Estimating Derivatives for Samples of Sparsely Observed Functions, with Application to On-line Auction Dynamics

Bitao Liu and Hans-Georg Müller

Oct 19, 2008

ABSTRACT

It is often of interest to recover derivatives of a sample of random functions from sparse and noise-contaminated measurements, especially when the dynamics of underlying processes is of interest. We propose a novel approach based on estimating derivatives of eigenfunctions and expansions of random functions into their eigenfunctions to obtain a representation for derivatives. In combination with estimates for functional principal component scores for sparse data, this leads to a viable solution of the challenging problem to recover derivatives for sparsely observed functions. We establish consistency results and demonstrate in simulations that the method is superior to alternative approaches (derivative estimation with random effects models based on B-spline bases, kernel smoothing, smoothing splines, or P-splines). Our study is motivated by an analysis of bidding histories for eBay auctions, where bids are typically very sparse in the middle and somewhat more frequent near the beginning and end of an auction. We demonstrate the estimation of derivatives of price curves for individual auctions from the sparsely observed bidding histories and also derive a model-free first order differential equation that applies in the case of Gaussian processes. This provides a data-driven dynamic model which we employ to elucidate auction dynamics.

KEY WORDS: Dynamic system, Eigenfunction, Functional Data Analysis, Price Curve, Smoothing.

Bitao Liu is a Ph.D. candidate, Department of Statistics, University of California, Davis, One Shields Ave., Davis, CA 95616 (E-mail: *btliu@wald.ucdavis.edu*). Hans-Georg Müller is Professor, Department of Statistics, University of California, Davis, One Shields Ave., Davis, CA 95616 (E-mail: *mueller@wald.ucdavis.edu*). The authors express their gratitude to the associate editor and two referees for insightful comments that led to many improvements in the paper and thank Wolfgang Jank for providing the Palm M515 PDA data. This research was supported by NSF grants DMS03-54448, DMS05-05537 and DMS08-06199.

1. INTRODUCTION

Our objective is to recover the derivatives of sparsely observed random trajectories. This poses a challenging problem since classical numerical differentiation (Ahnert and Abel 2007; Ling 2006) as well as statistical estimation of derivatives for functions observed with noise (Gasser and Müller 1984; Abramson and Müller 1994) typically require densely spaced adjacent observations, due to the fact that derivatives are defined as limits of difference quotients. An area where the problem of differentiation under noise is commonly encountered is the analysis of growth curves (Gasser et al. 1984).

In cases of reasonably dense measurement designs, difference quotients have traditionally been used as raw observations of derivatives, against which more sophisticated estimates can be gauged. Difference quotients are severely biased and therefore rather useless for sparse data. An instance where such data arise are online auctions. For example, eBay hosts millions of auctions of goods and services everyday. A better understanding of the dynamics of bidding is clearly of interest. For a typical auction, observed bid arrival times are very sparse during mid-auction, somewhat denser but still relatively sparse near the beginning and very dense at the end of an auction ("bid sniping"). The observed bid histories of auctions for the same product may be viewed as a sample of realizations of an underlying smooth stochastic process, the *price process*. To study auction dynamics, we target the derivatives of the price trajectories which requires methodology applicable for sparse data.

We note that all currently available statistical nonparametric methods for estimating derivatives that are supported by theory require densely observed data. These methods include various types of spline estimates, based on B-splines (de Boor 1972), smoothing splines (Chambers 1991; Zhou and Wolfe 2000) and P-splines (Ramsay and Silverman 2005), or kernel-based estimators, such as convolution-type kernel (Gasser and Müller 1984) or local polynomial estimators (Fan and Gijbels 1996). Dense designs are thus

an inherent requirement when estimating derivatives for single functions. The situation turns out to be quite different when sparse measurements on each of a sample of Gaussian random functions are available. As we will demonstrate, in this situation a reasonable approximation to the derivative of individual curves exists, which can be consistently estimated, even in a situation where for each curve one has only very few observations.

To achieve this goal, we extend a functional data analysis (FDA) method, where one substitutes functional principal components (FPC) by their conditional expectations, given the data (Yao et al. 2005), to the case of derivatives. The idea of functional principal component analysis (FPCA) is based on the spectral decomposition of the autocovariance operator of the underlying random functions and the FPCs serve as random effects in a representation of each random curve in terms of the orthonormal eigenfunctions. Early work in harnessing this expansion includes Rao (1958), Besse and Ramsay (1986) and Rice and Silverman (1991). Theoretical properties of FPCA have been studied in Silverman (1996), Hall and Hosseini-Nasab (2006) and Hall et al. (2006), while Ramsay and Silverman (2002, 2005) provide excellent overviews. Recent and ongoing work is establishing connections with the solutions of differential equations for densely observed functional data (Ramsay et al. 2007) and also with longitudinal data analysis.

Shi et al. (1996) and Rice and Wu (2001) proposed functional mixed effect models using B-spline basis functions. This approach can also be used for derivative estimation, and is included in our simulation comparisons; a variant targeting auction data was recently proposed in Reithinger et al. (2008). Apart from this, FDA methods for the recovery of derivatives of auction price functions were previously obtained by fitting high order P-spline functions (Jank and Shmueli 2005a; Reddy and Dass 2006; Bapna et al. 2008; Wang et al. 2008b). The sparse and irregularly spacing of the bid arrivals is addressed in this setting through a pre-processing step to interpolate the sparse data. Subsequently, FDA procedures are applied that are identical to those used in functional frameworks of fully

available noise-free trajectories. This approach appears problematic, as the interpolation step may lead to biases and inconsistency.

Our proposed method for derivative estimation is based on a preliminary step in which we obtain derivatives of the eigenfunctions. Obtaining these derivatives is a key step, as direct observations are not available. The only requirements for the proposed methodology are smoothness of mean, covariance and eigenfunctions and that the pooled locations of the measurements are dense in the one-dimensional domain and their pairs are dense in the square of the domain. This approach works because of the pooling of observations across all subjects (or auctions in the application), thus drawing strength from all observed data and not only relying on the sparse data that are available for a specific trajectory.

In section 2, we present the data model and estimation methods for recovering derivatives. Results on the asymptotic properties for estimating the first derivative are given in section 3, with proofs deferred to an appendix. In section 4, we compare our methods with existing approaches for derivative estimation, including functional linear mixed effect models, in a simulation study under different settings. Differentiation of sparsely sampled trajectories is illustrated in section 5 with 156 online auction bid histories for Palm M515 personal digital assistants (PDAs). In section 6 we describe alternative approaches to explore the dynamics of random processes and derive a differential equation that can be universally applied to describe the dynamics of Gaussian processes and is demonstrated for auction dynamics. Concluding remarks are in section 7.

2. FUNCTIONAL DATA ANALYSIS FOR DERIVATIVES

2.1 Background on Functional Principal Component Analysis

The sample of functions that give rise to the observed data are viewed as realizations of a smooth and square integrable random process $X(\cdot)$, defined on a finite interval

$\mathcal{T} = [0, T]$ with unknown mean function $EX(t) = \mu(t)$ and unknown covariance function $\text{cov}(X(t), X(s)) = G(t, s),\ s, t \in \mathcal{T}$. The function G is assumed to be symmetric, positive definite and continuous. The autocovariance operator A_G, defined by $(A_G f)(t) = \int_{\mathcal{T}} G(t, s) f(s) ds$ for any f in the Hilbert space L^2, is a Hilbert-Schmidt operator (see Ash and Gardner 1975), with eigenvalues λ_k and orthonormal eigenfunctions $\phi_k(t)$, such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and $G(t, s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(s)$. The random trajectories can be represented by the Karhunen-Loève expansion (see, e.g., Rice and Silverman 1991) $X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t)$, where the $\xi_{ik} = \int_{\mathcal{T}} (X_i(t) - \mu(t)) \phi_k(t) dt$, $k = 1, 2, \ldots$ are the functional principal component scores (FPCs) of X_i, for $k = 1, 2, \ldots$. These are uncorrelated random variables with $\text{E}(\xi_{ik}) = 0$, $E(\xi_{ik} \xi_{ik'}) = 0$ for $k \neq k'$, and $E\xi_{ik}^2 = \lambda_k$, with $\sum_k \lambda_k < \infty$. Upon differentiation, assuming derivatives exist as needed,

