument $u = (t, s)$, we consider here a scenario where t and s have inherently different meanings in terms of study design and scientific meaning, and where the component functions $\varphi_{kp}(t|s) = \psi_{kp}(s) \phi_k(t|s)$,

the principal surfaces of this model, have a direct interpretation. Even if one ignores the roles played by t and s , the proposed method is numerically faster and also has a faster rate of convergence compared to fitting a Karhunen-Loève expansion with two-dimensional argument $u = (t, s)$, especially for the case where the data are not regular and dense. In this case, a four-dimensional smoothing step for the covariance function $G(t_1, s_1, t_2, s_2)$ is needed to implement the two-dimensional Karhunen-Loève expansion, while the proposed method only involves a three-dimensional smoothing step $G(t_1, t_2|s)$. For further discussion see Section 7 and the online Supplement B.

A challenge for the theoretical study of this model and associated estimators is that for the second step FPCA in the double FPCA method, direct observations of the underlying random process are not available. Instead, one needs to target a working process that depends on estimates from the the first step FPCA. Addressing this challenge, we establish consistency results and show that when the repeated random functions are assumed to be densely observed along their argument t , the estimated model components obtained from double FPCA enjoy a one-dimensional convergence rate.

The paper is organized as follows. In Section 2, we provide further details for the proposed model for repeated functional data, followed by implementation details in Section 3. Theoretical properties and rates of convergence are studied in Section 4. Section 5 contains a simulation study, and Section 6 an illustration of the proposed approach to repeatedly observed human mortality data, followed by a discussion in Section 7. Assumptions are in the Appendix and detailed proofs as well as additional simulations in the online Supplement.

2. MODELING OF REPEATEDLY RECORDED FUNCTIONS

Denote the time coordinate for the repeatedly observed functions by $t \in \mathcal{T}$ and the longitudinal time domain over which multiple functions per subject are observed repeatedly by $s \in \mathcal{S}$. For fixed s , $X(\cdot|s)$ is the repeated function observed at time s . For the i -th

subject one then has the Karhunen-Loève representation of $X_i(\cdot|s)$,

$$X_i(t|s) = \mu(t|s) + \sum_{k=1}^{\infty} \xi_{ik}(s) \phi_k(t|s), \quad i = 1, \dots, n, \quad (3)$$

where upon fixing s , the $\phi_k(\cdot|s)$ are the eigenfunctions of the covariance operator with kernel $G(t_1, t_2|s) = \text{cov}(X(t_1|s), X(t_2|s))$ and $\xi_{ik}(s)$ the corresponding functional principal components (FPCs), which have zero mean and are uncorrelated for fixed s , with $\text{var}(\xi_k(s)) = \lambda_k(s)$. When viewing $\phi_k(t|s)$ as a two-dimensional surface in s and t , we assume that it is smooth in both t and s .

By a second application of the Karhunen-Loève representation, the FPCs as functions in s are further represented as

$$\xi_{ik}(s) = \sum_{p=1}^{\infty} \zeta_{ikp} \psi_{kp}(s), \quad (4)$$

with eigenfunctions ψ_{kp} and zero mean “second level” FPCs ζ_{ikp} . Here the ψ_{kp} are the eigenfunctions of the covariance operator with kernel $R_k(s_1, s_2) = \text{cov}(\xi_k(s_1), \xi_k(s_2))$, thus defining the double FPCA model. Combining these representations leads to

$$\begin{aligned} X_i(t|s) &= \mu(t|s) + \sum_{k=1}^{\infty} \xi_{ik}(s) \phi_k(t|s) \\ &= \mu(t|s) + \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \zeta_{ikp} \psi_{kp}(s) \phi_k(t|s) \\ &= \mu(t|s) + \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \zeta_{ikp} \varphi_{kp}(t|s), \end{aligned} \quad (5)$$

where $\varphi_{kp}(t|s) = \psi_{kp}(s) \phi_k(t|s)$.

Thus, the variance structure of individual random surfaces X_i is decomposed into a series of smooth two-dimensional surfaces, the principal surfaces, with random components ζ_{ikp} . These surfaces consist of the products of eigenfunctions of the repeatedly observed functions, conditional on longitudinal time s , and of the eigenfunctions of the longitudinal dynamics in s . This corresponds to a decomposition of total variation into products of variation conditional on longitudinal time s and of variation along longitudinal time s .

3. ESTIMATING PROCEDURES

The proposed two-step FPCA approach proceeds as follows: (i) Obtain estimators for $\mu(t|s)$ and $G(t_1, t_2|s)$. (ii) Obtain $\phi_k(t|s_{ij})$ from the estimates of $G(t_1, t_2|s_{ij})$, for all s_{ij} . We note that the order $k = 1, \dots, K$ of $\phi_k(t|s)$ may be determined according to the order of the corresponding eigenvalues for a fixed pre-selected time s_0 , but then in the case of crossing eigenvalue functions, the order for other s is determined by the smooth continuation of the functions $\phi_k(t|s)$ as s varies, so that $\phi_k(t|s)$ remains smooth in s , even if the order of eigenvalues changes for different s . (iii) For each subject i and $k = 1, \dots, K$, obtain the estimates of functional scores $\xi_{ik}(s_{ij}) = \int (X_i(t|s_{ij}) - \mu(t|s_{ij}))\phi_k(t|s_{ij})dt$, $j = 1, \dots, m_i$. (iv) For each $k = 1, \dots, K$, perform a second FPCA, applied to the working processes $\{\hat{\xi}_{ik}(s_{ij}), i = 1, \dots, n, j = 1, \dots, m_i\}$, which are obtained in the previous step.

Estimating procedures for the mean surface $\mu(t|s)$ and the covariance operator $G(t_1, t_2|s)$ for specific design situations are as follows.

Scenario A (*Dense Regular Design*). The design grids for both t and s are regular and dense, i.e., the observations are $\{X_i(t_l, s_j), 1 \leq i \leq n, 1 \leq j \leq m, 1 \leq l \leq L\}$, where i is the subject index, multiple functions $X_i(\cdot|s_j)$ are repeatedly recorded for subject i at a series of dense and regular time points or locations s_j , and t_l are the time coordinates where the measurements of the function $X_i(\cdot|s_j)$ are taken. For this case, we use *empirical estimators*, i.e., cross-sectional sample mean and sample covariance on a grid of t_l and s_j , and interpolate linearly between grid points,

$$\hat{\mu}(t_l|s_j) = \frac{1}{n} \sum_{i=1}^n X_i(t_l|s_j), \quad (6)$$

$$\hat{G}(t_l, t_k|s_j) = \frac{1}{n} \sum_{i=1}^n X_i(t_l|s_j)X_i(t_k|s_j) - \hat{\mu}(t_l|s_j)\hat{\mu}(t_k|s_j), \quad (7)$$

where $1 \leq l, k \leq L$ and $1 \leq j \leq m$. Theorem 1 in Section 4 establishes uniform strong consistency for these estimates when $\max_l(t_l - t_{l-1}) = O(n^{-1})$ and $\max_j(s_j - s_{j-1}) = O(n^{-1})$ with convergence rate $(\log n/n)^{1/2}$.

Scenario B (*Dense Random Design*) and Scenario C (*Sparse Random Design in s*). Here multiple dense functions $X_i(\cdot|s_{ij})$ are repeatedly recorded for subject i at a series of time points or locations $s_{ij}, j = 1, \dots, m_i$. If m_i increases as n increases, we refer to this as Scenario B, and in the “longitudinal case”, where the m_i are bounded, as Scenario C. In random designs, we pool all data together to take advantage of the information contained in the entire sample and adopt a local-linear smoothing approach (Yao et al. 2005) to estimate $\mu(t|s)$ and $G(t_1, t_2|s)$. We refer to these as *smoothing estimators*.

