Statistica Sinica **28** (2018), 2497-2520
doi:https://doi.org/10.5705/ss.202016.0556

# SINGULAR ADDITIVE MODELS FOR FUNCTION TO FUNCTION REGRESSION

Byeong U. Park[1], Chun-Jui Chen[2], Wenwen Tao[2] and Hans-Georg Müller[2]

[1]*Seoul National University and* [2]*University of California at Davis*

*Abstract:* In various functional regression settings one observes i.i.d. samples of paired stochastic processes $(X, Y)$ and aims at predicting the trajectory of $Y$, given the trajectory $X$. For example, one may wish to predict the future segment of a process from observing an initial segment of its trajectory. Commonly used functional regression models are based on representations that are obtained separately for $X$ and $Y$. In contrast to these established methods, often implemented with functional principal components, we base our approach on a singular expansion of the paired processes $X, Y$ with singular functions that are derived from the cross-covariance surface between $X$ and $Y$. The motivation for this approach is that the resulting singular components may better reflect the association between $X$ and $Y$. The regression relationship is then based on the assumption that each singular component of $Y$ follows an additive regression model with the singular components of $X$ as predictors. To handle the inherent dependency of these predictors, we develop singular additive models with smooth backfitting. We discuss asymptotic properties of the estimates as well as their practical behavior in simulations and data analysis.

*Key words and phrases:* Additive model, cross-covariance operator, functional data analysis, singular decomposition, smooth backfitting.

## 1. Introduction

In various regression settings one observes i.i.d. samples of paired stochastic processes $(X, Y)$, and is interested in predicting the trajectory of $Y$, given the trajectory $X$. An example of such a function to function regression problem from nephrology, which will be explored further as an illustration of our methods, features longitudinal profiles of various blood proteins, where one wishes to predict the profile of one protein given the profile of another.

We assume here that both predictors $X$ and responses $Y$ are square integrable random functions on domains $\mathcal{S}$, resp. $\mathcal{T}$, with $E(\|X\|^2) < \infty$, $E(\|Y\|^2) < \infty$, and our goal is to regress $Y$ on $X$. Predictors $X(\cdot)$ are defined on a compact domain $\mathcal{S}$ and response functions $Y(\cdot)$ on a compact domain $\mathcal{T}$. Key quantities

are the mean functions

$$\mu_X(s) = EX(s), \quad \mu_Y(t) = EY(t), \tag{1.1}$$

as well as the auto-covariance and cross-covariance functions

$$\begin{aligned} G_{XX}(s_1, s_2) &= \operatorname{cov}(X(s_1), X(s_2)), \\ G_{YY}(t_1, t_2) &= \operatorname{cov}(Y(t_1), Y(t_2)), \\ G_{XY}(s, t) &= \operatorname{cov}(X(s), Y(t)), \quad s, s_1, s_2 \in \mathcal{S},\ t, t_1, t_2 \in \mathcal{T}. \end{aligned} \tag{1.2}$$

We denote centered processes by $X^c(s) = X(s) - \mu_X(s)$, $Y^c(t) = Y(t) - \mu_Y(t)$, $s \in \mathcal{S}$, $t \in \mathcal{T}$. The commonly used linear functional regression model for regressing $Y$ on $X$ is

$$E\{Y(t)|X\} = \mu_Y(t) + \int_{\mathcal{S}} \beta(s,t) X^c(s)\, ds, \quad t \in \mathcal{T}, \tag{1.3}$$

with a smooth and square integrable regression parameter function $\beta$ (Ramsay and Silverman (2005); Morris (2015); Wang, Chiou and Müller (2016)). A popular implementation of this model, as well as the simpler functional linear model with scalar response $E(Y|X) = \mu_Y + \int_{\mathcal{S}} \beta(s) X^c(s)\, ds$, is through functional principal component (FPC) expansions of both $X$ and $Y$ (Cardot et al. (2003); Yao, Müller and Wang (2005); Hall and Horowitz (2007)) that are given by

$$X^c(s) = \sum_{k=1}^{\infty} \eta_{Xk} \varphi_{Xk}(s), \quad Y^c(t) = \sum_{m=1}^{\infty} \eta_{Ym} \phi_{Ym}(t), \tag{1.4}$$

where $\varphi_{Xk}$, $\phi_{Yk}$, $k \geq 1$, are the orthonormal eigenfunctions of the auto-covariance operators of $X$ and $Y$, respectively, and $\eta_{Xk} = \int_{\mathcal{S}} X^c(s) \varphi_{Xk}(s) ds$, $\eta_{Yk} = \int_{\mathcal{T}} Y^c(t)$ $\phi_{Yk}(t) dt$ are the functional principal components of $X$ and $Y$. Under certain regularity conditions, it can be shown that $\beta(s,t)$ in (1.3) can be represented as a limit,

$$\beta(s,t) = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{E(\eta_{Xk} \eta_{Ym})}{E(\eta_{Xk}^2)} \varphi_{Xk}(s) \phi_{Ym}(t). \tag{1.5}$$

An inherent drawback of functional principal component (FPC) based regression approaches is that they do not take into account the relationship between correlated processes $X$ and $Y$. While for any regression model with functional predictors some form of dimension reduction is needed, for which the FPC approach provides a convenient approach, the dimension reduction afforded by FPCs is likely suboptimal for regression. More specifically, the eigenbasis of $X$ that is used in (1.5) for the dimension reduction step may not provide an efficient representation of the regression parameter function $\beta$, as it ignores the dependency

between $X$ and $Y$. Other functional regression models such as functional additive models (FAM) (Müller and Yao (2008)) that utilize the FPCs of the predictor processes share the same weakness.

This motivates us to investigate an additive flexible model that uses singular components instead of principal components as arguments for the additive functions. Functional singular components are based on a functional singular value decomposition (Yang, Müller and Stadtmüller (2011)) and thus are derived from the cross-covariance $G_{XY}$ rather than the auto-covariance $G_{XX}$ as is the case for FPCs. Using singular components of predictor processes $X$ as arguments for additive modeling is expected to yield more informative representations (Zhang and Wang (2016)). The price to be paid is that the singular components for predictor processes must be considered to be dependent. This is in contrast to FPCs, which are always uncorrelated and are independent in the Gaussian case. If the predictor FPCs are independent, this makes it possible to implement FAM in a series of simple smoothing steps (Müller and Yao (2008)).

A consequence of the dependence of the singular components is that fitting a model that is additive in the singular components cannot be implemented in the same fashion as FAM and requires extra scrutiny to take the dependence of the predictors into account. In recent work of Zhang and Wang (2016), the overall goal is essentially the same as in the present paper, namely to develop an additive regression model for functional data that is additive in the singular components of predictor processes. However, while the case of dependent predictors is briefly mentioned, a crucial assumption for both theory and implementation in Zhang and Wang (2016) is that the predictor components are independent. Under this assumption, the FAM approach is applicable and no backfitting or other consideration of dependence of predictor scores is needed. Contrary to the independence of the FPCs in the Gaussian case, the independence assumption generally does not hold for singular components as predictors, irrespective of the type of predictor process. Even uncorrelatedness of the singular components cannot be assumed to hold in general; it requires special conditions that we discuss in more detail in Section A.1 of the Supplementary Materials and that are unlikely to be satisfied in general. Therefore, if one aims to develop a model that is additive in the singular components, one needs to confront the dependency issue for the predictors, as we do here. Similar considerations apply when one considers additive models for the situation where one has more than one predictor process (Han, Müller and Park (2018)).

To take the dependence of the singular predictors properly into account,

we develop a smooth backfitting approach for fitting singular additive models (SAM). The idea of smooth backfitting was introduced by Mammen, Linton and Nielsen (1999) and studied further by Mammen and Park (2005, 2006); Yu, Park and Mammen (2008); Lee, Mammen and Park (2010, 2012), who considered structural models for cross-sectional data. More recently, Zhang, Park and Wang (2013) applied smooth backfitting to an additive model for longitudinal data. As we do not directly observe the covariates in our model, which are the singular components, an additional technical challenge is to assess the effect of estimating the singular components within the framework of smooth backfitting, for which we adapt arguments of Han, Müller and Park (2018). Ordinary backfitting and additive fitting by regression splines have several disadvantages. For example, splines require one to fit a very high-dimensional model, which makes this approach less accessible to theoretical analysis, while ordinary backfitting requires a strong set of conditions for convergence (including near independence of predictors, the singular components in our case) and the estimators are not well defined since they are given as the limit of the ordinary backfitting iteration. The marginal integration method suffers from the curse of dimensionality. Our smooth backfitting estimators are defined under much weaker conditions without near independence and, importantly, smooth backfitting is amenable to theoretical analysis for the complex situation that we face since predictors are not known but must be estimated.

