Modeling Repeated Functional Observations

Kehui Chen
Department of Statistics, University of Pittsburgh,
Hans-Georg Müller
Department of Statistics, University of California, Davis

Supplemental Material

Supplement A: Auxiliary Results and Proofs

We first state two auxiliary results that are useful for the theoretical arguments. One of these is a Bernstein-type concentration inequality (Del Barrio et al. 2007; Geer 2006): Let $X_1, \ldots, X_n$ be independent real-valued random variables with expectation zero. Suppose that for all i,

$$E|X_i|^\ell \le \frac{\ell!}{2} C^{\ell-2} B^2, \ \ell = 2, 3, \ldots$$

Then for any $a > 0$,

$$P(\sum_{i=1}^{n} X_i \ge a) \le \exp[-\frac{a^2}{2(aC + nB^2)}]. \tag{26}$$

A second auxiliary result relates uniform convergence of eigenfunctions to the uniform convergence of corresponding covariance functions.

Lemma 1. *Under assumption L.3, if* $\sup_{t_1,t_2\in\mathcal{T},s\in\mathcal{S}} |\hat{G}(t_1,t_2|s) - G(t_1,t_2|s)| = O(a_n), \quad a.s.,$ *then*

$$\sup_{t\in\mathcal{T},s\in\mathcal{S}} |\hat{\phi}_k(t|s) - \phi_k(t|s)| = O(a_n), \quad a.s.$$

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 $R_k(s_1,s_2) = \mathrm{cov}(\xi_k(s_1),\xi_k(s_2))$, $k \ge 1$.

In the following proofs, we do not consider measurement errors for dense designs, as then pre-smoothing can be used to construct continuously observed functions with greatly reduced errors.

Proof of Theorem 1. Define $\mathcal{G} = (t_l, s_j)$ for $1 \le l \le L$ and $1 \le j \le m$. Under the assumptions $\max_l(t_l - t_{l-1}) = O(n^{-1})$ and $\max_j(s_j - s_{j-1}) = O(n^{-1})$, we only need to consider the case $L = n$ and $m = n$.

Since $\mathcal{T} \times \mathcal{S}$ is compact, and $X_i(t|s)$ is Lipschitz continuous by (L.2), and the values between grid points are obtained by linear interpolation, we have

$$\sup_{t\in\mathcal{T},s\in\mathcal{S}} |\hat{\mu}(t|s) - \mu(t|s)| = \sup_{(t,s)\in G} |\hat{\mu}(t|s) - \mu(t|s)| + O(\frac{1}{n}).$$

Now considering a fixed $(t, s) \in G$, using condition (L.1.a), by (26), for any $a > 0$,

$$P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i(t|s) - \mu(t|s)) \geq (\frac{\log n}{n})^{1/2}a\right) \leq \exp\left(-\frac{n\log n a^2}{2((n\log n)^{1/2}aC + nB^2)}\right) = n^{-B^*},$$

where $B^* = a^2/[2((\frac{\log n}{n})^{1/2}aC + B^2)]$. Thus,

$$P(\sup_{(t,s)\in G} |\hat{\mu}(t|s) - \mu(t|s)| \geq (\frac{\log n}{n})^{1/2}a) \leq 2n^2 \times n^{-B^*} = 2n^{2-B^*}. \tag{27}$$

We can find large enough a such that $B^* > 2$ for large n. Then,

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

Defining $Z_i(t_1, t_2|s) = X_i(t_1|s)X_i(t_2|s)$ on $\mathcal{T}^2 \times \mathcal{S}$, and using condition (L.1.b), a similar argument gives the result for $\hat{G}(t_1, t_2|s)$, which completes the proof.

Proof of Theorem 2. Write

$$\delta_{n1} = \{[1 + (h_T L_n)^{-1} + (h_S m_n)^{-1} + (h_S m_n h_T L_n)^{-1}]\log n/n\}^{1/2},$$

$$\delta_{n2} = \{[1+(b_T L_n)^{-1}+(b_T L_n)^{-2}+(b_S m_n)^{-1}+(b_S m_n b_T L_n)^{-1}+(b_S m_n b_T^2 L_n^2)^{-1}]\log n/n\}^{1/2}.$$

We note that the local linear smoothing estimators defined in (8) and (9) are slightly different from the one used in Li and Hsing (2010), and modify m_n accordingly. The strong law of large numbers for weighted samples is needed to obtain the same results as those in Lemma 1 in Li and Hsing (2010) (details for this step are provided in a working paper of Zhang and Wang (2012)). Minor modifications of the proofs of Theorem 3.1 and Theorem 3.3, with correspondingly modified conditions (A.1) - (A.6), yield

$$\begin{aligned}
\sup_{t\in\mathcal{T},s\in\mathcal{S}} |\hat{\mu}(t|s) - \mu(t|s)| &= O(h_T^2 + h_S^2 + h_S h_T + \delta_{n1}) \quad a.s. \\
\sup_{t_1,t_2,s} |\hat{G}(t_1, t_2|s) - G(t_1, t_2|s)| &= O(h_T^2 + h_S^2 + h_S h_T + \delta_{n1} + b_T^2 + b_S^2 + b_T b_S + \delta_{n2}) \quad a.s.
\end{aligned}$$

Theorem 2 (a) and (b) now follow by using the respective conditions.

Proof of Lemma 1. Define $||G(\cdot,\cdot|s)|| = \{\int\int G(t_1,t_2|s)^2 dt_1 dt_2\}^{1/2}$, the Hilbert-Schmidt norm of $G(\cdot,\cdot|s)$ for any s, and $||\psi(\cdot|s)|| = \{\int \psi(t|s)^2 dt\}^{1/2}$. Then one has

$$\sup_{s\in\mathcal{S}} ||\hat{G}(\cdot,\cdot|s) - G(\cdot,\cdot|s)|| = O(a_n) \quad a.s. \tag{28}$$

For each s and for a fixed k, Lemma 4.3 in Bosq (2000) implies that

$$|\hat{\lambda}_k(s) - \lambda_k(s)| \le ||\hat{G}(\cdot,\cdot|s) - G(\cdot,\cdot|s)||, \quad ||\hat{\phi}_k(\cdot|s) - \phi_k(\cdot|s)|| \le 2\sqrt{2}\delta_k^{-1}||\hat{G}(\cdot,\cdot|s) - G(\cdot,\cdot|s)||, \tag{29}$$

where δ_k is defined in (25). Combining (28) and (29) yields

$$\sup_{s\in\mathcal{S}} |\hat{\lambda}_k(s) - \lambda_k(s)| = O(a_n) \quad a.s., \qquad \sup_{s\in\mathcal{S}} ||\hat{\phi}_k(\cdot|s) - \phi_k(\cdot|s)|| = O(a_n) \quad a.s.$$