Let $K(\cdot)$ be a symmetric probability density function on $[0, 1]$ and $K_h(t) = (1/h)K(t/h)$, where h is a bandwidth. The local-linear estimators of the mean function and covariance function are given by $\tilde{\mu}(t|s) = \hat{a}_0$, where

$$(\hat{a}_0, \hat{a}_1, \hat{a}_2) = \arg \min \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{l=1}^{L_{ij}} \{ [X_i(t_{ijl}|s_{ij}) - a_0 - a_1(s_{ij} - s) - a_2(t_{ijl} - t)]^2 \times K_{h_s}(s_{ij} - s) K_{h_t}(t_{ijl} - t) \}, \quad (8)$$

respectively, by $\tilde{G}(t_1, t_2|s) = \hat{a}_0 - \tilde{\mu}(t_1|s)\tilde{\mu}(t_2|s)$, where

$$(\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3) = \arg \min \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{l \neq k} \{ [X_i(t_{ijl}|s_{ij})X_i(t_{ijk}|s_{ij}) - a_0 - a_1(s_{ij} - s) - a_2(t_{ijl} - t_1) - a_3(t_{ijk} - t_2)]^2 \times K_{b_s}(s_{ij} - s) K_{b_t}(t_{ijl} - t_1) K_{b_t}(t_{ijk} - t_2) \}. \quad (9)$$

These estimates are consistent (see Theorem 2 below). For Scenario B, the smoothing estimates $\tilde{\mu}$ and \tilde{G} have the same convergence rate $(\log n/n)^{1/2}$ as the empirical estimates $\hat{\mu}$ and \hat{G} , under appropriate regularity conditions (Theorem 2 (a)).

We note that in practice, *empirical estimators* can also be used for dense data with random design by pre-smoothing for individual curves. As designs get denser, the overall convergence rate $(\log n/n)^{1/2}$ will remain the same. We further remark that according to our experience, the proposed method works well for doubly sparse settings (using *smoothing estimators*). If the data are sparsely sampled in the t direction, $\hat{\xi}_{ik}(s)$ will converge to $E(\xi_{ik}(s)|\text{Data})$, which in the sparse case differs from $\xi_{ik}(s)$ and therefore our

theory does not apply. In Section 4, consistency results are provided for the case that dense curves $X(t|s)$ are observed at randomly sampled longitudinal times s .

Note that when the measurements of underlying functions are contaminated with errors, regardless of the design, the method also works in practice. For the theoretical analysis, we do not consider additional errors in the dense data case for mathematical simplicity. For dense data, pre-smoothing of individual curves can always be used to reduce the error to a level so that it will not affect the overall convergence rate. For data with error, off-diagonal smoothing of the covariance surface is used to separate out the error variance, which is the reason that in eq. (9) the diagonal terms are excluded; for further details on this issue, we refer to Yao et al. (2005). After obtaining the estimates of $\mu(t|s)$ and $G(t_1, t_2|s)$, the estimates $\hat{\phi}_k(\cdot|s)$ for $\phi_k(\cdot|s)$ in Eq. (3) are derived numerically from $\hat{G}(t_1, t_2|s)$ or $\tilde{G}(t_1, t_2|s)$. The code for the proposed procedures is available in the PACE package at <http://anson.ucdavis.edu/~mueller/data/pace.html>.

Additional estimation details that were specifically adopted to address the issue of possibly crossing eigenvalues are as follows. We initially obtain $\phi_k(\cdot|s)$ for $s = s_0$, a left endpoint of the domain, where the order k is determined by the corresponding order of the eigenvalues $\lambda_k(s_0)$. Then for s_1 with $0 \leq s_1 - s_0 \leq \epsilon$, where $\epsilon > 0$ is a small number, we determine the indices k and the sign of the eigenfunctions $\hat{\phi}_k(\cdot|s_1)$, such that the L_2 distance to the already obtained neighboring function $\hat{\phi}_k(\cdot|s_0)$ is minimized. This is to ensure the smoothness of the eigenfunction evolution $\phi_k(\cdot|s)$ in s . As the eigenfunctions $\phi_k(\cdot|s_1)$ are orthonormal, smooth continuation of eigenfunctions uniquely determines $\phi_k(\cdot|s_1)$; the procedure is analogous for the subsequent grid points.

Subsequently, the values of $\hat{\phi}_k(\cdot|s)$ are discarded for those s where $\delta_{jk}(s) = \frac{\Delta_{jk}(s)}{\text{se}(\Delta_{jk}(s))}$ is smaller than a threshold ϑ for any j, k , with estimated gaps $\Delta_{jk}(s) = |\hat{\lambda}_j(s) - \hat{\lambda}_k(s)|$, $j \neq k$, as the estimated eigenfunctions $\hat{\phi}_k(\cdot|s)$ are not reliable for small gaps between the eigenvalues. A practical approach is to obtain the standard error $\text{se}(\Delta_{jk}(s))$ by bootstrapping;

a threshold ϑ corresponding to the 10th percentile of $\delta(s)$, denoted as $\tilde{\vartheta}$, worked quite well in our analysis. With this approach, the measure of the set of discarded values of s can be kept quite small and the discarded values can be easily interpolated by smoothing the available $\hat{\phi}_k(\cdot|s_j)$ across s .

While processes $\{\xi_k(s)\}$, $k \geq 1$, are the targets for the second step FPCA, the quantities available are empirical working targets $\{\hat{\xi}_{ik}(s_{ij}), i = 1, \dots, n, j = 1, \dots, m_i\}$, where m_i may be small, and to which we apply a second FPCA. We first compute the estimates $\hat{R}_k(s_1, s_2)$ of $R_k(s_1, s_2) = \text{cov}(\xi_k(s_1), \xi_k(s_2))$ by local linear smoothing. Then the estimates $\hat{\psi}_{kp}(s)$ of the eigenfunctions $\psi_{kp}(s)$ in (4) are obtained numerically from $\hat{R}_k(s_1, s_2)$ and estimates for ζ_{kp} are obtained by numerical integration for dense designs in s , or by best linear predictors through conditional expectation (PACE) for the case of a sparse designs in s . It follows from results in Müller (2005) that PACE can also be used for dense design cases.

After selecting appropriate numbers of included components K and P_k , one obtains the estimated overall representation,

$$\begin{aligned} \hat{X}_i(t|s) &= \hat{\mu}(t|s) + \sum_{k=1}^K \sum_{p=1}^{P_k} \hat{\zeta}_{ikp} \hat{\psi}_{kp}(s) \hat{\phi}_k(t|s) \\ &= \hat{\mu}(t|s) + \sum_{k=1}^K \sum_{p=1}^{P_k} \hat{\zeta}_{ikp} \hat{\varphi}_{kp}(t|s), \end{aligned} \quad (10)$$

where $\hat{\varphi}_{kp}(t|s) = \hat{\psi}_{kp}(s) \hat{\phi}_k(t|s)$. The included number of components K can be selected by fraction of variance explained (FVE). We use the criterion that $\min_s \frac{\sum_{k=1}^K \hat{\lambda}_k(s)}{\sum_{k=1}^M \hat{\lambda}_k(s)} > 0.9$, where M is large. The numbers of included components P_k for the FPCA of random functions $\{\xi_{ik}(s)\}$ may also be determined by FVE, applied for each k .

4. THEORETICAL PROPERTIES

To study the asymptotics of the proposed two-step estimating procedure, recall that $\hat{\mu}(t|s)$ (Eq. 6), $\hat{G}(t_1, t_2|s)$ (7) are *empirical estimators* and $\tilde{\mu}(t|s)$ (8), $\tilde{G}(t_1, t_2|s)$ (9) are

smoothing estimators of the mean surface $\mu(t|s)$ and the covariance operator $G(t_1, t_2|s) = \text{cov}(X_i(t_1|s), X_i(t_2|s))$ and that $\hat{\phi}_k(t|s)$, $k \geq 1$, are estimates of the eigenfunctions $\phi_k(t|s)$ (3). We provide consistency results for these estimates in the following Theorems 1 and 2. Let h_T, h_S denote the smoothing bandwidths for the local linear smoothing estimation of $\mu(t|s)$, and b_T, b_S the smoothing bandwidths for $G(t_1, t_2|s)$, \tilde{b}_k the smoothing bandwidths for $R_k(s_1, s_2) = \text{cov}(\xi_k(s_1), \xi_k(s_2))$; we always assume $h_T = O(h_S)$, $b_T = O(b_S)$, and $h_S = O(b_S)$, $b_S = O(\tilde{b}_k)$.

Theorem 1. (*Scenario A, Dense Regular Design*) Under assumptions (L.1) and (L.2) in the Appendix, and if $\max_l(t_l - t_{l-1}) = O(n^{-1})$, $\max_j(s_j - s_{j-1}) = O(n^{-1})$, the empirical estimators $\hat{\mu}(t, s)$ (6) and $\hat{G}(t_1, t_2|s)$ (7) satisfy

$$\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |\hat{\mu}(t, s) - \mu(t, s)| = O((\log n/n)^{1/2}) \quad a.s., \quad (11)$$

$$\sup_{t_1, t_2 \in \mathcal{T}, s \in \mathcal{S}} |\hat{G}(t_1, t_2|s) - G(t_1, t_2|s)| = O((\log n/n)^{1/2}) \quad a.s. \quad (12)$$

Corresponding results for the smoothing estimates $\tilde{\mu}$ and \tilde{G} that are obtained by local linear smoothing are as follows.

