Functional Singular Component Analysis

Wenjing Yang*, Hans-Georg Müller* and Ulrich Stadtmüller**

*Department of Statistics, University of California, One Shields Ave., Davis, CA 95616, U.S.A.

**Abt. f. Zahlen- u. Wahrscheinlichkeitstheorie, Universität Ulm, 89069 Ulm, Germany

June 2010

Abstract

Aiming at quantifying the dependency of pairs of functional data (X, Y), we develop the concept of functional singular value decomposition for covariance and functional singular component analysis, building on the concept of "canonical expansion" of compact operators in functional analysis. We demonstrate the estimation of the resulting singular values, functions and components for the practically relevant case of sparse and noise-contaminated longitudinal data and provide asymptotic consistency results. Expanding bivariate functional data into singular functions emerges as a natural extension of the popular functional principal component analysis for single processes to the case of paired processes. A natural application of the functional singular value decomposition is a measure of functional correlation. Due to the involvement of an inverse operation, most previously considered functional correlation measures are plagued by numerical instabilities and strong sensitivity to the choice of smoothing parameters. These problems are acerbated for the case of sparse longitudinal data, on which we focus. The functional correlation measure derived from the functional singular value decomposition behaves well with respect to numerical stability and statistical error, as we demonstrate in a simulation study. Practical feasibility for applications to longitudinal data is illustrated with examples from a study on aging and online auctions.

Key words and phrases: bivariate stochastic process, functional correlation, functional data analysis, longitudinal data, principal component, singular component, singular value decomposition

1. Introduction

As data in the form of random functions and curves are becoming more common, so is the need for measures and estimates to capture the dependency between the components of multivariate functional data. An established method is functional canonical correlation analysis (FCCA), developed by Leurgans *et al.* (1993), extending the corresponding multivariate concept (**?**). FCCA is used in various applications (Ramsay and Silverman, 2005; Ghebreab *et al.*, 2007) involving functional data that are generated on densely sampled grids; it does not work well in situations where one is faced with irregular and sparse data, as typically encountered in longitudinal studies and other real world problems (He *et al.*, 2004). The theoretical foundations for FCCA have been laid in Hannan (1961); Silverman (1996); He *et al.* (2003), and Cupidon *et al.* (2008). An interesting approach, emphasizing the connections to reproducing kernel Hilbert spaces, has recently been developed in Eubank and Hsing (2008). Due to the involvement of an inverse operator, numerical stability of FCCA is not guaranteed, even in the best case where complete functions are observed; available implementations require rather complex regularization schemes.

As measuring dependency between pairs of underlying random processes (X, Y) is of special interest for data arising from longitudinal studies, there is a need for alternative approaches, geared towards the commonly encountered case of sparse and irregularly sampled measurements. Aiming at avoiding an inverse operation, an alternative and stable method for correlation estimation in the multivariate vector case was proposed in Service *et al.* (1998), where it was suggested to modify the target criterion to be maximized. This idea (see also **?**) and recent developments in functional partial least squares (Preda and Saporta, 2005) for scalar responses motivate a measure of functional correlation that is stable and well suited for irregular longitudinal vector data. The proposed approach can be viewed as an extension of multivariate partial least squares, where both predictors and responses are multivariate, to the functional case, more specifically as an extension of the version that has been denoted as PLS-SVD in the overview by Wegelin (2000). The proposed approach introduces singular value decompositions for bivariate functional data, leading to a measure of functional correlation that does not involve

an inverse operation, therefore does not depend critically on regularization and is numerically stable. These properties make the proposed measure well suited to substitute for the inherently unstable FCCA.

The proposed functional correlation measure is derived from the cross-covariance of the pair of processes (X, Y) and is therefore closely tied to the properties of cross-covariance operators. The singular functions and singular values of this operator play a crucial role in the analysis. The singular functions provide not only representations for the cross-covariance operators, but also orthonormal function systems in which one may expand X and Y; we refer to the corresponding expansion coefficients of X and Y as functional singular components. To obtain estimators for the singular values, functions and components for the case of sparse longitudinal data and to establish their asymptotic consistency, our starting point is a uniformly consistent estimate of the cross-covariance function between processes X and Y. This leads to consistent estimates of singular values and singular functions. Consistent estimates of the functional singular components are obtained by targeting the conditional expectations of these scores, given the observed data for the respective processes.

We illustrate the proposed functional correlation and functional singular component analysis (FSCA) with on-line auction data. There is a growing need for statistical methodology that is adequate for e-commerce data. In our example, the goal is to quantify the same-bidder effect, i.e. the dependency of bidding patterns across different auctions for bids that are submitted by the same bidder. To establish a dependency measure suitable for such data is a problem that cannot be tackled with classical multivariate methods, because each bidder places a random number of bids in each auction at random times and therefore the bids are not in vector form. On the other hand, established methods for quantifying functional correlation, which include FCCA, dynamical correlation (Dubin and Müller, 2005) and shape based methods (Heckman and Zamar, 2000), all require densely recorded data and are not geared towards situations with sparse and irregular measurements. In the functional approach to auction data, the recorded bid prices are viewed as manifestation of an underlying unobservable auction price process, which is only observed at the random times at which bids are submitted. By addressing

situations with sparsely and irregularly observed functional data, where available methods do not work, we provide a measure of dependency for longitudinal data in general.

A second example concerns the relationship of irregularly recorded body mass index with systolic blood pressure for a sample of longitudinally observed subjects. The problem then is to establish a correlation between the patterns of these measurements, the idea being that larger body mass index might be associated with higher systolic blood pressure. Again, common multivariate or standard functional methods are not amenable for defining a measure of time-dynamic correlation for such data, due to their irregular and sparse nature.

In functional canonical correlation analysis (FCCA) one assumes that $X, Y \in L_2(\mathcal{T})$ are jointly distributed L_2-processes on a bounded interval $\mathcal{T}$, where $L_2(\mathcal{T})$ is the Hilbert space of square integrable functions on $\mathcal{T}$, equipped with the inner product $\langle u, v\rangle = \int_{\mathcal{T}} u(s)v(s)\,ds$, for any functions $u, v \in L_2(\mathcal{T})$. FCCA encompasses functional canonical correlation coefficients and canonical weight functions. The first canonical correlation $\tilde{\rho}_1$ is defined as

$$\tilde{\rho}_1 = \sup_{\mathrm{var}(\langle u,X\rangle)=1\,,\mathrm{var}(\langle v,Y\rangle)=1} \mathrm{corr}(\langle u, X\rangle, \langle v, Y\rangle) = \mathrm{corr}(\langle u_1, X\rangle, \langle v_1, Y\rangle)\,, \tag{1}$$

where u_1 and v_1 are the weight functions associated with $\tilde{\rho}_1$.

Various tools of dimension reduction and augmentation (He *et al.*, 2004), such as smoothing methods and projection, have been applied in an attempt to stabilize functional canonical correlation. FCCA is inherently unstable as it involves inverse operators, which enter through the constraints in (1). Unregularized implementations routinely produce estimates of functional canonical correlation coefficients that are close to 1 (Leurgans *et al.*, 1993). For the multivariate p-dimensional case, Service *et al.* (1998) proposed an innovative approach to quantify correlation between random vectors. This approach avoids dealing with inverses by maximizing $\mathrm{cov}(u^{\mathrm{T}}X, v^{\mathrm{T}}Y)$, subject to the constraints

$$u^{\mathrm{T}}\Sigma_{XX}^{-1}u = v^{\mathrm{T}}\Sigma_{YY}^{-1}v = 1. \tag{2}$$

These replace the corresponding constraints for FCCA in the multivariate case, given by

$$u^{\mathrm{T}}\Sigma_{XX}u = v^{\mathrm{T}}\Sigma_{YY}v = 1, \tag{3}$$

where $\Sigma_{ZZ} = \mathrm{var}(Z)$ is the variance-covariance matrix of Z, $Z = X$ or Y.

Restating (2), respectively (3), for X in terms of the eigenvectors u_k and eigenvalues λ_k of Σ_{XX} leads to

$$u^{\mathrm{T}}\Sigma_{XX}^{-1}u = \sum_{k=1}^{p} \frac{(u^{\mathrm{T}}u_k)^2}{\lambda_k} = 1, \quad \text{respectively,} \quad u^{\mathrm{T}}\Sigma_{XX}u = \sum_{k=1}^{p} \lambda_k (u^{\mathrm{T}}u_k)^2 = 1, \tag{4}$$

and analogously for Y. Since the eigenvalues λ_k are ordered by size (and in the infinite-dimensional case $p \to \infty$ are summable and therefore converge to zero), one finds that the modified constraints in (2) corresponding to those on the l.h.s. of (4) force the weight vectors u to be primarily represented in terms of the first few, more stable eigenvectors, with larger eigenvalues that account for most of the variability in the data. Hence the modified constraints in (2) produce weight functions and associated correlations that are much more stable. The proposed notion of functional correlation is based on similar ideas for the functional case, where throughout we emphasize the case of longitudinal data with sparse and irregular observations. The case of traditional functional data collected on a regular and dense grid is always included.

The paper is organized as follows: The proposed functional singular value decomposition is introduced in Section 2. Estimation procedures are the focus of Section 3 and their asymptotic properties that of Section 4. This is followed by a report on simulation studies in Section 5 and a description of the two application examples, on-line bidding and a longitudinal study in the life sciences, in Section 6. Connections of the proposed functional singular component analysis (FSCA) with FCCA are explored in Section 7. The proofs are compiled in an appendix.

2. Functional Singular Value Decomposition

Longitudinal data of the type we consider here are generated by a sample of n pairs of random trajectories (X_i, Y_i), $i = 1, \ldots, n$, which are assumed to be realizations of smooth random processes (X, Y), with unknown smooth mean functions $EY(t) = \mu_Y(t)$, $EX(s) = \mu_X(s)$, and smooth covariance functions $\mathrm{cov}(Y(s), Y(t)) = C_{YY}(s,t)$, $\mathrm{cov}(X(s), X(t)) = C_{XX}(s,t)$ and cross-covariance function $\mathrm{cov}(X(s), Y(t)) = C_{XY}(s,t)$. In the following, we use the notations $Y^c(t) = Y(t) - \mu_Y(t)$, $X^c(t) = X(t) - \mu_X(t)$.

Assume that the domain $\mathcal{T}$ of both X and Y is a finite interval of length $|\mathcal{T}|$, and that X, Y are square integrable processes, i.e.,

$$E(\|X\|_2^2) = E\left(\int_{\mathcal{T}} X(s)^2\, ds\right) < \infty, \quad E(\|Y\|_2^2) < \infty. \tag{5}$$

Denoting the standard inner product in the Hilbert space $L_2(\mathcal{T})$ of square integrable functions on $\mathcal{T}$ by $\langle \cdot, \cdot \rangle$, we aim at a "functional covariance" corresponding to the first singular value

$$\sup_{\|u\|=\|v\|=1} \operatorname{cov}(\langle u, X\rangle, \langle v, Y\rangle), \tag{6}$$

which is attained at functions u_1, $v_1 \in L_2(\mathcal{T})$, say. As a standardized version that will serve as functional correlation we consider

$$\rho_1 = \frac{\operatorname{cov}(\langle u_1, X\rangle, \langle v_1, Y\rangle)}{\sqrt{\operatorname{var}(\langle u_1, X\rangle)\operatorname{var}(\langle v_1, Y\rangle)}}. \tag{7}$$

As discussed before, this quantity differs from the usual functional canonical correlation, which results from a different maximization problem, as defined in (1).