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

where $\mu^{(\nu)}(t)$ and $\phi_k^{(\nu)}(t)$ denote the νth derivative 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 side conditions of norm 1 and orthogonality on all previous eigenfunctions. Upon taking the νth derivative with respect to t from both sides of these eigen-equations,

$$\frac{d^\nu}{dt^\nu} \int_{\mathcal{T}} G(t, s) \phi_k(s) ds = \lambda_k \frac{d^\nu}{dt^\nu} \phi_k(t). \tag{2}$$

If $\frac{d^\nu}{dt^\nu} G(t, s) \phi_k(s)$ exists for all $t, s \in \mathcal{T}$ and $|\frac{d^\nu}{dt^\nu} G(t, s) \phi_k(s)|$ is bounded and integrable for all $t, s \in \mathcal{T}$, interchanging integration and differentiation leads to

$$\phi_k^{(\nu)}(t) = \frac{1}{\lambda_k} \int_{\mathcal{T}} \frac{\partial^\nu}{\partial t^\nu} G(t, s) \phi_k(s) ds. \tag{3}$$

This is a key relationship which ties the needed derivatives of eigenfunctions to the task of estimating a partial derivative of the covariance function. This strategy allows to reduce the difficult problem of obtaining derivatives for the eigenfunctions, which are only implicitly defined through the eigen-equations, to a variant of the much more straightforward

covariance surface estimation paradigm, taking full advantage of pooling all observed data from all trajectories and thus circumventing the sparse sampling problem.

Sparse sampling of trajectories is modeled as follows: Denote by T_{ij} the jth time at which trajectory X_i is observed, through a measurement Y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, N_i$. Then the number of repeated measurements N_i, is assumed to be a i.i.d. bounded random variable (independent of other random variables). The measurements Y_{ij} made at times T_{ij} for the ith trajectory are then

$$Y_{ij} = Y_i(T_{ij}) = X_i(T_{ij}) + \varepsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(T_{ij}) + \varepsilon_{ij}, \tag{4}$$

where the ε_{ij} are additional measurement errors, assumed to be i.i.d. with $E\varepsilon_{ij} = 0$ and $E\varepsilon_{ij}^2 = \sigma^2$, and such that ε_{ij} are independent of ξ_{ik} for all i, j, k.

2.2 Estimating Derivatives of Eigenfunctions

We aim at identifying the components of the derivative representation (1). To obtain $\mu^{(\nu)}(t)$, we apply local polynomial fitting to a pooled scatterplot where one aggregates the observed measurements from all sample trajectories. For each t, we minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} \kappa_1\left(\frac{T_{ij}-t}{h_{\mu,\nu}}\right)\{Y_{ij} - \sum_{m=0}^{\nu+1}\alpha_m(T_{ij}-t)^m\}^2 \tag{5}$$

with respect to α_m for $m = 0, \ldots, \nu+1$, and denote the minimizing value for α_ν by $\hat{\alpha}_\nu(t)$, leading to $\hat{\mu}^{(\nu)}(t) = \hat{\alpha}_\nu(t)\nu!$, where the factor $\nu!$ comes from a Taylor expansion, as described in more detail in Fan and Gijbels (1996), equation (3.2) of Chapter 3. The kernel κ_1 is a univariate density function and $h_{\mu,\nu}$ is a positive bandwidth.

Using from now on the notation $G^{(\nu,0)}(t,s) = \frac{\partial^\nu}{\partial t^\nu}G(t,s)$, estimates for the derivatives $\phi_k^{(\nu)}(t)$ of the eigenfunctions ϕ_k are based on the key relation eq. (3), which tells us that it suffices to find an estimator $\hat{G}^{(\nu,0)}(t,s)$ of $G^{(\nu,0)}(t,s)$,

$$\hat{\phi}_k^{(\nu)}(t) = \frac{1}{\hat{\lambda}_k}\int_{\mathcal{T}} \hat{G}^{(\nu,0)}(t,s)\hat{\phi}_k(s)ds. \tag{6}$$

Estimates $\hat{\lambda}_k$ and $\hat{\phi}_k$ for λ_k and ϕ_k are obtained from available procedures to conduct FPCA in the sparse data case, adopting implementations described in Yao et al. (2005).

To obtain $\hat{G}^{(\nu,0)}(t,s)$, we fix (t,s) and minimize

$$\sum_{i=1}^{n} \sum_{1\le j\neq l\le N_i} \kappa_2\left(\frac{T_{ij}-t}{h_{G_{1,\nu}}}, \frac{T_{il}-s}{h_{G_{2,\nu}}}\right)\{G_i(T_{ij},T_{il}) - (\sum_{m=0}^{\nu+1}\alpha_{1m}(T_{ij}-t)^m + \alpha_{21}(T_{il}-s))\}^2 \quad (7)$$

with respect to α_{1m} and α_{21} for $m = 1,\ldots,\nu+1$, where $G_i(T_{ij},T_{il}) = (Y_{ij}-\hat{\mu}(T_{ij}))(Y_{il}-\hat{\mu}(T_{il}))$, $j \neq l$, are the raw covariances, excluding the diagonal elements which are contaminated by the additional noise variance σ^2, κ_2 is a bivariate density function and $h_{G_{1,\nu}}$, $h_{G_{2,\nu}}$ are positive bandwidths. Denoting the minimizing value for $\alpha_{1\nu}$ by $\hat{\alpha}_{1\nu}(t,s)$, this leads to the estimate $\hat{G}^{(\nu,0)}(t,s) = \hat{\alpha}_{1\nu}(t,s)\nu!$, where we typically set $h_{G_{1,\nu}} = h_{G_{2,\nu}} = h_{G,\nu}$. For the case $\nu = 0$, the selection of the bandwidth $h_{G,0}$ through generalized cross-validation (GCV) worked well in practice, while for the cases $\nu > 0$ we apply a modifying factor to the GCV bandwidth, analogously to the factor method that has been proposed for bandwidth choice for derivatives (Müller et al. 1987). We note that in practice, one obtains $\hat{\lambda}_k$ and $\hat{\phi}_k(t)$ through spectral decomposition of the discretized smoothed covariance (Rice and Silverman 1991). Since the pooled data, which are typically plentiful, form the basis for estimating the covariance surface or its partial derivative, corresponding estimates are typically well behaved.

2.3 Estimation of Functional Principal Component Scores by Conditioning

Traditionally, the FPCs ξ_{ik} are estimated by numerical integration, which can be expected to work only if the observed data are dense and regularly spaced without measurement errors. Since this is rarely the case for longitudinal data and especially for the auction data that motivate our study, we apply instead the principal analysis by conditional expectation (PACE) approach of Yao et al. (2005).

Let $\mathbf{X}_i = (X_i(T_{i1}),\ldots,X_i(T_{iN_i}))^T$, $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{iN_i})^T$, $\boldsymbol{\mu}_i = (\mu(T_{i1}),\ldots,\mu(T_{iN_i}))^T$, $\boldsymbol{\phi}_{ik} = (\phi_k(T_{i1}),\ldots,\phi_k(T_{iN_i}))^T$, and assume ξ_{ik} and ε_{ij} in (4) are jointly Gaussian. The best

prediction of the k-th FPC score ξ_{ik} of X_i, given the observations for the specific trajectory is $\tilde{\xi}_{ik} = \mathrm{E}[\xi_{ik}|\mathbf{Y}_i] = \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$, where $\boldsymbol{\Sigma}_{\mathbf{Y}_i} = \mathrm{cov}(\mathbf{Y}_i, \mathbf{Y}_i) = \mathrm{cov}(\mathbf{X}_i, \mathbf{X}_i) + \sigma^2 \mathbf{I}_{N_i}$. If the data are not Gaussian, this is the best linear predictor. Substituting estimates for $\lambda_k, \boldsymbol{\phi}_{ik}, \boldsymbol{\Sigma}_{\mathbf{Y}_i}$, and $\boldsymbol{\mu}_i$, estimates are

$$\hat{\xi}_{ik} = \hat{\mathrm{E}}[\xi_{ik}|\mathbf{Y}_i] = \hat{\lambda}_k \hat{\boldsymbol{\phi}}_{ik}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i), \tag{8}$$

where the (j,l)th element of $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}$ is

$$(\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i})_{j,l} = \hat{G}(T_{ij}, T_{il}) + \hat{\sigma}^2 \delta_{jl}. \tag{9}$$

Here, $\hat{G}(T_{ij}, T_{il})$ can be estimated from (7) with $\nu = 0$ and $\hat{\sigma}^2$ is estimated as described in equation (2) of Yao et al. (2005).