Theorem 2. (*Scenario B and C*) Under assumptions (A.1) - (A.6) in the Appendix and if $h_T = O(b_T)$, and $b_T = O((\log n/n)^{1/4})$, the smoothing estimators $\tilde{\mu}(t|s)$ (8) and $\tilde{G}(t_1, t_2|s)$ (9) have the following properties.

(a) (*Scenario B, Dense Random Design*) If $\min_{i,j} L_{ij} \geq L_n$, $L_n^{-1} = O(n^{-1})$, $\min_i m_i \geq M_n$, $M_n^{-1} = O(n^{-1})$, then

$$\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |\tilde{\mu}(t|s) - \mu(t|s)| = O((\log n/n)^{1/2}) \quad a.s.,$$

$$\sup_{t_1, t_2 \in \mathcal{T}, s \in \mathcal{S}} |\tilde{G}(t_1, t_2|s) - G(t_1, t_2|s)| = O((\log n/n)^{1/2}) \quad a.s.$$

(b) (*Scenario C, Sparse Random Design in s*) If $\min_{i,j} L_{ij} \geq L_n$, $L_n^{-1} = O(n^{-1})$, $\max_i m_i \leq M$ for a fixed M , then

$$\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |\tilde{\mu}(t|s) - \mu(t|s)| = O(h_S^2 + [\log n / (nh_S)]^{1/2}) \quad a.s., \quad (13)$$

$$\sup_{t_1, t_2 \in \mathcal{T}, s \in \mathcal{S}} |\tilde{G}(t_1, t_2|s) - G(t_1, t_2|s)| = O(b_S^2 + [\log n / (nb_S)]^{1/2}) \quad a.s. \quad (14)$$

The next results show that the empirical working targets $\{\hat{\xi}_{ik}(s_{ij})\}$ converge uniformly to the target processes $\{\xi_{ik}(s), s \in \mathcal{S}\}$ (3) under the assumption that the repeated functions are densely sampled (in direction t), and with the rate of convergence depending on the longitudinal sampling design (dense or sparse in direction s). This convergence implies consistent estimation of the model components in the second FPCA step in (4) (see also Yao and Lee 2006). Estimates $\hat{R}_k(s_1, s_2)$, $k \geq 1$, of $R_k(s_1, s_2) = \text{cov}(\xi_k(s_1), \xi_k(s_2))$, obtained from the empirical working targets $\{\hat{\xi}_{ik}(s)\}$ by local linear smoothing, as described in Section 3, are seen to be consistent by combining the results in (15) and (19) below with arguments in Yao et al. (2005); Hall et al. (2006); Li and Hsing (2010).

Theorem 3. (*Scenario A Dense Regular Design and B Dense Random Design*) If empirical estimators $\hat{\mu}(t, s)$ and $\hat{G}(t_1, t_2|s)$ are used for Scenario A and the assumptions in Theorem 1 are satisfied, or smoothing estimators $\tilde{\mu}$ and \tilde{G} are used for Scenario B and the assumptions in Theorem 2 part (a) are satisfied, then under additional assumption (L.3), for fixed k and p ,

$$\frac{1}{n} \sum_{i=1}^n \sup_{1 \leq j \leq m_i} |\hat{\xi}_{ik}(s_{ij}) - \xi_{ik}(s_{ij})| = O((\log n/n)^{1/2}) \quad a.s., \quad (15)$$

$$\sup_{s_1, s_2 \in \mathcal{S}} |\hat{R}_k(s_1, s_2) - R_k(s_1, s_2)| = O((\log n/n)^{1/2}) \quad a.s., \quad (16)$$

$$\sup_{s \in \mathcal{S}} |\hat{\psi}_{kp}(s) - \psi_{kp}(s)| = O((\log n/n)^{1/2}) \quad a.s., \quad (17)$$

$$\|\hat{\psi}_{kp}(s) - \psi_{kp}(s)\|_{L_2} = O((\log n/n)^{1/2}) \quad a.s. \quad (18)$$

Similar convergence holds for the sparse design in s , but due to the necessary smoothing step, the rates of convergence are slower.

Theorem 4. (*Scenario C, Sparse Random Design in s*) *If smoothing estimators $\tilde{\mu}(t, s)$ and $\tilde{G}(t_1, t_2, |s)$ are used, and the assumptions in Theorem 2 are satisfied, then under the additional assumptions (L.3) and (B.1)-(B.3), we have for fixed k and p*

$$\frac{1}{n} \sum_{i=1}^n \sup_{1 \leq j \leq m_i} |\hat{\xi}_{ik}(s_{ij}) - \xi_{ik}(s_{ij})| = O(b_S^2 + [\log n / (nb_S)]^{1/2}) \quad a.s., \quad (19)$$

$$\sup_{s_1, s_2 \in \mathcal{S}} |\hat{R}_k(s_1, s_2) - R_k(s_1, s_2)| = O(\tilde{b}_k^2 + [\log n / (n\tilde{b}_k^2)]^{1/2}) \quad a.s., \quad (20)$$

$$\sup_{s \in \mathcal{S}} |\hat{\psi}_{kp}(s) - \psi_{kp}(s)| = O(\tilde{b}_k^2 + [\log n / (n\tilde{b}_k)]^{1/2}) \quad a.s., \quad (21)$$

$$\|\hat{\psi}_{kp}(s) - \psi_{kp}(s)\|_{L_2} = O(\tilde{b}_k^2 + [\log n / (n\tilde{b}_k)]^{1/2}) \quad a.s. \quad (22)$$

Finally, the overall representation $\hat{X}_i(t|s)$ in (10), with all component estimates plugged in, is shown to converge to $X_i(t|s)$ in Corollaries 1 (dense designs) and 2 (sparse designs).

Corollary 1. (*Scenario A and B*) *Under the assumptions of Theorem 3,*

(a) *for fixed k and p ,*

$$|\hat{\zeta}_{ikp} - \zeta_{ikp}| \xrightarrow{P} 0,$$

(b) *for all $s \in \mathcal{S}$ and $t \in \mathcal{T}$,*

$$|\hat{X}_i(t|s) - X_i(t|s)| \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty, \quad K(n), P(n) \rightarrow \infty,$$

where $\hat{X}_i(t|s)$ is given in (10).

Corollary 2. (*Scenario C*) *Under the assumptions of Theorem 4,*

(a) *for fixed k and p ,*

$$|\hat{\zeta}_{ikp} - \tilde{\zeta}_{ikp}| \xrightarrow{P} 0,$$

(b) for all $s \in \mathcal{S}$ and $t \in \mathcal{T}$,

$$|\hat{X}_i(t|s) - \tilde{X}_i(t|s)| \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty, \quad K(n), P(n) \rightarrow \infty,$$

where $\tilde{\zeta}_{ikp} = E(\zeta_{ikp} | \bar{\xi}_{ik})$, with $\bar{\xi}_{ik}$ defined as $(\xi_{ik}(s_{i1}), \dots, \xi_{ik}(s_{i,m_i}))$, and $\tilde{X}_i(t, s)$ is defined by substituting ζ_{ikp} with $\tilde{\zeta}_{ikp}$ in (5).

We note that the assumption (B.3) of non-crossing eigenvalue functions (see first paragraph of section 3) is used for theorem 4 and the corollaries. This assumption can be relaxed, as discussed in the remark in the Appendix. The proofs of the theoretical results are in the online Supplement A.

5. SIMULATION RESULTS

We conducted two simulation studies, one to investigate the scenario where eigenvalues cross along s , with results reported in the online Supplement D, and another one to evaluate the performance of the proposed double FPCA method for varying sample sizes n in a scenario that resembles a real data application, with results reported below.

For this simulation, we generate data that closely resemble the country mortality data, described in Section 6, by using the estimated mean function $\hat{\mu}(t|s)$ from the data application as mean function and the first four estimated two-dimensional functions as base functions $\varphi_{kp}(t, s)$, $k, p = 1, 2$, in (5). The random scores ζ_{kp} , $k, p = 1, 2$, are generated as independent normal random variables with variances $\{9, 4, 4, 1\}$ for $\{\zeta_{11}, \zeta_{12}, \zeta_{21}, \zeta_{22}\}$, respectively. The error variance was chosen as $\sigma = 0.001$, for a signal-to-noise ratio close to the observed value for the morality data.