To study functional singular values, we define auto-covariance operators $\mathcal{C}_{XX}$ and $\mathcal{C}_{YY}$ and cross-covariance operators $\mathcal{C}_{XY}$ and $\mathcal{C}_{YX}$ as follows,

$$\begin{aligned}
\mathcal{C}_{XX} : L_2(\mathcal{T}) \to L_2(\mathcal{T}), \quad f \mapsto g, \quad g(s) &= \int_{\mathcal{T}} C_{XX}(s,t) f(t) dt, \quad C_{XX}(s,t) = E(X^c(s)X^c(t)), \\
\mathcal{C}_{YY} : L_2(\mathcal{T}) \to L_2(\mathcal{T}), \quad f \mapsto g, \quad g(s) &= \int_{\mathcal{T}} C_{YY}(s,t) f(t) dt, \quad C_{YY}(s,t) = E(Y^c(s)Y^c(t)), \\
\mathcal{C}_{XY} : L_2(\mathcal{T}) \to L_2(\mathcal{T}), \quad f \mapsto g, \quad g(s) &= \int_{\mathcal{T}} C_{XY}(s,t) f(t) dt, \quad C_{XY}(s,t) = E(X^c(s)Y^c(t)), \\
\mathcal{C}_{YX} : L_2(\mathcal{T}) \to L_2(\mathcal{T}), \quad f \mapsto g, \quad g(s) &= \int_{\mathcal{T}} C_{YX}(s,t) f(t) dt, \quad C_{YX}(s,t) = E(Y^c(s)X^c(t)).
\end{aligned}$$

Similar operators have been discussed previously in Baker (1974) and Gualtierotti (1979). While the cross-covariance operators are not self-adjoint, $\mathcal{C}_{YX}$ is the adjoint operator of $\mathcal{C}_{XY}$. The compound operators $\mathcal{A}_{XY} = \mathcal{C}_{XY} \circ \mathcal{C}_{YX}$ and $\mathcal{A}_{YX} = \mathcal{C}_{YX} \circ \mathcal{C}_{XY}$ are self-adjoint, and moreover are linear Hilbert-Schmidt operators with L_2-kernels (see section V.3 of Kato, 1995),

$$A_{XY}(s,t) = \int_{\mathcal{T}} C_{XY}(s,u)\, C_{YX}(u,t)\, du = \int_{\mathcal{T}} C_{XY}(s,u)\, C_{XY}(t,u)\, du, \tag{8}$$

$$A_{YX}(s,t) = \int_{\mathcal{T}} C_{YX}(s,u)\, C_{XY}(u,t)\, du = \int_{\mathcal{T}} C_{XY}(u,s)\, C_{XY}(u,t)\, du\,. \tag{9}$$

According to the spectral theorem for Hilbert-Schmidt operators, $\mathcal{A}_{YX}$ and $\mathcal{A}_{XY}$ have a discrete spectrum with shared eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq 0$ and orthonormal eigenfunctions $\phi_1, \phi_2, \ldots$ for $\mathcal{A}_{YX}$ and $\psi_1, \psi_2, \ldots$ for $\mathcal{A}_{XY}$, respectively. Furthermore one has $\psi_k = \frac{1}{\sigma_k}\mathcal{C}_{XY}(\phi_k)$, $k = 1, 2, \ldots$. We refer to the sequences $\phi_1, \phi_2, \ldots$ and $\psi_1, \psi_2, \ldots$ as singular functions and to the sequence $\sigma_1, \sigma_2, \ldots$ as singular values. The sequence of pairs of singular functions $(\phi_1, \psi_1), (\phi_2, \psi_2), \ldots$ plays an analogous role for pairs of observed processes (X, Y) to that of the sequence of eigenfunctions when one observes repeated realizations of one process X (see, e.g. Rice and Silverman, 1991).

The following singular representation of cross-covariance is a key tool. It is a direct consequence of well-known results, which can be found e.g. in section V.3 of Kato (1995).

Proposition. *The cross-covariance operators can be represented as*

$$\mathcal{C}_{XY}(f)(t) = \sum_{k=1}^{\infty} \sigma_k \langle f, \psi_k \rangle \phi_k(t), \qquad \mathcal{C}_{YX}(f)(t) = \sum_{k=1}^{\infty} \sigma_k \langle f, \phi_k \rangle \psi_k(t), \qquad t \in \mathcal{T}. \tag{10}$$

Representation (10) leads to a straightforward solution of the maximization problem (6),

$$\begin{aligned} \sup_{\|u\|=\|v\|=1} \operatorname{cov}(\langle u, X \rangle, \langle v, Y \rangle) &= \sup_{\|u\|=\|v\|=1} \iint u(t)v(s)E(X^c(t)Y^c(s))dsdt \\ &= \sup_{\|u\|=\|v\|=1} \langle v, \mathcal{C}_{YX}(u) \rangle = \sup_{\|u\|=\|v\|=1} \sum_{k=1}^{\infty} \sigma_k \langle u, \phi_k \rangle \langle v, \psi_k \rangle = \sigma_1, \end{aligned}$$

where the maximum is attained at $u = \phi_1$ and $v = \psi_1$.

Repeating the maximization procedure on sequences of orthogonal complements, one is led to a sequence of singular values $(\sigma_1, \sigma_2, \ldots)$, along with the associated singular functions $u_k = \phi_k$; $v_k = \psi_k$, $k = 1, 2, \ldots$, and representations of cross-covariance surfaces

$$C_{XY}(s,t) = \sum_{k=1}^{\infty} \sigma_k \psi_k(s) \phi_k(t), \qquad C_{YX}(s,t) = \sum_{k=1}^{\infty} \sigma_k \phi_k(s) \psi_k(t). \tag{11}$$

According to the Hilbert-Schmidt theorem, one may represent processes X by the orthonormal singular functions $\{\phi_1, \phi_2, \ldots\}$, respectively processes Y by the orthonormal singular functions $\{\psi_1, \psi_2, \ldots\}$. These singular decompositions are

$$X(s) = \mu_X(s) + \sum_{m=1}^{\infty} \zeta_m \phi_m(s) + \nu_X(s), \quad Y(t) = \mu_Y(t) + \sum_{k=1}^{\infty} \xi_k \psi_k(t) + \nu_Y(t), \tag{12}$$

where $\nu_X \in L^2$ is a function in the kernel of the operator $\mathcal{A}_{YX}$, i.e., a function with $\mathcal{A}_{YX}(\nu_X) = 0$, and analogously $\nu_Y \in L^2$ is in the kernel of $\mathcal{A}_{XY}$. The random functions ν_X and ν_Y are residual functions with vanishing means and cross-covariances that are unrelated to the cross-covariance operators. Therefore these functions do not play a role in the representation (10) of cross-covariance operators or in the estimation procedures, on which our approaches and notions of functional correlation are based. Important quantities in (12) are $\zeta_m = \int X^c(t)\phi_m(t)dt,\ m \geqslant 1$, the coefficients of X with respect to its expansion in the orthonormal system $\{\phi_m\}_{m\geq 1}$ and $\xi_k = \int Y^c(t)\psi_k(t)dt,\ k \geqslant 1$, the coefficients of Y with respect to its expansion in the orthonormal system $\{\psi_k\}_{k\geq 1}$. We refer to these random coefficients as singular components. Using (11), it is easy to see that the singular components satisfy

$$E\zeta_m = 0, \quad E\xi_k = 0, \quad E(\zeta_m\xi_k) = 0 \quad \text{for} \quad m \neq k, \quad \text{and} \quad E(\zeta_m\xi_m) = \sigma_m. \tag{13}$$

Particularly noteworthy is the covariance structure of the singular components and the uncorrelatedness of different components between the two processes. One may interpret the proposed functional correlation coefficient as $\text{corr}(\zeta_1, \xi_1)$, the correlation of the first singular components of X and Y, according to the Proposition. We next discuss estimation of the singular values and singular functions $\{(\sigma_1, \phi_1, \psi_1), (\sigma_2, \phi_2, \psi_2), \dots\}$ and of the corresponding functional correlation ρ_1 in (7), as well as estimation of the singular components.

3. Estimation Methods

3.1 Sparsely observed functional data

In many data analysis situations, for example most longitudinal studies, the observed data can be thought of as generated by smooth random processes. However, the underlying functional trajectories are not directly observed. Instead, the available data correspond to noisy, sparse and irregularly timed measurements of these trajectories. We model this situation (following Yao *et al.*, 2005b) by assuming that the measurements are made at random time points $S_{i1}, \dots, S_{iL_i}$ for processes X_i, resp. $T_{i1}, \dots, T_{iR_i}$ for processes Y_i, where the numbers of measurements L_i resp. R_i are i.i.d random variables. The data (S_{il}, U_{il}, L_i) and

(T_{ij}, V_{ij}, R_i), $l = 1, \ldots, L_i$, $j = 1, \ldots, R_i$, $i = 1, \ldots, n$, are assumed to be independent for different i, where U_{il} (respectively, V_{ij}) denote the observations of the random trajectory X_i (respectively, Y_i) at the random times S_{il} (respectively, T_{ij}), contaminated with measurement errors ε_{il} (respectively, ϵ_{ij}), which are assumed to be i.i.d. with $E\varepsilon_{il} = 0$, $E[\varepsilon_{il}^2] = \sigma_X^2$ (respectively, $E\epsilon_{ij} = 0$, $E[\epsilon_{ij}^2] = \sigma_Y^2$), and independent of the trajectories.

One then may represent the observed data for processes (X_i, Y_i) as follows,

$$\begin{aligned} U_{il} &= X_i(S_{il}) + \varepsilon_{il}, \quad S_{il} \in \mathcal{T}, \ \ 1 \le l \le L_i, \ 1 \le i \le n, \\ V_{ij} &= Y_i(T_{ij}) + \epsilon_{ij}, \quad T_{ij} \in \mathcal{T}, \ \ 1 \le j \le R_i, \ \ 1 \le i \le n, \end{aligned} \tag{14}$$

and with (12),

$$U_{il} = \mu_X(S_{il}) + \sum_{m=1}^{\infty} \zeta_{im}\phi_m(S_{il}) + \nu_X(S_{il}) + \varepsilon_{il}, \tag{15}$$

$$V_{ij} = \mu_Y(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\psi_k(T_{ij}) + \nu_Y(T_{ij}) + \epsilon_{ij}. \tag{16}$$

3.2 Estimating singular values and singular functions

Define the local linear scatterplot smoothers for $\mu_X(s)$ (and analogously for $\mu_Y(t)$ by substituting the corresponding terms) through minimizing

$$\sum_{i=1}^{n}\sum_{l=1}^{L_i} K_1(\frac{S_{il} - s}{b_X})\{U_{il} - \beta_0^X - \beta_1^X(s - S_{il})\}^2, \tag{17}$$

with respect to β_0^X, β_1^X, leading to $\hat{\mu}_X(s) = \hat{\beta}_0^X(s)$.