Practical implementation of $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1}$ often requires a regularization step, as $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}$ may be close to singular for some i. Details can be found in Appendix A.1. Including K components then leads to estimates

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

for the derivatives of the trajectories. The choice of the number of included components K can be based on the leave-one-curve-out score (Rice and Silverman 1991). Faster and often better selectors are AIC (Shibata 1981) and BIC (Schwarz 1978) type of criteria, based on either a marginal pseudo-Gaussian log-likelihood (Rice and Wu 2001),

$$\hat{L}_1 = \sum_{i=1}^{n} \left\{ -\frac{N_i}{2} \log(2\pi) - \frac{1}{2} \log(\det \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}) - \frac{1}{2} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i) \right\},$$

or a conditional pseudo-Gaussian log-likelihood (Yao et al. 2005)

$$\begin{aligned} \hat{L}_2 \;=\; & \sum_{i=1}^{n} \left\{ -\frac{N_i}{2} \log(2\pi) - \frac{N_i}{2} \log \hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2} \left(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i - \sum_{k=1}^{K} \hat{\xi}_{ik} \widehat{\boldsymbol{\phi}}_{ik} \right)^T \right. \\ & \left. \times \left(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i - \sum_{k=1}^{K} \hat{\xi}_{ik} \widehat{\boldsymbol{\phi}}_{ik} \right) \right\}. \end{aligned}$$

One then chooses K through minimizing either $AIC_r(K) = -2\hat{L}_r + 2K$ or $BIC_r(K) = -2\hat{L}_r + K\log(N)$ where $r = 1, 2$, $N = \sum_{i=1}^{n} N_i$. Another fast and stable option is to choose the smallest K such that a preset level of variation is explained when including K components, where a level of 85% gives good practical results in many situations.

3. ASYMPTOTIC CONSISTENCY

We give a consistency result for the first derivative of the curve estimator by obtaining uniform convergence rates for local quadratic derivative estimators of the mean function, of the partial derivative of covariance function, and of the derivative of eigenfunctions. These results can be analogously extended to higher order derivatives. Additional assumptions are in Appendix A.2 and proofs in Appendix A.3.

Given data (T_{ij}, Y_{ij}), $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, where N_i is the number of observations for the ith trajectory, the following are basic assumptions.

(A1) N_i is a random variable with $N_i \overset{iid}{\sim} N$, where $N > 0$ is a positive discrete random variable with $\mathrm{E}N < \infty$ and $P\{N > 1\} > 0$, and $(\{T_{ij}, j \in J_i\}, \{Y_{ij}, j \in J_i\})$ are independent of N_i, for $J_i \subseteq \{1, \ldots, N_i\}$.

The (T_{ij}, Y_{ij}) are assumed to have the same distribution as (T, Y), with joint density $g(t, y)$, where we require $\mathrm{E}[(Y - \mu(T))^4] < \infty$, and the T_{ij} are i.i.d. with marginal density f; Y_{ij} and Y_{il} generally will be dependent. For kernels and underlying density functions, we need assumptions (B1)-(B3) in Appendix A.2. For the bandwidths, as $n \to \infty$,

(A2) $h_{\mu,0} \to 0$, $nh_{\mu,0}^4 \to \infty$, $nh_{\mu,0}^6 < \infty$, $h_{G,0} \to 0$, $nh_{G,0}^6 \to \infty$, and $nh_{G,0}^8 < \infty$.

(A3) $h_{\mu,1} \to 0$, $nh_{\mu,1}^8 \to \infty$, $h_{G,1} \to 0$, $nh_{G,1}^{10} \to \infty$.

Theorem 1. *Under (A1)-(A3) and (B1)-(B3),*

$$\sup_{t \in \mathcal{T}} |\hat{\mu}'(t) - \mu'(t)| = O_p(\frac{1}{\sqrt{n}h_{\mu,1}^2} + h_{\mu,1}^4), \tag{11}$$

$$\sup_{t,s\in\mathcal{T}} |\hat{G}^{(1,0)}(t,s) - G^{(1,0)}(t,s)| = O_p(\frac{1}{\sqrt{n}h_{G,1}^3} + h_{G,1}^4). \tag{12}$$

Uniform convergence of $\hat{\phi}'_k(t)$ is an immediate consequence.

Theorem 2. *Under (A1)-(A3) and (B1)-(B4),*

$$\sup_{t\in\mathcal{T}} |\hat{\phi}'_k(t) - \phi'_k(t)| = O_p(\frac{1}{\sqrt{n}h_{G,0}^2} + \frac{1}{\sqrt{n}h_{G,1}^3} + h_{G,1}^4), k \in \mathcal{I}' \tag{13}$$

where $\mathcal{I}'$ denote the set of indices of the eigenfunctions ϕ_k corresponding to eigenvalues λ_k of multiplicity 1. Finally, consistency of $\hat{X}'_{i,K}(t)$ for $\widetilde{X}'_i(t)$ is obtained under the two additional assumptions (B4) and (B5) (Appendix A.2).

Theorem 3. *Under (A1)-(A3) and (B1)-(B5) for all $t \in \mathcal{T}$,*

$$\lim_{K\to\infty} \lim_{n\to\infty} \hat{X}'_{i,K}(t) = \widetilde{X}'_i(t) \quad \textit{in probability} \tag{14}$$

where $\widetilde{X}'_i(t) = \mu'(t) + \sum_{k=1}^{\infty} \tilde{\xi}_{ik}\phi'_k(t)$.

Theorem 3 provides convergence of the proposed derivative estimates to the conditional derivative process, conditioning on the observed data. Under additional regularity assumptions, extending arguments used in the proof of Theorem 4 of Yao et al. (2005), the relationship between the estimated derivatives and the unconditional derivatives of process X can be studied further. Let $\hat{\mathbf{H}} = (\hat{\lambda}_1\hat{\boldsymbol{\phi}}_{i1}, \ldots, \hat{\lambda}_K\hat{\boldsymbol{\phi}}_{iK})^T$, $\hat{\mathbf{\Lambda}}$ =diag$\{\hat{\lambda}_1, \ldots, \hat{\lambda}_K\}$ and $\hat{\mathbf{\Omega}}_K = \hat{\mathbf{\Lambda}} - \hat{\mathbf{H}}\hat{\mathbf{\Sigma}}_{\mathbf{Y}_i}^{-1}\hat{\mathbf{H}}$, with $\hat{\mathbf{\Sigma}}_{\mathbf{Y}_i}^{-1}$ as in (9), furthermore $\hat{\omega}_K(t,s) = \hat{\boldsymbol{\phi}}'^T_{K,t}\hat{\mathbf{\Omega}}_K\hat{\boldsymbol{\phi}}'_{K,s}$ where $\hat{\boldsymbol{\phi}}'^T_{K,t} = (\hat{\phi}'_1(t), \ldots, \hat{\phi}'_K(t))^T$, for $t, s \in \mathcal{T}$. Then the difference between $\hat{X}'_{i,K}(t)$ and $X'_i(t)$ satisfies $\lim_{K\to\infty}\lim_{n\to\infty} P\left\{\frac{\hat{X}'_{i,K}(t)-X'_i(t)}{\sqrt{\hat{\omega}_K(t,t)}} \le x\right\} = \Phi(x)$, where Φ is the standard Gaussian cdf. This result can be used for pointwise inference about derivative estimates.

4. SIMULATION STUDIES

The following methods were studied: Two versions of the PACE approach, one based on differentiation of eigenfunctions and eq. (10), referred to as PACE-EIG, and a variant

PACE-QUO, which is a two-step procedure based on an initial PACE step, followed by forming first or second order difference quotients from the smooth curves obtained in the first step which are then interpolated to obtain estimates for first or second derivative; ME-SPLINE – mixed effects model with B-spline basis functions; KERNEL – convolution-type kernel estimation; CUBIC SPLINE – cubic smoothing spline and P-SPLINE – penalized spline. The aim at each simulation was to recover derivatives of random trajectories $X_i^{(\nu)}(t)$ for $\nu = 0, 1, 2$, $i = 1, \ldots, n$, from noise-contaminated discrete measurements that were generated according to model (4). Each simulation consisted of 500 Monte Carlo samples and $n = 156$ random trajectories per sample, mimicking the auction data.