Estimated and true principal surfaces φ_{kp} , obtained for one sample run with $n = 100$ are shown in Figure 1. Applying the FVE method as described in Section 3 with threshold 0.95 led to $K = 2$ and $P_1 = P_2 = 2$ as selected number of components. These data-based choices match the true number of components. Figure 1 demonstrates nearly perfect recovery of the true basis functions $\varphi_{kp}(t|s)$, $k, p = 1, 2$.

To quantify the quality of the estimates of $\mu(t|s)$, we use the relative squared error

$$RSE = \frac{\|\mu(t|s) - \hat{\mu}(t|s)\|^2}{\|\mu(t|s)\|^2}, \quad (23)$$

where $\|\mu(t|s)\|^2 = \int \int \mu(t|s)^2 ds dt$, analogously for $\hat{X}_i(t|s)$ and $\hat{\varphi}_{kp}(t|s)$. The average relative squared errors over 100 simulation runs, reported in Table 1, are found to be quite small for μ , X_i and φ_{kp} for $n = 25$, and to decline with increasing sample size n .

6. APPLICATION TO MORTALITY DATA

The analysis of trends in human mortality is important to assess the future demographic prospects of societies, quantify differences between countries with regard to this overall public health measure, and to appraise biological limits of longevity (Oeppen and Vaupel 2002; Vaupel et al. 1998). Various approaches have been designed for the modeling of mortality data as derived from period lifetables (Lee and Carter 1992; Currie et al. 2004), including functional approaches (Hyndman and Ullah 2005; Chiou and Müller 2009) that use the assumption of i.i.d. functional data. To study mortality across countries and calendar time, we applied the proposed double FPCA method to period life tables for 32 countries (listed in the online Supplement C), with a lifetable and corresponding rates of mortality available for each of the calendar years from 1960 to 2006. The data were obtained from the Human Mortality Database (www.mortality.org; Wilmoth et al. 2007).

Whereas cohort life tables reflect the life histories of specific groups of individuals, usually of birth cohorts, period life tables represent the mortality conditions at a specific moment in time (Yang et al. 2008) and are available up to current years. We assume that the observed period death rates for a given calendar year correspond to realizations of an underlying random process with age as argument, quantifying mortality as a function of age of subjects and laying the foundation for the study of country-specific changes of these functional relationships, as calendar time progresses from past to present.

Following the notation introduced in Section 2, $X_i(t|s)$ denotes the mortality rate for the i th country for subjects at age t for calendar year s , where $60 \leq t \leq 100$, focusing on the death rates of older individuals, and on a recent block of 47 years, $1960 \leq s \leq 2006$. Applying the proposed methodology to the entire data set, we obtained the random scores corresponding to principal surface components ζ_{ikp} in (5) for $n = 32$ countries. Visualizing the first two random scores ζ_{i11} and ζ_{i12} , contributing 57.2% and 35.4% to the total variance respectively, clearly indicates that there are two clusters of countries (Figure 2); five countries, namely Belarus, Bulgaria, Russia, Ukraine and Lithuania, form one cluster and the remaining countries the second cluster. The Eastern European countries that are grouped together in the small cluster display an increase in mortality rates starting at around 1980, across all ages, which is most pronounced at older ages, see Figure 3. All other countries share a declining trend of mortality rates with increasing calendar years.

For further analysis, we focus on the majority group of 27 countries. Fitted eigenfunctions $\hat{\phi}_k(t|s)$ and estimated FPCs $\hat{\xi}_{ik}(s)$ in (10) resulting from the first FPCA step are displayed in Figure 4, while the fitted mean function $\hat{\mu}(t|s)$ and the most important three principal surfaces $\hat{\varphi}_{kp}(t|s)$, for $(k, p) = (1, 1), (1, 2), (2, 1)$ in (10) resulting from the second FPCA step are visualized in Figure 5. The numbers of included components $K = 2$, $P_1 = 2$, $P_2 = 3$ were chosen by the FVE criterion with threshold 0.9; see Section 3.

The first step FPCA, as visualized in Figure 4, reveals several interesting features. As a family of one-dimensional eigenfunctions, the $\phi_k(t|s)$ provide the “modes of variation” of the repeated functions sampled at time s . Specifically, the family of first eigenfunctions $\phi_1(t|s)$ indicates the direction of largest variation in function space of the mortality functions $X_i(t|s)$ around their mean. These eigenfunctions and also the second eigenfunctions $\phi_2(t|s)$ are seen to be quite homogeneous across s , implying that the “modes of variation” do not change much for varying calendar years. The corresponding FPC scores as functions of calendar year $\xi_{i1}(s)$ display less homogeneity, indicating an increase in variance

from around $s = 1980$ on, followed by a subsequent decline. As the $\xi_{i1}(s)$ are multiplied with the $\phi_1(t|s)$, this points to a temporarily increased divergence along the main direction of variation during this calendar period, likely a consequence of societal changes in Europe. The family of second eigenfunctions $\phi_2(t|s)$ characterizes variation of mortality along a contrast between old (with a peak around 85 years) and oldest-old (post-90) mortalities (negative values). As the functions $\xi_{i2}(s)$ diverge in more recent years, this implies that the differences in the contrast between earlier and later mortalities across countries are rising, indicating increasing variability of the old versus oldest-old mortality differential. Further information on the eigenanalysis of the functions $\xi_{i1}(s)$ and $\xi_{i2}(s)$, including their eigenfunctions ψ_{kp} , is in online Supplement E.

The fitted mean surface in the upper left panel of Figure 5 reflects the well-known overall declining trend in mortality rates. Further scrutiny shows that for ages between 80 and 90, the age at which a given mortality rate level is reached increases by approximately one year per decade. The three principal surfaces depicted in Figure 5 provide a graphical overall decomposition of the variation of mortality curves across countries.

The first principal surface $\varphi_{11}(t|s)$, corresponding to the first principal component in both age coordinate s and calendar year coordinate t , explains most of the variation between countries. A sharp change is seen to occur at around age 85, after which individual country's mortality rates deviate from the mean surface with a declining or increasing rate. In the calendar year direction, a prominent “U- shape” feature is present at the oldest ages. To investigate this feature, we plot the fitted surfaces for the two countries with the largest and the smallest value of the multiplicative random factor $\hat{\zeta}_{11}$ (see (5)), which turn out to be Canada and the Czech Republic (Figure 6). One finds that the mortalities for these two countries exhibit a “U notch” and reversed “U-notch” for the oldest age mortality rates, viewed against calendar time. This indicates that oldest-old mortality rates across countries temporarily diverged in the 1980-1990s.

The second principal surface $\hat{\varphi}_{12}(t|s)$ is flat at lower age groups 60-70 and for older groups shows an increasingly prominent slope, providing a contrast between present and past calendar years. This component thus differentiates between countries according to the speed of decline of oldest old mortality. This speed of decline is an indicator of health of the oldest cohorts, as well as an important predictor for the size of elderly cohorts and the societal resources that will be needed to care for the oldest-old cohorts.

The third principal surface $\hat{\varphi}_{21}(t|s)$ complements the second principal surface: It has least curvature at past calendar years and increasing concave curvature towards more recent calendar years. This represents an increasing contrast between oldest-old mortality and mortality around ages 80-85, differentiating effects of mortality reduction before 85 and after 90. This may indicate that increasing mortality reduction at older ages before age 85 incurs a cost by being accompanied with decreased mortality reduction after age 90. One interpretation is that this reflects a selection effect: If mortality at ages prior to 85 is lowered, then larger numbers of frail individuals survive beyond this age range, and as a consequence, the decline of mortality of the oldest-old group is relatively smaller.