We review Peter Hall's contributions to functional regression and the connection of his work to our approach in Section 2, followed by a brief review of functional singular components and introduction to the singular additive model (SAM) in Section 3. Estimation of the functional singular components and additive functions in SAM with smooth backfitting is the topic of Section 4, with consistency results in Section 5. In Section 6 we report the results of a simulation study that shows the advantages of using SAM in comparison to a FPCA based linear model implementation, and in Section 7 we present a data illustration for a data set from nephrology, followed by a brief discussion in Section 8. Theoretical derivations and proofs are provided in an online Supplement.

## 2. Peter Hall and Functional Regression

We dedicate this article to the memory of Peter Hall. The work we report here is closely related to his research in functional linear models and functional principal component analysis (FPCA). Peter was a leader in nonparamet-

ric statistics, and he contributed to many areas, notably the bootstrap, the area where he made his name in the earlier stages of his career. In his later years, he wrote a substantial body of influential papers in Functional Data Analysis (FDA) and was a major force in the rapid development of this area since 2006 (Müller (2016)). His first paper in FDA appeared in 1998, with a focus on the estimation of modes of the distribution of functional data (Gasser, Hall and Presnell (1998)).

FDA is among the last research areas in which Peter made seminal contributions before his premature death in early 2016. It was a good fit for him, as it presents complex theoretical issues at the interface of smoothing, multivariate analysis, functional analysis and stochastic processes in Hilbert spaces (Hsing and Eubank (2015)), all fields in which Peter had accumulated substantial experience and a large and sophisticated toolbox. FDA presented (and still presents) challenging problems that enabled Peter and his various collaborators to solve some tough problems. Peter was a dedicated problem solver and his productivity was phenomenal. He usually wrote a paper in record time, sometimes substituting the original problem for one that was solvable, and often deriving results and writing the paper in one step.

Peter's major contributions to FDA were in the subareas of FPCA, functional linear regression and single index models, as well as densities and modes for functional data and functional classification. In the area of functional linear models, Peter and his collaborators focused on the case of a continuous scalar response variable coupled with a functional predictor, distinguishing between the prediction problem where the goal is to estimate linear predictors $\theta = \int \beta(s)X(s)\,ds$ that correspond to projections on regression slope functions $\beta$ for the scalar response case and the regression problem. In the latter, the goal is to estimate the function $\beta$. In this work, Peter and his collaborators adopted a traditional approach and used FPC expansions of the predictors to expand the function $\beta$ in the eigenbasis (Cai and Hall (2006); Hall and Horowitz (2007); Delaigle, Hall and Apanasovich (2009)).

This led to precise convergence rates and shed light on the differences between prediction and estimation tasks in well defined scenarios, where prediction was revealed to be an easier task, associated with faster rates of convergence, relative to the estimation of the regression parameter function $\beta$, a consequence of the smoothing effect of the integral in the predictors $\theta$.

Peter and co-authors contributed also to other aspects of functional regression models with linear predictors, specifically single index models (Chen, Hall and Müller (2011)), predictor component selection (Hall and Yang

(2010)) and domain selection for functional predictors (Hall and Hooker (2016)). Peter's paper with Yang (Hall and Yang (2010)) is especially relevant for our approach. In addition to developing theory for the cross-validation choice of the number of principal components to be included in a functional regression, this paper contains a nice discussion of the pros and cons of the FPCA-based implementation of functional linear models, as the FPCs are only derived from predictor processes and are not influenced in any way by the responses, which can be a downside.

The method we discuss here is based on singular components that are derived from the covariance of $X$ and $Y$ and therefore reflect the dependence between predictor and response processes. Also related to our approach is the partial least squares method that has been developed for the case of functional predictors in Delaigle and Hall (2012). Partial least squares is notoriously difficult to analyze, due to its iterative nature, which makes the analysis of the functional case particularly complex. A point of connection with the singular additive model that we study here is that partial least squares also aims to maximize covariance between predictor and response, rather than maximizing correlation, as basic linear regression does. This approach has the major benefit that it avoids the inverse problem associated with functional linear regression (Yang, Müller and Stadtmüller (2011)).

## 3. Singular Components and Singular Additive Model

To define the singular components for pairs of random functions $(X, Y)$, we discuss special linear operators in Hilbert spaces $L^2(\mathcal{S})$ and $L^2(\mathcal{T})$. Specifically, singular decompositions are based on auto-covariance operators $\mathcal{C}_{XX}$ and $\mathcal{C}_{YY}$ and cross-covariance operators $\mathcal{C}_{XY}$ and $\mathcal{C}_{YX}$ as follows (Gualtierotti (1979); Preda and Saporta (2005); Yang, Müller and Stadtmüller (2011)):

$$\mathcal{C}_{XX}: L^2(\mathcal{S}) \to L^2(\mathcal{S}), f \mapsto g, g(s) = \int_{\mathcal{S}} C_{XX}(s,t) f(t) dt, C_{XX}(s,t) = E\{X^c(s) X^c(t)\},$$

$$\mathcal{C}_{YY}: L^2(\mathcal{T}) \to L^2(\mathcal{T}), f \mapsto g, g(s) = \int_{\mathcal{T}} C_{YY}(s,t) f(t) dt, C_{YY}(s,t) = E\{Y^c(s) Y^c(t)\},$$

$$\mathcal{C}_{XY}: L^2(\mathcal{T}) \to L^2(\mathcal{S}), f \mapsto g, g(s) = \int_{\mathcal{T}} C_{XY}(s,t) f(t) dt, C_{XY}(s,t) = E\{X^c(s) Y^c(t)\},$$

$$\mathcal{C}_{YX}: L^2(\mathcal{S}) \to L^2(\mathcal{T}), f \mapsto g, g(s) = \int_{\mathcal{S}} C_{YX}(s,t) f(t) dt, C_{YX}(s,t) = E\{Y^c(s) X^c(t)\}.$$

Here $\mathcal{C}_{YX}$ is the adjoint operator of $\mathcal{C}_{XY}$, while the compound operators $\mathcal{A}_{XYX} = \mathcal{C}_{XY} \circ \mathcal{C}_{YX}$ and $\mathcal{A}_{YXY} = \mathcal{C}_{YX} \circ \mathcal{C}_{XY}$ are self-adjoint Hilbert-Schmidt

operators with $L^2$-kernels

$$A_{XYX}(s,t) = \int_{\mathcal{T}} C_{XY}(s,u)\, C_{YX}(u,t)\, du = \int_{\mathcal{T}} C_{XY}(s,u)\, C_{XY}(t,u)\, du, \quad (3.1)$$

$$A_{YXY}(s,t) = \int_{\mathcal{S}} C_{YX}(s,u)\, C_{XY}(u,t)\, du = \int_{\mathcal{S}} C_{XY}(u,s)\, C_{XY}(u,t)\, du\,. \quad (3.2)$$

The operators $\mathcal{A}_{XYX}$ and $\mathcal{A}_{YXY}$ 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 $\mathcal{A}_{XYX}$ and $\psi_1, \psi_2, \ldots$ for $\mathcal{A}_{YXY}$, respectively, which satisfy $\phi_k = (1/\sigma_k)\mathcal{C}_{XY}(\psi_k)$, $k = 1, 2, \ldots$. The $\phi_j$ and $\psi_j$ are the singular functions and the $\sigma_j$ are the singular values, for $j \geq 1$.

The singular functions usually will form genuine subspaces of $L^2$, that might be finite-dimensional, depending on the nature of the relation between $X$ and $Y$, with the unexplained parts of the infinite-dimensional processes $X$ and $Y$ contained in remainder processes $\nu_X$ and $\nu_Y$ as in (3.5) below; these remainder processes are unrelated to the regression relation to the extent it is determined by $C_{XY}$. The decomposition of both predictor and response processes into a part that is spanned by $M$ singular functions and a second part that corresponds to the remainder process motivates to model regression relations between $X$ and $Y$ by using only the first $M$ singular components, where we make the assumption that $M$ is finite but unknown. So while the functions $X$ and $Y$ are infinite-dimensional as is commonly assumed in FDA, the regression relation is assumed to only involve finitely many singular components.