For all simulations, mean function $\mu(t) = 4t + \frac{1}{\sqrt{2\pi}(.1)} \exp\{-\frac{(t-0.5)^2}{2(.1)^2}\}$, eight eigenfunctions $\phi_k(t) = \sqrt{2}\sin(\pi k t)$, for $t \in \mathcal{T} = [0, 1]$ and eigenvalues $\lambda_k = 1/2^{k-1}$, $k = 1, \ldots, 8$ were chosen. FPC scores ξ_{ik} were generated from $\mathcal{N}(0, \lambda_k)$, $k = 1, \ldots, 8$ and measurement errors ε_{ij} from $\mathcal{N}(0, \sigma^2)$. For simulations with irregular measurement times, these were assumed to be randomly located according to a Beta distribution with mean 0.4 and variance 0.3, to reflect the bathtub shape of the density of bid arrival times as observed for auction data. Three scenarios were considered, where SIMULATION B comes closest to the designs observed in the auction data.

SIMULATION A – IRREGULAR AND SPARSE OBSERVATIONS: The number N_i of available observations per trajectory was generated from a Poisson distribution with mean 8, under the constraint that $N_i \geq 2$, and choosing $\sigma^2 = 0.5$.

SIMULATION B – IRREGULAR BUT DENSER OBSERVATIONS: Same setting as SIMULATION A, requiring $N_i \geq 8$ and choosing $\sigma^2 = 0.1$.

SIMULATION C – REGULAR AND DENSE OBSERVATIONS: Each trajectory is observed at 50 equidistant time points on the domain $\mathcal{T} = [0, 1]$, $\sigma^2 = 0.1$.

The specifications for PACE included univariate kernels $\kappa_1(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ for (5), bivariate kernels $\kappa_2(x, y) = \frac{1}{2\pi} \exp(-\frac{x^2+y^2}{2})$ for (7) and bandwidths $h_{\mu,\nu}$ and $h_{G,\nu}$

determined by GCV and multiplied with empirically determined factors (1.1, 1.2, 2) for irregular and sparse data and (1.1, 0.8, 0.8) for regular and dense data, for $\nu = 0, 1, 2$, respectively. The number of included components was determined by the BIC_1 criterion, as defined at the end of Section 2, which is also used for ME-SPLINE. We found that PACE is not overly sensitive to minor misspecifications of bandwidths or number of included components. Implementations for ME-SPLINE were based on the Matlab code provided by the authors, using default settings; for KERNEL on the glkerns() function from the lokern R package using a convolution-type kernel estimator with adaptive global plug-in bandwidth; for CUBIC SPLINE on the smooth.spline() function from the stats R package with GCV bandwidth choice; and for P-SPLINE on the smooth.Pspline() function from the pspline R package (with order of spline 2 and CV bandwidths).

To evaluate the performance of the various methods, we use mean and median of the relative mean integrated square error,

$$\mathsf{RMISE} = \frac{\sum_{i=1}^{n} \int_0^1 (\hat{X}_i^{(\nu)}(t) - X_i^{(\nu)}(t))^2 dt/n}{\sum_{i=1}^{n} \int_0^1 (X_i^{(\nu)}(t))^2 dt/n}$$

for $\nu = 0, 1, 2$. Besides RMISE, an important criterion is the fraction of simulations for which the various methods produced estimates, and the reported values of RMISE are based on these simulations only. The results for SIMULATION A are in Table 1. The sparsity of the observations in this scenario prevented all methods from producing any estimates, except for PACE and ME-SPLINE, which worked for all simulation runs. The two PACE variants have lower values of RMISE for both trajectories and derivatives compared to ME-SPLINE , where PACE-QUO is slightly better than PACE-EIG for the first derivative, while PACE-EIG is somewhat better than PACE-QUO for the second derivative.

For SIMULATION B, with results in Table 2, only PACE and ME-SPLINE produced estimates for all 500 simulations, while KERNEL, CUBIC SPLINE and P-SPLINE produced estimates for 496, 311 and 477 runs, respectively. As the number of measurements per trajectory increases, the performance of both PACE and ME-SPLINE improves compared

to the sparser case reported in Table 1. The leading method for all outcomes is PACE-EIG, with PACE-QUO a close second for first derivatives and ME-SPLINE the second best method for second derivatives, and a close third for first derivatives. The worst performer is CUBIC SPLINE, which is clearly plagued by outliers but even in terms of median is not performing that well. KERNEL and P-SPLINE behave overall similarly for first derivatives, while P-SPLINE is clearly better for second derivatives. An advantage of KERNEL is that it it is more likely to produce an estimate than P-SPLINE.

For the case of a dense grid of measurements (SIMULATION C), the results are in Table 3. There is no problem for any of the methods to compute estimates, and also outliers do not play a noticeable role. In all metrics considered, PACE, particularly the version PACE-QUO, with PACE-EIGclosely behind, shows better performance than the other methods. For fitting trajectories, P-SPLINE ranks next, followed by CUBIC SPLINE and KERNEL, while ME-SPLINE performs relatively poorly. For first derivatives, KERNEL ranks third, followed by P-SPLINE, CUBIC SPLINE and ME-SPLINE. For second derivatives, KERNEL is tied with PACE-EIG, followed by ME-SPLINE and P-SPLINE which are tied. CUBIC SPLINE performs poorly. Summarizing the results from all simulations, the PACE variants consistently outperform the other methods. For the sparse case, ME-SPLINE is the next best method, while for the dense case this is KERNEL.

Motivated by the fact that patterns of bid arrivals lead to a natural segmentation of auctions into an early, middle and late phase, with few bids in the middle, we investigate the behavior of the various methods on corresponding subsegments $\mathcal{T}_1 = [0, 1/4], \mathcal{T}_2 = [1/4, 3/4]$ and $\mathcal{T}_3 = [3/4, 1]$ for the scenario of SIMULATION B which comes closest to the auction setting (Table 4). As expected, the overall sizes of the errors differ across segments, with large errors for trajectory and derivative estimation in the first period, small errors in the second and small trajectory but large derivative errors in the third period. PACE-EIG emerges as the best method across all three segments.

PACE-QUO is typically in second or third position and for second derivatives in fourth position, while ME-SPLINE is overall the third best method. KERNEL and P-SPLINE behave similarly, where KERNEL generally outperforms P-SPLINE for trajectory estimation while P-SPLINE is better for second derivatives. CUBIC SPLINE consistently is the worst performer.

The relative worse performance of PACE-QUO in comparison to PACE-EIG in the setting of SIMULATION B, in reversal of some of the trends observed for the very sparse and very dense settings in SIMULATIONS A and SIMULATION C, is noteworthy. This is likely due to the fact that in SIMULATION B the specific feature of the auction data of relatively dense recordings in the beginning and at the end is fully reflected, and drives the automatic bandwidth choice to a larger extent than in the ultra-sparse setting of SIMULATION A and the equidistant setting of SIMULATION C. Since PACE-EIG involves estimation of the partial derivative of the covariance surface, it is more sensitive to the smoothing parameter choice in the boundary regions of the data domain, and the automatic bandwidth choice therefore is associated with improved results when there is a relative abundance of data near the boundaries.

More generally, in the analysis of auction data or similar functional data it may be of interest to emphasize some areas of the domain, for example by giving extra weight to the subdomain near the end of an auction. This may improve the performance of bandwidth choices and the resulting derivative estimators for such targeted subareas. Subdomain targeting can be easily achieved by choosing a non-uniform measure on the domain (defined by a smooth positive weight function), thereby moving from the space L^2 to a space $L^2(d\mu)$ with a well-chosen measure $d\mu$. In this more general approach, inner products and eigenfunctions will depend on the selected measure $d\mu$; the theoretical results can be easily adapted and the convergence results do not change.

