TEST manuscript No.
(will be inserted by the editor)

Dynamic Relations for Sparsely Sampled Gaussian Processes

Hans-Georg Müller · Wenjing Yang

Received: date / Accepted: date

Abstract In longitudinal studies, it is common to observe repeated measurements data from a sample of subjects where noisy measurements are made at irregular times, with a random number of measurements per subject. Often a reasonable assumption is that the data are generated by the trajectories of a smooth underlying stochastic process. In some cases one observes multivariate time courses generated by a multivariate stochastic process. To understand the nature of the underlying processes, it is then of interest to relate the values of a process at one time with the value it assumes at another time, and also to relate the values assumed by different components of a multivariate trajectory at the same time or at specific times selected for each trajectory. In addition, an assessment of these relationships will allow to predict future values of an individual's trajectories.

Derivatives of the trajectories are frequently more informative than the time courses themselves, for instance in the case of growth curves. It is then of great interest to study the estimation of derivatives from sparse data. Such estimation procedures permit the study of time-dynamic relationships between derivatives and trajectory levels within the same trajectory and between the components of multivariate trajectories. Reviewing and extending recent work, we demonstrate the estimation of corresponding empirical dynamical systems and demonstrate asymptotic consistency of predictions and dynamic transfer functions. We illustrate the resulting prediction procedures and empirical first order differential equations with a study of the dynamics of longitudinal functional data for the relationship of blood pressure and body mass index.

Keywords Derivatives · Eigenfunction · Functional Data Analysis · Gaussian Process

Hans-Georg Müller · Wenjing Yang
Department of Statistics, University of California, Davis, CA 95616 USA
E-mail:

1 Introduction

Functional data analysis (FDA) deals with data samples that consist of curves or other infinite-dimensional data objects. Original approaches in FDA were devised to target data that correspond to a sample of fully observed random trajectories, such as obtained in automatic monitoring systems; for a systematic and accessible introduction to this field, see Ramsay and Silverman (2002, 2005); the review articles Rice (2004); Zhao et al. (2004); Müller (2005, 2008) provide additional perspectives.

A common approach is to view the observed curves as realizations of a random process which produces smooth trajectories. Often, however, it is more realistic to assume that the curves are only intermittently or even only sparsely observed and that the measurements are contaminated by noise. Inclusion of sparsely observed data broadens the reach of FDA considerably, as it allows to apply this powerful methodology to all kinds of repeated measures and longitudinal data. Such data are found in e-commerce and auction bid prices (Jank and Shmueli, 2006), genetic trait models (Kirkpatrick and Heckman, 1989), gene expression time courses (Müller et al., 2008), growth studies (Gervini and Gasser, 2005), and many other types of longitudinal studies in the social and life sciences.

The future is likely to see further widening of the scope of FDA by making functional approaches applicable to an ever increasing class of data-driven problems. A common thread of such extensions is the notion of an underlying smooth stochastic process coupled with some random mechanism that generates the data, given the process. The data for each subject or experimental unit may then be viewed as resulting from a realization of the underlying process, and the relation of the observations to these trajectories is determined by the nature of the data-generating mechanism. It is of interest and at the same time may be statistically challenging to infer the overall properties of the underlying process from the available (often sparse) data. An additional task is to recover an individual's underlying but unobserved trajectory to the extent possible.

Often it is convenient to assume that the underlying processes are Gaussian so that the properties of the underlying process are determined by mean and covariance functions. This assumption leads to various additional benefits. As has been demonstrated in various contexts, many of the FDA techniques are fairly robust in terms of violations of the Gaussian assumption (see, e.g. Müller et al., 2008), so that this assumption is less restrictive than it might seem. In terms of practical performance, the analysis of non-Gaussian data is therefore expected to also benefit from the application of these methods.

Key techniques for the representation and modeling of functional data include smoothing, warping methods (also known as alignment or curve registration), functional principal component analysis (FPCA) (Rice and Silverman, 1991; Yao et al., 2005a) and functional regression (Ramsay and Dalzell, 1991; Yao et al., 2005b). These methods generally provide a "global" approach in that the entire random function is included in the modeling approach. Here we are interested in studying the underlying random functions at "local" times by quantifying the relationships of the values assumed by one or several random functions at the same or at different times.

Consider the case where the observed data may be modeled as being generated by a pair $\{X(t), Y(t)\}$ of underlying random functions. One then can study the

relationship between the two functions X and Y by an extension of correlation analysis such as functional canonical correlation (Leurgans et al., 1993) or functional linear regression (Ramsay and Dalzell, 1991). In such standard functional analyses, in order to predict Y from X, a typical procedure is to infer X and Y over their entire time domains from the sparse measurements, as exemplified in Yao et al. (2005b), and then to use the resulting representations of the trajectories to implement these methods. These implementations require inversion of a compact operator and therefore correspond to solving an ill-posed problem, necessitating the construction of a generalized inverse. Especially in the case of functional canonical correlation, straightforward approaches are not stable and are highly sensitive to the choice of smoothing and regularization parameters (Dauxois et al., 1982; He et al., 2003, 2004).

In the following, we will always refer to the argument of the underlying processes as "time", for the sake of simplicity. In some application settings, a major drawback of the "global" approaches (which include functional canonical correlation, functional linear, functional nonparametric and functional additive regression) is that the entire predictor process is associated with the entire response process, i.e., future values of X will influence current and past observations of Y in functional regression models. Especially in longitudinal studies, but also in many other applications of FDA, one would like to disallow the influence of future predictor values on current or past responses. For example, it often will make sense to relate $Y(t)$ to $X(t)$ such as in a varying coefficient model (Şentürk and Müller, 2009) or $Y(t)$ to $X(s)$, $s < t$, i.e., a current response to a specific previous predictor value. If one studies situations with one underlying process X per subject, one also may wish to relate $X(t)$ to $X(s)$, $s < t$.

Even more intriguing are the possibilities provided by the estimation of derivatives of underlying random trajectories for sparse longitudinal data, as recently developed in Liu and Müller (2009), where the concept of a transfer function was introduced which relates $X'(t)$, the first derivative of the process X at t, to $X(t)$, the current value at t, for Gaussian processes. Alternatively, one could consider $X(s)$, $s < t$, as predictor for $X'(t)$. Studying such relationships may illuminate the dynamics of the underlying processes and shed light on the typically unknown mechanisms that govern the generation of the longitudinal observations.

The nonparametric analysis of sparsely sampled trajectories which we consider here rests on two principles: Dimension reduction of the assumed underlying smooth trajectories, which are infinite dimensional; and borrowing strength from the entire sample, since there are not enough data available to estimate individual trajectories nonparametrically. The dimension reduction is achieved by projecting on a finite-dimensional space, usually assumed to be generated by a finite set of basis functions. These functions can be prespecified, e.g., as trigonometric functions, B-splines (Shi et al., 1996; Rice and Wu, 2001; Zhou et al., 2008) or P-splines (Yao and Lee, 2006), in which case one can connect the analysis to common random effects modeling and linear mixed effects models.

Alternatively, one can determine the basis functions on which to project data-adaptively, where the selection of the basis is guided by some target criterion. A common approach is to use as target maximization of the variance explained by each finite set of the first K selected basis functions, $K \geq 1$, which leads to the eigenbasis. This basis consists of the orthonormal eigenfunctions of the auto-covariance operator of the underlying stochastic process. The estimation of

the eigenfunctions (Rice and Silverman, 1991; Staniswalis and Lee, 1998) is stable even for very sparse designs and by now there are various implementations available (James, 2002; Yao et al., 2005a; Liu and Müller, 2009).

Another issue is the (even better behaved) estimation of the overall mean function, for which one may pool all data together to form one "big" scatterplot. In the subsequent smoothing step, one then can either take the dependency of the measurements originating from the same subject into account (Lin and Carroll, 2001a,b; Wang et al., 2005) or ignore it in the implementation, however not in the theoretical analysis (Yao et al., 2005a). Ignoring these dependencies usually works surprisingly well in practice, although their presence poses some theoretical difficulties that need to be dealt with in the asymptotic analysis. It is noteworthy that the methods we discuss in the following require no more than square integrability and smoothness of the underlying processes. In particular, stationarity is not needed.

The paper is organized as follows: In the following Section 2 we summarize some models for functional regression with emphasis on the case where both predictors and responses are functions and on related topics and discuss some related topics. In Section 3 we give a brief account of recent developments for the estimation of derivatives in the difficult situation where the data are sparse and irregular. Algorithmic aspects are discussed in Section 4. Extensions of the recent concept of dynamic transfer functions for one process are the theme of Section 5, and further extensions to multivariate processes are described in Section 6. Estimation methods and a result on asymptotic consistency can be found in Section 7, while data illustrations of these methods are the topic of Section 8, followed by concluding remarks in 9.

2 Brief Overview on Selected Topics in Functional Data Analysis

In Functional Data Analysis (FDA) modeling one assumes that observed data for n subjects are generated by a sample of random curves $X_i(t)$, $t \in \mathcal{T}$, $i = 1, \ldots, n$, corresponding to i.i.d. realizations of an underlying stochastic process. Alternatively, for each subject one may have observations generated from multivariate random curves, where e.g. for the bivariate case, which we consider in the following, the underlying random curves will be denoted by $(X_i(t), Y_i(t))$, $t \in \mathcal{T}$, $i = 1, \ldots, n$. The domain $\mathcal{T}$ is an interval, and the trajectories X_i, Y_i are smooth (twice differentiable) and square integrable on $\mathcal{T}$.

FDA provides methodology for the estimation of mean and covariance functions of the underlying processes, their functional principal components, and the analysis of the relationship between Y and X through suitable correlation and regression approaches. As has been demonstrated in various settings (James et al., 2000; Yao et al., 2005a; Hall et al., 2008), the FDA paradigm to extend classical statistical methodology to the case of functional data can be worked out even when one has available only sparse and irregularly spaced measurements for each subject or item in the sample. A topic that has been of much interest lately is modeling variation in time in addition to variation in amplitude. Such time variation is addressed by warping or alignment methods (Gasser and Kneip, 1995; Ramsay and Li, 1998; Gervini and Gasser, 2004, 2005; Kneip and Ramsay, 2008; Tang and Müller, 2009) and is often implemented as a pre-processing step before

methods such as functional principal component analysis or functional regression are carried out. We assume in the following that warping has been addressed and the functional data have been aligned.

We are interested in modeling the relation between the values of a random function or of multiple random functions and their derivatives at specified time points. For relating multivariate functions with each other globally, various concepts of functional correlation and regression have been developed over the years. The extension of the classical concept of canonical correlation (Hotelling, 1936) to functional data was proposed in Leurgans et al. (1993). This approach requires the inversion of a compact linear operator in Hilbert space and therefore corresponds to an ill-posed problem. Various alternative functional correlation measures were therefore proposed which avoid this problem to various degrees (Service et al., 1998; Heckman and Zamar, 2000; Dubin and Müller, 2005; Eubank and Hsing, 2008). An interesting application of such alternative measures is the computation of functional partial correlations for a gene dependency network, where the partial correlations are derived from multivariate gene expression time courses (Opgen-Rhein and Strimmer, 2006).