If the cross-covariance operators are of rank $M$, where $M < \infty$, we obtain the representations (Yang, Müller and Stadtmüller (2011)),

$$\begin{aligned} \mathcal{C}_{XY}(f)(t) &= \sum_{k=1}^{M} \sigma_k \langle f, \psi_k \rangle \phi_k(t), \quad t \in \mathcal{S}, \\ \mathcal{C}_{YX}(f)(t) &= \sum_{k=1}^{M} \sigma_k \langle f, \phi_k \rangle\, \psi_k(t), \quad t \in \mathcal{T}, \end{aligned} \quad (3.3)$$

whence

$$\sup_{\|u\|=\|v\|=1} \operatorname{cov}(\langle u, X \rangle, \langle v, Y \rangle) = \sigma_1,$$

and the maximum is attained at $u = \phi_1$ and $v = \psi_1$. Repeating the maximization on sequences of orthogonal complements generates the singular values $(\sigma_1, \sigma_2, \ldots)$ and associated singular functions $u_k = \phi_k$; $v_k = \psi_k, k = 1, 2, \ldots,$ leading to representations of cross-covariance surfaces

$$C_{XY}(s,t) = \sum_{k=1}^{M} \sigma_k \phi_k(s) \psi_k(t),$$
$$C_{YX}(s,t) = \sum_{k=1}^{M} \sigma_k \psi_k(s) \phi_k(t), \tag{3.4}$$

The Hilbert-Schmidt theorem then implies the singular decompositions

$$X(s) = \tilde{X}(s) + \nu_X(s),\ \tilde{X}(s) = \mu_X(s) + \sum_{m=1}^{M} \zeta_m \phi_m(s),$$
$$Y(t) = \tilde{Y}(t) + \nu_Y(t),\ \tilde{Y}(t) = \mu_Y(t) + \sum_{k=1}^{M} \xi_k \psi_k(t), \tag{3.5}$$

where $\nu_X \in L^2$ is in the kernel of the operator $\mathcal{A}_{XYX}$, a function with $\mathcal{A}_{XYX}(\nu_X) = 0$, and analogously $\nu_Y \in L^2$ is in the kernel of $\mathcal{A}_{YXY}$. The random functions $\nu_X$ and $\nu_Y$ are remainder functions with zero means and zero cross-covariance that are unrelated to the cross-covariance operators and cannot be represented by the singular functions $\phi_j$ and $\psi_j$.

The singular components of $X$ are $\zeta_m = \int X^c(s)\phi_m(s)ds$, $1 \leqslant m \leqslant M$, the coefficients of $X$ with respect to its expansion in the orthonormal functions $\{\phi_m\}_{1 \leq m \leq M}$ and $\xi_k = \int Y^c(t)\psi_k(t)dt$, $1 \leqslant k \leqslant M$, the components of $Y$ with respect to its expansion in the orthonormal functions $\{\psi_k\}_{1 \leq k \leq M}$. From (3.4), one finds

$$E\zeta_m = 0,\ E\xi_k = 0,\ E(\zeta_m \xi_k) = 0 \text{ for } m \neq k, \text{ and } E(\zeta_m \xi_m) = \sigma_m. \tag{3.6}$$

While the remainder processes $\nu_X$ and $\nu_Y$ are uncorrelated, $\nu_X$ is uncorrelated with the $\xi_k$ and $\nu_Y$ is uncorrelated with the $\zeta_j$, we make the stronger assumption that $\nu_X$ is independent of $Y$ and $\nu_Y$ is independent of $X$. We obtain $E\{Y(t)|X\} = E\{\tilde{Y}(t)|X\}$, while $\nu_X$ plays the role of an additional error in the predictor that is unrelated to the response. Like in some errors-in-variables regression approaches it is then more meaningful to replace the original regression target $E(Y|X)$ by the target $E(Y|\tilde{X})$, a model where predictors are not contaminated by unrelated errors. Here it is fortuitous that denoised predictors $\tilde{X}$ can be readily obtained from the data by the singular representation. This provides the motivation to consider the functional regression model $E(Y|\tilde{X})$ as our target, which can be written as

$$E(Y|\zeta_1, \ldots, \zeta_M) = \mu_Y + \sum_{k=1}^{M} E(\xi_k|\zeta_1, \ldots, \zeta_M)\psi_k. \tag{3.7}$$

We further postulate an additive structure for each response process singular component. With $M$ included components, this leads to the model

$$E(\xi_k|\zeta_1,\ldots,\zeta_M) = f_{k0} + \sum_{j=1}^{M} f_{kj}(\zeta_j), \quad k = 1,\ldots,M, \tag{3.8}$$

where $f_{k0}$ is an unknown constant and $f_{kj}$ for $1 \le j \le M$ are unknown univariate functions, so that

$$E\{Y(t)|\zeta_1,\ldots,\zeta_M\} = \mu_Y(t) + \sum_{k=1}^{M} \left\{ f_{k0} + \sum_{j=1}^{M} f_{kj}(\zeta_j) \right\} \psi_k(t). \tag{3.9}$$

Let $I_{kj}$ denote given intervals on which one aims to estimate the component functions $f_{kj}$. The univariate functions $f_{kj}$ at (3.8) are then subject to the constraints

$$\int_I f_{kj}(u_j) p(\mathbf{u})\, d\mathbf{u} = 0, \quad 1 \le j,k \le M, \tag{3.10}$$

which are necessary for identifiability of the $f_{kj}$, where $I = I_{k1} \times \cdots \times I_{kM}$ and $p$ is the joint density function of $(\zeta_1,\ldots,\zeta_M)$. The constant $f_{k0}$ and the component functions $f_{kj}$ depend on the intervals $I_{kj}$ and the associated constraints in such a way that they differ only by a constant if the intervals and constraints change. Importantly, their sum $f_{k0} + \sum_{j=1}^{M} f_{kj}(\zeta_j)$ does not depend on the choice of these intervals and $I$. To see this, express $E(\xi_k|\zeta_1,\ldots,\zeta_M)$ as $f^*_{k0} + \sum_{j=1}^{M} f^*_{kj}(\zeta_j)$ with $f^*_{kj}$ satisfying $\int_{-\infty}^{\infty} f^*_{kj}(u_j) p_j(u_j) = 0$ for all $1 \le j \le M$, where the $p_j$ denote the marginal densities of $\zeta_j$. Expressing $E(\xi_k|\zeta_1,\ldots,\zeta_M)$ also as $f_{k0} + \sum_{j=1}^{M} f_{kj}(\zeta_j)$ with $f_{kj}$ now satisfying $\int_I f_{kj}(u_j) p(\mathbf{u})\, d\mathbf{u} = 0$ for all $1 \le j \le M$, it holds that

$$f_{kj}(u_j) = f^*_{kj}(u_j) - \left\{ \int_I p(\mathbf{u})\, d\mathbf{u} \right\}^{-1} \int_I f^*_{kj}(u_j) p(\mathbf{u})\, d\mathbf{u}, \quad 1 \le j \le M,$$

$$f_{k0} = f^*_{k0} + \sum_{j=1}^{M} \left\{ \int_I p(\mathbf{u})\, d\mathbf{u} \right\}^{-1} \int_I f^*_{kj}(u_j) p(\mathbf{u})\, d\mathbf{u}.$$

## 4. Estimation

We assume throughout that the sample of realizations $(X_i, Y_i)_{i=1,\ldots,n}$ of functional processes $X$ and $Y$ consists of random trajectories that are either fully observed, or are sampled at a dense and regular grid. In the latter case, the estimates described in the following require an additional interpolation step.

### 4.1. Estimation of singular functions and singular components

For the estimation of the singular values and singular functions $\{(\sigma_j, \phi_j, \psi_j) :$

$1 \le j < M\}$ as well as singular components $\zeta_j$ and $\xi_k$, the starting point are cross-sectional averages to estimate the cross-covariance surfaces,

$$\hat{C}_{YX}(s,t) = n^{-1}\sum_{i=1}^{n}\{Y_i(s) - \hat{\mu}_Y(s)\}\{X_i(t) - \hat{\mu}_X(t)\},$$

$$\hat{C}_{XY}(s,t) = n^{-1}\sum_{i=1}^{n}\{X_i(s) - \hat{\mu}_X(s)\}\{Y_i(t) - \hat{\mu}_Y(t)\},$$

where $\hat{\mu}_X(t) = n^{-1}\sum_{i=1}^{n} X_i(t)$ and $\hat{\mu}_Y(t) = n^{-1}\sum_{i=1}^{n} Y_i(t)$. These are the building blocks for the estimation of singular functions and singular values. Then the shared eigenvalues and eigenfunctions of the integral operators $\hat{\mathcal{A}}_{XYX}$ and $\hat{\mathcal{A}}_{YXY}$, based on estimated kernels $\hat{A}_{XYX}(s,t) = \int_{\mathcal{T}} \hat{C}_{XY}(s,u)\hat{C}_{YX}(u,t)\,du$ and $\hat{A}_{YXY}(s,t) = \int_{\mathcal{S}} \hat{C}_{YX}(s,u)\hat{C}_{XY}(u,t)\,du$, respectively, are obtained by numerical eigen-decomposition of suitably discretized versions of these estimated kernels. The resulting shared eigenvalue estimates $\hat{\sigma}_1^2 \ge \hat{\sigma}_2^2 \ge \cdots$, which correspond to the singular value estimates, are then ordered in declining order.