Let $C_i(S_{il}, T_{ir}) = (U_{il} - \hat{\mu}_X(S_{il}))(V_{ir} - \hat{\mu}_Y(T_{ir}))$, and define the local linear surface smoother for the cross-covariance surface $C(s,t)$ through minimizing

$$\sum_{i=1}^{n}\sum_{l=1}^{L_i}\sum_{r=1}^{R_i} K_2(\frac{S_{il} - s}{h_1}, \frac{T_{ir} - t}{h_2})\{C_i(S_{il}, T_{ir}) - f(\beta, (s,t), (S_{il}, T_{ir}))\}^2, \tag{18}$$

where $f(\beta, (s,t), (S_{il}, T_{ir})) = \beta_0 + \beta_{11}(s - S_{il}) + \beta_{12}(t - T_{ir})$, with respect to $\beta = (\beta_0, \beta_{11}, \beta_{12})$, yielding minimizers $\hat{\beta}_0(s,t), \hat{\beta}_{11}(s,t)$ and $\beta_{12}(s,t)$ and estimates

$$\hat{C}(s,t) = \hat{C}_{XY}(s,t) = \hat{\beta}_0(s,t). \tag{19}$$

Estimates of $\mathcal{A}_{XY}$ and $\mathcal{A}_{YX}$ can be constructed through estimating the operator kernels

$$\begin{aligned} \hat{A}_{XY}(t,s) = & \int_{\mathcal{T}} \hat{C}(s,u)\hat{C}(t,u)du, \\ \hat{A}_{YX}(t,s) = & \int_{\mathcal{T}} \hat{C}(u,s)\hat{C}(u,t)du \end{aligned} \tag{20}$$

by numerical integration.

The estimates of $\{\sigma_1, \phi_1, \psi_1\}$ correspond to the solutions $\{\hat{\sigma}_1, \hat{\phi}_1, \hat{\psi}_1\}$ of the eigenequations

$$\begin{aligned} \hat{\mathcal{A}}_{XY}(\hat{\psi}_1)(t) = & \int_{\mathcal{T}} \hat{A}_{XY}(s,t)\hat{\psi}_1(s)ds = \hat{\sigma}_1^2\hat{\psi}_1(t), \\ \hat{\mathcal{A}}_{YX}(\hat{\phi}_1)(t) = & \int_{\mathcal{T}} \hat{A}_{YX}(s,t)\hat{\phi}_1(s)ds = \hat{\sigma}_1^2\hat{\phi}_1(t), \end{aligned} \tag{21}$$

which are obtained numerically by discretization and then matrix spectral decomposition. Then one may obtain $\{\hat{\sigma}_k, \hat{\phi}_k, \hat{\psi}_k\}$ for $k \geq 2$ by defining $\hat{A}_{XY}^{(\ell+1)}(t,s) = \hat{A}_{XY}^{(\ell)}(t,s) - \hat{\sigma}_\ell^2\hat{\psi}_\ell(s)\hat{\psi}_\ell(t)$, $\ell = 1, 2, \ldots,$ $\hat{A}_{XY}^{(1)}(t,s) = \hat{A}_{XY}(t,s)$. The estimates of $\{\sigma_\ell, \psi_\ell\}$ correspond to the solutions $\{\hat{\sigma}_\ell, \hat{\psi}_\ell\}$ of the eigenequations associated with $\hat{A}_{XY}^{(\ell)}$. Analogous considerations apply to ϕ_ℓ. In our implementation of this method, we used an equidistant grid of 51 time points for the numerical discretization step; simulations showed that the estimates are not sensitive to this choice.

3.3 Estimating the functional correlation

For estimation of C_{XX}, C_{YY}, we use local linear smoothers. With $C_{i,XX}(S_{il}, S_{ir}) = (U_{il} - \hat{\mu}_X(S_{il}))(U_{ir} - \hat{\mu}_X(S_{ir}))$ and $C_{i,YY}(T_{il}, T_{ir}) = (V_{il} - \hat{\mu}_Y(T_{il}))(V_{ir} - \hat{\mu}_Y(T_{ir}))$, we obtain local linear surface smoothers for the covariance surfaces $C_{XX}(s_1, s_2)$ and $C_{YY}(t_1, t_2)$ through minimizing

$$\sum_{i=1}^{n} \sum_{1 \leq l \neq j \leq L_i} K_2\left(\frac{S_{il} - s_1}{h_X}, \frac{S_{ij} - s_2}{h_X}\right)\{C_{i,XX}(S_{il}, S_{ij}) - f(\beta, (s_1, s_2), (S_{il}, S_{ij}))\}^2, \tag{22}$$

$$\sum_{i=1}^{n} \sum_{1 \leq r \neq j \leq R_i} K_2\left(\frac{T_{ir} - t_1}{h_Y}, \frac{T_{ij} - t_2}{h_Y}\right)\{C_{i,YY}(T_{ir}, T_{ij}) - f(\beta, (t_1, t_2), (T_{ir}, T_{ij}))\}^2, \tag{23}$$

respectively, omitting the diagonal elements, where f is defined as in (18), with respect to $\beta = (\beta_0, \beta_{11}, \beta_{12})$, yielding minimizers $(\hat{\beta}_{0,X}(s_1, s_2), \hat{\beta}_{11,X}(s_1, s_2), \hat{\beta}_{12,X}(s_1, s_2))$ for (22) and $(\hat{\beta}_{0,Y}(t_1, t_2), \hat{\beta}_{11,Y}(t_1, t_2), \hat{\beta}_{12,Y}(t_1, t_2))$ for (23), and then

$$\hat{C}_{XX}(s_1, s_2) = \hat{\beta}_{0,X}(s_1, s_2), \quad \hat{C}_{YY}(t_1, t_2) = \hat{\beta}_{0,Y}(t_1, t_2). \tag{24}$$

To obtain the adjusted estimate of C_{XX} on the diagonal, i.e., $\tilde{C}_{XX}(s) = \hat{C}_{XX}(s,s) + \hat{\sigma}_X^2$, one may judiciously smooth data only along the diagonal (for details, see Yao *et al.*, 2005a).

Observe that the functional correlation ρ_1 in (7) can be written as

$$\rho_1 = \sigma_1[\int_{\mathcal{T}} \phi_1(s_1)C_{XX}(s_1,s_2)\phi_1(s_2)ds_1ds_2]^{-\frac{1}{2}}\,[\int_{\mathcal{T}} \psi_1(t_1)C_{YY}(t_1,t_2)\psi_1(t_2)dt_1dt_2]^{-\frac{1}{2}}, \qquad (25)$$

motivating the plug-in estimator

$$\hat{\rho}_1 = \hat{\sigma}_1[\int_{\mathcal{T}} \hat{\phi}_1(s_1)\hat{C}_{XX}(s_1,s_2)\hat{\phi}_1(s_2)ds_1ds_2]^{-\frac{1}{2}}\,[\int_{\mathcal{T}} \hat{\psi}_1(t_1)\hat{C}_{YY}(t_1,t_2)\hat{\psi}_1(t_2)dt_1dt_2]^{-\frac{1}{2}}, \qquad (26)$$

where estimates $\hat{\sigma}_1, \hat{\phi}_1, \hat{\psi}_1$ are obtained as after (21) and estimates $\hat{C}_{XX}, \hat{C}_{YY}$ as in (24). The integration step may be carried out by any suitable numerical integration method.

3.4 Estimating singular components

In the situation of sparsely observed trajectories, estimates of the singular components aim at the conditional components, given the observations for a specific trajectory, building on a similar idea in functional principal component analysis (Yao *et al.* (2005a)). Assume that in (15) and (16), ν_X, ν_Y, ζ_{im}, ξ_{ik}, ε_{il} and ϵ_{ij} are jointly Gaussian. The following arguments are always conditional on the given observation times S_{il} and T_{ij}, $i = 1,\ldots,n$, $l = 1,\ldots,L_i$, $j = 1,\ldots,R_i$. Denote the vectors of X_i, U_i, μ and ϕ_k at the observation grid $\{S_{il}\}_{l=1}^{L_i}$ by $\boldsymbol{X}_i$, $\boldsymbol{U}_i$, $\mu_{i,X}$ and ϕ_{im}, i.e. $\boldsymbol{X}_i = (X_i(S_{i1}),\ldots,X_i(S_{iL_i}))^{\mathrm{T}}$, $\boldsymbol{U}_i = (U_i(S_{i1}),\ldots,U_i(S_{iL_i}))^{\mathrm{T}}$, $\mu_{i,X} = (\mu_X(S_{i1}),\ldots,\mu_X(S_{iL_i}))^{\mathrm{T}}$, $\phi_{im} = (\phi_m(S_{i1}),\ldots,\phi_m(S_{iL_i}))^{\mathrm{T}}$, and analogously $\boldsymbol{Y}_i$, $\boldsymbol{V}_i$, $\mu_{i,Y}$ and ψ_{ik} for Y, i.e. $\boldsymbol{Y}_i = (Y_i(T_{i1}),\ldots,Y_i(T_{iR_i}))^{\mathrm{T}}$, $\boldsymbol{V}_i = (V_i(T_{i1}),\ldots,V_i(T_{iR_i}))^{\mathrm{T}}$, $\mu_{i,Y} = (\mu_Y(T_{i1}),\ldots,\mu_Y(T_{iR_i}))^{\mathrm{T}}$, $\psi_{ik} = (\psi_k(T_{i1}),\ldots,\psi_k(T_{iR_i}))^{\mathrm{T}}$.

The best linear unbiased predictor (BLUP) of the singular components for the ith subject, given the data available for that individual, is the conditional expectation, obtained under Gaussian assumptions,

$$\tilde{\eta} = E[\eta|\,\boldsymbol{U}_i,\,\boldsymbol{V}_i] = \Sigma_\eta \Sigma^{-1}_{\boldsymbol{U}_i,\boldsymbol{V}_i} \begin{bmatrix} \boldsymbol{U}_i - \mu_{i,X} \\ \boldsymbol{V}_i - \mu_{i,Y} \end{bmatrix}, \quad \eta = \zeta_{im} \text{ or } \xi_{ik}, \qquad (27)$$

where

$$\Sigma_{\zeta_{im}} \quad = \quad [\mathrm{cov}(\zeta_m,\,\boldsymbol{U}_i^{\mathrm{T}},),\mathrm{cov}(\zeta_m,\,\boldsymbol{V}_i^{\mathrm{T}})] = \left[\mathrm{cov}\left(\int_{\mathcal{T}} X(s)\phi_m(s)ds,\,\boldsymbol{X}_i^{\mathrm{T}}\right),\sigma_m\psi_{im}^{\mathrm{T}}\right]$$

$$= \quad [P_{im}, \sigma_m \psi_{im}^{\mathrm{T}}]\,,$$

$$\text{and} \quad P_{im} = \left[\int_{\mathcal{T}} C_{XX}(s, S_{i1})\phi_m(s)ds, \int_{\mathcal{T}} C_{XX}(s, S_{i2})\phi_m(s)ds, \ldots, \int_{\mathcal{T}} C_{XX}(s, S_{iL_i})\phi_m(s)ds\right]^{\mathrm{T}}.$$

Analogously, $\Sigma_{\xi_{ik}} = [\sigma_k \phi_{ik}^{\mathrm{T}}, Q_{ik}]$, where

$$Q_{ik} = \left[\int_{\mathcal{T}} C_{YY}(s, T_{i1})\psi_k(s)ds, \int_{\mathcal{T}} C_{YY}(s, T_{i2})\psi_k(s)ds, \ldots, \int_{\mathcal{T}} C_{YY}(s, T_{iR_i})\psi_k(s)ds\right]^{\mathrm{T}}.$$

Also $\Sigma_{\boldsymbol{U}_i, \boldsymbol{V}_i} = \mathrm{var}([\boldsymbol{U}_i^{\mathrm{T}}, \boldsymbol{V}_i^{\mathrm{T}}]^{\mathrm{T}}) = \mathrm{var}([\boldsymbol{X}_i^{\mathrm{T}}, \boldsymbol{Y}_i^{\mathrm{T}}]^{\mathrm{T}}) + \Sigma_i$, with $\Sigma_i = \mathrm{diag}(\sigma_X^2, \ldots, \sigma_X^2, \sigma_Y^2, \ldots, \sigma_Y^2)$ denoting the $(L_i + R_i) \times (L_i + R_i)$ diagonal measurement error variance matrix.