Regarding functional regression, one can distinguish several classes of models. Both predictors and responses can be scalars, vectors or functions and this allows for various combinations. A functional regression model is characterized by the inclusion of a functional predictor or a functional response or both. In this paper, we focus on the case where both predictors and responses are random functions; however, the most studied functional linear model connects a scalar response with a functional predictor and can be written as

$$E(Y|X) = \mu_Y + \int_{\mathcal{T}} (X(s) - \mu_X(s))\alpha(s)\, ds,$$

where Y is a scalar response, $E(Y) = \mu_Y$, X is the predictor process and α the so-called regression parameter function. This function is generally assumed to be smooth and square integrable.

Various implementations and asymptotic results have been developed for this functional predictor model (Faraway, 1997; Cuevas et al., 2002; Cardot et al., 2003a,b, 2007; Mas and Pumo, 2008), including optimality considerations (Cai and Hall, 2006; Hall and Horowitz, 2007). An extension is the generalized functional linear model (GFLM) (James, 2002; Escabias et al., 2004; Müller and Stadtmüller, 2005), one application of which is the classification of functional data (Müller, 2005). In the GFLM, the responses are scalars with general, often discrete distributions, such as binomial or Poisson, while the predictors are functional. With a monotone and invertible link function g, the GFLM is given by

$$E(Y|X) = g\left(\mu + \int_{\mathcal{T}} X(s)\beta(s)\, ds\right),$$

coupled with a variance function $\mathrm{var}(Y|X) = V(E(Y|X))$. This model has been implemented with both known or unknown link/variance function in Müller and Stadtmüller (2005) and is included in the PACE 2.9 package.

The extension of the functional linear model to the case of functional responses is more relevant to the goals we pursue in this paper. This model is given by

$$E(Y(t)|X) = \mu_Y(t) + \int_{\mathcal{T}} (X(s) - \mu_X(s))\alpha(s,t)\, ds, \tag{1}$$

where α is the regression parameter function, which is assumed to be a smooth and square integrable surface. This functional linear regression model, which includes both functional responses and functional predictors, was introduced in Ramsay and Dalzell (1991), a seminal paper that contains many interesting concepts in addition to (1). An extension to the case of sparse and irregularly observed predictor and response processes has been discussed in Yao et al. (2005b) and the extension to an additive version (for more details, see below) in Müller and Yao (2008).

The functional regression model (1) may be viewed as an extension of the multivariate linear model $E(Y|X) = BX$, relating random vectors X and Y through a parameter matrix B, to the case of functional predictors and responses. In the multivariate linear regression model, for $X \in \mathbf{R}^p$, $Y \in \mathbf{R}^q$, the least squares normal estimating equation is $\mathrm{cov}(X, Y) = \mathrm{cov}(X)B$, where $\mathrm{cov}(X, Y)$ is the $p \times q$ matrix with elements $a_{jk} = \mathrm{cov}(X_j, Y_k)$. This equation can be solved for the parameter matrix B if the $p \times p$ covariance matrix $\mathrm{cov}(X)$ is invertible; otherwise one needs to resort to a generalized inverse. For the functional extension, one can define a corresponding "Functional Normal Equation" (He et al., 2000) for the regression parameter surface α,

$$r_{XY} = R_{XX}\alpha, \quad \text{for} \quad \alpha \in L_2,$$

where $R_{XX} : L^2 \to L^2$ is the auto-covariance operator of X, defined by

$$(R_{XX}\alpha)(s,t) = \int r_{XX}(s,w)\alpha(w,t)dw,$$

with $r_{XX}(s,t) = \mathrm{cov}[X(s), X(t)]$, $r_{XY}(s,t) = \mathrm{cov}[X(s), Y(t)]$. As R_{XX} is a compact operator in L^2, it is not invertible. This leads to an inverse problem, requiring regularization.

The more recent functional additive model is a variant for which one assumes

$$E(Y(t)|X) = \mu_Y(t) + \sum_{k=1}^{\infty} f_k \left\{ \int_{\mathcal{T}} (X(s) - \mu_X(s))\phi_k(s)\, ds, t \right\} \tag{2}$$

with suitable smooth link functions f_k and projection functions ϕ_k, which form a basis and are most conveniently chosen as the (orthonormal) eigenfunctions of predictor processes X. This model strikes a good balance between the more structured and thus more restrictive functional linear model (1) and a fully nonparametric approach (Ferraty and Vieu, 2006; Ferraty et al., 2007). The latter is subject to the "curse of dimension", due to the infinite-dimensional nature of the predictors. A consequence is that under reasonable assumptions, such nonparametric approaches cannot attain polynomial rates of convergence, no matter how slow the rate.

Other classes of "functional response models" (Chiou et al., 2004) are of interest in functional dose-response models and similar applications. Such models relate a vector predictor Z with a functional response Y,

$$E\{Y(t)|Z = z\} = \mu(t) + \sum_{k=1}^{\infty} E(\xi_k|Z = z)\phi_k(t).$$

Here ϕ_k, $k = 1, 2, \ldots$ is a basis of function space and ξ_k are random coefficients. A convenient specification of the functions $E(\xi_k|Z = z)$ can be obtained through nonparametric regression for low-dimensional Z (Müller and Yao, 2006), parametric models where warranted, or through nonparametric structured models such as

additive or single index models. For example, in the latter case one would specify $E(\xi_k|Z = z) = \alpha_k(\gamma_k^T z)$ for suitable link functions α_k and parameter vectors γ_k, $k = 1, 2, \ldots$.

For specific applications, sometimes models with specific structure such as a "Multiplicative Effects Model" (Chiou et al., 2003) are useful. This model is given by

$$\mu(t, z) = \mu_0(t)\theta(z),\ E\{Y(t)\} = \mu_0(t),\ E(\theta(Z)) = 1,$$

for a suitable function $\theta(\cdot)$. Fitting this model requires no more than estimating two one-dimensional nonparametric regression functions and is very straightforward. The motivation for this model is boosted by the fact that in many applications one finds that the first eigenfunction of the processes being studied is more or less proportional to the mean function. It is easy to see that indeed the Multiplicative Effects Model will apply when this is the case and the first eigenfunction explains a large fraction of the total variation in the functional data.

A key tool for FDA is functional principal component analysis (FPCA). This method aims to obtain the spectrum of the auto-covariance operator of underlying processes and has been developed over the last decades, beginning with the seminal work of Karhunen (1946) and Grenander (1950). The study of functional principal components involved increasingly sophisticated tools, based on various results in perturbation theory for linear operators (Kato, 1995). Perturbation theory is needed as the operator itself has to be estimated from the data. Various aspects of FPCA, including asymptotic convergence of eigenfunctions and eigenvalues, are discussed e.g. in Rao (1958); Dauxois et al. (1982); Besse and Ramsay (1986); Castro et al. (1986); Hall et al. (2006). FPCA implements dimension reduction of the initially infinite dimensional data to a finite number of functional principal components (FPCs). It also provides representations of individual random trajectories and can give meaningful representations even in the case of longitudinal, i.e., sparse and irregular measurements (Yao et al., 2005a). A version of this approach can also be used to obtain derivatives of irregularly measured trajectories, to be described in the following section.

3 Estimating Trajectories and Their Derivatives from Sparse Irregular Data

Estimation of trajectories and even more so of derivatives is a difficult proposition when only sparsely observed data are available for each trajectory. Derivatives are of interest in many applications, as they allow to better describe the dynamics of the underlying processes. Applications where derivatives provide useful insights into underlying dynamics include growth curves (Gasser et al., 1984) and e-commerce, especially online auction data (Reddy and Dass, 2006; Wang et al., 2008). Derivatives are limits of difference quotients as the span converges to 0 and therefore require dense data when computed for a single nonparametric regression function (Gasser and Müller, 1984). If one has a sample of subjects for each of which the observations are generated by an underlying random trajectory, the situation can be improved by borrowing information across subjects. For the case of derivatives this was demonstrated in Liu and Müller (2009). We review these results in the following.

For the underlying (but unobserved) random trajectories X that generate the available sparse observations, we assume that they are square integrable and smooth (twice differentiable) on the domain $\mathcal{T} = [0, T]$. The basic quantities on which the "borrowing information" principle rests are the mean function $EX(t) = \mu(t)$ and the auto-covariance function $\mathrm{cov}(X(t), X(s)) = G(t, s)$, $s, t \in \mathcal{T}$, which is smooth, symmetric and non-negative definite. Using G as kernel in a linear operator leads to the Hilbert-Schmidt operator $(A_G f)(t) = \int_{\mathcal{T}} G(t, s) f(s) ds$. We denote the ordered eigenvalues (in declining order) of this operator by $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and the corresponding orthonormal eigenfunctions by $\phi_k(t)$.

One has the well-known representations $G(t, s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(s)$ and the Karhunen-Loève representation (Ash and Gardner, 1975) $X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t)$, where the $\xi_{ik} = \int_{\mathcal{T}} (X_i(t) - \mu(t)) \phi_k(t) dt$, $k = 1, 2, \ldots$ are the functional principal components (FPCs) of the random trajectories X_i, for $k = 1, 2, \ldots$. The ξ_k are uncorrelated random variables with $\mathrm{E}(\xi_k) = 0$ and $E\xi_k^2 = \lambda_k$, with $\sum_k \lambda_k < \infty$. Differentiating both sides,

$$X_i^{(\nu)}(t) = \mu^{(\nu)}(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k^{(\nu)}(t), \quad \nu = 0, 1, \ldots. \tag{3}$$

where $\mu^{(\nu)}(t)$ and $\phi_k^{(\nu)}(t)$ are the νth derivatives of mean and eigenfunctions.

The eigenfunctions ϕ_k are the solutions of the eigen-equations $\int G(s, t) \phi_k(s)\, ds = \lambda_k \phi_k(t)$, under the constraint of orthonormality. Differentiating both sides ν times, under regularity conditions that allow to exchange differentiation and expectation, one obtains

$$\frac{d^\nu}{dt^\nu} \int_{\mathcal{T}} G(t, s) \phi_k(s) ds = \lambda_k \frac{d^\nu}{dt^\nu} \phi_k(t) \tag{4}$$

and for the eigenfunction derivatives

$$\phi_k^{(\nu)}(t) = \frac{1}{\lambda_k} \int_{\mathcal{T}} \frac{\partial^\nu}{\partial t^\nu} G(t, s) \phi_k(s) ds. \tag{5}$$

This relation can then be used to devise estimation approaches for the fitting of processes X_i according to (3).

We now describe the data model for longitudinal observations, which consist of sparse, irregular and noise corrupted measurements of a random trajectory for each subject. Given realizations X_i of the underlying process X and N_i of an integer-valued bounded random variable N, we assume that N_i measurements Y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, are obtained at random times T_{ij}, according to

$$Y_{ij} = Y_i(T_{ij}) = X_i(T_{ij}) + \varepsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(T_{ij}) + \varepsilon_{ij}, \tag{6}$$

where ε_{ij} are zero mean i.i.d. measurement errors, with $\mathrm{var}(\varepsilon_{ij}) = \sigma^2$, independent of all other random components.

Turning to estimation, in a first step we aggregate all measurements across all subjects into one “big” scatterplot and apply a smoothing method that allows to obtain the $\nu-$th derivative of a regression function from scatterplot data. For

example, in the case of local polynomial fitting, given a univariate density function κ_1 and bandwidth $h_{\mu,\nu}$, one would minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i}\kappa_1\left(\frac{T_{ij}-t}{h_{\mu,\nu}}\right)\{Y_{ij}-\sum_{m=0}^{\nu+1}\alpha_m(T_{ij}-t)^m\}^2 \tag{7}$$

for each t with respect to α_m for $m = 0, \ldots, \nu+1$, from which one obtains $\hat{\mu}^{(\nu)}(t) = \hat{\alpha}_\nu(t)\nu!$ (Fan and Gijbels, 1996).