Denoting the corresponding orthonormal eigenfunctions of $\hat{\mathcal{A}}_{XYX}$ by $\hat{\phi}_j$, and those of $\hat{\mathcal{A}}_{YXY}$ by $\hat{\psi}_j$, the resulting singular components are $(\hat{\sigma}_j, \hat{\phi}_j, \hat{\psi}_j)$. The singular components $\zeta_{ij} = \int X_i^c(s)\phi_j(s)ds$ and $\xi_{ij} = \int Y_i^c(s)\psi_j(s)ds$ for $X_i$ and $Y_i$ are then obtained by numerically approximating the integrals $\hat{\zeta}_{ij} = \int\{X_i(s) - \hat{\mu}_X(s)\}\hat{\phi}_j(s)ds$ and $\hat{\xi}_{ij} = \int\{Y_i(s) - \hat{\mu}_Y(s)\}\hat{\psi}_j(s)dt$.

### 4.2. Estimation in the singular additive model

We implement the smooth backfitting idea of Mammen, Linton and Nielsen (1999) to fit model (3.8) for each singular component $\xi_k$ of $Y$, aiming to solve the integral equations

$$f_{kj}(u_j) = E(\xi_k|\zeta_j = u_j) - \sum_{l\neq j}^{M}\int f_{kl}(u_l)\frac{p_{jl}(u_j,u_l)}{p_j(u_j)}\,du_l, \quad 1 \le j \le M, \tag{4.1}$$

where $p_j$ and $p_{jl}$ are the marginal and joint density functions, respectively, of $\zeta_j$ and $(\zeta_j, \zeta_l)$. The main idea is to estimate the unknown functions in (4.1), $E(\xi_k|\zeta_j = u_j)$, $p_{jl}(u_j,u_l)$ and $p_j(u_j)$, plug the estimators into (4.1) and then solve the estimated integral equations.

The singular components $\zeta_j$ will usually have unbounded supports. We consider estimating the additive regression function $E(\xi_k|\zeta_1,\ldots,\zeta_M)$ at (3.8) only on a compact subset of the support of $\boldsymbol{\zeta} = (\zeta_j : 1 \le j \le M)$, however. This is in the same spirit as the usual practice in nonparametric regression, namely to estimate the nonparametric regression function on a compact set. Since the

predictors $\zeta_{ij}$ and $\xi_{ik}$ are not available but need to be estimated, and since the domains where the $f_{kj}$ are estimated are different from the supports of $\zeta_j$, it is necessary to modify the existing methodology and theory of smooth backfitting for the current setting.

Let $I_{kj}$ denote the intervals where one wants to estimate the component functions $f_{kj}$. The univariate functions $f_{kj}$ at (3.8) are subject to the constraints

$$\int_I f_{kj}(u_j)p(\mathbf{u})\,d\mathbf{u} = 0, \quad 1 \le j,k \le M, \tag{4.2}$$

which are necessary for identifiability of $f_{kj}$, where $I = I_{k1} \times \cdots \times I_{kM}$ and $p$ denotes the joint density function of $(\zeta_1, \ldots, \zeta_M)$. Here the constant $f_{k0}$ and the component functions $f_{kj}$ change if the intervals $I_{kj}$ in the constraints change. However, the corresponding component functions differ from each other only by constants, and their sum $f_{k0} + \sum_{j=1}^M f_{kj}(\zeta_j)$ does not depend on the choice of $I$ as was demonstrated at the end of Section 3.

In the following, we omit the index $k$ for the singular component of $Y$ in $I_{kj}$, writing $I_j \equiv I_{kj}$. With $p_j^I(u_j) = \{\int_I p(\mathbf{u})\,d\mathbf{u}\}^{-1} \int_{I_{-j}} p(\mathbf{u})\,d\mathbf{u}_{-j}$, $I_{-j} = \prod_{l \ne j} I_l$, (4.2) is equivalent to

$$\int_{I_j} f_{kj}(u_j)p_j^I(u_j)\,du_j = 0, \quad 1 \le j \le M. \tag{4.3}$$

In the current setting one cannot adopt the usual constraints $\int_{-\infty}^{\infty} f_{kj}(u)$ $p_j(u)\,du = 0$, where $p_j$ is the marginal density of $\zeta_j$, since this requires the estimation of $f_{kj}$ on the entire support of $\zeta_j$. One can employ constraints other than (4.2), for example, $\int_{I_j} f_{kj}(u)w_j(u)\,du = 0$ for some known weight functions $w_j$. One technical advantage of the latter is that it leads to the constraint $\int_{I_j} \hat{f}_{kj}(u)w_j(u)\,du = 0$ for the estimator $\hat{f}_{kj}$ that uses the same known weight $w_j$, so that one need not carry out an additional asymptotic analysis of the constraints for the estimators. Because of this advantage, the latter approach was adopted in Lee, Mammen and Park (2010, 2012). In contrast, (4.3) leads to a constraint for the estimator whereby the density $p_j^I$ is replaced by an estimated density. This requires asymptotic analysis of the effects of estimating $p_j^I$ on the statistical properties of the estimator of $f_{kj}$. Nevertheless, we choose the constraint (4.3) since it is natural and yields simpler forms for $f_{k0}$ and its estimator. The methods and theory that we describe below can be modified accordingly if one uses a different constraint.

To derive an analogue of (4.1), we define $p_0^I = \int_I p(\mathbf{u})\,d\mathbf{u}$ and $p_{jl}^I(u_j, u_l) = \int_{I_{-jl}} p(\mathbf{u})\,d\mathbf{u}_{-jl}/p_0^I$, where $\mathbf{u}_{-jl}$ is the vector $\mathbf{u}$ with $(u_j, u_l)$ deleted and $I_{-jl} =$

$\prod_{l' \neq j,l} I_{l'}$. Then

$$f_{kj}(u_j) = \frac{1}{p_0^I \cdot p_j^I(u_j)} \int_{I_{-j}} E(\xi_k|\boldsymbol{\zeta} = \mathbf{u}) p(\mathbf{u})\, d\mathbf{u}_{-j} - f_{k0} - \sum_{l \neq j}^{M} \int_{I_l} f_{kl}(u_l) \frac{p_{jl}^I(u_j, u_l)}{p_j^I(u_j)}\, du_l, \quad 1 \leq j \leq M, \tag{4.4}$$

where $f_{k0} = \int_I E(\xi_k|\boldsymbol{\zeta} = \mathbf{u}) p(\mathbf{u})\, d\mathbf{u}/p_0^I$ under the constraints (4.3). For the estimation of the integral equation (4.4), if the singular components $\zeta_{ij}$ and $\xi_{ik}$ were available, we could use these for solving the backfitting equation, in which case the asymptotics would be a straightforward extension of the existing theory of smooth backfitting. Since the singular functions $\phi_j$ and $\psi_k$ are also unknown, we replace them by corresponding estimators as defined in Section 4.1. A major technical challenge is to find suitable bounds to control the effect of estimating the singular components $\zeta_{ij}$ and $\xi_{ik}$ on the estimation of the additive functions $f_{kj}$.

Define a scaled kernel function

$$K_{h_j}(u, v) = \mathbb{I}(u \in I_j) \frac{K_{h_j}(u - v)}{\int_{I_j} K_{h_j}(t - v)\, dt} \tag{4.5}$$

whenever $\int_{I_j} K_{h_j}(t - v)\, dt \neq 0$, and $K_{h_j}(u, v) = 0$ otherwise, where $K_{h_j}(u - v) = h_j^{-1} K(h_j^{-1}(u - v))$ for a baseline kernel $K$, and a bandwidth $h_j$. Observing that $p_j^I$ and $p_{jl}^I$ are conditional densities of $\zeta_j$ and $(\zeta_j, \zeta_l)$, respectively, given that the event $\boldsymbol{\zeta} \in I$ occurs, suggests the estimates

$$\begin{aligned} \hat{p}_j^I(u) &= \frac{n^{-1} \sum_{i=1}^n K_{h_j}(u, \hat{\zeta}_{ij}) \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I)}{\hat{p}_0^I}, \\ \hat{p}_{jl}^I(u, v) &= \frac{n^{-1} \sum_{i=1}^n K_{h_j}(u, \hat{\zeta}_{ij}) K_{h_l}(v, \hat{\zeta}_{il}) \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I)}{\hat{p}_0^I}, \end{aligned} \tag{4.6}$$

where $\hat{p}_0^I = n^{-1} \sum_{i=1}^n \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I)$ and $\mathbb{I}$ is the indicator. The definition of the scaled kernel function entails

$$\int_{I_j} \hat{p}_j^I(u)\, du = 1, \quad \int_{I_l} \hat{p}_{jl}^I(u, v)\, dv = \hat{p}_j^I(u).$$