5. APPLICATION TO E-BAY ONLINE AUCTION DATA

eBay uses a proxy bidding system which means that the real-time current price or "live bid" (Jank and Shmueli 2005b) is often lower than the bid amount entered by a bidder, the "willingness to pay" (WTP) price. We base our analysis on these live bids, which represent the price process, the dynamics of which is of interest. The live bids are converted from WTP values which are publicly available (for details see Liu and Müller 2008).

The bid histories entering our analysis reflect 156 eBay online auctions of Palm M515 Personal Digital Assistants (PDA) that were transacted between March and May, 2003. All of these auctions are 7-day second-price auctions transacted in US dollars. Auction time was transformed from calendar time to the interval $[0, 168]$ in units of hours, with 0 corresponding to the beginning of an auction. Bid prices were log-transformed. We view the live bids Y_{ij}, observed at bid arrival times T_{ij}, where $i = 1, \ldots, 156$ is the auction index and $j = 1, \ldots, N_i$ refers to the bids within one auction, as being generated from an underlying smooth price process, where individual bids are subject to random aberrations from the continuous price trajectory. We therefore assume that model (4) applies. The number of observed bids N_i ranges from 9 to 52 per auction. Our objective is to explore the auction price dynamics through the estimation of derivatives. To apply the PACE method, one needs to check that the pooled data are dense so that mean and covariance functions can be reasonably well estimated. This assumption can be checked through a design plot in which all pairs of bid arrival times that are observed within the same auction are displayed. This plot (Figure 1) confirms that the assumption is met and also demonstrates the striking non-uniformity of the bid arrivals.

For estimation of mean function, covariance function and their derivatives, we chose the same univariate and bivariate kernel functions as in the simulations. Bandwidths were chosen visually by slightly adjusting the GCV bandwidths (with values $h_{\mu,0} = 6.3$, $h_{\mu,1} = 13.6$, $h_{\mu,2} = 15.2$, $h_{G,0} = h_{G,1} = h_{G,2} = 17.8$). Estimated mean function and its

first two derivatives are shown in Figure 2, along with the pooled bids. The increase of average log-bid prices is clearly nonlinear in time. The rate of increase decreases for the first two days and then remains almost constant between day 3 to somewhere between days 5 and 6, after which it climbs slowly. Maximum acceleration is reached near the end of an auction. The end-auction gain in momentum is likely related to the "bid sniping" phenomenon. The estimated covariance surface and the estimated first partial derivative of the covariance surface are displayed in Figure 3. The fitted correlation surface

$$\hat{C}(t,s) = \sum_{\hat{\lambda}_k>0} \hat{\lambda}_k \hat{\phi}_k(t)\hat{\phi}_k(s) / [\sum_{\hat{\lambda}_k>0} \hat{\lambda}_k \hat{\phi}_k^2(t) \sum_{\hat{\lambda}_k>0} \hat{\lambda}_k \hat{\phi}_k^2(s)]^{1/2} \tag{15}$$

shown in Figure 4 demonstrates that log bid prices are very highly correlated for all times between approximately 55 and 128 hours, across the mid-range of the auction, while correlation of log prices in the early and final phases of the auction falls off rapidly with increasing time difference. This suggests increased price volatility during these phases.

The number of optimal functional principal component scores K is selected as 2, which is the result from fraction of variance explained (FVE) method with threshold 0.85 and it accounts for 95.48% of the total variation. BIC_1 selects $K = 3$, which accounts for 98.93% of the total variation. We opt for $K = 2$ as the more parsimonious model. The first two eigenfunctions account for 72.79% and 22.69% of total variation. The estimated eigenfunctions and their first two derivatives are shown in Figure 5. From the top panel, it appears that the first eigenfunction (solid line) increases over time in a similar fashion as the mean function, while the second eigenfunction (dashed line) behaves differently. The derivatives of the eigenfunctions are shown in the lower panels. Four randomly selected fitted log price functions, obtained by eq. (10), are visualized in Figure 6. The gaps between bid arrivals are seen to be often very large indeed. One finds that detailed patterns of price dynamics can be much more clearly discerned from the derivatives than from the price trajectories themselves.

6. DYNAMICS FOR GAUSSIAN RANDOM PROCESSES

6.1 Transfer functions

Consider mean-centered process $X(t) - \mu(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t)$ and derivative process $X^{(\nu)}(t) - \mu^{(\nu)}(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu)}(t)$. If the FPC scores are jointly Gaussian, then $X^{(\nu)}(t) - \mu^{(\nu)}(t)$ and $X(t) - \mu(t)$ are jointly Gaussian for each $t \in \mathcal{T}$,

$$\begin{pmatrix} \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(\nu)}(t) \\ \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t) \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{k=1}^{\infty} \lambda_k \left(\phi_k^{(\nu)}(t)\right)^2 & \sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu)}(t)\phi_k(t) \\ \sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu)}(t)\phi_k(t) & \sum_{k=1}^{\infty} \lambda_k \phi_k^2(t) \end{pmatrix} \right),$$

implying

$$E(X^{(\nu)}(t) - \mu^{(\nu)}(t) | X(t) - \mu(t)) = \beta_\nu(t)(X(t) - \mu(t)), \quad t \in \mathcal{T}, \tag{16}$$

for a function $\beta_\nu(t) = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu)}(t)\phi_k(t) / \sum_{k=1}^{\infty} \lambda_k \phi_k^2(t)$. The *transfer functions* β_ν describe the dynamics of the stochastic system. We note that beyond the Gaussian assumption no further assumptions are needed to establish these empirical differential equations which are therefore widely applicable. Identifying these equations requires estimation of the derivatives of the eigenfunctions for which we use the proposed estimators; alternative estimators could be used as well. Illustration and interpretation for the auction data is provided in subsection 6.2. Alternative dynamic analysis of auction data is described in Wang et al. (2008a).

According to (16) for $\nu = 1$, for each fixed $t \in \mathcal{T}$, we expect the mean-centered derivative process $X' - \mu'$ at time t to increase or decrease by an amount of $\beta_1(t)$ for a unit increase in the mean-centered process $X - \mu$ at t. For the covariance function of the derivative process $\mathrm{cov}(X^{(\nu)}(t), X^{(\nu)}(s)) = G_\nu(t, s)$, (16) implies $G_\nu(t, s) = \beta_\nu(t)\beta_\nu(s)G(t, s)$. Including K components and substituting estimates for λ_k, $\phi_k(t)$ and $\phi_k^{(\nu)}(t)$ as discussed in section 3, we obtain estimates

$$\hat{\beta}_\nu(t) = \frac{\sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k^{(\nu)}(t)\hat{\phi}_k(t)}{\sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k^2(t)}, \qquad \hat{G}_\nu(t, s) = \hat{\beta}_\nu(t)\hat{\beta}_\nu(s)\hat{G}(t, s), \quad t, s \in \mathcal{T}, \tag{17}$$

where $\hat{G}(t,s)$ can be obtained from (7) for $\nu = 0$.

Another way to explore the dynamics is through functional principal component analysis of derivative processes. Under obvious regularity conditions, the autocovariance operator for the ν-th derivative process is defined as $(A_{G,\nu}f)(t) = \int_{\mathcal{T}} G_\nu(t,s)f(s)ds$ for any $f \in L^2$. The operator $A_{G,\nu}$ has eigenvalues $\lambda_{1,\nu} \geq \lambda_{2,\nu} \geq \ldots \geq 0$ and eigenfunctions $\phi_{k,\nu}$, so that $G_\nu(t,s) = \sum_{k=1}^{\infty} \lambda_{k,\nu}\phi_{k,\nu}(t)\phi_{k,\nu}(s)$. The derivative process can be represented as $X^{(\nu)}(t) = \mu^{(\nu)}(t) + \sum_{k=1}^{\infty} \xi_{k,\nu}\phi_{k,\nu}(t)$, where $\xi_{k,\nu}$ are the uncorrelated FPC scores for the ν-th derivative of the random process. This approach provides an alternative implementation for data analysis, by using plug-in estimates for trajectories of the derivative process, $\hat{X}_{L,\nu}(t) = \hat{\mu}^{(\nu)}(t) + \sum_{k=1}^{L} \hat{\xi}_{k,\nu}\hat{\phi}_{k,\nu}(t)$, for suitably chosen L. Here one obtains $\hat{\mu}^{(\nu)}(t)$ as described in section 3, and then performs a spectral decomposition of a suitably discretized version of $\hat{G}_\nu$ obtained as in in eq. (17) which then yields $\hat{\lambda}_{k,\nu}$ and $\hat{\phi}_{k,\nu}$. Derivative estimates (10), as obtained in section 3, coupled with numerical integration, then lead to estimates of the FPC scores $\hat{\xi}_{ik,\nu} = \int_{\mathcal{T}} (\hat{X}_{i,K}^{(\nu)}(t) - \hat{\mu}^{(\nu)}(t))\hat{\phi}_{ik,\nu}(t)\, dt$.