By substituting corresponding estimates in (27), assuming that K components are selected, the estimated singular components are

$$\hat{\eta} = \hat{E}[\eta | \boldsymbol{U}_i, \boldsymbol{V}_i] = \hat{\Sigma}_{\eta} \hat{\Sigma}_{\boldsymbol{U}_i, \boldsymbol{V}_i}^{-1} \begin{bmatrix} \boldsymbol{U}_i - \hat{\mu}_{i,X} \\ \boldsymbol{V}_i - \hat{\mu}_{i,Y} \end{bmatrix}, \quad \eta = \zeta_{im} \text{ or } \xi_{ik}, \tag{28}$$

where $\hat{\Sigma}_{\zeta_{im}} = \left[\hat{P}_{im}, \hat{\sigma}_m \hat{\psi}_{im}^{\mathrm{T}}\right]_{1 \times (L_i + R_i)}$, $\hat{\Sigma}_{\xi_{ik}} = \left[\hat{\sigma}_k \hat{\phi}_{ik}^{\mathrm{T}}, \hat{Q}_{ik}\right]_{1 \times (L_i + R_i)}$ and $\hat{P}_{im} = [\int_{\mathcal{T}} \hat{C}_{XX}(s, S_{i1})\hat{\phi}_m(s)ds, \ldots,$ $\int_{\mathcal{T}} \hat{C}_{XX}(s, S_{iL_i})\hat{\phi}_m(s)ds]_{1 \times L_i}$, $\hat{Q}_{ik} = [\int_{\mathcal{T}} \hat{C}_{YY}(s, T_{i1})\hat{\psi}_k(s)ds, \ldots, \int_{\mathcal{T}} \hat{C}_{YY}(s, T_{iR_i})\hat{\psi}_k(s)ds]_{1 \times R_i}$, $\hat{\Sigma}_{\boldsymbol{U}_i \boldsymbol{V}_i} = \begin{bmatrix} \hat{\Sigma}_{i,XX} & \hat{\Sigma}_{i,XY} \\ \hat{\Sigma}_{i,YX} & \hat{\Sigma}_{i,YY} \end{bmatrix} + \begin{bmatrix} \hat{\sigma}_X^2 I_{L_i} & 0 \\ 0 & \hat{\sigma}_Y^2 I_{R_i} \end{bmatrix}$. Here $\hat{\mu}_{i,X}$ and $\hat{\mu}_{i,Y}$ are obtained from the local linear smoother in (17), the (j, j') element of the matrix $\hat{\Sigma}_{i,XX}$ is $\hat{C}_{XX}(S_{ij}, S_{ij'})$ as in (24) and the (j, j') element of the matrix $\hat{\Sigma}_{i,YY}$ is $\hat{C}_{YY}(T_{ij}, T_{ij'})$ as in (24), while the (j, j') element of the matrix $\hat{\Sigma}_{i,XY} = \hat{\Sigma}_{i,YX}^{\mathrm{T}}$ is $\hat{C}_{XY}(S_{ij}, T_{ij'})$ as in (19).

4. Asymptotic Properties

We provide consistency results for estimated singular values $\hat{\sigma}_1$ in (21), singular functions $\hat{\phi}_1, \hat{\psi}_1$ in (21) and the functional correlation $\hat{\rho}_1$ in (26). These results can be analogously extended to higher order singular values and singular functions. The necessary assumptions, which must hold in addition to basic regularity conditions, can be found in the Appendix. The first result concerns estimates (20) of the two-dimensional kernels A_{XY} (8) and A_{YX} (9), which define operators $\mathcal{A}_{XY}$ and $\mathcal{A}_{YX}$. Seeking the spectral decomposition of these operators is central to the proposed functional singular component analysis.

Theorem 1. *Under assumptions* (A1.1)-(A4) in the Appendix,

$$\sup_{s,t\in\mathcal{T}}\left|\hat{A}_{XY}(s,t)-A_{XY}(s,t)\right|=O_p(\frac{1}{\sqrt{n}h_1h_2}),\quad \sup_{s,t\in\mathcal{T}}\left|\hat{A}_{YX}(s,t)-A_{YX}(s,t)\right|=O_p(\frac{1}{\sqrt{n}h_1h_2}),$$

where h_1, h_2 *are the bandwidths used for the estimate* (18) *of the cross-covariance surfaces* $C = C_{XY}$.

Since $\{\sigma_k, \phi_k, \psi_k\}_{k\geqslant 1}$ are the eigenvalues and eigenfunctions of A_{XY} and A_{YX}, respectively, the convergence rates for their estimates $\{\sigma_k, \phi_k, \psi_k\}_{k\geqslant 1}$, obtained according to (21), follow by perturbation arguments from Theorem 1. Specifically, we obtain uniform rates of convergence.

Theorem 2. *Under assumptions (A1.1)-(A5) in the Appendix, for every* $k \geq 1$,

$$|\hat{\sigma}_k-\sigma_k| = O_p(\frac{1}{\sqrt{n}h_1h_2}), \tag{29}$$

$$\sup_{s\in\mathcal{T}}\left|\hat{\phi}_k(s)-\phi_k(s)\right| = O_p(\frac{1}{\sqrt{n}h_1h_2}), \tag{30}$$

$$\sup_{s\in\mathcal{T}}\left|\hat{\psi}_k(s)-\psi_k(s)\right| = O_p(\frac{1}{\sqrt{n}h_1h_2}). \tag{31}$$

An important consequence of Theorem 2 is the following result, which establishes consistency of the proposed estimate $\hat{\rho}_1$ (26) of the functional correlation coefficient ρ_1 (25).

Theorem 3. *Under the assumptions of Theorem 2, with bandwidths* h_X *and* h_Y *for estimates* $\hat{C}_{XX}$ *(22) and* $\hat{C}_{YY}$ *(23), it holds that*

$$|\hat{\rho}_1-\rho_1|=O_p(\frac{1}{\sqrt{n}h_1h_2}+\frac{1}{\sqrt{n}h_X^2}+\frac{1}{\sqrt{n}h_Y^2}).$$

Under Gaussian assumptions, the following result establishes the consistency of the estimates $\hat{\zeta}_{im}$ and $\hat{\xi}_{ik}$ in (28), with respect to their targets in (27).

Theorem 4. *Under assumptions (A1.1)-(A7),*

$$\lim_{n\to\infty}\hat{\zeta}_{im}=\tilde{\zeta}_{im},\ \text{and}\ \lim_{n\to\infty}\hat{\xi}_{ik}=\tilde{\xi}_{ik}\ \text{in probability for any fixed}\ m,\ k\in\mathbb{N}\,. \tag{32}$$

5. Simulation studies

5.1 Estimation accuracy

We discuss here how to choose the needed auxiliary parameters (bandwidths and number of included singular components) and demonstrate the performance of the proposed estimates of the singular values and singular functions in a simulation study. Random processes with given singular components were generated by

$$X(t) = \sum_{j=1}^{K} \zeta_j \phi_j(t) \text{ and } Y(t) = \sum_{j=1}^{K} \xi_j \psi_j(t)\,,\ t \in \mathcal{T} = [0, T], \tag{33}$$

with $K = 3, T = 10$, and $\{\phi_j(t), \psi_j(t)\}$ as Fourier basis functions on $[0, T]$, with $\phi_1(t) = \psi_2(t) = \sqrt{2/T}\sin(2\pi t/T)$, $\phi_2(t) = \psi_3(t) = -\sqrt{2/T}\cos(4\pi t/T)$, and $\phi_3(t) = \psi_1(t) = -\sqrt{2/T}\cos(2\pi t/T)$, and zero-mean Gaussian vectors $\zeta = \{\zeta_j\}$ and $\xi = \{\xi_j\}$ with covariances

$$\text{cov}(\zeta) = \begin{bmatrix} 8 & 3 & -2 \\ 3 & 4 & 1 \\ -2 & 1 & 3 \end{bmatrix}, \quad \text{cov}(\xi) = \begin{bmatrix} 6 & -2 & 1 \\ -2 & 4.5 & 1.5 \\ 1 & 1.5 & 3.25 \end{bmatrix}, \quad \text{cov}(\zeta, \xi) = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}.$$

In this setting X and Y have singular values $\sigma_1 = 3$, $\sigma_2 = 1.5$, and $\sigma_3 = 0.5$, and singular functions $\{\phi_j, \psi_j\}$. Observations were obtained from processes X, Y by adding measurement errors, $X_{il} = X_i(S_{il}) + \epsilon_{il}$, $Y_{il} = Y_i(T_{il}) + \varepsilon_{il}$, where $\epsilon_{il}, \varepsilon_{il}$ were i.i.d. $N(0, 0.25)$.

Aiming for the analysis of sparsely sampled longitudinal data, sampling designs were chosen from the following scenarios (i)-(v). In each scenario, the sampling locations were uniformly distributed for each subject on the domain $[0, 10]$.

(i) Sparse2: Each random function is sampled at only two random times.

(ii) Sparse5: Each random function is sampled at five random times.

(iii) Sparse10: Each random function is sampled at ten random times.

(iv) Dense15: Each random function is sampled at fifteen equidistant non-random time points.

(v) Dense20: Each random function is sampled at twenty equidistant non-random time points.

For each scenario, the sample consists of the (sparse or dense) observations made on 200 pairs of random processes. We apply generalized cross-validation (GCV) to determine the

bandwidths for smoothing the cross-covariance surface, chosen to be the same in both directions. While we found empirically that GCV-based bandwidth choices are working well for most data settings, an empirical adjustment is often advantageous, taking into account the observed level of sparsity in the data. This motivates a bandwidth selector, where GCV-based bandwidths are multiplied by an empirically derived sparsity-adjusting factor. These factors are empirically determined as 0.8 for the very sparse case with only two measurements, and as 1.2 for dense cases, then interpolating between these for intermediate levels of sparseness, leading to the factor $\alpha = 0.8 + (m-2)/45$, m being the observed average number of measurements per subject.

The results for 400 Monte Carlo runs are in Table 1. We find that even for very sparse cases the singular values can be estimated with reasonably small errors. These errors become smaller as designs get denser, as expected. The improvement of estimates when transitioning to denser designs is especially large for the estimation of the singular functions.