We also estimate $f_{k0}$ by $\hat{f}_{k0} = n^{-1} \sum_{i=1}^n \hat{\xi}_{ik} \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I)/\hat{p}_0^I$, and the first term on the right hand side of (4.4) by

$$\tilde{f}_{kj}(u) = \left\{ n^{-1} \sum_{i=1}^n K_{h_j}(u, \hat{\zeta}_{ij}) \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I) \right\}^{-1} n^{-1} \sum_{i=1}^n \hat{\xi}_{ik} K_{h_j}(u, \hat{\zeta}_{ij}) \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I). \tag{4.7}$$

Our smooth backfitting estimator $(\hat{f}_{kj} : 1 \leq j \leq M)$ of $(f_{kj} : 1 \leq j \leq M)$ is

defined as the solution of the backfitting system of equations

$$\hat{f}_{kj}(u) = \tilde{f}_{kj}(u) - \hat{f}_{k0} - \sum_{l \neq j}^{M} \int_{I_l} \hat{f}_{kl}(v) \frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)}\, dv, \quad 1 \leq j \leq M, \tag{4.8}$$

subject to the constraints

$$\int_{I_j} \hat{f}_{kj}(u) \hat{p}^I_j(u)\, du = 0, \quad 1 \leq j \leq M. \tag{4.9}$$

An iteration scheme to obtain the solution of the equation (4.8) starts with an initial tuple $(\hat{f}^{[0]}_{kj} : 1 \leq j \leq M)$, updating $(\hat{f}^{[r]}_{kj} : 1 \leq j \leq M)$ in the $r$th cycle by

$$\begin{aligned} \hat{f}^{[r]}_{kj}(u) = \tilde{f}_{kj}(u) - \hat{f}_{k0} &- \sum_{l=1}^{j-1} \int_{I_l} \hat{f}^{[r]}_{kl}(v) \frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)}\, dv \\ &- \sum_{l=j+1}^{M} \int_{I_l} \hat{f}^{[r-1]}_{kl}(v) \frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)}\, dv. \end{aligned} \tag{4.10}$$

Once the estimators of all components are in hand, we predict the response process $Y$ by

$$\hat{Y}(t) = \hat{\mu}_Y(t) + \sum_{k=1}^{M} \sum_{j=1}^{M} \hat{f}_{kj}(\hat{\zeta}_j) \hat{\psi}_k(t). \tag{4.11}$$

We demonstrate in Section 5 that with probability tending to one the backfitting equation (4.8) has a unique solution and the iterative algorithm (4.10) converges to the solution exponentially fast, under weak conditions.

## 5. Theoretical Results

### 5.1. Consistency of the singular functions and singular components

The convergence rates of the estimators of the singular functions $\hat{\phi}_j$ and $\hat{\psi}_j$ and of the singular components are key auxiliary results and are based on

$$E\|\hat{\mathcal{A}}_{XYX} - \mathcal{A}_{XYX}\|_{\text{op}} = O(n^{-1/2}), \quad E\|\hat{\mathcal{A}}_{YXY} - \mathcal{A}_{YXY}\|_{\text{op}} = O(n^{-1/2}), \tag{5.1}$$

where $\|\cdot\|_{\text{op}}$ denotes the operator norm. These results hold if $E\|X\|^2\|Y\|^2 < \infty$. Under the additional assumptions that the eigenvalues $\sigma_1^2, \ldots, \sigma_{j+1}^2$ are separated and that $E\|X\|^{2\alpha}\|Y\|^{2\alpha} < \infty$ for some $\alpha \geq 2$, properties (5.1) ultimately lead to the following results for the maximal errors of the estimated singular components,

$$\max_{1 \leq i \leq n} |\hat{\zeta}_{ij} - \zeta_{ij}| = O_p(n^{-(\alpha-1)/2\alpha}), \quad \max_{1 \leq i \leq n} |\hat{\xi}_{ij} - \xi_{ij}| = O_p(n^{-(\alpha-1)/2\alpha}), \tag{5.2}$$

with further details provided in Section A.2 of the Supplementary Materials.

### 5.2. Consistency of the estimated singular additive model

Without loss of generality, assume that $I_j = [0,1]$ for all $1 \le j \le M$ and let $p_j^{(1)}(\mathbf{u}) = \partial p(\mathbf{u})/\partial u_j$ with $f'_{kj}$, $f''_{kj}$, respectively, denoting first and second derivatives of $f_{kj}$. Define

$$\tilde{\beta}_{kj}(u) = \int v^2 K(v)\,dv \sum_{l=1}^{M} c_l^2 E\left\{ f'_{kl}(\zeta_l)\frac{p_l^{(1)}(\boldsymbol{\zeta})}{p(\boldsymbol{\zeta})}\Big|\zeta_j = u, \boldsymbol{\zeta} \in I\right\},$$

$$\tau_{kj}^2(u) = \frac{1}{p_0^I p_j^I(u)} c_j^{-1}\mathrm{Var}(\xi_k|\zeta_j = u, \boldsymbol{\zeta} \in I)\int K(v)^2\,dv,$$

$$\beta_{kj}(u) = \beta_{kj}^*(u) + \frac{1}{2}c_j^2 f''_{kj}(u)\int u^2 K(u)\,du,$$

where constants $c_j$ are as in condition (A2) below, and the tuple $(\beta_{kj}^* : 1 \le j \le M)$ is the solution of the system of equations

$$\beta_{kj}^*(u) = \tilde{\beta}_{kj}(u) - \sum_{l\neq j}^{M}\int_{I_l} \beta_{kl}^*(v)\frac{p_{jk}^I(u,v)}{p_j^I(u)}\,dv, \quad 1 \le j \le M,$$

subject to the constraints

$$\int_0^1 \beta_{kj}^*(u)p_j^I(u)\,du = c_j^2\int_0^1 f'_{kj}(u)\frac{\partial}{\partial u}p_j^I(u)\,du\int u^2K(u)\,du.$$

We need the following assumptions.

(A1) The baseline kernel function $K$ is bounded, has compact support $[-1,1]$, is symmetric around zero, differentiable and its derivative is Lipschitz continuous.

(A2) The bandwidths $h_j$ satisfy $n^{1/5}h_j \to c_j$ for some positive constants $c_j$.

(A3) The joint density $p$ of $\boldsymbol{\zeta}$ is bounded away from zero and infinity on $I$.

(A4) The additive functions $f_{kj}$ are twice continuously differentiable and the densities $p_j$ and $p_{jk}$ are (partially) continuously differentiable on $[0,1]$.

(A5) $E|\xi_k|^c < \infty$ for $c > 5/2$ and $\mathrm{Var}(\xi_k|\zeta_j = \cdot, \boldsymbol{\zeta} \in I)$ are continuous on $[0,1]$.

(A6) $E\|X\|^{2\alpha} < \infty$ and $E\|Y\|^{2\alpha} < \infty$ for some $\alpha > 5$ and the eigenvalues $\sigma_1^2, \ldots, \sigma_{M+1}^2$ are separated.

Assumptions (A1)–(A4) are widely assumed in kernel smoothing theory. The moment condition (A5) is also typical for response variables in regression models, which is $\xi_k$ in our case, while (A6) is used to prove (5.2) for $\alpha > 5$, which entails

that both $\max_{1\le i\le n}|\hat{\zeta}_{ij}-\zeta_{ij}|$ and $\max_{1\le i\le n}|\hat{\xi}_{ij}-\xi_{ij}|$ are of smaller order than the univariate rate $n^{-2/5}$.

**Theorem 1.** *Assume* (A1)–(A6). *Then,* (i) *with probability tending to one, there exists a unique solution* $(\hat{f}_{kj}: 1\le j\le M)$ *of* (4.8) *subject to the constraints* (4.9); (ii) *there exist constants* $0<\gamma<1$ *and* $c>0$ *such that with probability tending to one*

$$\int_0^1 \left\{\hat{f}_{kj}^{[r]}(u)-\hat{f}_{kj}(u)\right\}^2 p_j^I(u)\,du \le c\cdot\gamma^{2r}\left\{1+\sum_{j=1}^{M}\int_0^1 \hat{f}_{kj}^{[0]}(u)^2 p_j^I(u)\,du\right\};$$

(iii) *for a given vector* $(\mathbf{u}: 0<u_j<1, 1\le j\le M)$, *the estimators* $\hat{f}_{kj}(u_j)$ *for* $1\le j\le M$ *are asymptotically independent and*

$$n^{2/5}\left\{\hat{f}_{kj}(u_j)-f_{kj}(u_j)\right\}\xrightarrow{d} N\left(\beta_{kj}(u_j),\tau_{kj}^2(u_j)\right).$$

If the true singular components $\zeta_{ij}$ and $\xi_{ik}$ are used in the estimators $\tilde{f}_{kj}$, $\hat{p}_j^I$ and $\hat{p}_{jl}^I$, then Theorem 1 is a straightforward extension of the existing theory of smooth backfitting, as one only needs to take care of the truncation $\mathbb{I}(\hat{\boldsymbol{\zeta}}_i\in I)$ in $\tilde{f}_{kj}$, $\hat{p}_j^I$ and $\hat{p}_{jl}^I$. Thus, the main step in the proof of Theorem 1 is to show that the estimation of the singular components $\zeta_{ij}$ and $\xi_{ik}$ has a negligible effect on the convergence of $\hat{f}_{kj}^{[r]}$ and on the first-order asymptotic properties of the estimators $\hat{f}_{kj}$. The proof of the theorem is in Section A.3 of the Supplementary Materials.