According to (5), we will also need estimates of $\frac{\partial^\nu}{\partial t^\nu}G(t,s) = G^{(\nu,0)}$. There are various techniques available for this task. Following Liu and Müller (2009), to which we refer for further details, using again local polynomial fitting, one approach is to minimize the pooled scatterplot of pairwise raw covariances

$$\sum_{i=1}^{n}\sum_{1\le j\ne l\le N_i}\kappa_2\left(\frac{T_{ij}-t}{h_{G,\nu}},\frac{T_{il}-s}{h_{G,\nu}}\right)\left\{G_i(T_{ij},T_{il})\right.$$
$$\left.-\left(\sum_{m=0}^{\nu+1}\alpha_{1m}(T_{ij}-t)^m+\alpha_{21}(T_{il}-s)\right)\right\}^2, \tag{8}$$

for fixed (t,s) with respect to α_{1m} and α_{21} for $m = 1, \ldots, \nu+1$, where $G_i(T_{ij},T_{il}) = (Y_{ij}-\hat{\mu}(T_{ij}))(Y_{il}-\hat{\mu}(T_{il}))$, $j \ne l$, κ_2 is a kernel chosen as a bivariate density function and $h_{G,\nu}$ is a bandwidth. This leads to $\hat{G}^{(\nu,0)}(t,s) = \hat{\alpha}_{1\nu}(t,s)\nu!$.

The pooling that takes place in the scatterplots for estimating the derivatives of μ and of G is the means to accomplish the borrowing of information across the sample which is needed to overcome the sparse sampling designs. We note that the case of no derivative $\nu = 0$ is always included, and solving the eigenequations for that case yields the required estimates $\hat{\lambda}_1, \hat{\lambda}_2, \ldots$ of the eigenvalues. One key feature of the covariance surface smoothing step in (8) is the exclusion of the diagonal elements (for which $j = l$); the expected value for these elements includes the measurement error variance σ^2 in addition to the variance of the process. The difference between a smoother that uses the diagonal elements only and the resulting diagonal from the smoothing step (8) when no derivatives are involved can then be used to find consistent estimates for the error variance σ^2 (Yao et al., 2005a).

To obtain estimates for the derivatives of the trajectories X_i, a realistic target is the conditional expectation $E(X_i^\nu(t)|Y_{i1},\ldots,Y_{iN_i})$. It turns out that this conditional estimation can be consistently estimated in the case of Gaussian processes by applying principal analysis by conditional expectation (PACE) (Yao et al., 2005a). For $\mathbf{X}_i = (X_i(T_{i1}),\ldots,X_i(T_{iN_i}))^T$, $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{iN_i})^T$, $\boldsymbol{\mu}_i = (\mu(T_{i1}),\ldots,\mu(T_{iN_i}))^T$, $\boldsymbol{\phi}_{ik} = (\phi_k(T_{i1}),\ldots,\phi_k(T_{iN_i}))^T$, if ξ_{ik} and ε_{ij} in (6) are jointly Gaussian, then by standard properties of the Gaussian distribution,

$$\tilde{\xi}_{ik} = \mathrm{E}[\xi_{ik}|\mathbf{Y}_i] = \lambda_k\boldsymbol{\phi}_{ik}^T\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}_i), \tag{9}$$

where $\boldsymbol{\Sigma}_{\mathbf{Y}_i} = \mathrm{cov}(\mathbf{Y}_i,\mathbf{Y}_i) = \mathrm{cov}(\mathbf{X}_i,\mathbf{X}_i)+\sigma^2\mathbf{I}_{N_i}$.

The quantities on the r.h.s. of (9) can all be consistently estimated (Yao et al., 2005a; Liu and Müller, 2009). Regularity conditions include assumptions for the smoothing kernels, the distribution of the design points, behavior of eigenfunctions and eigenvalues and the large sample behavior of the bandwidths $h_{\mu,0}, h_{\mu,1}$ for

the estimation of the mean function μ and its first derivative $\mu'(t)$, and $h_{G,0}, h_{G,1}$ for the estimation of the covariance surface and its partial derivative, where the bandwidth conditions are given by

$$\begin{aligned} h_{\mu,0} \to 0, \quad nh_{\mu,0}^4 \to \infty, \quad nh_{\mu,0}^6 < \infty \\ h_{G,0} \to 0, \quad nh_{G,0}^6 \to \infty, \quad nh_{G,0}^8 < \infty, \\ h_{\mu,1} \to 0, \quad nh_{\mu,1}^8 \to \infty, \quad h_{G,1} \to 0, \quad nh_{G,1}^{10} \to \infty. \end{aligned}$$

For the first derivative, denote the resulting estimates by $\hat{\mu}'$ for the derivative $\mu'(t)$ of the mean function and by $\hat{\phi}'_k$ for the derivative $\phi'_k(t)$ of the k-th eigenfunction, leading to the estimate for the derivative $X'_i(t)$ given by $\hat{X}'_{i,K}(t) = \hat{\mu}'(t) + \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}'_k(t)$. If $\mathcal{I}'$ denotes the set of indices with eigenvalues of multiplicity 1, one may then obtain the following results (Liu and Müller, 2009)

$$\begin{aligned} \sup_{t\in\mathcal{T}} |\hat{\mu}'(t) - \mu'(t)| &= O_p(\frac{1}{\sqrt{n}h_{\mu,1}^2} + h_{\mu,1}^4) \\ \sup_{t,s\in\mathcal{T}} |\hat{G}^{(1,0)}(t,s) - G^{(1,0)}(t,s)| &= O_p(\frac{1}{\sqrt{n}h_{G,1}^3} + h_{G,1}^4) \\ \sup_{t\in\mathcal{T}} |\hat{\phi}'_k(t) - \phi'_k(t)| &= O_p(\frac{1}{\sqrt{n}h_{G,0}^2} + \frac{1}{\sqrt{n}h_{G,1}^3} + h_{G,1}^4),\ k \in \mathcal{I}' \\ \lim_{K\to\infty} \lim_{n\to\infty} \hat{X}'_{i,K}(t) &= \widetilde{X}'_i(t) \quad \text{in probability,} \end{aligned}$$

where $\widetilde{X}'_i(t) = \mu'(t) + \sum_{k=1}^{\infty} \tilde{\xi}_{ik}\phi'_k(t)$, and similar results can be obtained for higher derivatives. It is noteworthy that the only assumption needed for the number of observations N_i, which is the number of repeated measurements assumed to be available per subject, is $P(N_i \geq 2) > 0$. This weak assumption demonstrates that extremely sparse designs are covered by these results.

4 Algorithmic Implementations – Principal Analysis by Conditional Expectation (PACE)

The numerical implementation of Step (9) of the PACE estimation scheme involves inversion of individual covariance matrices $\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}$ for all subjects. A good implementation requires an efficient regularization scheme, since these matrices are of random dimension $N_i \times N_i$ and need to be evaluated at random times. One also needs to choose a finite number of included components K to represent fitted trajectory derivatives by

$$\hat{X}_{i,K}^{(\nu)}(t) = \hat{\mu}^{(\nu)}(t) + \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}_k^{(\nu)}(t). \tag{10}$$

As $K \to \infty$ with increasing sample size n, choice of K corresponds to a model selection problem. Various criteria are available for this choice for the case $\nu = 0$, such as fraction of variance explained (FVE), AIC and BIC (for details, see Liu and Müller, 2009). In the applications reported below, we adopt the use of FVE, which is related to the scree plot known from multivariate analysis. We implement

FVE by selecting the smallest number K of components which explain at least 85% of the variation in the data.

A continuing challenge is the construction of a comprehensive package of programs which implement the described procedures and other FDA methods. One such package emphasizing functional methods for longitudinal data is under development by a team based at UC Davis. A current version is available as PACE 2.9 (released in May 2009) from `http://anson.ucdavis.edu/~mueller/data/programs.html`. PACE 2.9 implements a variety of FDA procedures, including warping, functional principal component analysis and various functional regression models.

Generally, the PACE package addresses the case of sparse and irregularly measured functional data. For such data, as well as densely recorded data, it provides programs for the following tasks: Determination of eigenfunctions and eigenvalues; fitting of individual trajectories and their derivatives via PACE (10); functional linear regression analysis (1), including diagnostics and bootstrap inference; fitting of the functional additive model (2); warping and registration through pairwise alignment as described in Tang and Müller (2008); generalized functional linear regression, where the response is a scalar generalized variable such as binary or Poisson; this approach can also be used for classification of functional data via binary regression. Also there is a variant where the response is a series of generalized (binary, Poisson etc.) responses, which are modeled by a latent Gaussian process (Hall et al., 2008).

It is worthwhile to note that PACE does not use pre-smoothing to obtain smooth trajectories from noisy data. While pre-smoothing is a standard preprocessing step in many alternative approaches to FDA, it is quite problematic if functional data are sparsely sampled or measurements are corrupted with noise, since the resulting distortions cannot anymore be removed from the data. These problems are magnified for the case of derivatives.

5 Dynamic Transfer Functions

"Transfer function" in the following has a different meaning from that in control theory and in particular applies to the time domain, quantifying the influence of the value of a stochastic process at a given time t on the value of another or the same process at a time s. Throughout we assume that we are in the situation of sparse longitudinal measurements. All processes considered are assumed to be Gaussian and in the case of multivariate processes jointly Gaussian, but not stationary. Gaussianity for one process implies that the FPCs are jointly Gaussian and independent. In practice, violation of the Gaussian assumption is common. While one might try to transform data to enhance their Gaussianity, violations from Gaussianity often do not seriously affect the practical performance of the methods we discuss in the following, according to our experience.

Sparse longitudinal observations Y_{ij} are generated by the trajectories X_i of a continuous Gaussian process, following (6). The dynamics of a single process can be explored by simply relating values of the process at a later time to those observed earlier. We emphasize here the case where the predictor value corresponds to the value of the process at one fixed time. A straightforward approach is to consider the conditional expectations $E(X(t)|X(s))$, $s < t$.

The dynamics of the underlying stochastic system that generates the observations was studied in Liu and Müller (2009) by relating the derivative process $X'(t)$ to the process $X(t)$ at the same time t. The conditional expectation of the derivative process $X'(t)$, given the level of the original process $X(t)$ gives rise to a population-average differential equation. The Gaussianity of the underlying processes implies the joint Gaussianity of $X^{(\nu)}(t)$ and $X(t)$ for any order of derivative $\nu \geq 0$. This guarantees the existence of a *dynamic transfer function* β_ν such that

$$E(X^{(\nu)}(t) - \mu_X^{(\nu)}(t)|X(t) - \mu_X(t)) = \beta_\nu(t)(X(t) - \mu_X(t)). \tag{11}$$

This result can be used for instance to predict trends in the slope $X'(t)$ of the process at t from current levels of the process X at t. For the case $\nu = 1$, (11) suggests to explore the approximation to the observations provided by processes $\tilde{X}$ that are generated by the equation

$$\tilde{X}'(t) = \beta_1(t)(\tilde{X}(t) - \mu_X(t)), \quad \tilde{X}(0) = X(0). \tag{12}$$

Approximations (12) have been explored in Liu and Müller (2009) in the context of auction bid price data and in this special example case lead to surprisingly good approximations, implying that a large part of the variability in the data is explained by this simple dynamic scheme.