Note that for any $1 \le k \le K_0$, $\lambda_k(s)\phi_k(t|s) = \int G(t',t|s)\phi_k(t'|s)dt'$, and therefore

$$\begin{aligned}
&\sup_{s\in\mathcal{S},t\in\mathcal{T}} |\lambda_k(s)\hat{\phi}_k(t|s) - \lambda_k(s)\phi_k(t|s)| \\
\le &\sup_{s,t} |\hat{\lambda}_k(s)\hat{\phi}_k(t|s) - \lambda_k(s)\phi_k(t|s)| + \sup_{s,t} |\hat{\lambda}_k(s)\hat{\phi}_k(t|s) - \lambda_k(s)\hat{\phi}_k(t|s)|,
\end{aligned}$$

where the second term is $O(a_n)$ $a.s.$ and the first term is bounded by

$$\begin{aligned}
&\sup_{s,t} |\int (\hat{G}(t',t|s) - G(t',t|s))\hat{\phi}_k(t'|s)dt'| + \sup_{s,t} |\int G(t',t|s)(\hat{\phi}_k(t'|s) - \phi_k(t'|s))dt'| \\
&= O(\sup_{t_1,t_2,s} |\hat{G}(t_1,t_2|s) - G(t_1,t_2|s)| \sup_{s,t} |\hat{\phi}_k(s|t)| + \sup_{t_1,t_2|s} |G(t_1,t_2|s)| \sup_{s} ||\hat{\phi}_k(\cdot|s) - \phi_k(\cdot|s)||) \\
&= O(a_n) \quad a.s.,
\end{aligned} \tag{30}$$

using (L.3). Further noting that $\lambda_k = \inf_s |\lambda_k(s)|$ is a positive constant, Lemma 1 follows.

Proof of Theorem 3. Define

$$\theta_{in} = \big(\sup_{t,s} |X_i(t|s)| + \sup_{t,s} |\mu(t|s)| + \sup_{t,s} |\phi_k(t|s)| + 1\big)(\log n/n)^{1/2}.$$

As $L_n^{-1} = O(n^{-1})$, we can neglect the error in the numerical integration, so may consider $\hat{\xi}_{ik}(s_{ij}) = \int \big(X_i(t|s_{ij}) - \hat{\mu}(t|s_{ij})\big)\hat{\phi}_k(t|s_{ij})dt$. Noting the target is

$$\xi_{ik}(s_{ij}) = \int \big(X_i(t|s_{ij}) - \mu(t|s_{ij})\big)\phi_k(t|s_{ij})dt,$$

one finds, for any k, using (11), (12) and Lemma 1,

$$\sup_{1\le j\le m_i}|\hat{\xi}_{ik}(s_{ij})-\xi_{ik}(s_{ij})| \le \sup_{1\le j\le m_i}|\int (X_i(t|s_{ij})-\mu(t|s_{ij}))(\phi(t|s_{ij})-\hat{\phi}(t|s_{ij}))dt|$$
$$+\sup_{1\le j\le m_i}|\int (\mu(t|s_{ij})-\hat{\mu}(t|s_{ij}))\phi(t|s_{ij})dt|$$
$$+\sup_{1\le j\le m_i}|\int (\hat{\mu}(t|s_{ij})-\mu(t|s_{ij}))(\phi(t|s_{ij})-\hat{\phi}(t|s_{ij}))dt| = O(\theta_{in}) \quad a.s., \tag{31}$$

where $O(\cdot)$ is uniform over i. Then (L.3) and the strong law of large numbers imply (15).

For each k, if the target working processes $\xi_{ik}(s)$ are used, one can easily derive that

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

as in Yao et al. (2005), Hall et al. (2006), and Li and Hsing (2010). Scrutinizing the estimating procedure for $R(s_1,s_2)$, we find that if the empirical working data are used, then

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

where $\theta_n = \frac{1}{n}\sum_1^n \theta_{in}$; see also Yao and Lee (2006). By (15), $\theta_n = O(\log n/n)^{1/2}) \quad a.s.$. This means that the rate of convergence for $\hat{R}_k(s_1,s_2)$ and $\{\hat{\psi}_{kp}(s), p=1,\ldots,P_0\}$ remains the same as for the true targets.

Proof of Theorem 4. For the case of a design that is sparse in s, if we were to observe $X_i(t_{ijl}|s_{ij})$, the proof of part (a) would be the same as that of Theorem 3. However, the additional measurement errors for the sparse case cannot be alleviated by individual curve smoothing. Starting with noisy observations $U_{ijl} = X_i(t_{ijl},s_{ij})+\epsilon_{ijl}$, we decompose

$$\begin{aligned}
\hat{\xi}_{ik}(s_{ij}) &= \tilde{\xi}_{ik}(s_{ij}) + R_{ik}(s_{ij}) + \tilde{\epsilon}_{ijk},\\
\tilde{\xi}_{ik}(s_{ij}) &= \sum_{l=2}^{L_{ij}}\{X_{ij}(t_{ijl}|s_{ij})-\hat{\mu}(t_{ijl}|s_{ij})\}\hat{\phi}_k(t_{ijl}|s_{ij})(t_{ijl}-t_{ij,l-1}),\\
R_{ik}(s_{ij}) &= \sum_{l=2}^{L_{ij}}\epsilon_{ijl}\{\hat{\phi}_k(t_{ijl}|s_{ij})-\phi_k(t_{ijl}|s_{ij})\}(t_{ijl}-t_{ij,l-1}),\\
\tilde{\epsilon}_{ijk} &= \sum_{l=2}^{L_{ij}}\epsilon_{ijl}\phi_k(t_{ijl}|s_{ij})(t_{ijl}-t_{ij,l-1}).
\end{aligned} \tag{32}$$

As $L_n^{-1} = O(n^-1)$, one may neglect the numerical integration error in $\tilde{\xi}_{ik}(s_{ij})$ and consider $\tilde{\xi}_{ik}(s_{ij}) = \int \big(X_i(t|s_{ij})-\hat{\mu}(t|s_{ij})\big)\hat{\phi}_k(t|s_{ij})dt$. For the targets $\xi_{ik}(s_{ij}) = \int \big(X_i(t|s_{ij})-$

$\mu(t|s_{ij})\big)\phi_k(t|s_{ij})dt$, for any $k \geq 1$, using (13), (14) and Lemma 1,

$$\max_{1\leq j\leq m_i} |\tilde{\xi}_{ik}(s_{ij}) - \xi_{ik}(s_{ij})|$$
$$= O(\big(\sup_{t,s} |X_i(t|s)| + \sup_{t,s} |\mu(t|s)| + \sup_{t|s} |\phi_k(t|s)| + 1\big)(b_s^2 + \log n/nb_s)^{1/2})) \quad a.s., \quad (33)$$

where $O(\cdot)$ is uniform over i. Then $\frac{1}{n}\sum_{i=1}^n \max_{1\leq j\leq m_i} |\tilde{\xi}_{ik}(s_{ij}) - \xi_{ik}(s_{ij})| = O(b_s^2 + \log n/nb_s)^{1/2}) \quad a.s.$, using (L.3). For the remainders $R_{ik}(s_{ij})$, by (14) and Lemma 1,

$$\max_{1\leq j\leq m_i} |R_{ik}(s_{ij})| = O((b_s^2 + \log n/nb_s)^{1/2}[\sum_{j=1}^{m_i} \frac{1}{L_{ij}} \sum_{l=1}^{L_{ij}} |\epsilon_{ijl}|]) \ \ a.s., \quad (34)$$