## 6. Simulation Results

We generated paired random processes $X_i, Y_i$ with given singular components by

$$X_i(s)=\sum_{j=1}^{K}\zeta_{ij}\phi_j(s)+\mu_x(s)\quad\text{and}$$

$$Y_i(t)=\sum_{j=1}^{K}\xi_{ij}\psi_j(t)+\mu_y(t),\quad s\in\mathcal{S}=[0,S],\quad t\in\mathcal{T}=[0,T],$$

with $K=4$, $S=10$, $T=5$ and $\{\phi_j(s),\psi_j(t)\}$, $j=1,\dots,4$, $s\in\mathcal{S}$, $t\in\mathcal{T}$, chosen as Fourier basis with $\phi_1(s)=\sqrt{2/S}\sin(2\pi s/S)$, $\phi_2(s)=-\sqrt{2/S}\cos(4\pi s/S)$, $\phi_3(s)=\sqrt{2/S}\sin(6\pi s/S)$, $\phi_4(s)=-\sqrt{2/S}\cos(8\pi s/S)$, $\mu_x(s)=\sin(s)+s$, and $\psi_1(t)=-\sqrt{2/T}\cos(2\pi t/T)$, $\psi_2(t)=\sqrt{2/T}\sin(4\pi t/T)$, $\psi_3(t)=-\sqrt{2/T}\cos(6\pi t/T)$, $\psi_4(t)=\sqrt{2/T}\sin(8\pi t/T)$, $\mu_y(t)=\sin(t)+t$. The random predictor vector $\zeta=\{\zeta_j\}$, $j=1,\dots,4$, was generated by a normal distribution with zero mean and covariance matrix

$$\mathrm{cov}(\zeta) = \begin{bmatrix} 16 & 6 & 2 & -2 \\ 6 & 8 & 3 & 1 \\ 2 & 3 & 4 & 1.5 \\ -2 & 1 & 1.5 & 2 \end{bmatrix},$$

and the vector of response coefficients $\xi = \{\xi_j\}$, $j = 1, \ldots, 4$, using the additive functions $f_{11}(x) = -1.98 - 0.11x + 0.12x^2$, $f_{12}(x) = -0.58 + 0.13x + 0.07x^2$, $f_{13}(x) = -0.30 + 0.11x + 0.08x^2$, $f_{14}(x) = 2.03 + 0.16x - 1.01x^2$, $f_{21}(x) = -1.85 + 0.08x + 0.12x^2$, $f_{22}(x) = 1.35 + 0.11x - 0.17x^2$, $f_{23}(x) = -0.26 - 0.12x + 0.07x^2$, $f_{24}(x) = 0.04 + 0.47x - 0.02x^2$, $f_{31}(x) = 0.12 + 0.04x - 0.01x^2$, $f_{32}(x) = 1.10 + 0.14x - 0.14x^2$, $f_{33}(x) = -0.81 - 0.71x + 0.20x^2$, $f_{34}(x) = 0.05 + 0.28x - 0.02x^2$, $f_{41}(x) = -0.79 + 0.06x + 0.05x^2$, $f_{42}(x) = 0.08 - 0.01x^2$, $f_{43}(x) = -0.12 + 0.20x + 0.03x^2$, $f_{44}(x) = -0.32 + 0.14x + 0.16x^2$. The random coefficient vectors $\xi_i$ for the $i$-th subject were then obtained as $\xi_{ik} = \sum_{j=1}^{4} f_{kj}(\zeta_{ij})$, $k = 1, \ldots, 4$. The additive functions $f_{jk}$ were constructed to satisfy the constraints

$$E\{f_{ij}(\zeta_j)\} = 0, \quad E(\zeta_m \xi_k) = 0 \quad \text{for} \quad m \neq k, \quad E(\zeta_m \xi_m) = \sigma_m.$$

The design points $s_l, t_l$ on $[0, S]$ and $[0, T]$ where functions are sampled were chosen as 100 equidistant points, respectively, and observations were generated as

$$\tilde{X}(s_l) = X(s_l) + \nu_X(s_l), \quad \tilde{Y}(t_l) = Y(t_l) + \nu_Y(t_l),$$

where $\nu_X, \nu_Y$ are remainder processes as in (3.5), obtained as $\nu_X(s_l) = z_{x1}\rho_{x1}(s_l) + z_{x2}\rho_{x2}(s_l)$ and $\nu_Y(t_l) = z_{y1}\rho_{y1}(t_l) + z_{y2}\rho_{y2}(t_l)$, with $z_{x1}, z_{y1} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $z_{x2}, z_{y2} \overset{i.i.d.}{\sim} N(0, 0.5\sigma^2)$ and

$$\rho_{x1}(s_l) = \sqrt{2/S}\sin(12\pi s_l/S), \qquad \rho_{x2}(s_l) = -\sqrt{2/S}\cos(10\pi s_l/S),$$
$$\rho_{y1}(t_l) = -\sqrt{2/T}\cos(12\pi t_l/T), \quad \rho_{y2}(t_l) = \sqrt{2/T}\sin(10\pi t_l/T).$$

As in (3.8) we consider the number of components $M$ to be the same for predictor and response processes and report simulation results for combinations of sample sizes $n = 100, 500$ and factors $\sigma = 1, 5$ by which $\nu_X, \nu_Y$ are multiplied, allowing for different magnitudes of the remainder processes. We report the values of Integrated Squared Prediction Error, ISPE $= \int\{\tilde{Y}(t) - \hat{Y}(t)\}^2 dt$, for the proposed singular additive modeling (SAM) approach, and also for the functional linear model (FLM) as in (1.3), (1.5) and the functional additive model (FAM). The latter is additive in the functional principal component (FPC) scores of predictor processes $X$, $E(\eta_{Ym}|X) = \sum_{j=1}^{\infty} g_{mj}(\eta_{Xj})$, using the principal component scores $\eta_{Ym}$ of response processes $Y$ and $\eta_{Xj}$ of predictor processes $X$ as defined in (1.4) (Müller and Yao (2008)).

Table 1. The 25th, 50th and 75th percentiles of scaled integrated squared prediction error comparing the proposed singular additive modeling (SAM), the functional additive model (FAM) (Müller and Yao, 2008) and the functional linear model (FLM), as in (1.3), (1.5). Results are based on 400 simulation runs for sample size $n = 100$. Model training and prediction is done by 5-fold cross-validation, where $M$ is the number of singular components for both predictor and response processes, where $\sigma^2$ indicates the strength of the residual processes.

| | | $n = 100$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 25th | | | 50th | | | 75th | | |
| $M$ | $\sigma^2$ | SAM | FAM | FLM | SAM | FAM | FLM | SAM | FAM | FLM |
| 3 | 1 | 3.60 | 3.89 | 5.22 | 7.03 | 8.56 | 11.09 | 15.15 | 22.70 | 29.22 |
| 4 | 1 | 3.56 | 4.26 | 5.41 | 6.93 | 9.26 | 11.49 | 14.56 | 24.00 | 29.70 |
| 5 | 1 | 3.53 | 4.70 | 5.60 | 6.88 | 10.10 | 11.94 | 14.47 | 25.46 | 30.39 |
| 3 | 5 | 7.84 | 9.09 | 10.36 | 14.59 | 17.64 | 20.01 | 26.70 | 35.02 | 40.21 |
| 4 | 5 | 7.68 | 9.15 | 10.33 | 14.19 | 17.76 | 19.87 | 25.86 | 35.41 | 40.05 |
| 5 | 5 | 7.56 | 9.36 | 10.26 | 14.07 | 18.34 | 19.85 | 25.48 | 36.68 | 40.03 |

Table 2. Same as Table 1, but for $n = 500$.

| | | $n = 500$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 25th | | | 50th | | | 75th | | |
| $M$ | $\sigma^2$ | SAM | FAM | FLM | SAM | FAM | FLM | SAM | FAM | FLM |
| 3 | 1 | 3.23 | 3.32 | 4.67 | 6.13 | 7.35 | 9.76 | 13.27 | 19.60 | 26.64 |
| 4 | 1 | 3.15 | 3.41 | 4.72 | 5.95 | 7.53 | 9.84 | 12.72 | 19.85 | 26.70 |
| 5 | 1 | 3.11 | 3.51 | 4.76 | 5.91 | 7.69 | 9.95 | 12.52 | 20.06 | 26.82 |
| 3 | 5 | 6.88 | 8.26 | 9.53 | 12.99 | 16.05 | 18.51 | 23.93 | 31.38 | 37.56 |
| 4 | 5 | 6.76 | 7.79 | 9.22 | 12.76 | 15.28 | 17.79 | 23.30 | 30.12 | 36.41 |
| 5 | 5 | 6.66 | 7.58 | 8.96 | 12.59 | 15.01 | 17.40 | 22.97 | 29.84 | 35.88 |

When implementing SAM, here and in our data analysis in Section 6, we standardized each of the estimated singular components $\hat{\zeta}_{ij}$ and subsequently chose the intervals $I_{jk}$ in (3.10) as $I_{kj} = [-2, 2]$ for the standardized values, then after fitting transformed back to the original scale when reporting the results. The tuning parameters for SAM, except for $M$, which was fixed at various levels, were chosen by 5-fold cross-validation.