6.2 Dynamics of online auctions

If we assume that the FPC scores for the log-bid price functions are jointly Gaussian, applying (16) for $\nu = 1$ provides a prediction of the derivative, given the price curve, through the empirical differential equation

$$\hat{X}_i^{*\prime}(t) = \hat{\mu}'(t) + \hat{\beta}_1(t)(\hat{X}_{i,K}(t) - \hat{\mu}(t)), \tag{18}$$

where $\hat{\beta}_1$ is the estimated dynamic transfer function given in (17) for $\nu = 1$ (displayed in the top left panel of Figure 7). The interpretation of this transfer function is that those trajectories which are situated high above the overall mean function will experience smaller than average increases near the end of the auction and vice versa. The dynamic system underlying auctions thus has the effect of maintaining an equilibrium, whereby excessive price levels are eliminated by subsequent slowing down of further increases, and

by the same token low price levels at the end of an auction are rendered less likely by higher than average increases being associated with auctions that enter the final stretch at lower than expected price levels. The dynamics thus reflect a correction mechanism for price imbalances that can be described as a *dynamic regression to the mean* effect.

While this effect is in force throughout the auction, the shape of the transfer function β_1 indicates that it is most pronounced near the end of an auction. The other panels of Figure 7 display the implementation of eq. (18) for three auctions. Here the fitted price curves and the estimated transfer function are used to obtain estimates of expected derivatives of these price curves according to (18), and these are compared with the derivative estimates (10). The targets of these two functions differ as the derivative obtained from the transfer function is an average prediction while the derivative estimate (10) targets the actual derivative of the price curve. The agreement of the predicted with the observed derivatives is surprisingly good, given that the predictions from the transfer function reflect no more than an average relationship. This provides additional validation of the usefulness of the dynamic equation (16).

7. CONCLUSIONS AND DISCUSSION

In this paper, we have shown how one can efficiently obtain derivatives from sparsely and irregularly observed repeated measurements that are possibly corrupted by measurement errors. Beyond exploring the dynamics of auctions, this has applications for the study of the dynamics of many longitudinal data, especially in the life sciences. As demonstrated by our simulation results, the method also works well for the case of densely sampled and regularly spaced functional data.

Besides providing a method for obtaining derivatives for sparse data when one has a sample of curves, we demonstrate how one can develop a general and model-free dynamic equation through (16) in the case of underlying Gaussian processes. This empirical differ-

ential equation lends itself to interpretation: For the auction data, we find a regression-to-the mean dynamics, where larger current trajectory values $X(t)$ are associated with smaller positive derivatives $X'(t)$. There are many alternative dynamics of interest. In one simple scenario, derivatives $X'(t)$ might tend to increase with increasing levels $X(t)$, which would then correspond to explosive behavior of the trajectories, reinforcing initial differences. This empirical differential equation has the potential to provide valuable insights into the time-dynamics of random trajectories.

APPENDIX

A.1 Thresholding algorithm for estimating predicted FPC scores

Step 1: Compute the residual sum of squares (RSS)

$$\hat{\sigma}^2_{new,1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}(Y_{ij} - \hat{Y}_{ij})^2,$$

where $\hat{Y}_{ij} = \hat{\mu}(T_{ij}) + \sum_{k=1}^{K}\hat{\xi}_{ik}\hat{\phi}_k(T_{ij})$. Here, the estimation of ξ_{ik} is based on (8) and $\hat{\sigma}^2$ in $(\hat{\mathbf{\Sigma}}_{\mathbf{Y}_i})_{j,l}$ is based on (2) in Yao et al. (2005), and K is chosen as described in section 2.3.

Step 2: Repeat Step 1 and compute an updated RSS

$$\hat{\sigma}^2_{new,2} = \frac{1}{n}\sum_{i=1^n}\frac{1}{N_i}\sum_{j=1}^{N_i}(Y_{ij} - \hat{Y}_{ij,new})^2,$$

where $\hat{Y}_{ij,new}$ is derived from (9) by using $\hat{\sigma}_{new,1}$ in $(\hat{\mathbf{\Sigma}}_{Y_i})_{j,l}$ instead of $\hat{\sigma}^2$.

Step 3: If $\hat{\sigma}^2_{new,2} < \rho$, reset $\hat{\sigma}^2_{new,2} = \rho$, where ρ is chosen by a cross-validation approach, leaving out one randomly chosen measurement per trajectory and minimizing RSS as a function of η, where $\rho = \rho(\eta) = \eta\mathcal{R}$, for

$$\mathcal{R} = \left[E\frac{1}{|\mathcal{T}|}\int_{\mathcal{T}}\{\mu(t) + \sum_{k=1}^{K}\xi_k\phi_k(t)\}^2dt\right]^{\frac{1}{2}} = \frac{1}{|\mathcal{T}|^{\frac{1}{2}}}\left[\{\int_{\mathcal{T}}\mu^2(t)dt + \sum_{k=1}^{K}\lambda_k\}\right]^{\frac{1}{2}},$$

substituting the unknown quantities by their obvious estimates. Then for $J = \{i : N_i \geq 2\}$ obtain

$$\eta_{opt} = \underset{a_1 \le \eta \le a_2}{\operatorname{argmin}} CV(\eta) = \sum_{i \in J} \left(Y_{ij} - \hat{\mu}(T_{ij}) - \sum_{k=1}^{K} \hat{\xi}_{ik}^{(-j)} \hat{\phi}_k(T_{ij}) \right)^2 ,$$

where for subject i we randomly select one observation (T_{ij}, Y_{ij}) which is left out from the data sampled for this subject, obtaining the FPC scores $\hat{\xi}_{ik}^{(-j)}$ by using (8) with $\hat{\sigma}^2_{new,2} = \max(\hat{\rho}(\eta), \hat{\sigma}^2_{new,2})$ for $(\hat{\mathbf{\Sigma}}_{Y_i})_{j,l}$. Choices $a_1 = 0.01$, $a_2 = 0.225$ worked empirically well.

Step 4: Estimate ξ_{ik} using (8) with $\hat{\sigma}^2_{new,2} = \max(\hat{\rho}(\eta_{opt}), \hat{\sigma}^2_{new,2})$ for $(\hat{\mathbf{\Sigma}}_{Y_i})_{j,l}$.

A.2 Assumptions

Writing $\mathbf{T}_i = (T_{i1}, \ldots, T_{iN_i})^T$ and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$, the triples $\{\mathbf{T}_i, \mathbf{Y}_i, N_i\}$ are iid. Regularity conditions for marginal and joint densities, $f(t)$, $g(t, y)$, $f_2(t_1, t_2)$ and $g_2(t_1, t_2, y_1, y_2)$ are as follows:

(B1) $\frac{d^5}{dt^5} f(t)$ exists and is continuous on $\mathcal{T}$ with $f(t) > 0$, $\frac{\partial^5}{\partial t^5} g(t, y)$ exists and is uniformly continuous on $\mathcal{T} \times \mathbb{R}$, $\frac{\partial^5}{\partial t_1^5} f_2(t_1, t_2)$ and $\frac{\partial^5}{\partial t_2^5} f_2(t_1, t_2)$ exist and are uniformly continuous on $\mathcal{T}^2$, $f_2(t_1, t_2) > 0$ and $\frac{\partial^2}{\partial t_2^2} f_2(t_1, t_2) \neq 0$, $\frac{\partial^5}{\partial t_1^{l_1} \partial t_2^{l_2}} g_2(t_1, t_2, y_1, y_2)$ exists and is uniformly continuous on $\mathcal{T}^2 \times \mathbb{R}^2$, for all integers l_1 and l_2 such that $l_1 + l_2 = 5$, $0 \le l_1, l_2 \le 5$.