It bears emphasizing that for the type of sparse data obtained from longitudinal studies that we consider here, derivatives are not directly available, since one will normally not have a measurement of X at a given time t or in the immediate neighborhood of t, and even if one does, it will be contaminated by noise. So while even $X(t)$ cannot be directly observed, the situation is much more dire for the case of derivatives where the usual approximation by difference quotients or through a smoothing based differentiation procedure is not feasible in the sparse case, and it is essential to use functional methods to borrow strength across subjects to study the underlying process dynamics.

Specifically, consider mean-centered process $X(s) - \mu_X(s) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(s)$ and derivative processes $X^{(\nu)}(t) - \mu_X^{(\nu)}(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu)}(t), \nu = \nu_1, \nu_2$. Since the FPC scores ξ_{ik} are jointly Gaussian (and independent), $X^{(\nu_1)}(s) - \mu_X^{(\nu_1)}(s)$ and $X^{(\nu_2)}(t) - \mu_X^{(\nu_2)}(t)$ are jointly Gaussian for each $(s,t) \in \mathcal{T} \times \mathcal{T}$,

$$\begin{pmatrix} \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu_1)}(s) \\ \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu_2)}(t) \end{pmatrix}$$

$$\sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{k=1}^{\infty} \lambda_k(\phi_k^{(\nu_1)}(s))^2 & \sum_{k=1}^{\infty} \lambda_k\phi_k^{(\nu_1)}(s)\phi_k^{(\nu_2)}(t) \\ \sum_{k=1}^{\infty} \lambda_k\phi_k^{(\nu_1)}(s)\phi_k^{(\nu_2)}(t) & \sum_{k=1}^{\infty} \lambda_k(\phi_k^{(\nu_2)}(t))^2 \end{pmatrix}\right),$$

whence

$$E(X^{(\nu_2)}(t) - \mu_X^{(\nu_2)}(t)|X^{(\nu_1)}(s) - \mu_X^{(\nu_1)}(s)) = \beta_{\nu_1\nu_2}(s,t)(X^{(\nu_1)}(s) - \mu_X^{(\nu_1)}(s)), \tag{13}$$

with

$$\begin{aligned}\beta_{\nu_1\nu_2}(s,t) &= \frac{\mathrm{cov}(X^{(\nu_2)}(t), X^{(\nu_1)}(s))}{\mathrm{var}(X^{(\nu_1)}(s))} \\ &= \frac{\mathrm{cov}\left(\sum_{k=1}^{\infty}\xi_{ik}\phi_k^{(\nu_1)}(s), \sum_{k=1}^{\infty}\xi_{ik}\phi_k^{(\nu_2)}(t)\right)}{\mathrm{var}\left(X^{(\nu_1)}(s)\right)} \\ &= \frac{\sum_{k=1}^{\infty}\lambda_k\phi_k^{(\nu_2)}(t)\phi_k^{(\nu_1)}(s)}{\sum_{k=1}^{\infty}\lambda_k(\phi_k^{(\nu_1)}(s))^2}. \end{aligned} \tag{14}$$

The transfer function $\beta_{\nu_1\nu_2}(s,t)$ defines the best predictor of $X^{(\nu_2)}(t)$, given $X^{(\nu_1)}(s)$, where we assume $s \leq t$.

On average, $X^{(\nu_2)}$ at time t will increase by an amount of $\beta_{\nu_1\nu_2}(s,t)$ for a unit increase in the level of process $X^{(\nu_1)}$ at s. If we set $s = t$, $\nu_1 = 0$ and $\nu_2 = 1$, then (13) corresponds to (11) and the population-average differential equation described in Liu and Müller (2009),

$$\begin{aligned} E(X'(t)|X(t)) &= \mu_X'(t) + \beta_1(t)(X(t) - \mu(t)), \\ \beta_1(t) &= \sum_{k=1}^{\infty}\lambda_k\phi_k'(t)\phi_k(t) / \sum_{k=1}^{\infty}\lambda_k\phi_k^2(t). \end{aligned} \tag{15}$$

This modeling approach leads to useful interpretations: In subdomains of $\mathcal{T}$ where β_1 is positive, a time course X that is above (below) the mean function μ_X will tend to be associated with larger positive (negative) derivatives, which means the difference to the mean will increase. Such behavior may lead to trajectories with explosive growth (or decline), if this behavior persists throughout the domain. For subdomains where β_1 is negative, a time course X that is above (below) the mean function μ_X will tend to be associated with negative (positive) derivatives, so that the distance to the mean will tend to decrease. This situation can be described as "dynamic regression to the mean".

Even without the Gaussianity assumption, (14) provides the best linear predictor and is the minimizer of a least squares problem. This is easily seen: For any function $\omega(s,t), (s,t) \in \mathcal{T} \times \mathcal{T}$,

$$\begin{aligned}\mathrm{MSE}(\omega) &= E[(X^{(\nu_2)}(t) - \mu_X^{(\nu_2)}(t)) - \omega(s,t)(X^{(\nu_1)}(s) - \mu_X^{(\nu_1)}(s))]^2 \\ &= \mathrm{var}(X^{(\nu_2)}(t)) - 2\omega(s,t)\mathrm{cov}(X^{(\nu_2)}(t), X^{(\nu_1)}(s)) + \omega^2(s,t)\mathrm{var}(X^{(\nu_1)}(s)) \\ &= \left[\omega(s,t) - \frac{\mathrm{cov}(X^{(\nu_2)}(t), X^{(\nu_1)}(s))}{\mathrm{var}(X^{(\nu_1)}(s))}\right]^2 + c(s,t),\end{aligned}$$

where $c(s,t)$ is independent of $\omega(s,t)$, and one finds $\mathrm{argmin}_{\omega \in L^2}\mathrm{MSE}(\omega) = \beta_{\nu_1\nu_2}(s,t)$, as given in (14).

6 Dynamics of Multivariate Processes

An important task in FDA is to study the relationships between the components of a multivariate process. We concentrate here on the bivariate case (considering just two components in case of a higher-dimensional process). Standard functional

approaches such as functional canonical correlation or regression models that relate a process Y to a process X such as (1) and (2) are based on the assumption that the entire predictor process X influences the entire response process Y, i.e., the prediction for $Y(t)$ at any fixed time t depends on the levels $X(s)$ both for $s < t$ and $s > t$. This applies to functional regression as well as common functional correlation measures, including canonical correlation, dynamical correlation (Dubin and Müller, 2005) and also a recently proposed B-spline based FPC approach for sparse data obtained from two correlated processes (Zhou et al., 2008).

Modeling the dependency of the outcome Y at time t on both past and future levels of X is sometimes not adequate. Often one is interested in predicting the value of an unobserved process $Y(t)$ only from past observations of a related process, i.e., only from data generated by $X(s)$, $s \leq t$, up to time t. As time t increases, the prediction needs to be continuously updated. Models which relate the entire functional history of the process X up to time s to a real-valued outcome that is observed later (such as $Y(t)$ or some other outcome which will be observed after time s and the distribution of which will change as s increases) have been studied in Malfait and Ramsay (2003) and Müller and Zhang (2005).

Suppose $X(t)$ and $Y(t)$ for $t \in \mathcal{T}$ are jointly Gaussian random processes with (eigenvalues, eigenfunctions) given by (λ_k, ϕ_k) for X and (ρ_l, ψ_l) for Y, respectively, with Karhunen-Loéve decompositions $X(t) - \mu_X(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t)$, $\quad Y(t) - \mu_Y(t) = \sum_{l=1}^{\infty} \zeta_{il}\psi_l(t)$. As before, these representations can be extended to derivatives of orders $\nu_1 \geq 0$ for X and $\nu_2 \geq 0$ for Y,

$$\begin{aligned} X^{(\nu_1)}(t) - \mu_X^{(\nu_1)}(t) &= \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu_1)}(t), \\ Y^{(\nu_2)}(t) - \mu_Y^{(\nu_2)}(t) &= \sum_{l=1}^{\infty} \zeta_{il}\psi_l^{(\nu_2)}(t), \end{aligned} \tag{16}$$

and the corresponding auto- and cross-covariance functions are given by

$$\begin{aligned} G_{XX}^{(\nu_1,\nu_1)}(s,t) &= \mathrm{cov}(X^{(\nu_1)}(s), X^{(\nu_1)}(t)) = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu_1)}(s)\phi_k^{(\nu_1)}(t), \\ G_{YY}^{(\nu_2,\nu_2)}(s,t) &= \mathrm{cov}(Y^{(\nu_2)}(s), Y^{(\nu_2)}(t)) = \sum_{l=1}^{\infty} \rho_l \psi_l^{(\nu_2)}(s)\psi_l^{(\nu_2)}(t), \\ G_{XY}^{(\nu_1,\nu_2)}(s,t) &= \mathrm{cov}(X^{(\nu_1)}(s), Y^{(\nu_2)}(t)) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} E(\xi_k\zeta_l)\phi_k^{(\nu_1)}(s)\psi_l^{(\nu_2)}(t). \end{aligned} \tag{17}$$

To define local relationships between jointly Gaussian processes X and Y and their derivatives, we observe that (17) implies for each fixed $(s,t) \in \mathcal{T} \times \mathcal{T}$,

$$\begin{pmatrix} \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu_1)}(s) \\ \sum_{l=1}^{\infty} \zeta_{il}\psi_l^{(\nu_2)}(t) \end{pmatrix} \sim$$
$$N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{k=1}^{\infty} \lambda_k (\phi_k^{(\nu_1)}(s))^2 & \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} E(\xi_k\zeta_l)\phi_k^{(\nu_1)}(s)\psi_l^{(\nu_2)}(t) \\ \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} E(\xi_k\zeta_l)\phi_k^{(\nu_1)}(s)\psi_l^{(\nu_2)}(t) & \sum_{l=1}^{\infty} \rho_l (\psi_l^{(\nu_2)}(t))^2 \end{pmatrix} \right).$$

As before, we conclude from the joint Gaussianity that there exists a dynamic transfer function $\gamma_{\nu_1\nu_2}$ such that

$$E(Y^{(\nu_2)}(t) - \mu_Y^{(\nu_2)}(t)|X^{(\nu_1)}(s) - \mu_X^{(\nu_1)}(s)) = \gamma_{\nu_1\nu_2}(s,t)(X^{(\nu_1)}(s) - \mu_X^{(\nu_1)}(s)), \quad (18)$$

where the transfer function is given by

$$\gamma_{\nu_1\nu_2}(s,t) = \frac{\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sigma_{kl}\phi_k^{(\nu_1)}(s)\psi_l^{(\nu_2)}(t)}{\sum_{k=1}^{\infty}\lambda_k(\phi_k^{(\nu_1)}(s))^2}, \quad (19)$$

with $\sigma_{kl} = E(\xi_k\zeta_l)$.