Noting that $\max_{1\leq i\leq n} m_i \leq M$ for finite M, and as all $|\epsilon_{ijl}|$ are i.i.d variables with finite mean, we have $\frac{1}{n}\sum_{i=1}^n \max_{1\leq j\leq m_i} |R_{ik}(s_{ij})| = O((b_s^2 + \log n/nb_s)^{1/2}), \quad a.s.$

For $\tilde{\epsilon}_{ijk}$, we know that $E(\tilde{\epsilon}_{ijk}) = 0$ and $E|\tilde{\epsilon}_{ijk}| \leq \text{var}(\tilde{\epsilon}_{ijk})^{1/2} = O(L_n^{-1/2})$. Since $m_i \leq M$ for finite M, we have $E\sup_{1\leq j\leq m_i} |\tilde{\epsilon}_{ijk}| = O(L_n^{-1/2}) = O((\log n/n)^{1/2})$, where $O(\cdot)$ is uniform over i for fixed k. Then by the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n \sup_{1\leq j\leq m_j} |\tilde{\epsilon}_{ijk}| = O((\log n/n)^{1/2}) \quad a.s. \quad (35)$$

Combining (33), (34) and (35), (19) follows. For each k, if the true target processes $\xi_{ik}(s)$ were used, then by the Corollary 3.5 in Li and Hsing (2010),

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

Results (20) (21) and (22) then follow by the same arguments as used for Theorem 3.

Proofs of Corollary 1 and Corollary 2. Considering the case of a dense design in s, for each fixed k and p, the proof for Theorem 3 implies that $\sup_{s\in\mathcal{S}} |\hat{\xi}_{ik}(s) - \xi_{ik}(s)| \xrightarrow{P} 0$, and also $\sup_{s\in\mathcal{S}} |\hat{\psi}_{kp}(s) - \psi_{kp}(s)| \xrightarrow{P} 0$. Part (a) directly follows.

For part (b), by the Karhunen-Loève expansion, for fixed s and t,

$$X_i^K(t|s) = \mu(t,s) + \sum_{k=1}^K \xi_{ik}(s)\phi_k(t|s) \xrightarrow{P} X_i(t|s) \ \text{ as } \ K \to \infty,$$

and also for any fixed k and s,

$$\xi_{ik}^P(s) = \sum_{p=1}^P \zeta_{ikp}\psi_{kp}(s) \xrightarrow{P} \xi_{ik}(s) \ \text{ as } \ P_k \to \infty.$$

Define the truncated version $X_i^{K,P}(t|s) = \mu(t|s) + \sum_{k=1}^{K}\sum_{p=1}^{P}\zeta_{ikp}\phi_{kp}(s)\psi_k(t|s)$. For any integer $m \geq 1$, and $\delta > 0$, one can find $K(m)$ and $P(m)$, such that

$$P(|X_i^{K,P}(t|s) - X_i(t|s)| > \delta/2) \leq \frac{1}{2m}.$$

Note that $\hat{X}_i(t|s)$ as in (10) is obtained by plugging in the estimates of $\phi_k(t|s)$, $\psi_{kp}(s)$ and ζ_{ikp} in $X_i^{K,P}(t|s)$. Using the consistency results for $\hat{\phi}_k(t|s)$, $\hat{\psi}_{kp}(s)$ and $\hat{\zeta}_{ikp}$, one can find large enough $n(m)$ such that

$$P(|\hat{X}_i(t|s) - X_i^{K,P}(t|s)| > \delta/2) < \frac{1}{2m}.$$

So for any $m \geq 1$ and $\delta > 0$, there are large $n(m)$, $K(n(m))$, and $P(n(m))$ such that $P(\hat{X}_i(t|s) - X_i(t|s)| > \delta) < \frac{1}{m}$, which implies $\hat{X}_i(t|s) \xrightarrow{P} X_i(t|s)$.

For the case of a sparse design in s, note that

$$\tilde{\zeta}_{ikp} = \gamma_{kp}\bar{\psi}_{ikp}^T\bar{R}_{ik}^{-1}\bar{\xi}_{ik}, \tag{36}$$

where $\tilde{\psi}_{ikp} = (\psi_{kp}(s_{i1}), \ldots, \psi_{kp}(s_{i,m_i}))$, $\bar{R}_{ik}$ is a m_i by m_i matrix with (j,l)-th element $R_k(s_{ij}, s_{il})$, $\bar{\xi}_{ik}$ is defined as $(\xi_{ik}(s_{i1}), \ldots, \xi_{ik}(s_{i,m_i}))$, and γ_{kp} is the pth largest eigenvalue of the covariance $R_k(s_1, s_2)$; see also Theorem 3 in Yao et al. (2005). Estimates $\hat{\zeta}_{ikp}$ are obtained from (36), by substituting estimates of γ_{kp}, $\bar{\psi}_{ikp}$, $\bar{R}_{ik}$ and $\bar{\xi}_{ik}$, leading to

$$\hat{\zeta}_{ikp} = \hat{\gamma}_{kp}\hat{\bar{\psi}}_{ikp}^T\hat{\bar{R}}_{ik}^{-1}\hat{\bar{\xi}}_{ik}.$$

The uniform convergence of $\psi_{kp}(s)$, $R_k(s_1, s_2)$ and $\hat{\xi}_{ik}(s_{ij})$ then implies $\hat{\zeta}_{ikp} \xrightarrow{P} \tilde{\zeta}_{ikp}$. Part (b) then follows by similar arguments as used in the cases of a dense design in s.

REFERENCES

Bosq, D. (2000), *Linear Processes in Function Spaces: Theory and Applications*, New York: Springer-Verlag.

Del Barrio, E., Deheuvels, P., and Geer, S. (2007), *Lectures on empirical processes: theory and statistical applications.*, European Mathematical Society.

van de Geer, S. A. (2006), *Empirical Process Theory and Applications*, handout WS 2006, ETH Zürich.

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.

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.

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.

Zhang, X. and Wang, J.L. (2012), "Nonparametric Estimation for Longitudinal and Functional Data: A Unified Theory."

Supplement B: Comparisons With the Karhunen-Loève Expansion

As an alternative approach to the proposed two-step FPCA, we also implemented and fitted the two-dimensional Karhunen-Loève expansion as in Eq. (24). This implementation is illustrated with the mortality data that have been introduced and discussed in Section 6.

These data are regular and quite dense, which means that one can use the sample covariance as an estimate of the four-dimensional covariance function $G(t_1, t_2, s_1, s_2) = \mathrm{cov}(X(t_1, s_1), X(t_2, s_2))$, where the covariance function is a necessary ingredient for the Karhunen-Loève implementation. Since the data contain measurement errors, the eigenfunctions $\rho_k(t, s)$ resulting from this approach are then smoothed, with the resulting estimates shown in Figure 7. Alternatively, one may estimate the covariance function $G(t_1, t_2, s_1, s_2)$ by four-dimensional local linear smoothing. The eigenfunction estimates resulting from this smoothing approach are in Figure 8. We find that in the dense regular case, these two estimating methods for the covariance function $G(t_1, t_2, s_1, s_2)$ lead to almost identical estimates of the eigenfunctions $\rho_k(t, s)$. Not surprisingly, the four-dimensional smoothing method is computationally much slower.