5.2 Comparisons with existing methods

In additional simulations, we compared the performance of the proposed functional singular component analysis (FSCA) with alternative measures of correlation that can be computed for sparse and irregularly sampled functional data, which include a version of simple empirical Pearson correlation and multivariate canonical correlation. Simulating processes as in (33) and section 5.1 with covariance matrices $\mathrm{cov}(\zeta) = \mathrm{cov}(\xi) = \mathrm{diag}(8, 4.5, 3.5)$, $\mathrm{cov}(\zeta, \xi) = \mathrm{diag}(6, 4, 3)$, processes X and Y are found to possess singular values $\sigma_1 = 6$, $\sigma_2 = 4$, and $\sigma_3 = 3$, and singular functions $\{\phi_j, \psi_j\}$, as defined above. Observations were generated from processes X, Y by adding measurement errors, $X_{il} = X_i(S_{il}) + \epsilon_{il}$, $Y_{il} = Y_i(T_{il}) + \varepsilon_{il}$, $\epsilon_{il}, \varepsilon_{il} \sim N(0, 0.64)$ i.i.d..

In a first simulation study, we compared correlation estimates based on first functional singular component (FSC), (multivariate) canonical correlation (CAN) and empirical correlation (COR). Here CAN and COR are implemented in a straightforward way, using the corresponding multivariate procedures, where one calculates the canonical correlation coefficient and the Pearson correlation coefficient, respectively, directly for the observed bivariate data vectors. The observed data were generated according to scenarios (ii)-(v), with each sample consisting of the

observations made for 200 pairs of random processes. The results for 400 Monte Carlo runs are in Table 2 for the correlation as measured by the three methods. The FSC derived estimator is seen to be much closer to its target than CAN, while COR is not yielding meaningful results. This is not too surprising, as there is no clear functional target for this measure, especially for the case of sparse irregular designs.

We note that functional canonical correlation in the version proposed in Leurgans *et al.* (1993) and discussed in Ramsay and Silverman (2005) does not apply to sparse irregular designs, as it requires fully observed random trajectories. Neither can the simple multivariate canonical correlation and simple Pearson correlation approaches, that serve as comparisons for the proposed method, be applied in unbalanced situations, when the number of observations is not the same across subjects or differs between the pair of trajectories for the same subject. This is a strong restriction in practical longitudinal applications, where missing data and other discrepancies in observed numbers of observations are common occurrences. Our proposed method, in contrast, is geared towards handling such unbalanced situations, and is seen in Table 2 to perform better also in balanced sparse and irregular situations.

In a second simulation study we compare the performance of FSCA with that of functional principal component analysis through conditional expectation (PACE, Yao *et al.*, 2005b). Both methods aim to recover sparsely observed random trajectories. Whereas the PACE method directly targets the trajectories, separately for each trajectory within the pair of trajectories, ignoring their mutual dependence structure, FSCA aims at modeling both covariance structure and trajectories simultaneously and thus provides additional information.

We fit random functions X and Y with functional singular decomposition as in (12), i.e.

$$\hat{X}_i(s) = \hat{\mu}_X(s) + \sum_{m=1}^{K} \hat{\zeta}_{im}\hat{\phi}_m(s), \quad \hat{Y}_i(t) = \hat{\mu}_Y(t) + \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\psi}_k(t),$$

where the number of singular components K is chosen as the minimum number of singular components that explains at least 80% of the cross covariance between X and Y. The results for estimates obtained with FSCA and those obtained with PACE are in Table 3. The two methods are seen to have overall comparable MISEs, which confirms that the criterion for the

choice of the number of included components works reasonably well for FSCA.

6. Applications

6.1 Functional correlation of systolic blood pressure and body mass index

It is well known that systolic blood pressure (SBP) and body mass index (BMI) are two important indices of human health. While the dependency between these variables has been studied cross-sectionally, the correlation in the longitudinal or functional sense is harder to elucidate, although it is of great interest since it reflects the time-dynamic nature of these functional variables. To establish longitudinal correlation, one needs to take into account that SBP and BMI are random processes and follow subject-specific trajectories. In a longitudinal study on aging (Shock *et al.*, 1984), SBP (in mm Hg) and BMI (in kg/m^2) were recorded on each visit of 1590 male volunteers bi-annually. Due to missed scheduled visits and the high variance of the measurements, the collected data is sparse and noisy, with not only different observation times across subjects, but also varying numbers of available observations per subject. The $n = 507$ subjects included in our data analysis were known to survive beyond age 70, and for each subject at least two measurements were recorded. For additional details see Yao *et al.* (2005b).

In this situation, with subject-specific numbers and times of measurement, classical correlation measures, even vector-based measures, are not applicable, while the proposed method yields a consistent estimate of the functional correlation coefficient and therefore a quantification of the dependency. The starting point of the proposed approach via functional singular component analysis (FSCA) is the estimate of the cross-covariance function $\hat{C}_{XY}(s,t)$, depicted in Figure 1. This cross-covariance surface displays various peaks, primarily situated in regions where one of the ages is relatively low (around 40 years).

Using the estimation methods as described in Section 3, the resulting estimate of the functional correlation coefficient ρ_1 (25) is $\hat{\rho}_1 = 0.57$ (26). The corresponding estimated first singular functions $\hat{\phi}_1$ and $\hat{\psi}_1$ (obtained as described after (20)) for SBP and BMI are seen to be overall similar in shape (Figure 2). Both singular functions initially decrease rapidly between ages 40 and 47, then remain relatively constant until the late fifties, with a possible minor peak at

around 57 to 60, followed by further decline. These shapes suggest that the functional correlation is largely determined by the time-dynamic behavior of SBP and BMI before age 62, and particularly at ages before 45, but caution is advised when interpreting singular functions that are generated from sparse data.

We implemented a heuristic bootstrap method to test the null hypothesis of no correlation, by randomly reshuffling BMI and SBP data when resampling subjects with replacement, so that for resampled subjects BMI and SPB are uncorrelated. Under the resulting bootstrap null distribution, the p-value of the functional correlation is found to be $p = 0.025$, based on 400 bootstrap samples, confirming the significance of the correlation, and implying that subject-specific longitudinal trajectories of BMI and SBP tend to move synchronously in the same direction over time.

6.2 Functional correlation of repeated bidding in online auctions

Another example for sparse data are e-Bay auction bids as discussed by Jank and Shmueli (2006). The data (courtesy of W. Jank) consist of all recorded bids observed for 159 7-day second-price auctions of Palm M515 Personal Digital Assistants (PDA) that took place between March and May, 2003, at e-bay. These bids involved 1024 bidders, many of whom participated in more than one auction. For each bidder who participated in at least two auctions, we randomly selected two auctions at which this bidder submitted two or more bids. Here a common assumption is that the bids reflect measurements of a hidden price trajectory, that evolves as the auction moves forward in time; for more details about e-Bay auctions and this assumption, we refer to Shmueli and Jank (2005); Bapna *et al.* (2008); Liu and Müller (2008).

The available data then correspond to sparse observations of 99 paired bid price trajectories, where each pair of trajectories characterizes a particular bidder. Our goal is to quantify bidder-specific effects in bidding behaviors for online auctions, for which FSCA is well suited, due to the highly irregular observations for each auction. A symmetry constraint is natural in this setting, given the symmetry of the data, which is due to the fact that the pair of bidding price curves is randomly and arbitrarily assigned to X and Y, so that one expects that the pairs

of singular functions for X and Y are identical. The estimated singular values and singular functions are derived from the estimated cross-covariance surface (for which we used visually determined bandwidths) and lead to an estimate of the functional correlation coefficient ρ_1 (25) of $\hat{\rho}_1 = 0.68$ (26).

The associated bootstrap p-value was obtained as described in Section 6.1 and found to be $p = 0.035$, providing evidence for a positive relationship in the bidding dynamics of bids submitted by the same bidder in different auctions. This points to bidder-specific characteristics, reflected in similarities in the bid levels made across different auctions by the same bidder, and it is therefore of interest to profile bidders and to identify their "bidder signature". This symmetry constraint can be easily implemented by replacing the operator kernels $\hat{A}_{XY}$ and $\hat{A}_{YX}$ in eq. (20) and (21) by the symmetrized kernels $\tilde{A}_{XY} = \tilde{A}_{YX} = \frac{1}{2}(\hat{A}_{YX} + \hat{A}_{YX})$.

In Figure 3, the estimated first singular function $\hat{\phi}_1$ is found to remain relatively constant through the first 3 days of an auction, then to increase to a peak at around day 5, followed by a rapid decrease, then to assume slightly negative values at the end of an auction. The overall positive functional correlation is seen to be particularly associated with similarity in bidding patterns around day 5. The decrease of the singular function beyond day 5, near the end of the auction, is possibly due to the presence of highly volatile price changes near an auction's end, often caused by "bid sniping", i.e. by bidders trying to outbid each other before the auction is over. Even if a specific bidder is prone to participate in bid sniping, the price levels at this stage can vary substantially, with the effect of potentially obscuring bidder-specific effects.

The "bidder signature" as captured by the functional correlation manifests itself in the bidding behavior mostly during the first 6 days of a 7-day auction. Since in contrast to the BMI-SBP evolution paths for individual patients, the bidding behaviors of the participants in a given auction are likely to mutually influence each other, one needs to be careful with the interpretation of a "bidder signature". Since the composition of bidders in any given auction is random, even though the bidding behavior of a particular selected bidder is likely influenced by the bids placed by the other bidders, the specific fashion in which a bidder is influenced by others in itself is part of the "bidder signature". While this external influence, including

the reactions of other bidders to the bids placed by the selected bidder, may have a strong effect on the bids actually placed by the bidder in a given auction, the "bidder signature" can be understood as a co-mingling of the bidder's particular bidding strategy and these external influences; the "bidder signature" will reflect an average response to these influences over many random auctions.

7. Representing functional canonical correlation via singular components

This representation provides an example of how the proposed FSCA can be applied to analyze other functional approaches. Consider the spectral decomposition of the auto-covariance operator kernels C_{XX} and C_{YY} of processes X and Y

$$C_{XX}(s,t) = \sum_{j}^{\infty} \lambda_{X,j} \Phi_j(s) \Phi_j(t), \quad C_{YY}(s,t) = \sum_{l}^{\infty} \lambda_{Y,l} \Psi_l(s) \Psi_l(t), \tag{34}$$

where $\{\lambda_{X,j}, \Phi_j\}_{j=1}^{\infty}$ and $\{\lambda_{Y,l}, \Psi_l\}_{l=1}^{\infty}$ are the eigenvalues and eigenfunctions of X and Y, respectively, satisfying $\|\Phi_k\| = \|\Psi_k\| = 1$, $\lambda_{Z,1} > \lambda_{Z,2} > \cdots > 0$ for $Z = X$ or Y, $\langle \Phi_j, \Phi_{j'} \rangle = \delta_{jj'}$ and $\langle \Psi_l, \Psi_{l'} \rangle = \delta_{ll'}$.