As above, the dynamic transfer function can be derived as the solution of a simple minimization problem. This model relates the level of a "predictor process" $X(s)$, $s \leq t$ to $Y(t)$; most often one will have $\nu_1 = \nu_2 = 0$ and $s = t$, in which case the dynamic transfer function (19) becomes the varying coefficient function γ in the varying coefficient model

$$\begin{aligned} E(Y(t)|X(t)) &= \mu_Y(t) + \gamma(t)(X(t) - \mu_X(t)), \\ \gamma(t) &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sigma_{kl}\phi_k(t)\psi_l(t) / \sum_{k=1}^{\infty}\lambda_k\phi_k^2(t). \end{aligned} \quad (20)$$

This connection of functional models with varying coefficient models has been studied and exploited in Şentürk and Müller (2009). It leads to alternative approaches for the fitting of varying coefficient models. Especially for the case of sparse or noisy data, this is an attractive approach, since the components of the varying coefficient "slope" function γ in (20) can be easily estimated for such data (see next section).

Standard methods for varying coefficient model fitting (Fan and Zhang, 2000, 2008; Chiang et al., 2001; Hoover et al., 1998; Huang et al., 2004; Qu and Li, 2006; Wu et al., 2000) are based on fitting simple linear regression models of Y_{ij} on X_{ij}. Typically these are fitted in small slices of data, which include those measurements with measurements times $|T_{ij} - t| \leq h$, where h is a bin width. For each bin with midpoint t this yields intercept and slope estimates $\tilde{\gamma}_0(t), \tilde{\gamma}_1(t)$ which are obtained on the grid of $|\mathcal{T}|/(2h)$ bin midpoints and then smoothed over t to obtain the final varying coefficient function estimates. In the sparse or noisy measurements case, such standard estimates will not be consistent and alternative approaches using dynamical transfer functions are preferrable (Şentürk and Müller, 2008). The approach based on FDA turns out to be also surprisingly competitive in the dense and noise-free case, which is the standard situation that is usually considered in the varying coefficients literature. We note that even in the absence of Gaussianity, (20) provides a minimizer of expected least squared error and a best local linear approximation.

7 Estimation Procedures

In most longitudinal studies, the observed data can be thought of as generated by underlying smooth random processes, rather than corresponding to entirely observed functional trajectories. Complete trajectories are rarely observed, due

to additive noise in the measurements and irregular, often sparse, measurement designs. Extending the estimation procedures described in Section 3 and following Yao et al. (2005b), we assume that the measurements are made at random time points $S_{i1}, \ldots, S_{iQ_i}$ for processes X_i, resp. $T_{i1}, \ldots, T_{iR_i}$ for processes Y_i, where the numbers of measurements Q_i resp. R_i are i.i.d random variables. The data (S_{im}, U_{im}, Q_i) and (T_{ij}, V_{ij}, R_i), $i = 1, \ldots, n$, $m = 1, \ldots, Q_i$, $j = 1, \ldots, R_i$, are assumed to have the same distribution as (S, U, Q) and (T, V, R), where U_{im} (respectively, V_{ij}) denote the observations of the random trajectory X_i (respectively, Y_i) at the random times S_{im} (respectively, T_{ij}), contaminated with measurement errors ε_{im} (respectively, ϵ_{ij}).

The errors are assumed to be i.i.d. with $E\varepsilon_{im} = 0$, $E[\varepsilon_{im}^2] = \sigma_X^2$ (respectively, $E\epsilon_{ij} = 0$, $E[\epsilon_{ij}^2] = \sigma_Y^2$), and independent of the trajectories. Then we may represent the observed data for processes (X_i, Y_i) as follows,

$$\begin{aligned} U_{im} &= X_i(S_{im}) + \varepsilon_{im}, \qquad S_{im} \in \mathcal{T}, \quad 1 \le i \le n, \quad 1 \le m \le Q_i, \\ V_{ij} &= Y_i(T_{ij}) + \epsilon_{ij}, \qquad T_{ij} \in \mathcal{T}, \quad 1 \le i \le n, \quad 1 \le j \le R_i, \end{aligned} \tag{21}$$

and with (16),

$$\begin{aligned} U_{im} &= \mu_X(S_{im}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(S_{im}) + \varepsilon_{im}, \\ V_{ij} &= \mu_Y(T_{ij}) + \sum_{l=1}^{\infty} \zeta_{il}\psi_l(T_{ij}) + \epsilon_{ij}. \end{aligned} \tag{22}$$

The estimation of eigenvalues and eigenfunction derivatives $\lambda_k, \phi_k^{(\nu_1)}$, $k \ge 1$, for X and ρ_k, $\psi_k^{(\nu_2)}$ for Y, $l \ge 1$, then proceeds in complete analogy to the procedure described in Section 3, using scatterplot smoothers (7) and (8).

For the estimation of the transfer function $\gamma_{\nu_1\nu_2}$ (19) we also require an estimate of $\sigma_{kl} = E(\xi_k\zeta_l)$. Following a proposal in Yao et al. (2005b), such an estimate can be obtained from an estimate of the cross-covariance function $C(s,t) = G_{XY}^{(0,0)}(s,t)$. Starting with raw covariances $C_i(S_{im}, T_{ij}) = (U_{im} - \hat{\mu}_X(S_{im}))(V_{ij} - \hat{\mu}_Y(T_{ij}))$, we apply a local linear surface smoother for the cross-covariance surface $C(s,t)$ through minimizing

$$\sum_{i=1}^{n}\sum_{m=1}^{Q_i}\sum_{j=1}^{R_i} \kappa_2\Big(\frac{S_{im}-s}{h_1}, \frac{T_{ij}-t}{h_2}\Big)\{C_i(S_{im}, T_{ij}) - f(\alpha, (s,t), (S_{im}, T_{ij}))\}^2,$$

where $f(\alpha, (s,t), (S_{il}, T_{ir})) = \alpha_0 + \alpha_{11}(s - S_{il}) + \alpha_{12}(t - T_{ir})$, with respect to $\alpha = (\alpha_0, \alpha_{11}, \alpha_{12})$. This yields minimizers $\hat{\alpha}_0(s,t), \hat{\alpha}_{11}(s,t)$ and $\hat{\alpha}_{12}(s,t)$ and estimates $\hat{C}(s,t) = \hat{G}_{XY}^{(0,0)}(s,t) = \hat{\alpha}_0(s,t)$, where h_1,h_2 are positive bandwidths, usually chosen as $h = h_1 = h_2$.

From (17) for $\nu_1 = \nu_2 = 0$ one then obtains estimates

$$\hat{\sigma}_{kl} = \int_{\mathcal{T}}\int_{\mathcal{T}} \hat{\psi}_l(t)\hat{C}(s,t)\hat{\phi}_k(s)ds\,dt, \qquad k = 1, \ldots, K,\ l = 1, \ldots, L. \tag{23}$$

In practical applications one needs to choose numbers K and L of included eigenfunctions; options include by a pseudo-AIC or BIC type criteria as described in

Yao et al. (2005b). For asymptotic analysis, we require $K = K(n) \to \infty$ and $L = L(n) \to \infty$.

Plug-in estimates of time-dynamic transfer functions $\beta_{\nu_1\nu_2}$ (for the dynamics of a single process) and $\gamma_{\nu_1\nu_2}$ (for the dynamics of paired processes) are then obtained as follows,

$$\hat{\beta}^K_{\nu_1\nu_2}(s,t) = \frac{\sum_{k=1}^K \hat{\lambda}_k \hat{\phi}_k^{(\nu_2)}(t)\hat{\phi}_k^{(\nu_1)}(s)}{\sum_{k=1}^K \hat{\lambda}_k (\hat{\phi}_k^{(\nu_1)}(s))^2}, \tag{24}$$

$$\hat{\gamma}^{KL}_{\nu_1\nu_2}(s,t) = \frac{\sum_{k=1}^K \sum_{l=1}^L \hat{\sigma}_{kl} \hat{\phi}_k^{(\nu_1)}(s)\hat{\psi}_l^{(\nu_2)}(t)}{\sum_{k=1}^K \hat{\lambda}_k (\hat{\phi}_k^{(\nu_1)}(s))^2}. \tag{25}$$

These estimates are easy to obtain with the PACE implementations as described in Section 4.

We provide an asymptotic consistency result for the estimated transfer functions (24) and (25) when $\nu_1 = 0, \nu_2 \in \{0,1\}$, i.e. the transfer functions between the first derivative and the original curves. The result can be analogously extended to higher order derivatives. Some formal assumptions and outline of proof can be found in the Appendix.

Theorem. Under assumptions (A1)-(A7) in the Appendix, and regularity assumptions for kernels and underlying density functions, for fixed s, t, and $\nu \in \{0,1\}$,

$$\lim_{K\to\infty} \lim_{n\to\infty} \sup_{s,t\in\mathcal{T}\times\mathcal{T}} \left|\hat{\beta}^K_{0\nu}(s,t) - \beta_{0\nu}(s,t)\right| = 0, \quad \text{in probability,} \tag{26}$$

$$\lim_{K,L\to\infty} \lim_{n\to\infty} \sup_{s,t\in\mathcal{T}\times\mathcal{T}} \left|\hat{\gamma}^{KL}_{0\nu}(s,t) - \gamma_{0\nu}(s,t)\right| = 0, \quad \text{in probability.} \tag{27}$$

One can also obtain rates of convergence by adding an assumption about the asymptotic behavior of the covariance expansions. This behavior depends on the rate of decline of the eigenvalues and their spacings and on the shapes of the eigenfunctions of underlying processes.

8 Dynamics in Action

The following data application serves to illustrate the implementation of derivative estimation with PACE, as described in Section 4, and the estimates for the dynamic transfer functions that were introduced in Section 7. In a longitudinal study on aging (Shock et al., 1984; Pearson et al., 1997), SBP (Systolic Blood Pressure, in mm Hg) and BMI (Body Mass Index, in kg/m^2) were recorded on each visit of 1590 male volunteers bi-annually. There are many missed visits and the measurements are highly variable, so that these longitudinal data must be considered to be truly sparse and noisy. Both number of observations and observation times vary widely from subject to subject; see Yao et al. (2005b). We select subjects for whom at least two measurements were recorded between ages 40 and 70, the minimum number of repeated measurements needed for meaningful analysis, and who survived beyond 70, to avoid problems of selection bias due to non-surviving subjects. This led to a sample size of $n = 507$ subjects entering the analysis.

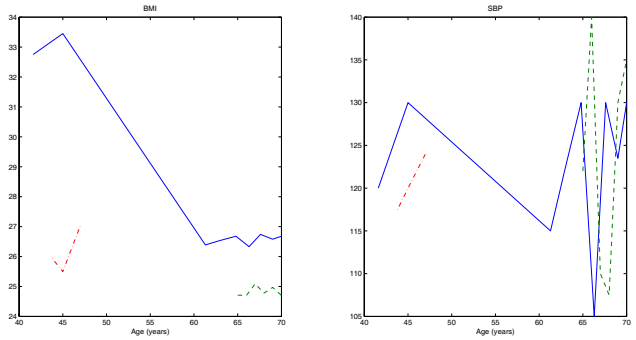

Fig. 1 The observed values of BMI (left panel) and SBP (right panel) for three randomly selected subjects, measurements are connected by straight lines.

The SBP and BMI data for three randomly selected subjects are shown in Figure 1. In spite of the visibly large noise in the measurements and the highly irregular sampling times, these data still may be viewed as being generated by underlying smooth random trajectories of blood pressure and body mass index. A basic assumption we make is that SBP and BMI measurements are generated by jointly Gaussian random processes, or equivalently, that all functional principal components are jointly Gaussian. However, even in the absence of this assumption our analysis can be meaningful, as in non-Gaussian situations the dynamic transfer functions at fixed times are the best linear approximations to the true relationships, which may be nonlinear.