From the results in Table 1 (for $n = 100$) and Table 2 (for $n = 500$) we find that SAM performs consistently better than FLM or FAM in these comparisons for all quantiles of ISPE that were considered. As expected, the ISPEs increase for larger values of $\sigma$ and decrease for larger sample size. The second best performer is FAM, followed by FLM.

In a second simulation we generated singular components for predictor pro-

Table 3. Same as Table 1, but for $\sigma = 1$ only in the second simulation scenario where data are generated with two rather than four singular components.

| | $n = 100$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 25th | | | 50th | | | 75th | | |
| $M$ | SAM | FAM | FLM | SAM | FAM | FLM | SAM | FAM | FLM |
| 3 | 1.74 | 2.91 | 5.99 | 3.39 | 4.94 | 9.16 | 7.07 | 10.51 | 15.82 |
| 4 | 1.55 | 3.07 | 5.91 | 3.08 | 5.51 | 9.40 | 6.36 | 11.63 | 16.88 |
| 5 | 1.49 | 3.29 | 5.90 | 2.98 | 6.11 | 9.63 | 6.20 | 12.96 | 17.84 |

cesses as

$$\begin{bmatrix} \zeta_{i1} \\ \zeta_{i2} \end{bmatrix} \sim N(\mathbf{0}, \Sigma), \qquad \text{where} \qquad \Sigma = \begin{bmatrix} 8 & 3 \\ 3 & 4 \end{bmatrix}$$

and for response processes as $\xi_{i1} = f_{11}(\zeta_{i1}) + f_{12}(\zeta_{i2})$, $\xi_{i2} = f_{21}(\zeta_{i1}) + f_{22}(\zeta_{i2})$, with additive functions $f_{11}(x) = -1.84 - 0.37x + 0.23x^2$, $f_{12}(x) = -1.21 + 0.27x + 0.30x^2$, $f_{21}(x) = 1.58 + 0.11x - 0.20x^2$, $f_{22}(x) = -2.05 - 0.28x + 0.51x^2$. The following singular functions and mean functions were chosen for predictor processes $X$, $\phi_1(s) = \sqrt{2/S}\sin(2\pi s/S)$, $\phi_2(s) = -\sqrt{2/S}\cos(6\pi s/S)$, $\mu_x(s) = \sin(s) + s$, and for response processes $Y$, $\psi_1(t) = -\sqrt{2/T}\cos(2\pi t/T)$, $\psi_2(t) = \sqrt{2/T}\sin(6\pi t/T)$, $\mu_y(t) = \sin(t) + t$. All other settings were as above. In Figure 1 we demonstrate the surface estimates of the additive regressions of $\xi_1$ on $\zeta_1, \zeta_2$, and of $\xi_2$ on $\zeta_1, \zeta_2$, for $n = 100$ and $n = 500$ for this second simulation scenario. We show estimates with close to median mean integrated squared errors (MISE) among 400 simulations and find that these surface estimates improve as the sample size gets larger. Table 3 indicates that in terms of ISPE, the results are similar to those in the first simulation scenario.

## 7. Data Analysis

We demonstrate the comparative performance of FLM and SAM for data that were obtained in a nephrological study for 32 hemodialysis patients (Kaysen et al. (2000)). For each patient the expression levels of acute phase blood proteins were collected longitudinally. Exploring the longitudinal relationship between the negative acute phase protein Albumin (alb) and positive acute phase protein $\alpha$-aminoglobulin (aag), we use aag as predictor and alb as response. To avoid biases resulting from non-uniform observation designs, we removed the observations falling within the first and last 5% of the design points. Since the measurement times were more and more spread out away from the origin with increasing spacings, we log-transformed them, which led to more regular designs.

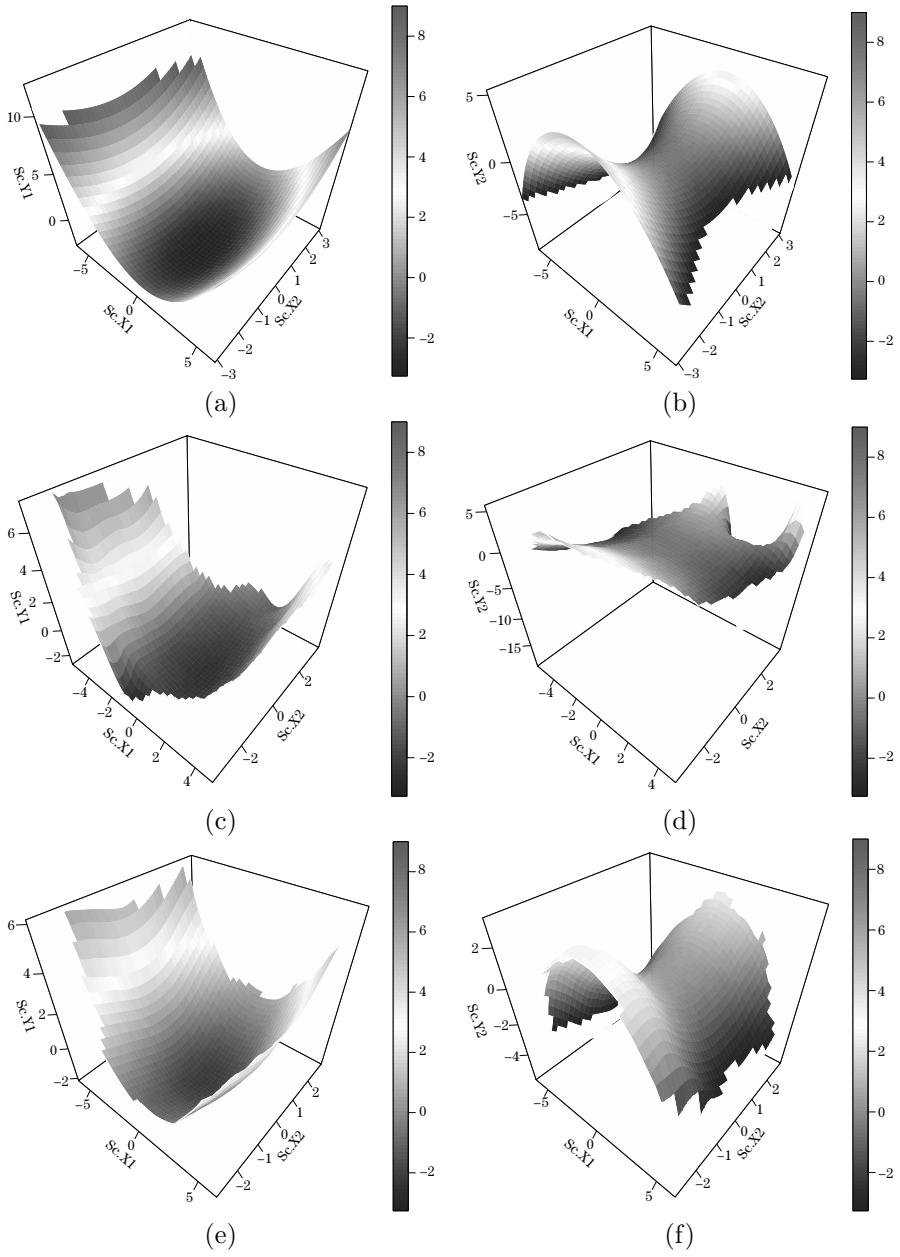


Figure 1. First row left for first (a) and right for second (b) singular component of response processes $Y$ versus first and second singular component of predictor processes $X$, true relationships. Second row depicts corresponding estimates (c) and (d) with SAM for $n = 100$ and third row in (e) and (f) for $n = 500$. These estimates were selected to have an MISE that is near the median over 400 Monte Carlo runs.