7. DISCUSSION

As an alternative to the proposed two-step FPCA, we also implemented and fitted the two-dimensional Karhunen-Loève expansion

$$X(t, s) = \mu(t, s) + \sum_{k=1}^{\infty} \eta_k \rho_k(t, s), \quad (24)$$

where $\rho_k(t, s)$, $k \geq 1$, are the eigenfunctions of the associated autocovariance operator A_G with kernel $G(t_1, s_1, t_2, s_2) = \text{cov}(X(t_1, s_1), X(t_2, s_2))$, i.e., $G(t_1, s_1, t_2, s_2)$ has an orthogonal expansion $G(t_1, s_1, t_2, s_2) = \sum_k \omega_k \rho_k(t_1, s_1) \rho_k(t_2, s_2)$ with nonincreasing eigenvalues ω_k and eigenfunctions $\rho_k(t, s)$, where $\eta_k = \int_{\mathcal{T}, \mathcal{S}} (X(t, s) - \mu(t, s)) \rho_k(t, s) dt ds$ are the functional principal components (FPCs). In practice, this representation can be implemented by (i) obtaining estimates of $G(t_1, s_1, t_2, s_2)$ on a fine regular grid of, say p^4 points; (ii)

defining an (arbitrary) ordering of the grid points (t_1, s_1) according to which they are vectorized, and vectorizing grid points (t_2, s_2) using the same ordering; (iii) using the resulting vectors to re-arrange the four-dimensional array into a symmetric $p^2 \times p^2$ matrix and performing a regular eigen-decomposition for this matrix; (iv) arranging the k th eigenvector of dimension p^2 back into a $p \times p$ matrix, leading to an estimate of the eigenfunction $\rho_k(t, s)$ on the same $p \times p$ grid.

For data such as the mortality data that are regular and relatively dense, the empirical or cross-sectional sample covariance is a feasible estimate for the four-dimensional covariance $G(t_1, s_1, t_2, s_2) = \text{cov}(X(t_1, s_1), X(t_2, s_2))$. Alternatively, one can estimate the covariance $G(t_1, s_1, t_2, s_2)$ by four-dimensional local linear smoothing. We found that in the dense regular case, these two approaches give nearly identical results. Not surprisingly, the four-dimensional smoothing method is much slower; more details on this can be found in the online Supplement B. We emphasize that we consider here a scenario where t and s have inherently different meanings and scales. Accordingly, the eigenfunctions ϕ_k and principal components $\xi_k(s)$ in the direction of repeated time, as obtained from the first step of the proposed two-step FPCA approach, are of interest in themselves, as seen in the simulation and data analysis. We note that these quantities are not available when using the two-dimensional Karhunen-Loève representation.

Even if we ignore the asymmetry of the roles played by t and s , we find that the proposed two-step FPCA method computationally is much faster than the two-dimensional Karhunen-Loève representation as in (24). Especially when the data are not regular and dense in either t or s , a four-dimensional smoothing step for the covariance $G(t_1, s_1, t_2, s_2)$ is mandatory to implement the Karhunen-Loève approach. This then renders it computationally inefficient, relative to the proposed two-step FPCA procedure, which at most requires three-dimensional smoothing and thus achieves faster rates of convergence under the same smoothness assumptions. In an illustration provided in the online Supplement B,

the proposed method worked well for sparse designs, while the Karhunen-Loève approach was not only very time consuming but also became unstable under sparsity.

The double FPCA approach for the analysis of repeated functions takes into account the asymmetry of the longitudinal and the functional time components and is a natural extension of the standard FPCA technique of functional data analysis. It can be implemented for both dense designs as well as sparse longitudinal designs for the times where the repeated functional measurements are taken. For both cases, the proposed methodology enjoys good asymptotic properties. In applications to cross-country comparisons of older age mortality, the proposed double FPCA method leads to interesting insights into mortality and longevity across countries, by decomposing the overall variability into a series of principal surfaces. The proposed method thus shows promise for the analysis of longitudinal studies, in which one observes a curve at each observation time.

Appendix: Assumptions and Notations

Let $m_n = \frac{1}{n} \sum_{i=1}^n m_i$, $L_n = \min_{i,j} \{L_{ij}\}$ and h_T, h_S denote the smoothing bandwidths to estimate $\mu(s, t)$, and b_T, b_S the smoothing bandwidths for $G(t_1, t_2|s)$ when smoothing is used. For the second FPCA step, let \tilde{b}_k denote the smoothing bandwidth for $R_k(s_1, s_2)$, where $R_k(s_1, s_2) = \text{cov}(\xi_k(s_1), \xi_k(s_2))$, $k \geq 1$.

In the following, $0 < B < C < \infty$ are generic constants that can take different values at different places. The following mild moment conditions will be used:

$$(L.1.a) \text{ For all } i, \text{ and all } s, t, E|X_i(t|s)|^\ell \leq \frac{\ell!}{2} C^{\ell-2} B^2, \ell = 2, 3, \dots$$

$$(L.1.b) \text{ For all } i, \text{ and all } s, t_1, t_2, E|X_i(t_1|s)X_i(t_2|s)|^\ell \leq \frac{\ell!}{2} C^{\ell-2} B^2, \ell = 2, 3, \dots$$

Here, (L.1.a) and (L.1.b) are needed for (??) below, which will be used in the proof of Theorem 1.

$$(L.2) \text{ Processes } X_i(t|s) \text{ are smooth in } t \text{ and } s \text{ and satisfy } |X_i(t|s) - X_i(t'|s')| \leq C(|t - t'| + |s - s'|).$$

(L.3) $E(\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |X_i(t|s)|) < \infty$, $\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |\mu(t|s)| < \infty$, $\sup_{t_1, t_2, s} |G(t_1, t_2|s)| < \infty$, $\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |\phi_k(t|s)| < \infty$ for each $k \geq 1$.

The following Conditions (A.1) - (A.6) are needed for the first FPCA step for functions $X_i(\cdot|s)$, and (B.1) - (B.3) for the second FPCA step for functions $\{\hat{\xi}_{ik}(s)\}$.

(A.1) t_{ijl} and s_{ij} are random observation points distributed with density $f_{\mathcal{T}}$, respectively, $f_{\mathcal{S}}$. We require $B \leq f_{\mathcal{T}}(t) \leq C$ for all $t \in \mathcal{T}$, and $B \leq f_{\mathcal{S}}(s) \leq C$ for all $s \in \mathcal{S}$. Further, $f_{\mathcal{T}}$ and $f_{\mathcal{S}}$ are differentiable with bounded derivatives.

(A.2) The kernel function K is a symmetric probability density function on $[-1, 1]$ with bounded variation $[-1, 1]$. Write $\nu_2 = \int_{[-1, 1]} t^2 K(t) dt$.

(A.3) All second-order partial derivatives of $\mu(t|s)$ exist and are uniformly bounded.

(A.4) All second-order partial derivatives of $G(t_1, t_2|s)$ exist and are uniformly bounded on $\mathcal{T}^2 \times \mathcal{S}$.

(A.5) $E(|\epsilon_{ijl}|^{\gamma_\mu}) < \infty$, $E(\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |X_i(t|s)|^{\gamma_\mu}) < \infty$ for some $\gamma_\mu > 2$; $h_s, h_t \rightarrow 0$ and $[(h_s^2 + h_s/m_n)(h_t^2 + h_t/L)]^{-1}(\log n/n^r)^{1-2/\gamma_\mu} \rightarrow \infty$, as $n \rightarrow \infty$, for some $r \in (0, 1]$.

(A.6) $E(|\epsilon_{ijl}|^{2\gamma_G}) < \infty$, $E(\sup_{t \in \mathcal{T}, s \in \mathcal{S}} |X_i(t|s)|^{2\gamma_G}) < \infty$ for some $\gamma_G > 2$; $b_s, b_t \rightarrow 0$ and $[(b_t^4 + b_t^3/L + b_t^2/L^2)(b_s^2 + b_s/m_n)]^{-1}(\log n/n^r)^{1-2/\gamma_G} \rightarrow \infty$ as $n \rightarrow \infty$, for some $r \in (0, 1]$.

Conditions similar to the above have been required in Hall et al. (2006) and Li and Hsing (2010). Here we assume that these properties hold uniformly over s . This is necessary so that the empirical working targets $\hat{\xi}_{ik}(s_{ij})$ are uniformly close to the theoretical working processes $\xi_{ik}(s)$, a fact that will be used in the second step FPCA.

(B.1) For any fixed k , all second-order partial derivatives of $R_k(s_1, s_2)$ exist and are bounded on \mathcal{S}^2 .

(B.2) For any fixed k , $E(\sup_{s \in \mathcal{S}} |\xi_{ik}(s)|^{\gamma_R}) < \infty$ for some $\gamma_R > 2$; $\tilde{b}_k \rightarrow 0$ and $(\tilde{b}_k^4 + \tilde{b}_k^3/m_n + \tilde{b}_k^2/(m_n^2))^{-1}(\log n/n)^{1-2/\gamma_R} \rightarrow \infty$ as $n \rightarrow \infty$.