Various relationships between processes and their derivatives at various fixed times are of interest. We begin by studying the influence exercised by past SBP levels at age s on current level at age t, where $s < t$. Applying (13) provides the relevant prediction of current process levels $X(t)$ from earlier levels $X(s)$, through

$$E(X(t)|X(s)) = \mu_X(t) + \beta_{00}(s,t)(X(s) - \mu_X(s)),\ s < t. \tag{28}$$

Here β_{00} is the dynamic transfer function given in (14), where one has $\nu_1 = \nu_2 = 0$. Since this dynamic transfer function is only of interest for $s < t$, the support of the estimated surface $\hat{\beta}_{00}$ is a triangular shaped subset of $\mathcal{T} \times \mathcal{T}$. The estimated surface for $s < t$ is displayed in Figure 2.

An immediate finding is that the entire transfer function surface is positive over the domain, which means that above average SBP levels at earlier age tend to be associated with higher levels SBP throughout later ages, and likewise for below average levels. Furthermore, the transfer function increases from 1 to a maximum, as current age t is increased for a fixed predictor age s, and then for further increasing t declines rapidly, falling below 1. This suggests that the deviation of SBP from the average will increase (i.e., an above average SBP will be even more above the average and a below average SBP will fall even further below the average) a few years into the future, and the amount of this increase is more pronounced at older ages s as compared to younger ages. However, the deviation from the mean, while not changing sign throughout, will generally decline as t increases even further into the older age ranges. This means eventually the effect of above

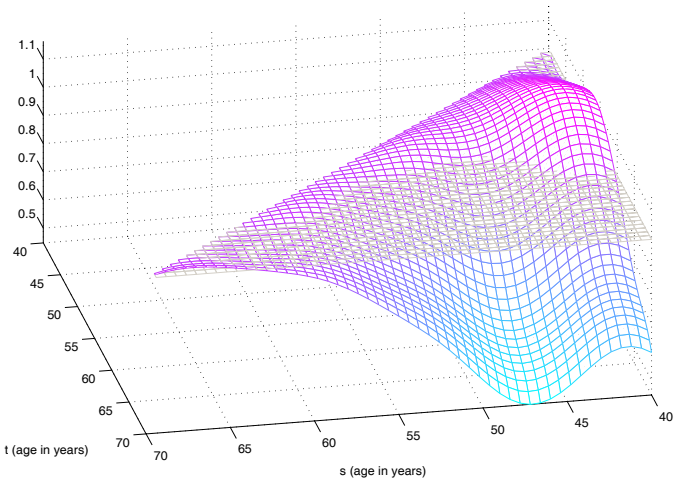

Fig. 2 Estimated dynamic transfer function $\hat{\beta}_{00}(s,t)$, $s < t$, linking current with past SBP values through (14), (28), with overlaid reference plane $z = 1$.

(below) average blood pressure will diminish in the more distant future, while a residual effect remains.

The time-dynamics of SBP can be further studied through relationships that involve derivatives. We fit the model $E(X'(t)|X(s)) = \mu'_X(t) + \beta_{01}(s,t)(X(s) - \mu_X(s))$, implementing (11) for $\nu_1 = 0, \nu_2 = 1$, by obtaining the estimated dynamic transfer function $\hat{\beta}_{01}(s,t)$, $s < t$. The resulting surface is illustrated in Figure 3. We find that for a subject whose SBP is above the average value by a certain amount, the derivative of SBP in the immediate future is also above the average, so the movement is further away from the population mean (the latter also applies to subjects whose SBP is below the average), so there is explosive growth of the deviation from the mean. For more distant future ages, the sign of the transfer function however becomes negative, which indicates that there will be a trend in the increase of SBP which is below the average derivative for subjects whose current SBP is above average, leading to a regression to the mean effect. For those subjects where SBP at current age is below average, the increase in the more distant future analogously will be above average. There is an exception to this at relatively young ages, where one notes an immediate short-lived regression to the mean effect, which is then followed by explosive growth and later by sustained regression to the mean.

To some extent these dynamics can also be visualized through the simpler dynamic transfer function β_1 (11) which relates $X'(t) - \mu'_X(t)$ to $X(t) - \mu_X(t)$. The estimated function is depicted in Figure 4. One finds that within the age range of about 43 to 63 years the deviation of derivatives to the mean derivative across the sample tends to move in the same way as the deviation of the current level to the mean level of SBP. This means that those subjects who are above or below the mean in this age range will tend to move even further above or below the mean as they age. For example, for a subject with blood pressure above the

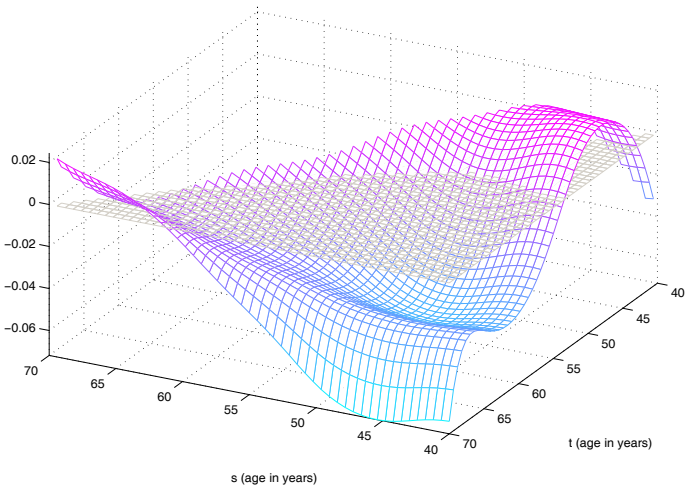

Fig. 3 Estimated dynamic transfer function $\hat{\beta}_{01}(s,t)$, $s \leq t$, relating the first derivative of SBP with the level of SBP through the dynamic relationship $E(X'(t)|X(s)) = \mu'_X(t) + \beta_{01}(s,t)(X(s) - \mu_X(s))$, see (14), overlaid with the reference plane $z = 0$.

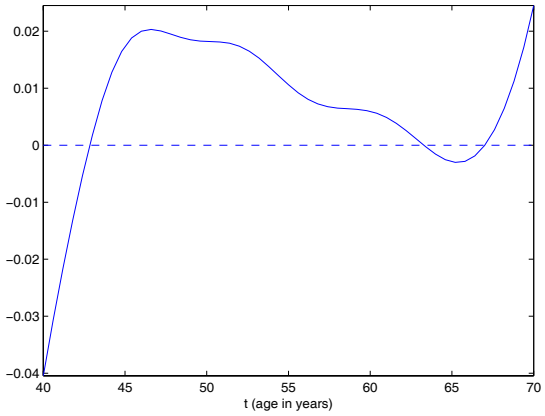

Fig. 4 Estimated transfer function $\hat{\beta}_1(t)$ (11), corresponding to the diagonal of the surface in Figure 3 (where $s = t$), relating the centered derivative of SBP to its current centered level.

mean, blood pressure will continue to rise faster than the rise for average subjects. Before age 43 and in the age range 64 to 66, there are short periods where the opposite happens, i.e., a subject with a blood pressure reading above the mean will on average experience rises that are below the rise of average blood pressure. This is the dynamic regression to the mean situation. Overall, however, the deviation from the mean is seen to reinforce itself and a subject with high or low blood pressure (measured as deviation from the norm) is likely to move most of the time

further away (above or below) from the norm, as time progresses and the subject ages.

Turning to the relationships between the bivariate process components BMI and SBP, we consider BMI as predictor process X and SBP as response process Y. The relationship of interest is that between $Y(t)$ and $X(s)$, where $s \leq t$. According to (19), $E(Y(t)|X(s)) = \mu_Y(t) + \gamma_{00}(s,t)(X(s) - \mu_X(s))$. The estimated transfer function $\hat{\gamma}_{00}$ can be found in Figure 5. As $\hat{\gamma}_{00}$ is positive everywhere, the deviation of BMI from the average at any time tends to be associated with a deviation in the same direction at a later time for SBP. This effect is most pronounced for younger ages s and for ages $t > s$ that are not too far away from s; the effect of BMI at a given age s is seen to attenuate for much higher ages t, but does not vanish or revert. This is similar to the effect observed for subjects with above or below the mean blood pressure. It indicates that above average BMI at any age is associated with larger than average SBP at current age and all subsequent ages, with diminishing effect for much later ages. We also note that the surface plot shows that for relatively small s around 40-45 years, the expected deviation of SBP from its mean is larger a few years later than in the immediate future.

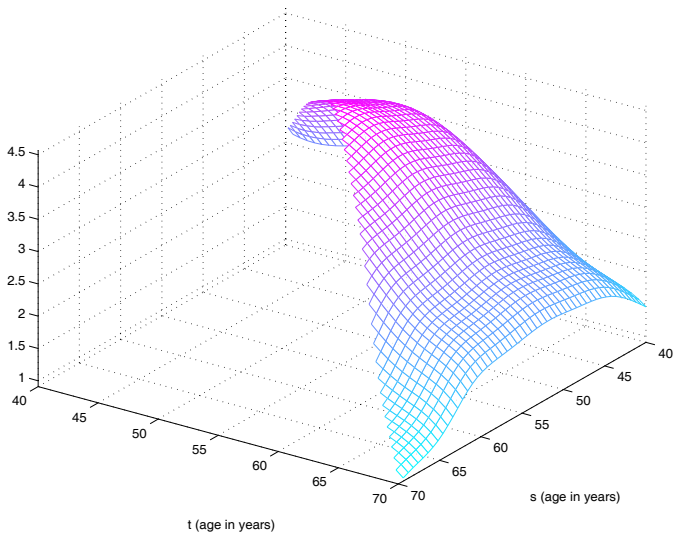

Fig. 5 Estimated transfer function $\gamma_{00}(s,t)$, $s < t$ (19), linking SBP levels (response process $Y(t)$) with BMI levels (predictor process $X(s)$) through the relation $E(Y(t)|X(s)) = \mu_Y(t) + \gamma_{00}(s,t)(X(s) - \mu_X(s))$.

Another angle on this relationship is provided by the prediction of the derivative $Y'(t)$ of SPB at t from $X(s)$, i.e., the level of BMI at s for $s < t$, based on the relation $E(Y'(t)|X(s)) = \mu'_Y(t) + \gamma_{01}(s,t)(X(s) - \mu_X(s))$. The surface corresponding to the estimate of the corresponding transfer function $\gamma_{01}(s,t)$, $s \leq t$, is shown in Figure 6. We find the anomaly for relatively small ages s that was discussed above in the negative left tail of $\hat{\gamma}_{01}(s,t)$. Here the immediate effect of

BMI on the derivative of SBP is negative, meaning that elevated BMI will lead to smaller slopes of SBP as compared to subjects with lower than average BMI. This effect is however very short-lived, and the longer-term effects are as expected: Above average BMI at time s is associated with above average SBP slope at t, especially for t around 55-60 years. There is a weakening of this effect at older ages, but it resumes its full force again at very old ages.