Note that the first two eigenfunction estimates $\rho_1(t, s)$ and $\rho_2(t, s)$ are quite similar to the corresponding first two surfaces of the proposed model, $\varphi_{11}(t|s)$ and $\varphi_{12}(t|s)$, which are shown in Figure 5. We note that we consider here a scenario where t and s have inherently different meanings in that s is a longitudinal time and t a functional time, and therefore the components ϕ_k and $\xi_k(s)$ obtained from the first step in our proposed two-step FPCA approach are of interest in themselves (see Figure 4). These components are however not available when one uses the Karhunen-Loève approach.

Even if one were to ignore the roles played by t and s, the proposed method runs much faster compared to the Karhunen-Loève expansion with a two-dimensional argument as in equation (24), especially when the data are not regular and dense. This is due to the fact that then a four-dimensional smoothing step for the covariance $G(s_1, t_1, s_2, t_2)$ is needed to implement the Karhunen-Loève expansion.

For further illustration, we sparsified the mortality data, retaining only one third of the measurements available per trajectory, where the measurements retained are randomly selected. Then we applied the proposed method, which involves a three-dimensional smoothing step for $G(t_1, t_2|s)$, as well as the Karhunen-Loève expansion with a four-dimensional smoothing step for $G(t_1, t_2, s_1, s_2)$. We found that the proposed method yields estimates of the model components that are very close to those obtained when observing the entire data set without missings (Figure 9). In contrast, the more complex Karhunen-Loève approach (24) proved to be extremely time consuming (it is 25 times slower than the proposed method) and the estimates resulting from the case of sparse data were rather poor (Figure 10).

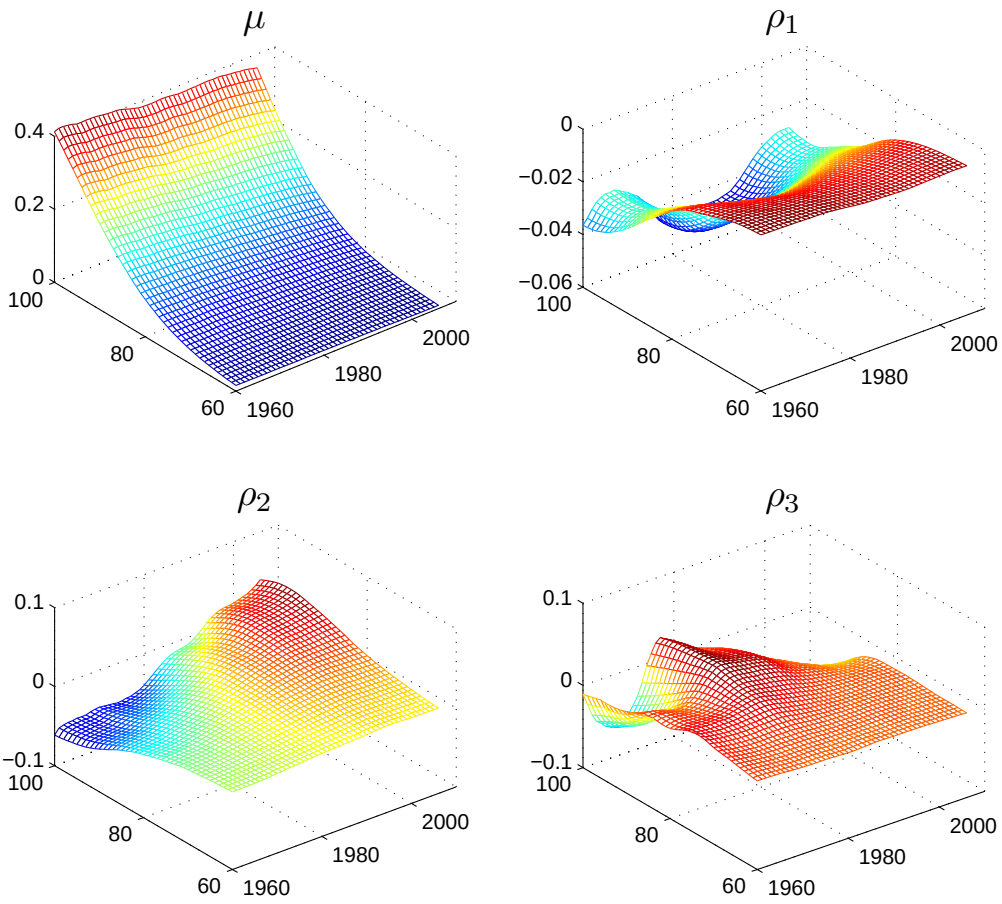

Figure 7: The estimated mean function and first three eigenfunctions $\hat{\rho}_k(t, s)$ for the Karhunen-Loève decomposition (24) for the mortality data, using sample covariances for estimation.