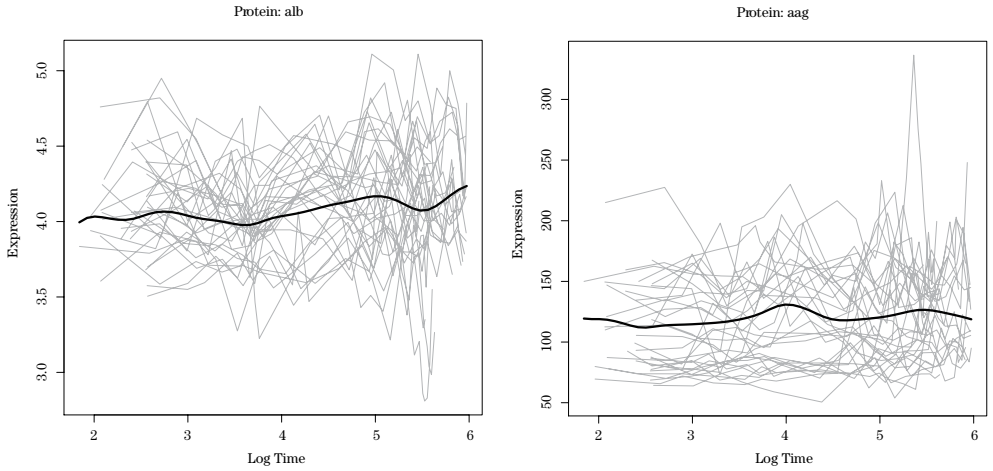


Figure 2. Longitudinal recordings of Albumin (left panel) and of $\alpha$-aminoglobulin (right panel) with log time scale for $n = 32$ subjects. The thick black lines indicate the corresponding group means obtained by local linear kernel smoothing.

The spaghetti plots of alb and aag are shown in Figure 2.

The quartiles of integrated squared prediction error ISPE obtained when applying SAM and FLM using five-fold cross-validation are reported in Table 4. The selection of the number of included components for SAM was based on the number of components selected by applying the BIC criterion for FLM (Yang, Müller and Stadtmüller (2011)). This is expected to favor the FLM. Nevertheless, the results in Table 3 show that the overall predictive performance of SAM is somewhat better than that of FLM. We conclude that one can often achieve improvements by implementing functional linear models via singular additive modeling.

## 8. Discussion

We did not fully investigate the choice of the number of included components $M$, for which we used the data-based BIC criterion that is geared towards the functional linear model but may provide a suboptimal choice for the singular additive model. A full investigation of this choice is left as a topic for future research.

For the derivation and implementation of the proposed singular additive model (SAM) with smooth backfitting, we have made several assumptions that are plausible but nevertheless restrictive. A basic assumption is that there are only finitely many singular components that are sufficient to explain the re-

Table 4. The 25th, 50th and 75th percentiles of integrated squared prediction error ISPE for the proposed singular additive model (SAM) and the functional linear model (FLM), when predicting Albumin trajectories from $\alpha$-aminoglobulin trajectories for $n = 32$ subjects. Model training and prediction was done by selecting separate tuning and test sets, using 5-fold cross-validation.

| | $n = 32$ | |
|---|---|---|
| | SAM | FLM |
| 25th | 0.24 | 0.25 |
| 50th | 0.38 | 0.39 |
| 75th | 0.57 | 0.85 |

gression relation between $X$ and $Y$ and that the remainders $\nu_X$ and $\nu_Y$ of the infinite-dimensional processes $X$ and $Y$ can be ignored. This assumption seems at least more plausible than a corresponding assumption on the FPCs, since the latter are constructed without taking into account the interaction between $X$ and $Y$. This is the main motivation for functional singular component analysis (Yang, Müller and Stadtmüller (2011)) and singular component based functional regression (Zhang and Wang (2016)). A second restrictive assumption is that the functional data are fully observed without errors, which is rarely if ever the case in practical applications. In situations where trajectories are observed with noise or on an irregular grid, a pre-smoothing step can be employed, but to get the requisite uniform bounds, additional restrictive assumptions on the underlying smooth processes $X$ and $Y$ are needed (see, e.g., Müller, Stadtmüller and Yao (2006)).

A third assumption is that additive regression models are reasonable and that the idea of dimension reduction through singular components carries forward to additive models. Additional assumptions are needed for the fitting of the additive model, as described in detail in the theory section. Also, we are only able to estimate the additive functions on compact intervals even though the predictor scores are typically not bounded.

On the other hand, we have strengthened the case that using singular components can be advantageous not only for functional correlation but also when using additive models, and presumably also other functional regression models when both predictors and responses are functional. As we demonstrate, the proposed smooth backfitting approach works reasonably well. By adopting smooth backfitting, we are able to deal with the dependency of the predictor scores that is an inherent feature of functional singular components.

## Supplementary Materials

These consist of Section A.1., where the dependency of the functional singular components is discussed, Section A.2, which contains a proof of the important auxiliary result given in (5.2), and Section A.3 with the proof of Theorem 1.

## Acknowledgment

We thank the Editor of the Special Issue, an associate editor and referees for helpful comments, which led to numerous improvements. The research of B. Park was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2015R1A2A1A05001753). The research of H.G. Müller was supported by National Science Foundation grants DMS-1407852 and DMS-1712864.

# References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34**, 2159–2179.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics* **30**, 241–255.

Chen, D., Hall, P. and Müller, H. G. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics* **39**, 1720–1747.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics* **40**, 322–352.

Delaigle, A., Hall, P. and Apanasovich, T. (2009). Weighted least squares methods for prediction in the functional data linear model. *Electronic Journal of Statistics* **3**, 865–885.

Gasser, T., Hall, P. and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **60**, 681–691.

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

Hall, P. and Hooker, G. (2016). Truncated linear models for functional data. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **78**, 637–653.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35**, 70–91.

Hall, P. and Yang,Y. J. (2010). Ordering and selecting components in multivariate or functional data linear prediction. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **72**, 93–110.

Han, K., Müller, H.-G. and Park, B. U. (2018). Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple

predictor functions. *Bernoulli* **24**, 1233–1265.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Kaysen, G. A., Dubin, J. A., Müller, H.-G., Rosales, L. M. and Levin, N. W. (2000). The acute-phase response varies with time and predicts serum albumin levels in hemodialysis patients. *Kidney International* **58**, 346–352.

Lee, Y. K., Mammen, E. and Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *The Annals of Statistics* **38**, 2857–2883.

Lee, Y. K., Mammen, E. and Park, B. U. (2012). Flexible generalized varying coefficient regression models. *The Annals of Statistics* **40**, 1906–1933.

Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics* **27**, 1443–1490.

Mammen, E. and Park, B. U. (2005). Bandwidth selection for smooth backfitting in additive models. *The Annals of Statistics* **33**, 1260–1294.

Mammen, E. and Park, B. U. (2006). A simple smooth backfitting method for additive models. *The Annals of Statistics* **34**, 2252–2271.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application* **2**, 321–359.

Müller, H.-G. (2016). Peter Hall, Functional data analysis and random objects. *The Annals of Statistics* **44**, 1867–1887.

Müller, H.-G., Stadtmüller, U. and Yao, F. (2006). Functional variance processes. *Journal of the American Statistical Association* **101**, 1007–1018.

Müller, H.-G. and Yao, F. (2008). Functional additive models. *Journal of the American Statistical Association* **103**, 1534–1544.

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

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd edition. Springer Series in Statistics, New York, Springer.

Wang, J.-L., Chiou, J. M. and Müller, H.-G. (2016). Functional Data Analysis. *Annual Review of Statistics and its Application* **3**, 257–295.

Yang, W., Müller, H.-G. and Stadtmüller, U. (2011). Functional singular component analysis. *Journal of the Royal Statistical Society: Series B Statistical Methodology* **73**, 303–324.

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

Yu, K., Park, B. U. and Mammen, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics* **36**, 228–260.

Zhang, T. and Wang, Q. (2016). Functional singular component analysis based functional additive models. *Science China Mathematics* **59**, 2443–2462.

Zhang, X., Park, B. U. and Wang, J.-L. (2013). Time-varying additive models for longitudinal data. *Journal of the American Statistical Association* **108**, 983–998.

Department of Statistics, Seoul National University, Seoul 151-747, Korea.
E-mail: bupark@stats.snu.ac.kr
Department of Statistics, University of California, Davis, CA 95616, USA.
E-mail: cjuchen@ucdavis.edu
Department of Statistics, University of California, Davis, CA 95616, USA.
E-mail: shirlytao@gmail.com
Department of Statistics, University of California, Davis, CA 95616, USA.
E-mail: hgmueller@ucdavis.edu

(Received December 2016; accepted March 2018)