As analyzing functional canonical correlation is equivalent to an eigenanalysis of the cross-correlation operator $\mathcal{R} = \mathcal{C}_{XX}^{-1/2} \mathcal{C}_{XY} \mathcal{C}_{YY}^{-1/2}$ (see e.g. Theorem 4.8 in He *et al.*, 2003), (11) and (34) imply for the kernel R of $\mathcal{R}$:

$$\begin{aligned} R(s,t) &= C_{XX}^{-1/2} C_{XY} C_{YY}^{-1/2}(s,t) = \left\langle C_{XX}^{-1/2}(s,u), \langle C_{XY}(u,v), C_{YY}^{-1/2}(v,t) \rangle \right\rangle \\ &= \left\langle \sum_j \lambda_{X,j}^{-1} \Phi_j(s) \Phi_j(u), \sum_{l,k} \sigma_k \lambda_{Y,l}^{-1} \phi_k(u) \langle \psi_k, \Psi_l \rangle \Psi_l(t) \right\rangle \\ &= \sum_{j,k,l} \left(\lambda_{X,j}^{-1} \, \sigma_k \, \lambda_{Y,l}^{-1} \right) \left(\langle \Phi_j, \phi_k \rangle \Phi_j(s) \right) \left(\langle \psi_k, \Psi_l \rangle \Psi_l(t) \right) = \sum_{j,k,l} \sigma_{j,k,l}^* \phi_{j,k}^* \psi_{k,l}^*, \end{aligned} \tag{35}$$

$$\sigma_{j,k,l}^* = \lambda_{X,j}^{-1} \, \sigma_k \, \lambda_{Y,l}^{-1}, \quad \phi_{j,k}^* = \langle \phi_k, \Phi_j \rangle \Phi_j(s), \quad \psi_{k,l}^* = \langle \psi_k, \Psi_l \rangle \Psi_l(t).$$

The components of this representation have the properties

$$\|\phi_{j,k}^*\| = \langle \phi_k, \Phi_j \rangle^2 \leq 1, \qquad \|\psi_{k,l}^*\| = \langle \psi_k, \Psi_l \rangle \leq 1,$$

$$\langle \phi_{j,k}^*, \phi_{j',k'}^* \rangle = \langle \phi_k, \Phi_j \rangle \langle \phi_{k'}, \Phi_{j'} \rangle \delta_{jj'}, \qquad \langle \psi_{k,l}^*, \psi_{k,l'}^* \rangle = \langle \phi_k, \Psi_l \rangle \langle \phi_{k'}, \Psi_{l'} \rangle \delta_{ll'}.$$

Therefore, the singular values and singular functions $(\sigma_k, \phi_k, \psi_k)_{k=1,2,...}$ and the eigencomponents $(\lambda_{X,j}, \Phi_j)_{j=1,2,...}$ of X and $(\lambda_{Y,l}, \Psi_l)_{l=1,2,...}$ of Y can be combined to provide an explicit representation (35) of the cross-correlation kernel R. This representation is useful to obtain an estimate of the functional canonical correlation for the case where the data consist of sparse and noisy measurements of X and Y. Estimates for the components σ_k, ϕ_k, ψ_k appearing in (35) were introduced in Section 3, and estimates for $\lambda_{X,j}, \lambda_{Y,l}, \Phi_j, \Psi_l$ in Yao *et al.* (2005a). Estimates $\hat{R}$ are then obtained by plugging in these estimates in (35). However, this approach will not alleviate the inherent instability of functional canonical correlation that is due to the inverse operators that are an integral part of this concept. This is also reflected in the inverse eigenvalues that appear in representation (35). Nevertheless, this representation provides a viable approach for the sparse data case. We note that other functional relations such as functional regression might also be represented via FSCA.

8. Concluding Remarks

We develop functional singular value decomposition of covariance operators as a tool to investigate dependency between paired random functions, leading to a functional singular component analysis (FSCA). This includes estimation procedures for singular values, singular functions and singular components, which are shown to be asymptotically consistent under weak conditions on the measurement designs. A stable version of the functional correlation coefficient that does not depend on inverse operators and is easy to obtain from the cross-covariance of paired processes emerges as a major application. This functional correlation is obtained as correlation of the first singular components, for which we also discuss heuristic inference procedures. Of interest are connections of FSCA with other functional methods. Functional singular components can be used for regularization and dimension reduction of infinite-dimensional functional data, taking the dependency structure of paired processes into account; FSCA provides a representation for pairs of stochastic processes that is useful for applications to functional data.

While scalar- or vector-based correlation measures require that the dimension of the random vectors as well as the meaning of the components of the random vectors remains the same across

subjects, this is not a requirement for the proposed methods. Viable and consistent estimators are available for FSCA even when one encounters numbers and locations of measurements that may vary substantially from subject to subject. This feature makes FSCA especially useful for longitudinal data with sparse, irregular and noisy measurements.

Appendix

The data (S_{il}, U_{il}) and $(T_{ij}, V_{ij}), i = 1, \ldots, n, l = 1, \ldots, L_i, j = 1, \ldots, R_i$ are assumed to be identically distributed with joint density $g_X(s, x)$ and $g_Y(t, y)$, respectively, the observation times S_{il} to be i.i.d. with marginal density $f_X(t)$ (T_{ij} i.i.d. with marginal density $f_Y(T)$). and $(S_{ij}, S_{il}, X_{ij}, X_{il})$, $1 \leq j, l \leq L_i$, $j \neq l$, to be distributed with joint density function $g_{2,X}(s_1, s_2, x_1, x_2)$ (analogously for Y with density $g_{2,Y}(t_1, t_2, y_1, y_2)$). The following assumptions are needed.

(A1.1) The number of observations L_i and R_i are random with $L_i \sim_{\text{i.i.d.}} L$, $R_i \sim_{\text{i.i.d.}} R$, where L and R are positive discrete random variables, with $P(L > 1) > 0$ and $P(R > 1) > 0$.

(A1.2) The observation times $(\{S_{il} : l \in \{1, \ldots, L_i\}\}$ and $(\{T_{ij} : j \in \{1, \ldots, R_i\}\}$ are independent of L_i, resp. R_i; $(d^\ell/dt^\ell) f_Z(t)$ exists and is continuous on $\mathcal{T}$ with $f_Z(t) > 0$ on $\mathcal{T}$; $(d^\ell/dt^\ell) g_Z(t, z)$ is uniformly continuous on $\mathcal{T} \times \mathbb{R}$; $(d^\ell/dt_1^{\ell_1} dt_2^{\ell_2}) g_{2,Z}(t_1, t_2, z_1, z_2)$ is uniformly continuous on $\mathcal{T}^2 \times \mathbb{R}^2$, for $\ell_1 + \ell_2 = \ell$, $0 \leq \ell_1, \ell_2 \leq \ell$, for $\ell = 0, 1, 2$ and $Z = X, Y$. Choosing kernels $K_1(\cdot)$ and $K_2(\cdot)$ as univariate and bivariate density functions for the local linear smoothers (17) and (18) of the mean functions μ_X, μ_Y, variance surface C_{XX}, C_{YY} and cross-covariance surface C_{XY} and b_X, b_Y, h_X and h_Y as bandwidths for obtaining $\hat{\mu}_X$, $\hat{\mu}_Y$, $\hat{C}_{XX}$ and $\hat{C}_{YY}$, and h_1 and h_2 as the bandwidths for estimating $\hat{C}$, the following assumptions are needed.

(A2.1) $b_X \to 0$, $b_Y \to 0$, $nb_X^4 \to \infty$, $nb_Y^4 \to \infty$, $nb_X^6 < \infty$ and $nb_Y^6 < \infty$.

(A2.2) $h_X \to 0$, $h_Y \to 0$, $nh_X^6 \to \infty$, $nh_Y^6 \to \infty$, $nh_X^8 < \infty$ and $nh_Y^8 < \infty$.

(A2.3) Without loss of generality, $h_1/h_2 \to 1$, $nh_1^6 \to \infty$ and $nh_1^8 < \infty$.

The Fourier transforms $\kappa_1(t) = \int e^{-iut} K_1(u) du$ and $\kappa_2(t, s) = \int e^{-(iut+ivs)} K_2(u, v) du dv$ of K_1, K_2, are required to satisfy:

(A3) κ_1 and κ_2 are absolutely integrable, i.e., $\int |\kappa_1(t)|dt < \infty$, $\int \int |\kappa_2(t,s)|dtds < \infty$.

Additional assumptions needed for some of the results are:

(A4) $E[(X - \mu_X(S))^4] < \infty$, and $E[(Y - \mu_Y(T))^4] < \infty$.

(A5) Conditions (B5) of Yao *et al.* (2005b) are satisfied.

(A6) The singular component scores ζ_{im} and ξ_{ik}, measurement errors ε_{il} and ϵ_{ij}, ν_X and ν_Y are jointly Gaussian.

(A7) The number, location and values of measurements for a given subject or cluster remain unaltered as $n \to \infty$.

Proof of Theorem 1. By Lemma A.1 of Yao *et al.* (2005a),

$$\sup_{s,\, t \in \mathcal{T}} \left| \hat{C}_{XY}(s,t) - C_{XY}(s,t) \right| = O_p(\frac{1}{\sqrt{n} h_1 h_2}). \tag{36}$$

By definition,

$$\left| \hat{A}_{XY}(s,t) - A_{XY}(s,t) \right| \leqslant \int_{\mathcal{T}} \left| \hat{C}_{XY}(s,u)\hat{C}_{XY}(t,u) - C_{XY}(s,u)C_{XY}(t,u) \right| du$$

and hence

$$\begin{aligned}
& \sup_{s,\, t \in \mathcal{T}} \left| \hat{A}_{XY}(s,t) - A_{XY}(s,t) \right| \\
\leqslant\ & \sup_{s,\, t \in \mathcal{T}} \int_{\mathcal{T}} \left| \hat{C}_{XY}(s,u) - C_{XY}(s,u) \right| \left| \hat{C}_{XY}(t,u) - C_{XY}(t,u) \right| du \\
& + \sup_{s,\, t \in \mathcal{T}} \int_{\mathcal{T}} \left| \hat{C}_{XY}(s,u) - C_{XY}(s,u) \right| |C_{XY}(t,u)| \, du \\
& + \sup_{s,\, t \in \mathcal{T}} \int_{\mathcal{T}} |C_{XY}(s,u)| \left| \hat{C}_{XY}(t,u) - C_{XY}(t,u) \right| du \\
=\ & O_p(\frac{1}{\sqrt{n} h_1 h_2}),
\end{aligned}$$

where the last step follows from (36) and the continuity assumption for C_{XY}, which implies boundedness on the compact domain $\mathcal{T} \times \mathcal{T}$.