Recall that $\lambda_k(s)$ is the k th largest eigenvalue of the covariance operator $G(\cdot, \cdot|s)$ at a fixed $s = s_0$, but is defined by continuous continuation of the functions $\phi_k(t|s)$ for other

s , as described at the beginning of section 3. Defining

$$\delta_k(s) = \min_{1 \leq j \leq k} (\lambda_j(s) - \lambda_{j+1}(s)), \quad \delta_k = \inf_{s \in \mathcal{S}} \{\delta_k(s)\}, \quad \lambda_k = \inf_{s \in \mathcal{S}} \{\lambda_k(s)\}, \quad (25)$$

we make the assumption

(B.3) For all k , $\delta_k > 0$ and $\lambda_k > 0$.

This means that the eigenvalue functions $\lambda_j(s)$, $1 \leq j \leq \infty$, do not cross and there is always a gap between $\lambda_j(s)$ and $\lambda_k(s)$ for any $j \neq k$.

Remark. The assumption (B.3) of non-crossing eigenvalue functions in s can be relaxed, as follows. If the eigenvalue functions cross N times, at times s_1^c, \dots, s_N^c , we define the set $D_\epsilon = \cup_{i=1}^N \{s \in (s_i^c - \epsilon, s_i^c + \epsilon)\}$ for an arbitrarily small $\epsilon > 0$. Then on the set $\mathcal{S} \cap D_\epsilon^c$ all theoretical results still hold. One can derive results for the case where $\epsilon = \epsilon_n \rightarrow 0$, so that the measure of the exception set D_ϵ converges to zero, but for this scenario the rates of convergence are slower than those reported in the theorems.

REFERENCES

- Ash, R. B. and Gardner, M. F. (1975), *Topics in Stochastic Processes*, New York: Academic Press.
- Bosq, D. (2000), *Linear Processes in Function Spaces: Theory and Applications*, New York: Springer-Verlag.
- Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986), “Principal Modes of Variation for Processes with Continuous Sample Curves,” *Technometrics*, 28, 329–337.
- Chiou, J.-M. and Müller, H.-G. (2009), “Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting,” *Journal of the American Statistical Association*, 104, 572–585.
- Crainiceanu, C. M., Staicu, A.-M., and Di, C.-Z. (2009), “Generalized Multilevel Functional Regression.” *Journal of the American Statistical Association*, 104, 1550–1561.

- Currie, I. D., Durban, M., and Eilers, P. H. C. (2004), “Smoothing and Forecasting Mortality Rates,” *Statistical Modelling*, 4, 279–298.
- Del Barrio, E., Deheuvels, P., and Geer, S. (2007), *Lectures on empirical processes: theory and statistical applications.*, European Mathematical Society.
- Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009), “Multilevel functional principal component models.” *Annals of Applied Statistics*, 3, 458–488.
- van de Geer, S. A. (2006), *Empirical Process Theory and Applications*, handout WS 2006, ETH Zürich.
- Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010), “Longitudinal functional principal component analysis,” *Electronic Journal of Statistics*, 4, 1022–1054.
- Hall, P., Müller, H.-G., and Wang, J.-L. (2006), “Properties of principal component methods for functional and longitudinal data analysis,” *Annals of Statistics*, 34, 1493–1517.
- Human Mortality Database*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data download on March 1, 2011).
- Hyndman, R. J. and Ullah, M. S. (2005), “Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach,” *Computational Statistics & Data Analysis*, 51, 4942–4956, unpublished manuscript.
- Lee, R. D. and Carter, L. R. (1992), “Modeling and Forecasting U.S. Mortality (Disc: P671-675),” *Journal of the American Statistical Association*, 87, 659–671.
- Li, Y. and Hsing, T. (2010), “Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data,” *Annals of Statistics*, 38, 3321–3351.
- Morris, J. S. and Carroll, R. J. (2006), “Wavelet-based Functional Mixed Models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 179–199.
- Morris, J. S., Vannucci, M., Brown, P. J., and Carroll, R. J. (2003), “Wavelet-based

- nonparametric modeling of hierarchical functions in colon carcinogenesis,” *Journal of the American Statistical Association*, 98, 573–597.
- Müller, H.-G. (2005), “Functional modelling and classification of longitudinal data,” *Scandinavian Journal of Statistics. Theory and Applications*, 32, 223–240.
- Oeppen, J. and Vaupel, J. W. (2002), “Broken Limits to Life Expectancy,” *Science*, 296, 1029–1031.
- Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer Series in Statistics, New York: Springer, 2nd ed.
- Rice, J. A. and Silverman, B. W. (1991), “Estimating the mean and covariance structure nonparametrically when the data are curves,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53, 233–243.
- Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., Longo, V. D., Zeng, Y., Manton, K. G., and Curtsinger, J. W. (1998), “Biodemographic Trajectories of Longevity,” *Science*, 280, 855–860.
- Wilmoth, J. R., Andreev, K., Jdanov, D., and Gleijeses, D. A. (2007), *Methods Protocol for the Human Mortality Database*, with the assistance of C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, P. Vachon (last modified May 31 2007).
- Yang, Y., Schulhofer-Wohl, S., Fu, W. J., and Land, K. C. (2008), “The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It,” *American Journal of Sociology*, 113, 1697–1736.
- Yao, F. and Lee, T. C. M. (2006), “Penalized Spline Models for Functional Principal Component Analysis,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 3–25.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005), “Functional data analysis for sparse longitudinal data,” *Journal of the American Statistical Association*, 100, 577–590.

RE	$n = 25$	$n = 50$	$n = 100$
μ	0.0126 (0.0113)	0.0057 (0.0047)	0.0031 (0.0028)
X	0.0056 (0.0054)	0.0053 (0.0046)	0.0043 (0.0045)
φ_{11}	0.0988 (0.1506)	0.0387 (0.0302)	0.0231 (0.0185)
φ_{12}	0.1403 (0.1894)	0.0673 (0.0588)	0.0431 (0.0443)
φ_{21}	0.1039 (0.1075)	0.0689 (0.0608)	0.0407 (0.0359)
φ_{22}	0.1353 (0.1433)	0.0901 (0.0812)	0.0555 (0.0547)

Table 1: Simulation results for data generated to mimic the mortality data of Section 6, described in Section 5, reporting relative error (RE) as defined in (23) (with standard deviations in parentheses), for various components of model (5) for $k, p = 1, 2$ and varying sample sizes n .