Among other things, this condition provides for dense designs in the pooled samples for estimating $\mu^{(\nu)}$ and $G^{(\nu,0)}(t, s)$, where the properties of the marginal density f of the T_{ij} and their joint density f_2 are relevant for the derivation of uniform convergence.

Let $\kappa_1(\cdot) : \mathbb{R} \to \mathbb{R}$ and $\kappa_2(\cdot, \cdot) : \mathbb{R}^2 \to \mathbb{R}$ be symmetric univariate and bivariate kernel density functions, as used in the smoothing steps for $\mu'(t)$ in (5) and $G^{(1,0)}(t, s)$ in (7), satisfying

(B2) Kernel κ_1 has $\|\kappa_1\|^2 = \int \kappa_1^2(u) du < \infty$, $\int \kappa_1(u) du = 1$, $\int \kappa_1(u) u^r du = 0$, for $r = 1, 3, 4$, $\int \kappa_1(u) u^2 du = \gamma_{\kappa_1} \neq 0$, and an absolutely integrable Fourier-transform.

An example for a kernel satisfying (B2) is $\kappa_1^*(u) = \varphi(u)(\frac{5}{4} - \frac{1}{4}u^2)$, where $\varphi(u) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})$, with $\gamma_{\kappa_1} = \int \kappa_1^*(u)u^2du = \frac{1}{2}$.

(B3) Kernel κ_2 has $\|\kappa_2\|^2 = \iint \kappa_2^2(u,v)dudv < \infty$, $\iint \kappa_2(u,v)dudv = 1$, $\iint \kappa_2(u,v)u^{l_1}v^{l_2} \times dudv = 0$ for $0 < l_1 + l_2 \leq 2$, $l_1 \neq 2$ and $l_2 \neq 0$, $\iint \kappa_2(u,v)u^2dudv \neq 0$, and an absolutely integrable Fourier transform.

An example of a kernel satisfying (B3) is $\kappa_2(u,v) = \kappa_1^*(u)\bar{\kappa}_1(v)$, where $\bar{\kappa}_1(v) = \varphi(v)(\frac{3}{2} - \frac{1}{2}v^2)$. Further assumptions needed for Theorem 2 and 3, respectively are

(B4) The autocovariance operator A_G generated by the continuously differentiable covariance function $G(t,s)$ is positive definite, such that $\sup_{t\in\mathcal{T}} \int G^{(1,0)}(t,s)\phi_k(s)ds < \infty$, $\int |G^{(1,0)}(t,s)ds| < \infty$ and with eigenfunctions satisfying $\sup_{s\in\mathcal{T}} |\phi_k(s)| = O_p(1)$.

(B5) The FPC scores ξ_{ik} and measurement errors ε_{ij} are jointly Gaussian. The number, location, and values of measurements for a given subject or cluster remain unaltered as $n \to \infty$ and $\sup_{t\in\mathcal{T}} \mathrm{Var}(\widetilde{X}_i'(t)) = \sum_{k=1}^{\infty} \lambda_k \phi_k'^2(t) < \infty$.

A.3 Proofs

Proof of theorem 1 We provide only the arguments for (12), noting that (11) follows analogously. We note that $\hat{G}^{(1,0)}(t,s) = \hat{\alpha}_{11}(t,s)$ can be obtained explicitly as

$$\begin{aligned}
\hat{\alpha}_{11}(t,s) &= (z_{10}x_{20}^2x_{02} - z_{01}x_{11}x_{20}^2 + z_{20}x_{01}x_{11}x_{20} + z_{01}x_{01}x_{20}x_{30} - 2z_{10}x_{01}x_{20}x_{21} \\
&\quad -z_{00}x_{20}x_{02}x_{30} + z_{01}x_{10}x_{20}x_{21} + z_{00}x_{11}x_{20}x_{21} - z_{20}x_{10}x_{20}x_{02} - z_{20}x_{01}^2x_{30} \\
&\quad +z_{10}x_{01}^2x_{40} - z_{01}x_{10}x_{01}x_{40} - z_{00}x_{01}x_{11}x_{40} + z_{20}x_{10}x_{01}x_{21} + z_{00}x_{01}x_{21}x_{30} \\
&\quad -z_{00}x_{10}x_{21}^2 - z_{01}x_{00}x_{21}x_{30} + z_{01}x_{00}x_{11}x_{40} + z_{00}x_{10}x_{02}x_{40} - z_{10}x_{00}x_{02}x_{40} \\
&\quad -z_{20}x_{00}x_{11}x_{21} + z_{20}x_{00}x_{02}x_{30} + z_{10}x_{00}x_{21}^2)/(x_{20}^3x_{02} + x_{00}x_{11}^2x_{40} + x_{01}^2x_{20}x_{40} \\
&\quad +x_{00}x_{02}x_{30}^2 + 2x_{10}x_{01}x_{21}x_{30} + x_{10}^2x_{02}x_{40} - x_{00}x_{20}x_{02}x_{40} + 2x_{10}x_{11}x_{20}x_{21} \\
&\quad -x_{11}^2x_{20}^2 + x_{00}x_{20}x_{21}^2 - x_{10}^2x_{21}^2 - x_{01}^2x_{30}^2 - 2x_{01}x_{20}^2x_{21} - 2x_{10}x_{20}x_{02}x_{30} \\
&\quad -2x_{00}x_{11}x_{21}x_{30} + 2x_{01}x_{11}x_{20}x_{30} - 2x_{10}x_{01}x_{11}x_{40}), \qquad (19)
\end{aligned}$$

where

$$\begin{aligned}
x_{pq} &= \sum_{i=1}^{n} \frac{1}{EN(EN-1)} \sum_{1\le j\ne l\le N_i} (T_{ij}-t)^p (T_{il}-s)^q W_{ijl}, \\
z_{pq} &= \sum_{i=1}^{n} \frac{1}{EN(EN-1)} \sum_{1\le j\ne l\le N_i} G_i(T_{ij},T_{il})(T_{ij}-t)^p (T_{il}-s)^q W_{ijl}, \\
W_{ijl} &= \frac{1}{nh_{G,1}^2}\kappa_2\left(\frac{t-T_{ij}}{h_{G,1}}, \frac{s-T_{il}}{h_{G,1}}\right).
\end{aligned}$$

Define $\widetilde{G}_i(T_{ij},T_{il}) = (Y_{ij}-\mu(T_{ij}))(Y_{il}-\mu(T_{il}))$ and $\tilde{\alpha}_{11} = \tilde{\alpha}_{11}(t,s)$ by replacing all z_{pq} in (19) by $\tilde{z}_{pq} = \sum_{i=1}^{n} \frac{1}{EN(EN-1)} \sum_{1\le j\ne l\le N_i} \widetilde{G}_i(T_{ij},T_{il})(T_{ij}-t)^p (T_{il}-s)^q W_{ijl}$. Then $G_i(T_{ij},T_{il}) = \widetilde{G}_i(T_{ij},T_{il}) + O_p(\frac{1}{\sqrt{n}h_{\mu,0}})$ for $n\to\infty$, where the O_p-term is uniform, implying $z_{pq} = \tilde{z}_{pq} + O_p(\frac{1}{\sqrt{n}h_{\mu,0}})$ and $\hat{\alpha}_{11} = \tilde{\alpha}_{11} + O_p(\frac{1}{\sqrt{n}h_{\mu,0}})$. It is sufficient to study uniform convergence of $\tilde{\alpha}_{11}$.