Proof of Theorem 2. Denote the separable Hilbert space of Hilbert-Schmidt operators on H by F, endowed by the inner product $\langle T_1, T_2 \rangle_F = \mathrm{tr}(T_1 T_2^*) = \sum_j \langle T_1 u_j, T_2 u_j \rangle_H$, where $T_1, T_2 \in F$, T_2^* is the adjoint of T_2, and $\{u_j : j \geqslant 1\}$ is a complete orthonormal basis in H. The operator $\mathcal{A}_{XY}$ is generated by the kernel A_{XY} in (8), and $\hat{\mathcal{A}}_{XY}(f) = \int_{\mathcal{T}} \hat{A}_{XY}(s,t) f(t) dt$ for

the corresponding estimators. Then Theorem 1 implies $\|\hat{\mathcal{A}}_{XY} - \mathcal{A}_{XY}\|_F = O_p(1/(\sqrt{n}h_1h_2))$. Since $\{\phi_k : k = 1, 2, \dots\}$ are the eigenfunctions of $\mathcal{A}_{XY}$, it follows analogously to the proof of Theorem 2 in Yao *et al.* (2005b) that for any fixed $k \in \mathbb{N}$

$$\|\hat{\phi}_k - \phi_k\|_H = O_p(\frac{1}{\sqrt{n}h_1h_2}), \tag{37}$$

and analogously for $\mathcal{A}_{YX}$, that

$$\|\hat{\psi}_k - \psi_k\|_H = O_p(\frac{1}{\sqrt{n}h_1h_2}).$$

Note that $\sigma_k = \langle \phi_k, \mathcal{C}_{XY}(\psi_k)\rangle_H$ and $\hat{\sigma}_k = \langle \hat{\phi}_k, \hat{\mathcal{C}}_{XY}(\hat{\psi}_k)\rangle_H$. Then (29) follows from Slutsky's Theorem. To obtain (30) and (31), by representation (10),

$$\begin{aligned}
&\left|\hat{\sigma}_k\hat{\phi}_k(t) - \sigma_k\phi_k(t)\right| = |\int_{\mathcal{T}} \hat{C}_{XY}(s,t)\hat{\psi}_k(s)ds - \int_{\mathcal{T}} C_{XY}(s,t)\psi_k(s)ds| \\
&\leqslant \int_{\mathcal{T}} |\hat{C}_{XY}(s,t) - C_{XY}(s,t)| \cdot |\hat{\psi}_k(s)|ds + \int_{\mathcal{T}} |C_{XY}(s,t)| \cdot |\hat{\psi}_k(s) - \psi_k(s)|ds \\
&\leqslant \left\{\int_{\mathcal{T}} (\hat{C}_{XY}(s,t) - C_{XY}(s,t))^2 ds\right\}^{1/2} + \left\{\int_{\mathcal{T}} C_{XY}(s,t)ds\right\}^{1/2} \|\hat{\psi}_k - \psi_k\|_H.
\end{aligned}$$

By (36) and (37), we have $|\hat{\sigma}_k\hat{\phi}_k(t)/\sigma_k - \phi_k(t)| = O_p(1/\sqrt{n}h_1h_2)$, uniformly for $t \in \mathcal{T}$. Then (30) follows with (29), and (31) follows analogously.

Proof of Theorem 3. By Theorem 1 in Yao *et al.* (2005a),

$$\sup_{s,\, t\in\mathcal{T}} \left|\hat{C}_{ZZ}(s,t) - C_{ZZ}(s,t)\right| = O_p(\frac{1}{\sqrt{n}h_Z^2}), \quad \text{for } Z = X \text{ and } Z = Y,$$

implying

$$\begin{aligned}
&\left|\langle\hat{\phi}_1, \hat{\mathcal{C}}_{XX}(\hat{\phi}_1)\rangle - \langle\phi_1, \mathcal{C}_{XX}(\phi_1)\rangle\right| \\
\leqslant\ & \iint \left|\hat{\phi}_1(s)\hat{C}_{XX}(s,t)\hat{\phi}_1(t) - \phi_1(s)C_{XX}(s,t)\phi_1(t)\right| dsdt \\
=\ & \iint |\left[(\hat{\phi}_1(s) - \phi_1(s)) + \phi_1(s)\right]\left[(\hat{C}_{XX}(s,t) - C_{XX}(s,t)) + C_{XX}(s,t)\right]\left[(\hat{\phi}_1(t) - \phi_1(t)) + \phi_1(t)\right] \\
& -\phi_1(s)C_{XX}(s,t)\phi_1(t)|dsdt \\
\leqslant\ & \left(\iint \left|(\hat{\phi}_1(s) - \phi_1(s))C_{XX}(s,t)\phi_1(t)\right| dsdt + \iint \left|\phi_1(s)(\hat{C}_{XX}(s,t) - C_{XX}(s,t))\phi_1(t)\right| dsdt \right. \\
& \left. + \iint \left|\phi_1(s)C_{XX}(s,t)(\hat{\phi}_1(t) - \phi_1(t))\right| dsdt\right) \left(1 + o_p(1)\right)
\end{aligned}$$

$$= \quad I + II + III + o_p\left(I + II + III\right).$$

Now

$$\begin{aligned} I &\leqslant \sup_{u\in\mathcal{T}} |\hat{\phi}_1(u) - \phi_1(u)| \iint |C_{XX}(s,t)\phi_1(t)|\, dsdt = O_p(\frac{1}{\sqrt{n}h_1h_2}), \\ II &\leqslant \sup_{u,v\in\mathcal{T}} |\hat{C}_{XX}(u,v) - C_{XX}(u,v)| \int |\phi_1(s)|ds \int |\phi_1(t)|dt = O_p(\frac{1}{\sqrt{n}h_X^2}). \end{aligned}$$

Analogously to I, $III = O_p(\frac{1}{\sqrt{n}h_1h_2})$. Then

$$\begin{aligned} \left|\langle\hat{\phi}_1, \hat{\mathcal{C}}_{XX}(\hat{\phi}_1)\rangle - \langle\phi_1, \mathcal{C}_{XX}(\phi_1)\rangle\right| &= O_p\left(\frac{1}{\sqrt{n}h_1h_2} + \frac{1}{\sqrt{n}h_X^2}\right), \\ \left|\langle\hat{\psi}_1, \hat{\mathcal{C}}_{YY}(\hat{\psi}_1)\rangle - \langle\psi_1, \mathcal{C}_{YY}(\psi_1)\rangle\right| &= O_p\left(\frac{1}{\sqrt{n}h_1h_2} + \frac{1}{\sqrt{n}h_Y^2}\right). \end{aligned}$$

This leads to

$$\begin{aligned} & \left|\langle\hat{\phi}_1, \hat{\mathcal{C}}_{XX}(\hat{\phi}_1)\rangle^{1/2}\langle\hat{\psi}_1, \hat{\mathcal{C}}_{YY}(\hat{\psi}_1)\rangle^{1/2} - \langle\phi_1, \mathcal{C}_{XX}(\phi_1)\rangle^{1/2}\langle\psi_1, \mathcal{C}_{YY}(\psi_1)\rangle^{1/2}\right| \\ = \quad & O_p\left(\frac{1}{\sqrt{n}h_1h_2} + \frac{1}{\sqrt{n}h_X^2} + \frac{1}{\sqrt{n}h_Y^2}\right) \end{aligned}$$

Thus, with (29),

$$\begin{aligned} |\hat{\rho}_1 - \rho_1| &= \left|\frac{\hat{\sigma}_1}{(\langle\hat{\phi}_1, \hat{\mathcal{C}}_{XX}(\hat{\phi}_1)\rangle\langle\hat{\psi}_1, \hat{\mathcal{C}}_{YY}(\hat{\psi}_1)\rangle)^{1/2}} - \frac{\sigma_1}{(\langle\phi_1, \mathcal{C}_{XX}(\phi_1)\rangle\langle\psi_1, \mathcal{C}_{YY}(\psi_1)\rangle)^{1/2}}\right| \\ &= O_p(\frac{1}{\sqrt{n}h_1h_2} + \frac{1}{\sqrt{n}h_X^2} + \frac{1}{\sqrt{n}h_Y^2}). \end{aligned}$$

Proof of Theorem 4. Under the Gaussian assumption (A5), the BLUPs of the singular component scores are immediately seen to be given by (27). By Lemma A.1, Theorem 1 and Corollary 1 of Yao *et al.* (2005a), for the estimated components in (28), $\hat{\Sigma}_{\boldsymbol{U}_i,\boldsymbol{V}_i}$ converges to $\Sigma_{\boldsymbol{U}_i,\boldsymbol{V}_i}$ in probability. For $\hat{\Sigma}_Z$, $Z = \zeta_{im}$ or ξ_{ik}, note that

$$\|\hat{\Sigma}_{\zeta_{im}} - \Sigma_{\zeta_{im}}\| \leq \|\hat{\sigma}_m\hat{\phi}_{im} - \sigma_m\phi_{im}\| + \|\hat{P}_{im} - P_{im}\| \tag{38}$$

The first r.h.s term converges to 0 in probability by Theorem 2. Define $\|f(\cdot)\|_\infty = \sup_{t\in\mathcal{T}}|f(t)|$. It follows that

$$\|\int_{\mathcal{T}} \hat{C}_{XX}(s,\cdot)\hat{\phi}_m(s)ds - \int_{\mathcal{T}} C_{XX}(s,\cdot)\phi_m(s)ds\|_\infty$$

$$
\begin{aligned}
&\leq \|\int_{\mathcal{T}}(\hat{C}_{XX}(s,\cdot)-C_{XX}(s,\cdot))\hat{\phi}_m(s)ds\|_\infty + \|\int_{\mathcal{T}} C_{XX}(s,\cdot)(\hat{\phi}_m(s)-\phi_m(s))ds\|_\infty \\
&\leq \|\int_{\mathcal{T}}(\hat{C}_{XX}(s,\cdot)-C_{XX}(s,\cdot))^2 ds\|_\infty^{1/2}[\int_{\mathcal{T}}\hat{\phi}_m^2(s)ds]^{1/2} \\
&\quad +\|\int_{\mathcal{T}} C_{XX}^2(s,\cdot)ds\|_\infty^{1/2}[\int_{\mathcal{T}}(\hat{\phi}_m(s)-\phi_m(s))^2 ds]^{1/2} \to 0, \text{ in probability,}
\end{aligned}
$$

which implies $\|\hat{P}_{im} - P_{im}\| \to 0$ in probability. Analogously to the proof of Theorem 3, $\hat{\Sigma}_Z$ is seen to converge to Σ_Z in probability, for $Z = \zeta_{im}$ or ξ_{ik}. Finally, (32) follows by the Continuous Mapping Theorem.

Acknowledgements

Thanks are due to two reviewers for helpful comments that led to numerous improvements. This research was supported in part by National Science Foundation grants DMS05-05537 and DMS08-06199.

References

Baker, C. R. (1974) Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, **186**, 273–289.

Bapna, R., Jank, W. and Shmueli, G. (2008) Price formation and its dynamics in online auctions. *Decis. Support Syst.*, **44**, 641–656.

Cupidon, J., Eubank, R., Gilliam, D. and Ruymgaart, F. (2008) Some properties of canonical correlations and variates in infinite dimensions. *Journal of Multivariate Analysis*, **99**, 1083–1104.

Dubin, J. A. and Müller, H.-G. (2005) Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association*, **100**, 872–881.

Eubank, R. L. and Hsing, T. (2008) Canonical correlation for stochastic processes. *Stochastic Processes and their Applications*, **118**, 1634–1661.

Ghebreab, S., Smeulders, A. W. M. and Adriaans, P. (2007) Predictive modeling of fmri brain states using functional canonical correlation analysis. In *Proc. AIME2007*, vol. 4594 of *Lecture Notes in Artificial Intelligence*, 1 – 8. Springer Verlag.

Gualtierotti, A. F. (1979) On cross-covariance operators. *SIAM Journal on Applied Mathematics*, **37**, 325–329.