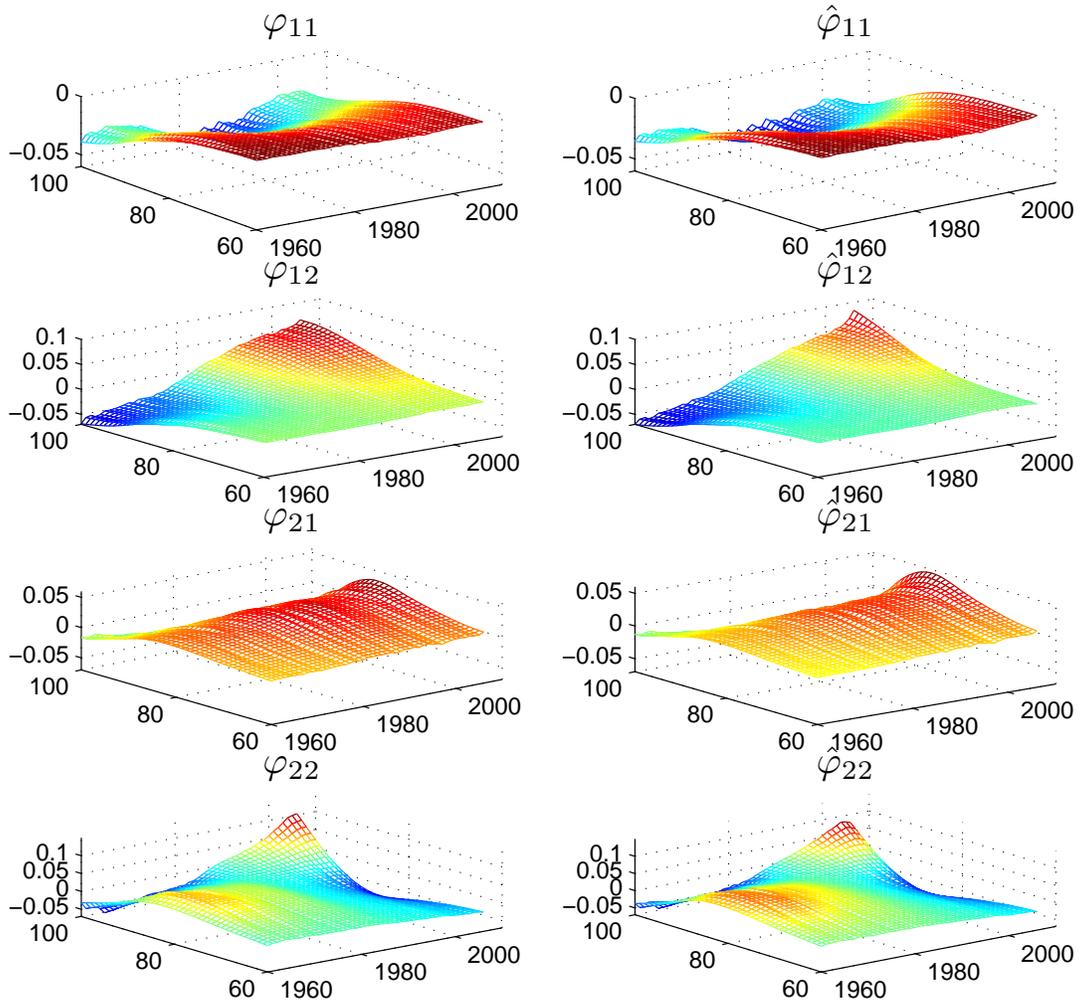


Figure 1: True (left) and estimated (right) principal surfaces φ_{kp} for $k = 1, 2$ and $p = 1, 2$, as obtained in one simulation run.

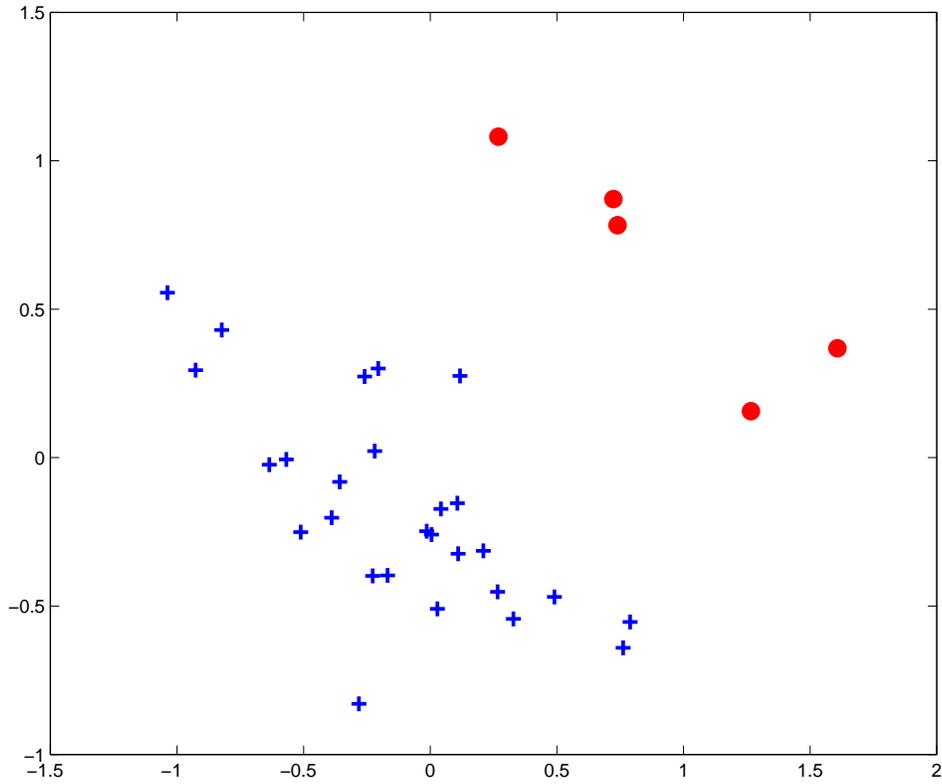


Figure 2: Plotting random scores ζ_{i12} against ζ_{i11} to summarize the mortality trajectories for $n = 32$ countries. The five countries in the upper right (marked by solid circles, corresponding to the Eastern European countries of Belarus, Bulgaria, Lithuania, Russia and Ukraine) exhibit quite different features of mortality evolution from the bulk of the countries (crosses).

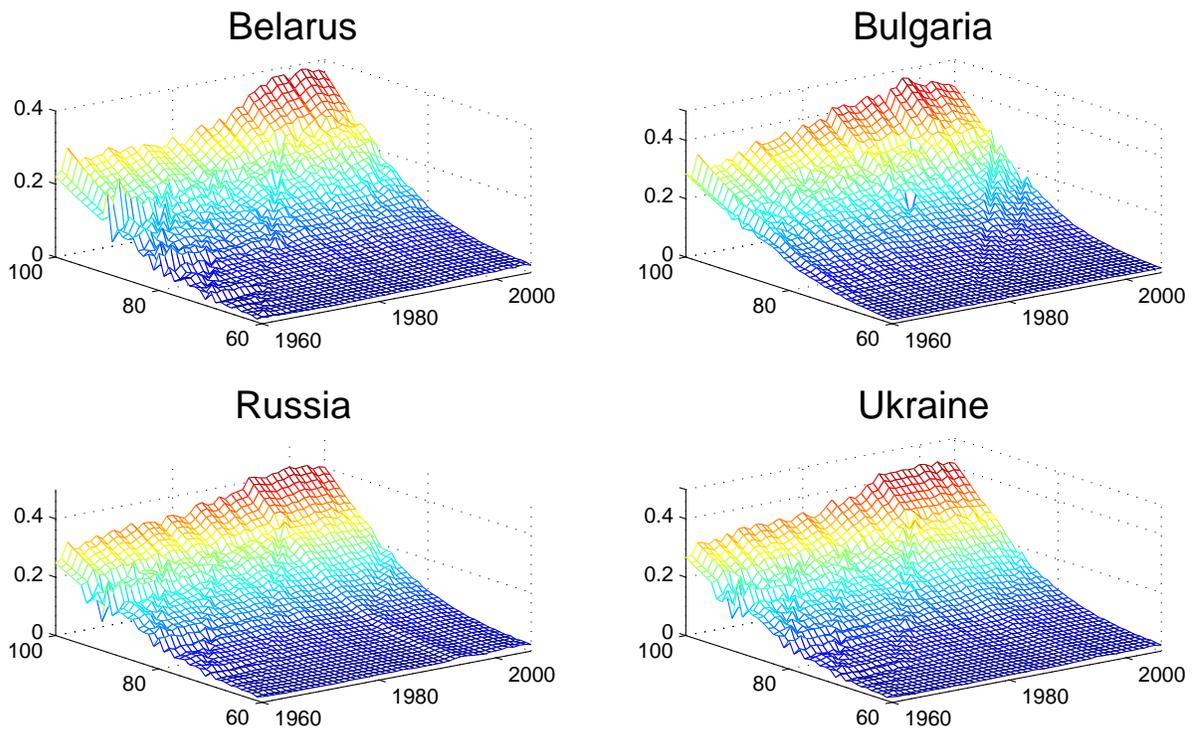


Figure 3: The raw mortalities for four countries that were identified to have increasing mortality rates after 1980, in contrast to the bulk of other countries.