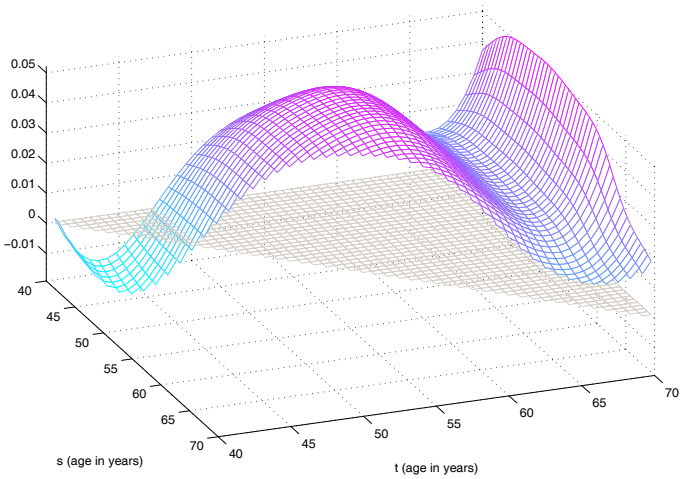

Fig. 6 Estimated transfer function $\gamma_{01}(s,t)$, $s < t$ (19), linking the first derivative $Y'(t)$ of SBP at t with the level $X(s)$ of BMI at s through $E(Y'(t)|X(s)) = \mu'_Y(t) + \gamma_{01}(s,t)(X(s) - \mu_X(s))$, with overlaid reference plane $z = 0$.

9 Conclusions

We found that functional approaches are of interest for the analysis of longitudinal data, even in the presence of highly irregular designs and sparse features. A basic assumption is that the observations are generated by stochastic processes with smooth sample paths. Dynamic transfer functions are a natural approach in longitudinal settings when one assumes underlying Gaussian processes, but they can also be used and lend themselves to interesting interpretations in the non-Gaussian case.

Dynamic transfer functions allow detailed studies of the relationships of function levels within one process at different times or across several processes. They may lead to interesting insights into the nature of the underlying mechanisms that manifest themselves in subject-specific trajectories. Transfer functions involving derivatives define empirical differential equations (in terms of conditional expectations). Applications include explorations of the underlying structure, prediction of future values for individuals and data-based checks for postulated equation systems. The shapes and features of the transfer functions are of interest in their own

right. Future research efforts are needed to explore this concept and associated inference further.

Developing new methods for the study and analysis of dynamic aspects of functional and longitudinal data poses many interesting challenges, for both theory and practice. Functional data analysis and more generally the assumption that the data are generated by smooth underlying but only partially observed stochastic processes provides a useful framework for such developments.

Acknowledgements H.G.M. wishes to thank María Ángeles Gil for her encouragement and warm hospitality during his stay at the University of Oviedo in 2008, during which the outline of this article was drafted. I am also grateful to Ana Colubi, Norberto Corral and Gil González-Rodríguez. This research was supported by NSF grant DMS-0806199.

Appendix

For $Z, W = X$ or Y, recall $G_{ZW}^{(\nu,\nu)}(s,t) = \mathrm{cov}(Z^{(\nu_1)}(s), W^{(\nu_2)}(t))$ as in (17). Then denote the smoothing bandwidth $h_{\mu,\nu}$ for estimating the mean function in (7) by $h_{Z,\nu}$, and the smoothing bandwidth $h_{G,\nu}$ for the covariance function in (8) by h_{Z,G_ν}. Given data (S_{im}, U_{im}, Q_i) and (T_{ij}, V_{ij}, R_i), $i = 1, \ldots, n$, $m = 1, \ldots, Q_i, j = 1, \ldots, R_i$, the following are basic assumptions we require for the consistency results.

(A1) Q_i is a random variable with $Q_i \sim Q$ i.i.d, where $Q > 0$ is a positive discrete random variable with $EQ < \infty$ and $P\{Q > 1\} > 0$, and $(\{S_{im}, m \in M_i\}, \{U_{im}, m \in M_i\})$ are independent of Q_i for $M_i \subseteq \{1, \ldots, Q_i\}$. Analogously for (T_{ij}, V_{ij}, R_i).
(A2) $h_{Z,0} \to 0, nh_{Z,0}^4 \to \infty, nh_{Z,0}^6 < \infty, nh_{Z,G_0}^6 \to \infty$, $h_{Z,G_0} \to 0$, and $nh_{Z,G_0}^8 < \infty$ for $Z = X$ or Y.
(A3) $h_{Z,1} \to 0$, $nh_{Z,1}^8 \to \infty$, $h_{Z,G_1} \to 0$ and $nh_{Z,G_1}^{10} \to \infty$ for $Z = X$ or Y.
(A4) $G_{ZZ}^{(\nu,\nu)}(s,t)$ is positive definite for $\nu \in \{0,1\}$, $Z = X$ or Y.
(A5) Finite fourth moments exist, $\mathrm{E}[(Z - \mu_Z(Z))^4] < \infty$, $Z = X$ or Y.
(A6) Without loss of generality, $h_1/h_2 \to 1$, $nh_1^6 \to \infty$, and $nh_1^8 < \infty$.
(A7) $\inf_s \mathrm{var}(X(s)) = \inf_s \sum_{k=1}^\infty \lambda_k \phi_k^2(s) = \varrho > 0$, $\sup_s \mathrm{var}(X(s)) = \sup_s \sum_{k=1}^\infty \lambda_k \phi_k^2(s) = \Delta < \infty$ and $\sup_t \sum_{l=1}^\infty \rho_l (\psi_k^{(\nu)}(t))^2 < \infty$.

Proof of Theorem.

$$|\hat\beta_{0\nu}^K(s,t) - \beta_{0\nu}(s,t)| = \left| \frac{\sum_{k=1}^K \hat\lambda_k \hat\phi_k(s) \hat\phi_k^{(\nu)}(t)}{\sum_{k=1}^K \hat\lambda_k (\hat\phi_k(s))^2} - \frac{\sum_{k=1}^\infty \lambda_k \phi_k(s) \phi_k^{(\nu)}(t)}{\sum_{k=1}^\infty \lambda_k (\phi_k(s))^2} \right|$$

$$\le \frac{1}{\sum_{k=1}^\infty \lambda_k \phi_k^2(s)} \left| \sum_{k=1}^K \hat\lambda_k \hat\phi_k(s) \hat\phi_k^{(\nu)}(t) - \sum_{k=1}^\infty \lambda_k \phi_k(s) \phi_k^{(\nu)}(t) \right|$$

$$+ \frac{\sum_{k=1}^K \hat\lambda_k \hat\phi_k(s) \hat\phi_k^{(\nu)}(t)}{\sum_{k=1}^K \hat\lambda_k (\hat\phi_k(s))^2 \sum_{k=1}^\infty \lambda_k (\phi_k(s))^2} \left| \sum_{k=1}^K \hat\lambda_k (\hat\phi_k(s))^2 - \sum_{k=1}^\infty \lambda_k (\phi_k(s))^2 \right| = I + II.$$

Here

$$\{\sum_{k=1}^\infty \lambda_k \phi_k^2(s)\} I = \left| \sum_{k=1}^K \hat\lambda_k \hat\phi_k(s) \hat\phi_k^{(\nu)}(t) - \sum_{k=1}^\infty \lambda_k \phi_k(s) \phi_k^{(\nu)}(t) \right|$$

$$\le \left| \sum_{k=1}^K \hat\lambda_k \hat\phi_k(s) \hat\phi_k^{(\nu)}(t) - \sum_{k=1}^K \lambda_k \phi_k(s) \phi_k^{(\nu)}(t) \right| + \left| \sum_{k=K+1}^\infty \lambda_k \phi_k(s) \phi_k^{(\nu)}(t) \right|.$$

From the basic square integrability assumptions and Mercer's theorem, the terms $|\sum_{k=K+1}^\infty \lambda_k \phi_k^2(s)|$ and $|\sum_{k=K+1}^\infty \lambda_k \phi_k(s) \phi_k^{(\nu)}(t)| \le [\sum_{k=K+1}^\infty \lambda_k \phi_k^2(s)]^{1/2} [\sum_{k=K+1}^\infty \lambda_k (\phi_k^{(\nu)}(t))^2]^{1/2}$

converge to 0 uniformly in $(s,t) \in \mathcal{T} \times \mathcal{T}$ as $K \to \infty$. On the other hand, for fixed K, under assumptions (A1)-(A4) and some additional regularity assumptions for kernels and underlying density functions, one has $|\hat{\lambda}_k - \lambda_k| = O_p(1/\sqrt{n}h^2_{X,G_0})$, $\sup_{t\in\mathcal{T}} |\hat{\phi}_k(t) - \phi_k(t)| = O_p(1/\sqrt{n}h^2_{X,G_0})$ and $\sup_{t\in\mathcal{T}} |\hat{\phi}'_k(t) - \phi'_k(t)| = O_p(1/\sqrt{n}h^2_{X,G_0} + 1/\sqrt{n}h^3_{X,G_1} + h^4_{X,G_1})$ (Liu and Müller, 2009). We conclude

$$\sup_{s,t\in\mathcal{T}\times\mathcal{T}} |\sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k(s)\hat{\phi}_k^{(\nu)}(t) - \sum_{k=1}^{K} \lambda_k \phi_k(s)\phi_k^{(\nu)}(t)| \to_p 0, \quad \sup_{s\in\mathcal{T}} |\sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k^2(s) - \sum_{k=1}^{K} \lambda_k \phi_k^2(s)| \to_p 0.$$

By (A7), $\lim_{K,L\to\infty} \lim_{n\to\infty} \sup_{s,t\in\mathcal{T}\times\mathcal{T}} I = \lim_{K,L\to\infty} \lim_{n\to\infty} \sup_{s,t\in\mathcal{T}\times\mathcal{T}} II = 0$, which leads to (26).

For transfer functions γ,

$$|\hat{\gamma}_{0\nu}^{KL}(s,t) - \gamma_{0\nu}(s,t)| = \left| \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} \hat{\sigma}_{kl}\hat{\phi}_k(s)\hat{\psi}_l^{(\nu)}(t)}{\sum_{k=1}^{K} \hat{\lambda}_k(\hat{\phi}_k(s))^2} - \frac{\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t)}{\sum_{k=1}^{\infty} \lambda_k(\phi_k(s))^2} \right|$$

$$\leq \frac{1}{\sum_{k=1}^{\infty} \lambda_k \phi_k^2(s)} \left| \sum_{k=1}^{K}\sum_{l=1}^{L} \hat{\sigma}_{kl}\hat{\phi}_k(s)\hat{\psi}_l^{(\nu)}(t) - \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t) \right|$$

$$+ \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} \hat{\sigma}_{kl}\hat{\phi}_k(s)\hat{\psi}_l^{(\nu)}(t)}{\sum_{k=1}^{K} \hat{\lambda}_k(\hat{\phi}_k(s))^2 \sum_{k=1}^{\infty} \lambda_k(\phi_k(s))^2} \left| \sum_{k=1}^{K} \hat{\lambda}_k(\hat{\phi}_k(s))^2 - \sum_{k=1}^{\infty} \lambda_k(\phi_k(s))^2 \right| = III + IV.$$

Similarly, as $K, L, n \to \infty$, $\sup_{s,t\in\mathcal{T}\times\mathcal{T}} III, IV \to 0$ holds, because

$$\left| \sum_{k=1}^{K}\sum_{l=1}^{L} \hat{\sigma}_{kl}\hat{\phi}_k(s)\hat{\psi}_l^{(\nu)}(t) - \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t) \right|$$

$$\leq \left| \sum_{k=1}^{K}\sum_{l=1}^{L} \hat{\sigma}_{kl}\hat{\phi}_k(s)\hat{\psi}_l^{(\nu)}(t) - \sum_{k=1}^{K}\sum_{l=1}^{L} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t) \right|$$

$$+ \left| \sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t) + \sum_{k=1}^{\infty}\sum_{l=L+1}^{\infty} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t) \right|,$$

and $|\sum_{k=K+1}^{\infty}\sum_{l=L+1}^{\infty} \sigma_{kl}\phi_k(s)\psi_l^{(\nu)}(t)| < |\sum_{k=K+1}^{\infty} \lambda_k\phi_k^2(s)|^{1/2} |\sum_{l=L+1}^{\infty} \rho_l(\psi_l^{(\nu)}(t))^2|^{1/2}$, in view of the consistency result for $\hat{\sigma}_{kl}$ that is provided in Yao et al. (2005b) under assumptions (A1)-(A6). This implies (27).