Hannan, E. J. (1961) The general theory of canonical correlation and its relation to functional analysis. *Journal of the Australian Mathematical Society: Series A - Pure Mathematics and Statistics*, **2**, 229–242.

He, G., Müller, H.-G. and Wang, J.-L. (2003) Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, **85**, 54–77.

— (2004) Methods of canonical analysis for functional data. *Journal of Statistical Planning and Inference*, **122**, 141–159.

Heckman, N. E. and Zamar, R. H. (2000) Comparing the shapes of regression functions. *Biometrika*, **87**, 135–144.

Jank, W. and Shmueli, G. (2006) Functional data analysis in electronic commerce research. *Statistical Science*, **21**, 155–166.

Kato, T. (1995) *Perturbation theory for linear operators*. Berlin: Springer-Verlag.

Leurgans, S. E., Moyeed, R. A. and Silverman, B. W. (1993) Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **55**, 725–740.

Liu, B. and Müller, H.-G. (2008) Functional data analysis for sparse auction data. In *Statistical Methods in eCommerce Research* (eds. W. Jank and G. Shmueli), 269–290. New York: Wiley.

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

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. Springer Series in Statistics. New York: Springer, second edn.

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.

Service, S. K., Rice, J. A. and Chavez, F. P. (1998) Relationship between physical and biological variables during the upwelling period in Monterey Bay. *Deep-Sea Research Part II: Topical Studies in Oceanography*, **45**, 1669–1685.

Shmueli, G. and Jank, W. (2005) Visualizing online auctions. *Journal of Computational and Graphical Statistics*, **14**, 299–319.

Shock, N. W., Greulich, R. C., Andres, R., Lakatta, E. G., Arenberg, D. and Tobin, J. D. (1984) Normal human aging: The Baltimore Longitudinal Study of Aging. In *NIH Publication No. 84-2450*. Washington, D.C.: U.S. Government Printing Office.

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

Wegelin, J. A. (2000) A survey of partial least squares (PLS) methods, with emphasis on the two-block case. *Tech. rep.*

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

— (2005b) Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, **33**, 2873–2903.

Table 1: Simulation results for estimation of the first and second singular values $\sigma_1 = 3$, $\sigma_2 = 1.5$, where m is the average number of measurements, α is the bandwidth factor described in Section 5.1 and MISE(Z) is the mean integrated square error for estimating the first singular function of Z, $Z = X, Y$.

	Design	m	Median	Mean	Bias	Std.Dev.	MSE	MISE(X)	MISE(Y)
σ_1	sparse	2	3.16	3.21	0.212	0.703	0.538	0.164	0.158
	sparse	5	2.981	3.03	0.034	0.606	0.368	0.155	0.157
	sparse	10	3.00	2.99	−0.011	0.532	0.282	0.154	0.148
	dense	15	2.96	2.99	−0.015	0.510	0.260	0.0054	0.013
	dense	20	3.02	3.01	0.008	0.495	0.244	0.0013	0.003
σ_2	sparse	2	1.86	1.91	0.413	0.418	0.345	0.339	0.323
	sparse	5	1.47	1.49	−0.013	0.323	0.104	0.323	0.297
	sparse	10	1.36	1.36	−0.143	0.284	0.100	0.321	0.269
	dense	15	1.30	1.30	−0.198	0.273	0.114	0.012	0.021
	dense	20	1.35	1.37	−0.131	0.268	0.089	0.004	0.006

Table 2: Simulation results for three correlation measures under four scenarios from Table 1. The target value for the proposed singular value in the functional singular component analysis (FSC) approach is .75 (Eq. 26); the target value for functional canonical correlation (CAN) is .79 (Eq. 1); COR is the empirical Pearson correlation described in section 5.2.

		FSC		COR		CAN	
Design	m	Mean	Median	Mean	Median	Mean	Median
sparse	5	0.812	0.830	0.0029	0.0093	0.515	0.515
sparse	10	0.756	0.754	−0.0001	0.0006	0.752	0.751
dense	15	0.768	0.768	0.0110	0.0108	0.850	0.849
dense	20	0.759	0.759	0.0084	0.0089	0.873	0.874

Table 3: Simulation results for MISE for fitted trajectories using the functional singular component analysis (FSCA) approach and principal analysis by conditional expectation (PACE) under the Sparse5 and Dense20 scenarios.

$\times 10^{-3}$	FSCA(X)	FSCA(Y)	PACE(X)	PACE (Y)
Sparse5	4.264	5.099	3.949	4.961
Dense20	2.3595	2.527	2.574	2.393

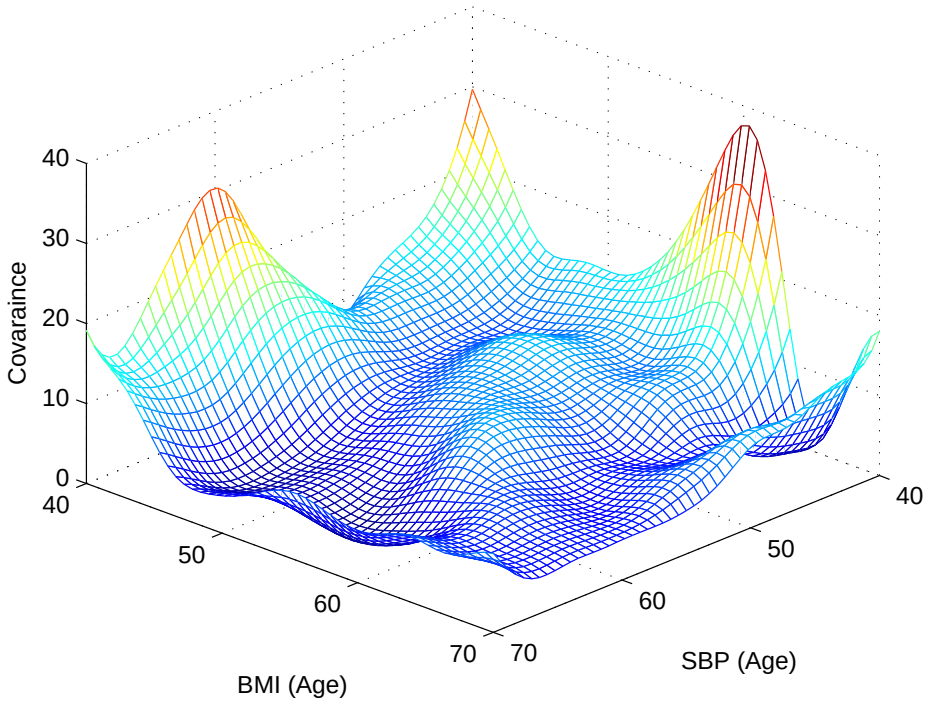

Figure 1: Smoothed cross-covariance surface between Systolic Blood Pressure (SBP) and Body Mass Index (BMI) for longitudinal aging data.

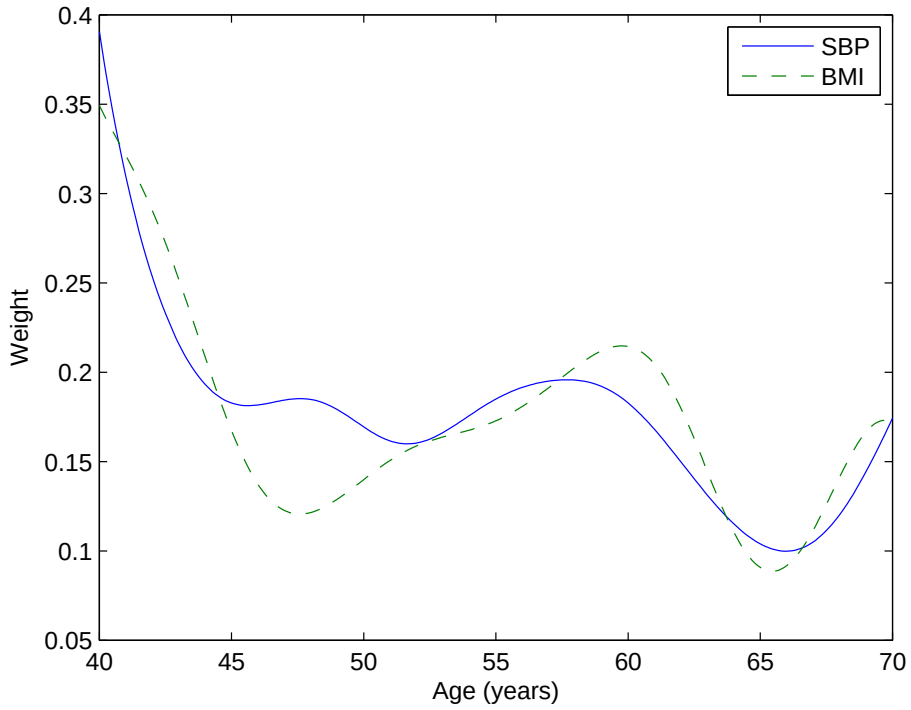

Figure 2: First pair of singular functions of Body Mass Index (BMI) and Systolic Blood Pressure (SBP) for longitudinal aging data.

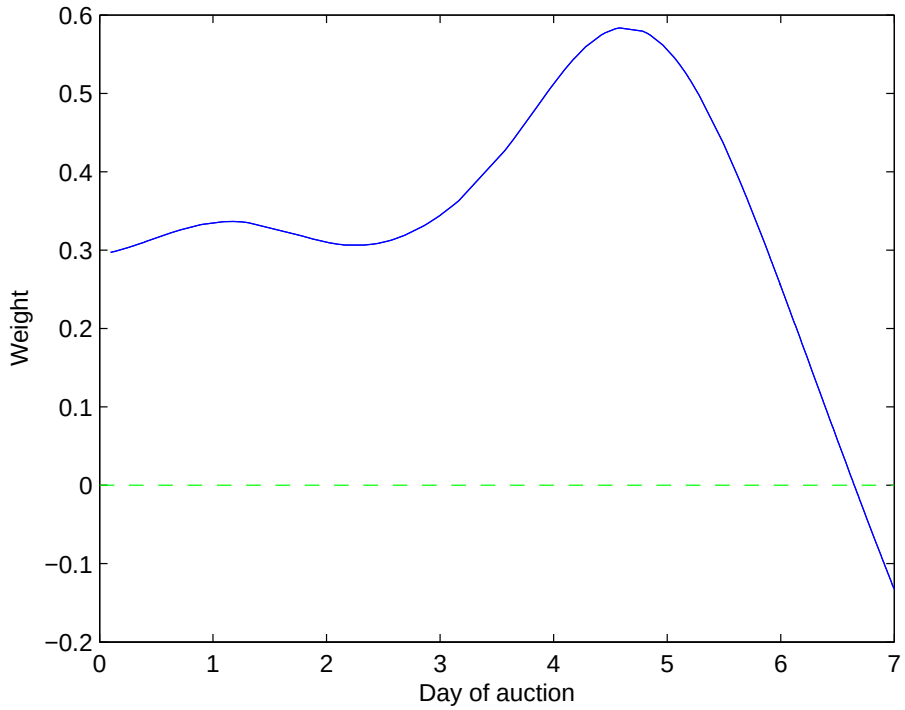

Figure 3: First singular function for e-bay auctions for functional within-bidder correlation of bid pricing.