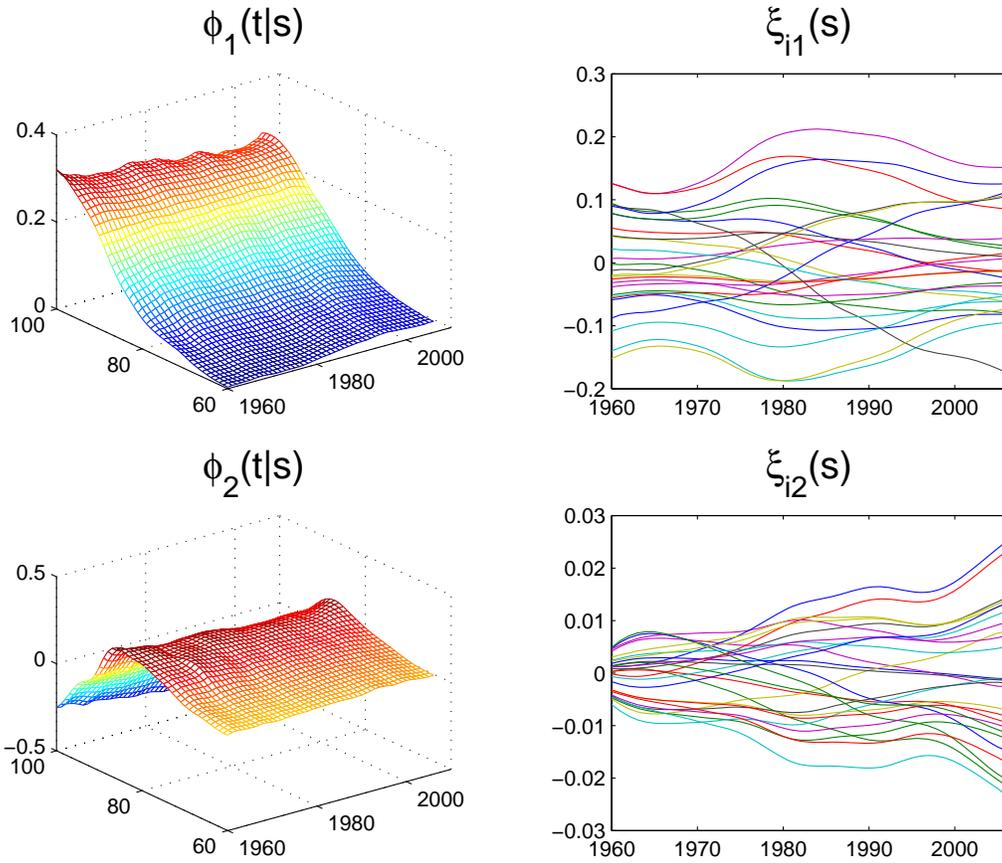


Figure 4: The estimated eigenfunctions and FPC scores from the first step FPCA, $\hat{\phi}_1(t|s)$, $\hat{\phi}_2(t|s)$, $\hat{\xi}_{i1}(s)$ and $\hat{\xi}_{i2}(s)$ in (10) for the mortality data.

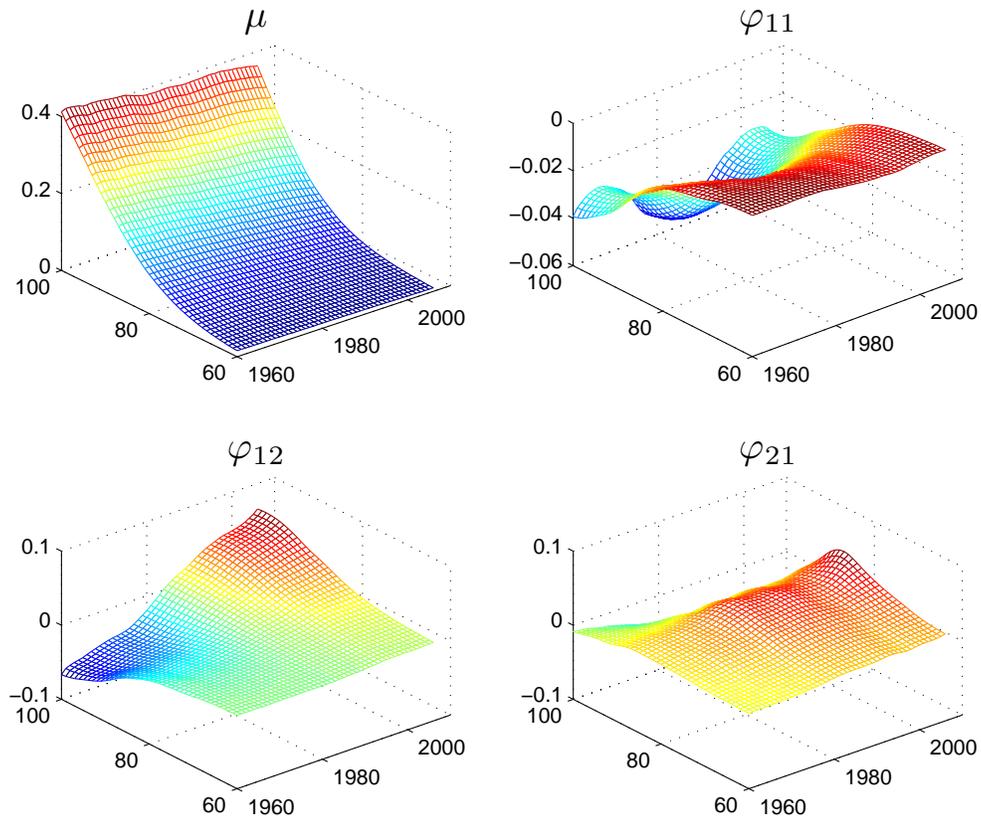


Figure 5: The estimated mean and first three principal surfaces $\hat{\varphi}_{11}$, $\hat{\varphi}_{12}$, $\hat{\varphi}_{21}$ in (10), explaining 57.2%, 35.4% and 5% of the variation, respectively, for the mortality data.

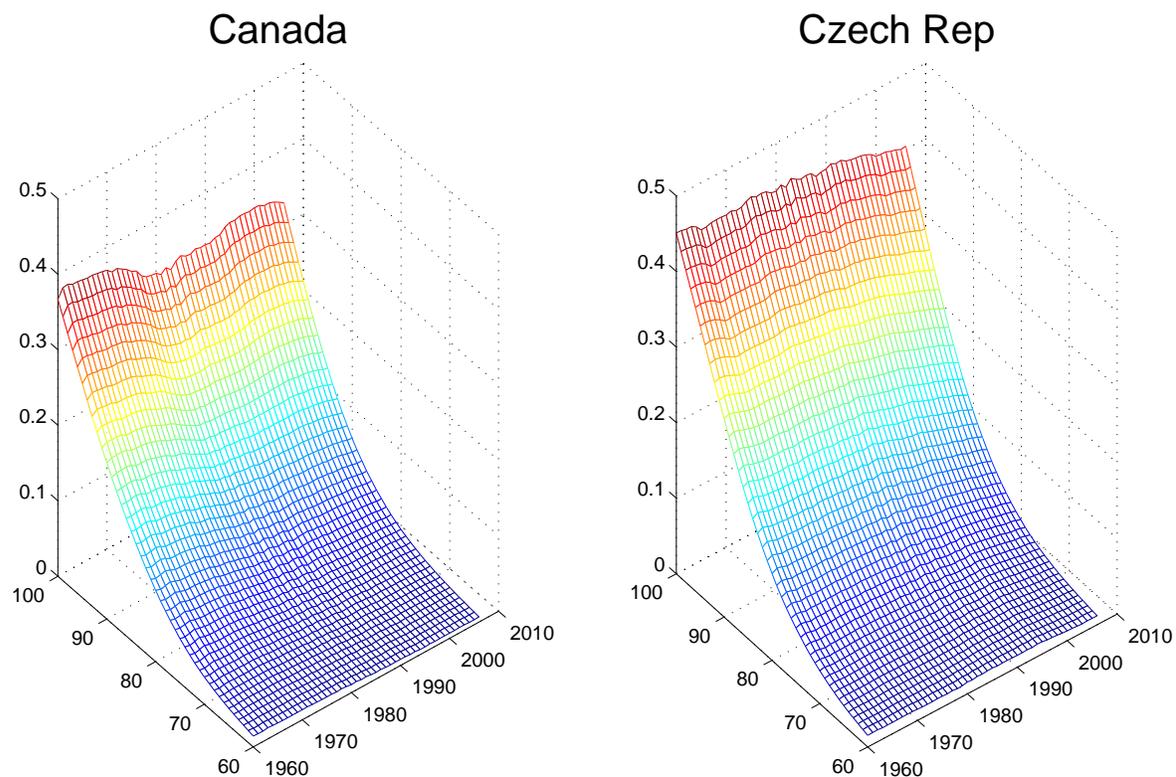


Figure 6: Fitted mortality rates for Canada with the most positive random scores ζ_{i11} (left) and Czech Republic with the most negative scores ζ_{i11} (right) in model (5). The trend over calendar years at older ages shows a “U notch” for Canada (left) and an inverse “U notch” for the Czech Republic (right). This effect is a consequence of the “U shape” feature found in the first principal surface φ_{11} in Figure 5 and demonstrates a contrast in old-age mortality between these two countries.