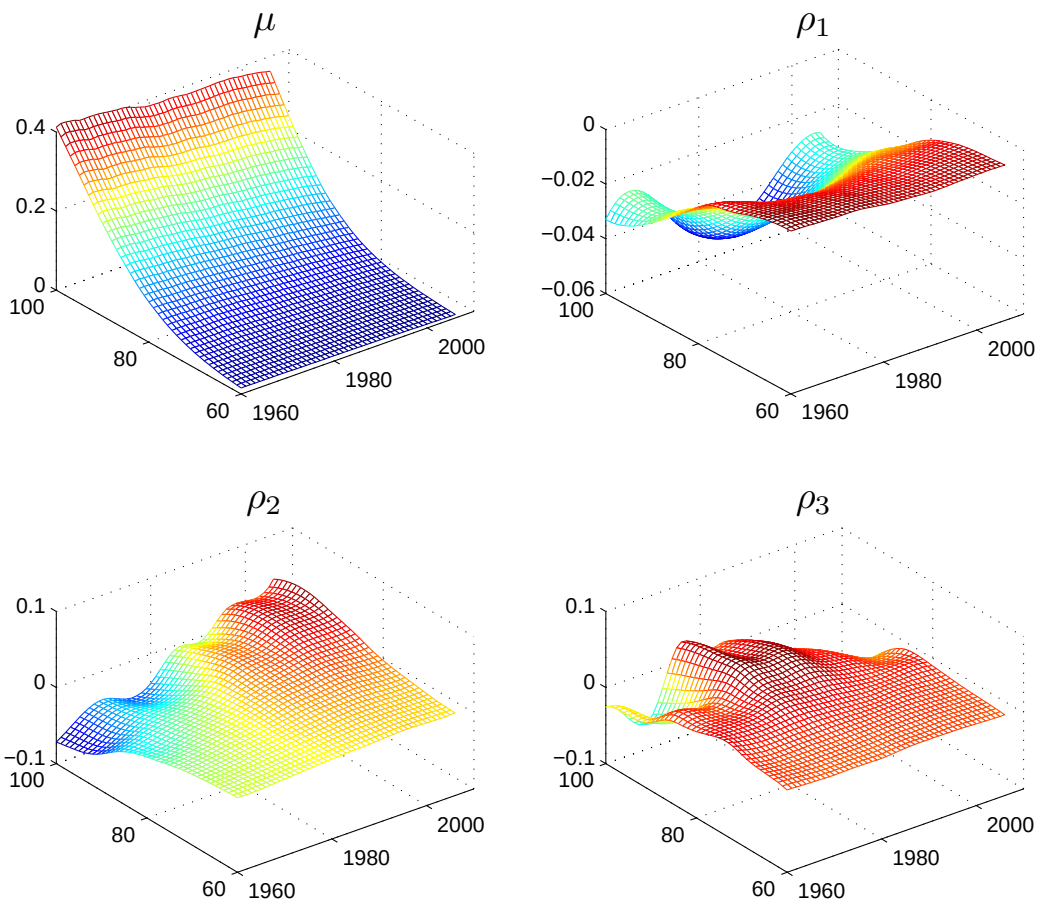

Figure 8: The estimated mean function and first three eigenfunctions $\hat{\rho}_k(t,s)$ for the Karhunen-Loève decomposition (24) for the mortality data, using four-dimensional smoothing for estimation.

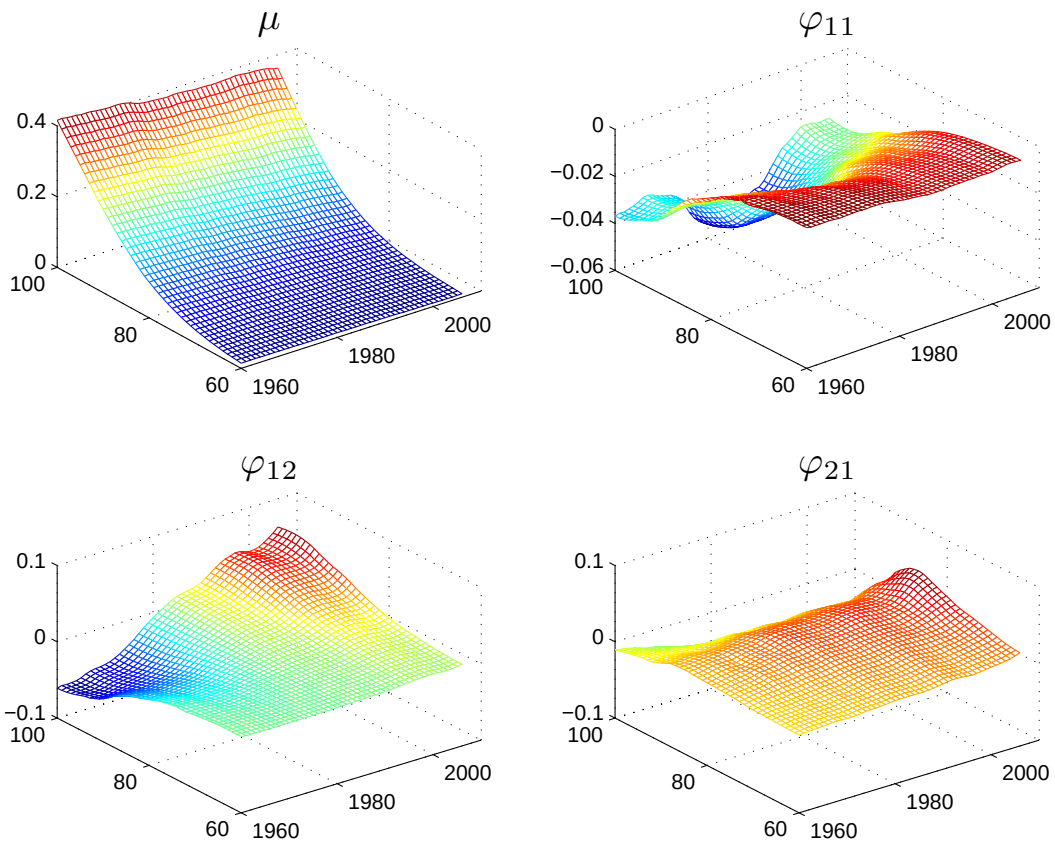

Figure 9: The estimated mean and first three principal surfaces $\hat{\varphi}_{11}$, $\hat{\varphi}_{12}$, $\hat{\varphi}_{21}$ for the proposed approach (10) for the sparsified mortality data (with one third of the available measurements randomly deleted).

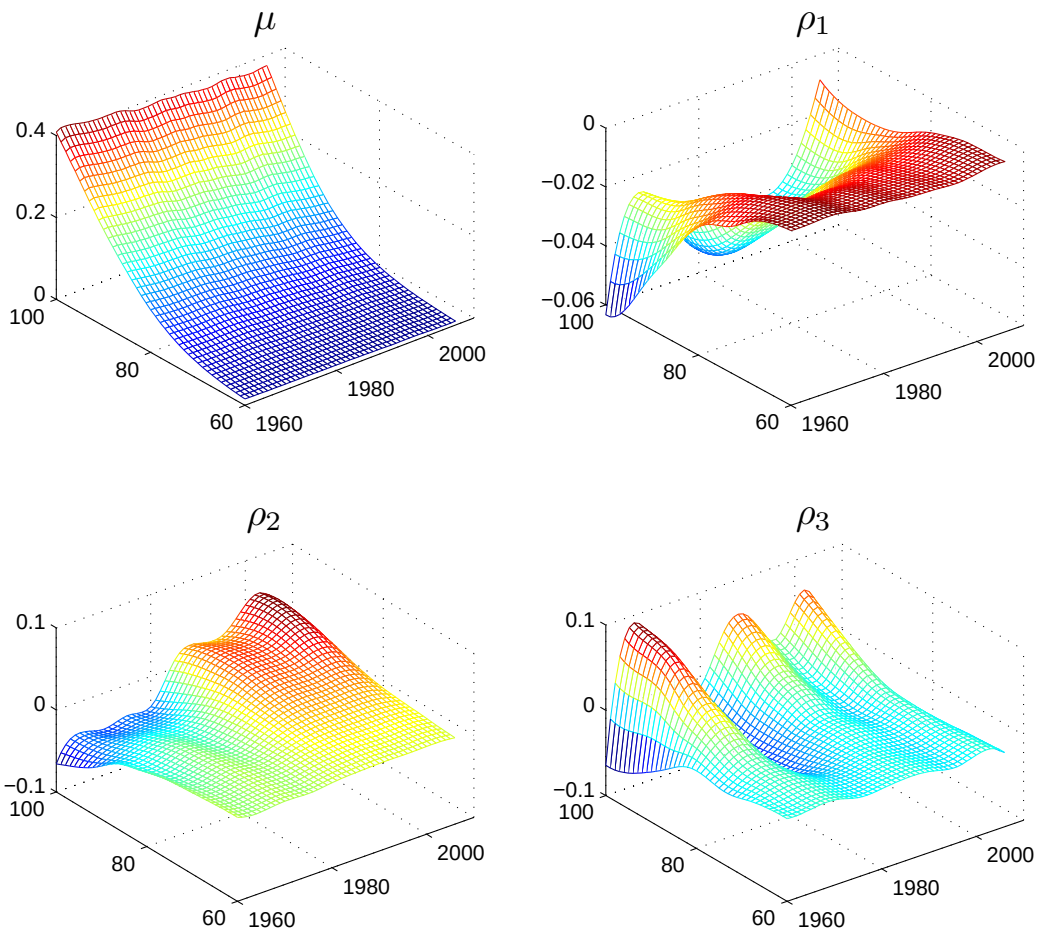

Figure 10: The estimated mean and first three eigenfunctions $\hat{\rho}_k(t,s)$ for the Karhunen-Loève decomposition (24) for the sparsified mortality data (one third of the available data), using four-dimensional smoothing for the covariance surface.