References

Ash RB, Gardner MF (1975) Topics in stochastic processes, vol. 27 of Probability and Mathematical Statistics. Academic Press [Harcourt Brace Jovanovich Publishers], New York.

Besse P, Ramsay JO (1986) Principal components analysis of sampled functions. Psychometrika 51:285–311

Cai T, Hall P (2006) Prediction in functional linear regression. Ann Statist 34:2159–2179

Cardot H, Crambes C, Kneip A, Sarda P (2007) Smoothing splines estimators in functional linear regression with errors-in-variables. Comput Statist Data Anal 51:4832–4848

Cardot H, Ferraty F, Mas A, Sarda P (2003a) Testing hypotheses in the functional linear model. Scand J Statist 30:241–255

Cardot H, Ferraty F, Sarda P (2003b) Spline estimators for the functional linear model. Statist Sinica 13:571–591

Castro PE, Lawton WH, Sylvestre EA (1986) Principal modes of variation for processes with continuous sample curves. Technometrics 28:329–337

Chiang C-T, Rice JA, Wu CO (2001) Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. J Amer Statist Assoc 96:605–619
Chiou J-M, Müller H-G, Wang J-L (2004) Functional response models. Statist Sinica 14:675–693
Chiou J-M, Müller H-G, Wang J-L, Carey JR (2003) A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. Statist Sinica 13:1119–1133
Cuevas A, Febrero M, Fraiman R (2002) Linear functional regression: the case of fixed design and functional response. Canad J Statist 30:285–300
Dauxois J, Pousse A, Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J Multivariate Anal 12:136–154
Dubin JA, Müller H-G (2005) Dynamical correlation for multivariate longitudinal data. J Amer Statist Assoc 100:872–881
Escabias M, Aguilera AM, Valderrama MJ (2004) Principal component estimation of functional logistic regression: discussion of two different approaches. J Nonparametr Stat 16:365–384
Eubank RL, Hsing T (2008) Canonical correlation for stochastic processes. Stoch Proc Appl 118:1634–1661
Fan J, Gijbels I (1996) Local polynomial modelling and its applications, vol. 66 of Monographs on Statistics and Applied Probability. Chapman & Hall, London
Fan J, Zhang J-T (2000) Two-step estimation of functional linear models with applications to longitudinal data. J Roy Statist Soc Ser B 62:303–322
Fan J, Zhang W (2008) Statistical methods with varying coefficient models. Stat Interface 1:179–195
Faraway JJ (1997) Regression analysis for a functional response. Technometrics 39:254–261
Ferraty F, Mas A, Vieu P (2007) Nonparametric regression on functional data: inference and practical aspects. Aust N Z J Stat 49:459–461
Ferraty F, Vieu P (2006) Nonparametric Functional Data Analysis. Springer, New York
Gasser T, Kneip A (1995) Searching for structure in curve samples. J Amer Statist Assoc 90:1179–1188
Gasser T, Müller H-G (1984) Estimating regression functions and their derivatives by the kernel method. Scand J Statist 11:171–185
Gasser T, Müller H-G, Köhler W, Molinari L, Prader A (1984) Nonparametric regression analysis of growth curves. Ann Statist 12:210–229
Gervini D, Gasser T (2004) Self-modeling warping functions. J Roy Statist Soc Ser B 66:959–971
Gervini D, Gasser T (2005) Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. Biometrika 92:801–820
Grenander U (1950) Stochastic processes and statistical inference. Arkiv för Matematik 1:195–277
Hall P, Horowitz JL (2007) Methodology and convergence rates for functional linear regression. Ann Statist 35:70–91
Hall P, Müller H-G, Wang J-L (2006) Properties of principal component methods for functional and longitudinal data analysis. Ann Statist 34:1493–1517
Hall P, Müller H-G, Yao F (2008) Modeling sparse generalized longitudinal observations with latent gaussian processes. J Roy Statist Soc Ser B 70:730–723
He G, Müller H-G, Wang J-L (2000) Extending correlation and regression from multivariate to functional data. In: Puri ML (ed) Asymptotics in statistics and probability. VSP International Science Publishers.
He G, Müller H-G, Wang J-L (2003) Functional canonical analysis for square integrable stochastic processes. J Multivariate Anal 85:54–77
He G, Müller H-G, Wang J-L (2004) Methods of canonical analysis for functional data. J Statist Plann Inference 122:141–159
Heckman NE, Zamar RH (2000) Comparing the shapes of regression functions. Biometrika 87:135–144
Hoover DR, Rice JA, Wu CO, Yang L-P (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85:809–822
Hotelling H (1936) Relations between two sets of variates. Biometrika 28:321–377
Huang JZ, Wu CO, Zhou L (2004) Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statist Sinica 14:763–788

James GM (2002) Generalized linear models with functional predictors. J Roy Statist Soc Ser B 64:411–432
James GM, Hastie TJ, Sugar CA (2000) Principal component models for sparse functional data. Biometrika 87:587–602
Jank W, Shmueli G (2006) Functional data analysis in electronic commerce research. Statistical Science 21:155–166
Karhunen K (1946) Zur Spektraltheorie stochastischer Prozesse. Annales Academiae Scientiarum Fennicae. Series A. I, Mathematica 1946:7
Kato T (1995) Perturbation theory for linear operators. Springer-Verlag, Berlin
Kirkpatrick M, Heckman N (1989) A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. J Math Biol 27:429–450
Kneip A, Ramsay JO (2008) Combining registration and fitting for functional models. J Amer Statist Assoc 103:1155–1165
Leurgans SE, Moyeed RA, Silverman BW (1993) Canonical correlation analysis when the data are curves. J Roy Statist Soc Ser B 55:725–740
Lin X, Carroll RJ (2001a) Semiparametric regression for clustered data. Biometrika 88:1179–1185
Lin X, Carroll RJ (2001b). Semiparametric regression for clustered data using generalized estimating equations. J Amer Statist Assoc 96:1045–1056
Liu B, Müller H-G (2009) Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics. J Amer Statist Assoc (In press)
Malfait N, Ramsay JO (2003) The historical functional linear model. Canad J Statist 31:115–128
Mas A, Pumo B (2008) Functional linear regression with derivatives. Manuscript
Müller H-G (2005) Functional modelling and classification of longitudinal data. Scand J Statist 32:223–240
Müller H-G (2008) Functional modeling of longitudinal data. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) Longitudinal Data Analysis (Handbooks of Modern Statistical Methods). Chapman & Hall/CRC, New York, 223–252
Müller H-G, Chiou J-M, Leng X (2008) Inferring gene expression dynamics via functional regression analysis. BMC Bioinformatics 9:60
Müller H-G, Stadtmüller U (2005) Generalized functional linear models. Ann Statist 33:774–805
Müller H-G, Yao F (2006) Regressing longitudinal response trajectories on a covariate. In: Fan J, Koul HL (eds) Frontiers in Statistics. Festschrift for Peter Bickel Imperial College Press, London
Müller H-G, Yao F (2008) Functional additive models. J Amer Statist Assoc 103:1534–1544
Müller H-G, Zhang Y (2005) Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. Biometrics 61:1064–1075
Opgen-Rhein R, Strimmer K (2006) Inferring gene dependency networks from genomic longitudinal data: A functional data approach. REVSTAT - Statistical Journal 4:53–65
Pearson JD, Morrell CH, Brant LJ, Landis PK (1997) Gender differences in a longitudinal study of age associated changes in blood pressure. J Gerontol A Biol Sci Med Sci 52:177–183
Qu, A, Li R (2006) Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics 62:379–391
Ramsay JO, Dalzell CJ (1991) Some tools for functional data analysis. J Roy Statist Soc Ser B 53:539–572
Ramsay JO, Li X (1998) Curve registration. J Roy Statist Soc Ser B 60:351–363
Ramsay JO, Silverman BW (2002) Applied functional data analysis. Springer Series in Statistics. Springer-Verlag, New York
Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd ed. Springer Series in Statistics, Springer, New York
Rao CR (1958) Some statistical methods for comparison of growth curves. Biometrics 14:1–17
Reddy SK, Dass M (2006) Modeling on-line art auction dynamics using functional data analysis. Statistical Sci 21:179–193
Rice JA (2004) Functional and longitudinal data analysis: Perspectives on smoothing. Statist Sinica 14:631–647.
Rice JA, Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. J Roy Statist Soc Ser B 53:233–243

Rice JA, Wu CO (2001) Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 57:253–259

Şentürk D, Müller H-G (2008) Generalized varying coefficient models for longitudinal data. Biometrika 95:653–666

Şentürk D, Müller H-G (2009) Functional varying coefficient models for longitudinal data. Preprint

Service SK, Rice JA, Chavez FP (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

Shi M, Weiss RE, Taylor JMG (1996) An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. Ann Statist 45:151–163

Shock NW, Greulich RC, Andres R, Lakatta EG, Arenberg D, Tobin JD (1984) Normal human aging: The Baltimore Longitudinal Study of Aging. In: NIH Publication No. 84-2450. U.S. Government Printing Office, Washington, D.C.

Staniswalis JG, Lee JJ (1998) Nonparametric regression analysis of longitudinal data. J Amer Statist Assoc 93:1403–1418

Tang R, Müller H-G (2008) Pairwise curve synchronization for functional data. Biometrika 95:875–889

Tang R, Müller H-G (2009) Time-synchronized clustering of gene expression trajectories. Biostatistics 10:32–45

Wang N, Carroll RJ, Lin X (2005) Efficient semiparametric marginal estimation for longitudinal/clustered data. J Amer Statist Assoc 100:147–157

Wang S, Jank W, Shmueli G, Smith P (2008) Modeling price dynamics in ebay auctions using principal differential analysis. J Amer Statist Assoc 103(483):1100-1118

Wu CO, Yu KF, Chiang C-T (2000) A two-step smoothing method for varying-coefficient models with repeated measurements. Ann Inst Statist Math 52:519–543

Yao F, Lee TCM (2006) Penalized spline models for functional principal component analysis. J Roy Statist Soc Ser B 68:3–25

Yao F, Müller H-G, Wang J-L (2005a) Functional data analysis for sparse longitudinal data. J Amer Statist Assoc 100:577–590

Yao F, Müller H-G, Wang J-L (2005b) Functional linear regression analysis for longitudinal data. Ann Statist 33:2873–2903

Zhao X, Marron JS, Wells MT (2004) The functional data analysis view of longitudinal data. Statist Sinica 14:789–808

Zhou L, Huang JZ, Carroll R (2008) Joint modelling of paired sparse functional data using principal components. Biometrika 95:601–619