Supplement C: List of Countries Included in Mortality Data Analysis

The 32 countries are: Australia, Austria, Belarus, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Estonia, Finland, France, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Russia, Slovakia, Spain, Sweden, Switzerland, United Kingdom, Ukraine, USA.

Supplement D: Additional Simulations

An additional simulation study was conducted specifically to assess the performance of the proposed method for a situation where eigenvalues cross at several locations s, with specifications as proposed by an anonymous reviewer. We generate data as in model (3)

$$X_i(t|s) = \mu(t|s) + \sum_{k=1}^{2} \xi_{ik}(s)\phi_k(t|s), \quad i = 1, \ldots, n, \quad s \in [0,1], \quad t \in [0,1],$$

where $\phi_1(t|s) = \sqrt{2}\sin\{2\pi(t-s)\}$, $\phi_2(t|s) = \sqrt{2}\cos\{2\pi(t-s)\}$ and sample size $n = 400$. The random functions $\xi_{i1}(s)$ and $\xi_{i2}(s)$ are generated as zero mean Gaussian processes with covariance structures $R_1(s_1, s_2) = 4\cos(4\pi s_1)\cos(4\pi s_2) + 2\sin(4\pi s_1)\sin(4\pi s_2)$ and $R_2(s_1, s_2) = 6\sin(2\pi s_1)\sin(2\pi s_2) + 2\cos(2\pi s_1)\cos(2\pi s_2)$. The grid for t consists of 100 equi-spaced points on [0,1], and the grid for s of 50 equi-spaced points on [0,1].

One of the challenges of this simulated data is that the two eigenvalue functions $\lambda_1(s) = \text{var}(\xi_{i1}(s))$ and $\lambda_2(s) = \text{var}(\xi_{i2}(s))$ cross four times. Following the method described in Section 3, we first estimate the functions $\mu(t, s)$ and $G(t_1, t_2|s)$ by their *empirical estimators*. For $s_0 = 0$, we determine $\phi_1(\cdot|s)$ to be the eigenfunction associated with the larger eigenvalue and $\phi_2(\cdot|s)$ to be the eigenfunction associated with the smaller eigenvalue. The average value of ϑ, chosen by the method described in Section 3, for 100 simulation runs was 3.01. The gaps from omitting some values of s as described in the method were small and easily filled by smoothing the available values $\hat{\phi}_k(\cdot|s_j)$ across s.

Figure 11 demonstrates nearly perfect recovery of the true basis functions $\phi_k(t|s)$ and the eigenvalues $\lambda_k(s)$, for $k = 1, 2$ obtained in the first step FPCA. Figure 12 demonstrates fairly good performance of the second FPCA, applied to the working processes $\hat{\xi}_{ik}(s)$, where $\psi_{kp}(s)$ are the eigenfunctions of $\xi_{ik}(s)$.

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

$$RSE = \frac{||\phi(t|s) - \hat{\phi}(t|s)||^2}{||\phi(t|s)||^2}, \tag{37}$$

where $||\phi(t|s)||^2 = \int\int \phi(t|s)^2 dsdt$, analogously for $\hat{\psi}_{kp}(s)$. The boxplots of the relative squared errors over 100 simulation runs as reported in Figure 13 are seen to be reasonably small for all $\phi_k(t|s)$ and $\psi_{kp}(s)$, except that there about 10 outliers, which occurred because the crossing of the eigenvalue functions was not correctly identified. We note that the number of outliers shrinks quickly with increasing sample size.

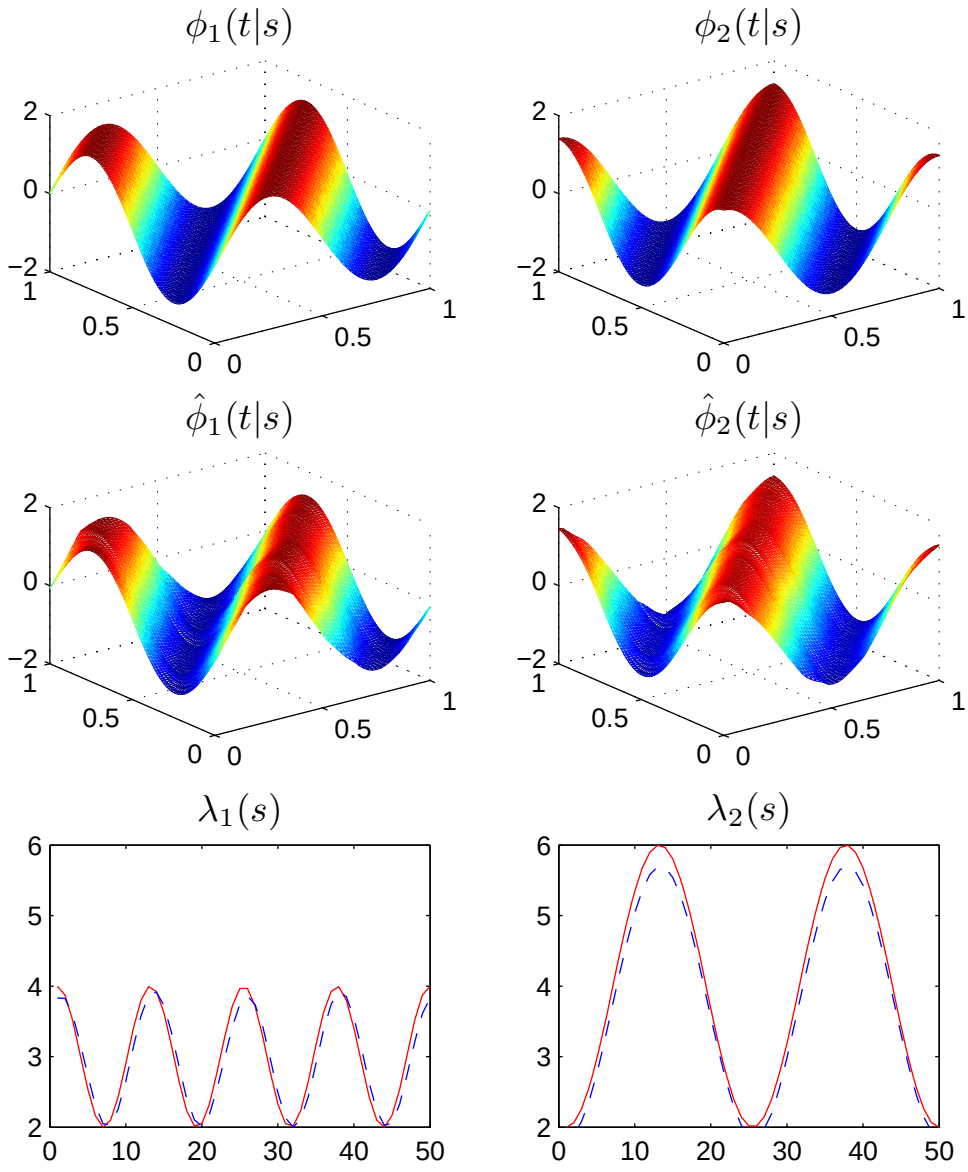

Figure 11: True and estimated $\phi_k(t|s)$ and eigenvalue functions $\lambda_k(s)$ for $k = 1$ (left) and $k = 2$ (right) from one simulation run, as described in Supplement D.

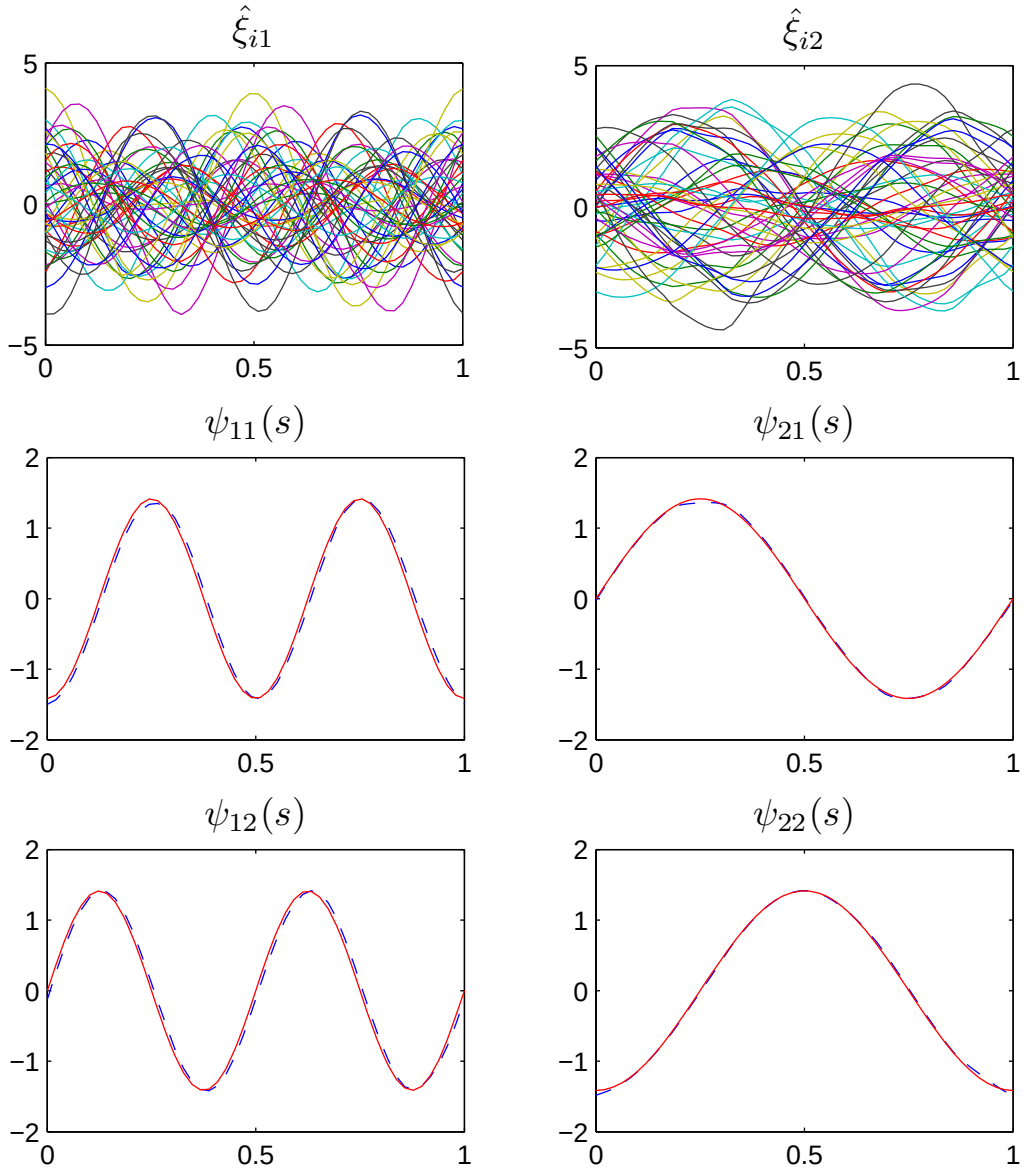

Figure 12: The estimated random functions $\xi_{ik}(s)$ for $k = 1, 2$, $i = 1, \ldots, 50$, and true and estimated eigenfunctions $\psi_{kp}(s)$ from one simulation run, as described in Supplement D.

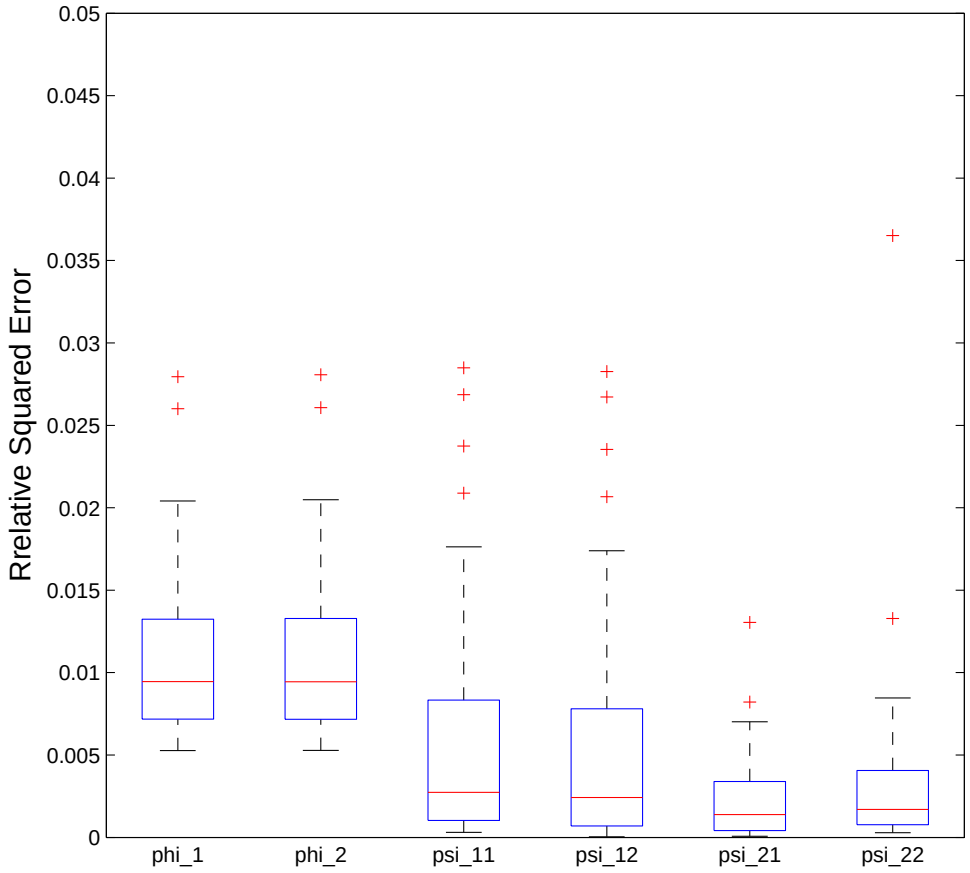

Figure 13: The relative squared errors for the eigenfunctions $\phi_k(t|s)$, $k = 1, 2$, and $\psi_{kp}(s)$, $k = 1, 2$, $p = 1, 2$, obtained from 100 simulation runs, as described in Supplement D.

Supplement E: Eigenanalysis of the random functions $\xi_1(s)$ and $\xi_2(s)$ for the mortality data

The functions $\xi_1(s)$ and $\xi_2(s)$ are obtained in the first stage of the proposed double FPCA method, as in the basic model in (1). For the mortality data application, these functions are plotted in Figure 4 and their characteristic features have been discussed in Section 6. Here we provide additional details on the second stage of the double FPCA method in the context of the mortality data analysis. The second stage consists of the eigenanalysis of the functions $\xi_1(s)$ and $\xi_2(s)$ that are shown in the right panels of Figure 4 and yields the eigenfunctions ψ_{1p} and ψ_{2p} of the covariance operators of these random processes, where $p = 1, 2$, as described in Section 6.

The eigenfunctions ψ_{kp} are key components of the principal surfaces $\varphi_k(t|s) = \psi_{kp}(s)\phi_k(t|s)$ and the corresponding FPC scores $\zeta_{ikp} = \int \xi_{ik}(s)\psi_{kp}(s)\,ds$ serve as random scores to represent the repeatedly observed functions $X_i(t|s)$, see eq. (5). We plot the eigenfunctions ψ_{11}, ψ_{12}, explaining 83.2 % and 13.4%, respectively, of the variation of random functions $\xi_1(s)$, in the upper panel of Figure 14, and eigenfunctions ψ_{21}, ψ_{22} explaining 73.2 % and 17.1%, respectively, of the variation of random functions $\xi_2(s)$ in the lower panel.

We find that the first eigenfunction ψ_{11} for processes $\xi_1(s)$ nicely reflects the main variance increase around 1980-1990, in accordance with the shapes in the top right panel of Figure 4. Observe here that the sign of the eigenfunctions is arbitrary. The second eigenfunction ψ_{12} indicates an additional increase in the variation of processes across countries with increasing calendar year. As processes $\xi_1(s)$ are tied to the basic age-increase in mortality, as seen in the top left panel of Figure 4, this implies an ongoing differentiation into higher- and lower-mortality countries.

Processes $\xi_2(s)$ are associated with the contrast between old and oldest-old mortality, as can be seen in the lower left panel of Figure 4. The first eigenfunction ψ_{21} reflects the increase in the variation across countries with increasing calendar date, including a recent slight acceleration of this increase, in accordance with the function shapes depicted in the lower right panel of Figure 4. The second eigenfunction ψ_{21} reflects a contrast between pre- and post-1980, indicating that there is a tendency for a reversal between the pre-1980 and the post-1980 old to oldest-old mortality differential. Indeed, taking a closer look at the right lower panel of Figure 4 shows that some countries exhibit a reversal in the shapes of $\xi_2(s)$.

To summarize, this analysis demonstrates the usefulness of the results of the second stage FPCA that is provided by the proposed method.

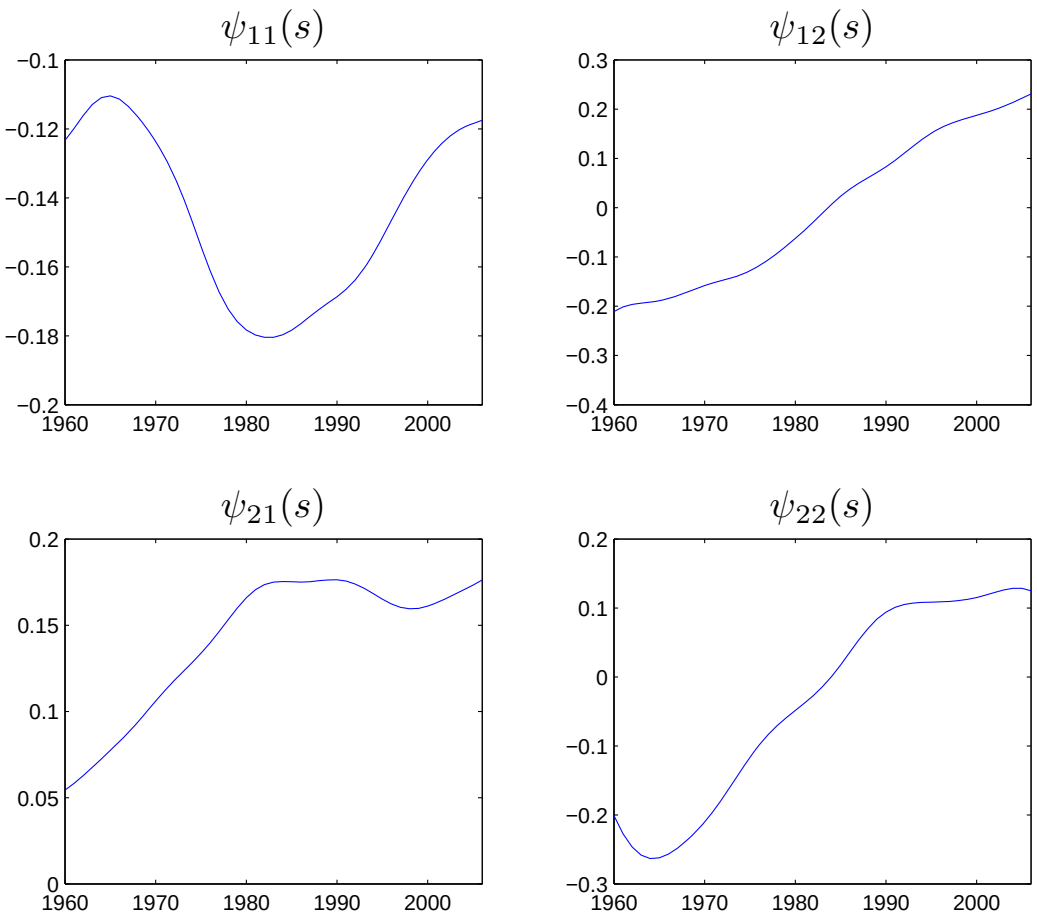

Figure 14: The estimated eigenfunctions $\psi_{kp}(s)$, $k = 1, 2$, $p = 1, 2$, for $\xi_1(s)$ (upper panels) and $\xi_2(s)$ (lower panels), obtained from the mortality data, as described in Supplement E.