For the quantities Θ and ϱ, defined as

$$\Theta = \Theta(t,s) = \frac{1}{nh_{G,\nu}^{|\widetilde{\nu}|+2}} \sum_{i=1}^{n} \frac{1}{EN(EN-1)} \sum_{1\le j\ne l\le N_i} \theta(T_{ij},T_{il},Y_{ij},Y_{il})\tilde{\kappa}\left(\frac{t-T_{ij}}{h_{G,\nu}}, \frac{s-T_{il}}{h_{G,\nu}}\right),$$

	$\tilde{\kappa}_1$, $\widetilde{\nu}=(0,0)$, $l=2$		$\tilde{\kappa}_3$, $\widetilde{\nu}=(0,2)$, $l=4$
	Θ_{11}	Θ_{12}	Θ_{31}
$\theta(u,v,y_1,y_2)$	$(y_1-\mu(u))(y_2-\mu(v))$	1	1
ϱ	$G(t,s)f_2(t,s)$	$f_2(t,s)$	$f_2^{(0,2)}(t,s)$
$\sup_{t,s\in\mathcal{T}}\lvert\Theta-\varrho\rvert$	$O_p(\frac{1}{\sqrt{n}h_{G,1}^2}+h_{G,1}^2)$		$O_p(\frac{1}{\sqrt{n}h_{G,1}^4}+h_{G,1}^2)$
as $n\to\infty$	$h_{G,1}\to 0$, $nh_{G,1}^4\to\infty$		$h_{G,1}\to 0$, $nh_{G,1}^8\to\infty$

Scheme 1: Uniform convergence results for three weighted averages related to terms $\tilde{z}_{00}$, x_{00} and z_{02} of $\tilde{\alpha}_{11}(t,s)$.

	$\tilde{\kappa}_2$, $\widetilde{\nu}=(1,0)$, $l=5$					
	Θ_{21}	Θ_{22}	Θ_{23}	Θ_{24}	Θ_{25}	Θ_{26}
$\theta(u,v,y_1,y_2)$	$(y_1-\mu(u))(y_2-\mu(v))$	$(y_1-\mu(u))(y_2-\mu(v))(u-t)$	1	$u-t$	$(u-t)^2$	$(u-t)^3$
ϱ	$G^{(1,0)}(t,s)f_2(t,s)$ $+G(t,s)f_2^{(1,0)}(t,s)$	$G(t,s)f_2(t,s)$	$f_2^{(1,0)}(t,s)$	$f_2(t,s)$	0	0
$\sup_{t,s\in\mathcal{T}}\lvert\Theta-\varrho\rvert$	$O_p(\frac{1}{\sqrt{n}h_{G,1}^3}+h_{G,1}^4)$					
as $n\to\infty$	$h_{G,1}\to 0$, $nh_{G,1}^6\to\infty$					

Scheme 2: Uniform convergence results for six weighted averages related to terms $\tilde{z}_{10}, \tilde{z}_{20}, x_{10}, x_{20}, x_{30}$ and x_{40} of $\tilde{\alpha}_{11}(t,s)$.

$$\varrho \;=\; \varrho(t,s)=\sum_{l_1+l_2=|\widetilde{\nu}|}\frac{d^{|\widetilde{\nu}|}}{dt^{l_1}ds^{l_2}}\iint\theta(t,s,y_1,y_2)g_2(t,s,y_1,y_2)dy_1dy_2,$$

where $\theta:\mathbb{R}^4\to\mathbb{R}$ is a real function, by Lemma A.2 in Yao et al. (2005), under regularity conditions for θ, (A1) and (B1), for a bivariate kernel $\tilde{\kappa}$ with absolutely integrable Fourier transform of order $(\widetilde{\nu},l)$, $\widetilde{\nu}=(\nu_1,\nu_2)$, it holds that

$$\sup_{t,s\in\mathcal{T}}|\Theta-\varrho|=O_p(\frac{1}{\sqrt{n}h_{G,\nu}^{\widetilde{\nu}+2}}+h_{G,\nu}^{l-|\widetilde{\nu}|}),\quad \text{provided } h_{G,\nu}\to 0, nh_{G,\nu}^{2|\widetilde{\nu}|+4}\to\infty,$$

where $|\widetilde{\nu}|=\nu_1+\nu_2$.

Setting $h_{G,\nu}=h_{G,1}$, and $\tilde{\kappa}_1(t,s)=\kappa_2(t,s)$ satisfying properties (B3) and the θ functions as indicated for terms Θ_{11} and Θ_{12} in Scheme 1, we obtain $\tilde{z}_{00}=\Theta_{11}$ and $x_{00}=\Theta_{12}$. As for $\tilde{z}_{10}, \tilde{z}_{20}, x_{10}, x_{20}, x_{30}$ and x_{40}, setting $\tilde{\kappa}_2(t,s)=-\frac{t\kappa_2(t,s)}{\gamma_1}$, $\gamma_1=\iint\kappa_2(u,v)u^2dudv$,

these six terms can be rewritten as $\Theta_{2p}h_{G,1}^2\gamma_1$, for $p = 1, \ldots, 6$, by choosing the θ functions as indicated for terms Θ_{21} to Θ_{26} in Scheme 2. For x_{02}, we set $\tilde{\kappa}_3(t,s) = \frac{2s^2\kappa_2(t,s)}{\gamma_2}$, $\gamma_2 = \iint \kappa_2(u,v)v^4 dudv$ and the θ function as indicated for term Θ_{31} in Scheme 1, leading to $z_{02} = \frac{1}{2}\Theta_{31}h_{G,1}^4\gamma_2$.

	$\tilde{\kappa}_4$, $\widetilde{\nu} = (0,3)$, $l = 5$			
	Θ_{41}	Θ_{42}	Θ_{43}	Θ_{44}
$\theta(u,v,y_1,y_2)$	$(y_1 - \mu(u))(y_2 - \mu(v))$	1	$u - t$	$(u-t)^2$
ϱ	$G^{(0,3)}(t,s)f_2(t,s) + 2G^{(0,2)}(t,s)f_2^{(0,1)}(t,s)$ $+2G^{(0,1)}(t,s)f_2^{(0,2)}(t,s) + G(t,s)f_2^{(0,3)}(t,s)$	$f_2^{(0,3)}(t,s)$	0	0
$\sup_{t,s\in\mathcal{T}} \lvert\Theta - \varrho\rvert$	$O_p(\frac{1}{\sqrt{n}h_{G,1}^5} + h_{G,1}^2)$			
as $n \to \infty$	$h_{G,1} \to 0$, $nh_{G,1}^{10} \to \infty$			

Scheme 3: Uniform convergence results for four weighted averages related to $\tilde{z}_{01}$, x_{01}, x_{11} and x_{21} of $\tilde{\alpha}_{11}(t,s)$.

Finally, for $\tilde{z}_{01}, x_{01}, x_{11}$ and x_{21}, setting $\tilde{\kappa}_4(t,s) = -\frac{6s\kappa_2(t,s)}{\gamma_2}$ and θ functions as indicated for terms Θ_{41} to Θ_{44} in Scheme 3, these four terms can be re-expressed as $\frac{1}{6}\Theta_{4p}h_{G,1}^4\gamma_2$ for $p = 1, \ldots, 4$.

The uniform convergence results for each of these thirteen weighted averages are summarized in the last two rows of Schemes 1, 2 and 3. Substituting these representations, one obtains

$$\tilde{\alpha}_{11}(t,s) = \left\{ H_1 + \frac{2H_3}{H_2} \right\} \frac{H_2}{H_2 + 2H_4}, \quad \text{with} \tag{20}$$

$$H_1 = \frac{\Theta_{21}\Theta_{24}^2 - \Theta_{22}\Theta_{23}\Theta_{24}}{\Theta_{24}^3 + \Theta_{23}^2\Theta_{26}}, \qquad H_2 = (\Theta_{24}^3 + \Theta_{23}^2\Theta_{26})\Theta_{31}\gamma_1^3\gamma_2,$$

$$H_3 = \frac{1}{36}(-\Theta_{24}^2\Theta_{41}\Theta_{43} + \Theta_{22}\Theta_{24}\Theta_{42}\Theta_{43} + \Theta_{24}\Theta_{25}\Theta_{41}\Theta_{42} - 2\Theta_{21}\Theta_{24}\Theta_{42}\Theta_{44} + \Theta_{23}\Theta_{24}\Theta_{41}$$
$$\times\Theta_{44} - \Theta_{22}\Theta_{25}\Theta_{42}^2 + \Theta_{21}\Theta_{26}\Theta_{42}^2 - \Theta_{23}\Theta_{26}\Theta_{41}\Theta_{42} + \Theta_{22}\Theta_{23}\Theta_{42}\Theta_{44})h_{G,1}^2\gamma_1^2\gamma_2^2 + \frac{1}{36}$$
$$(\Theta_{11}(\Theta_{24}\Theta_{43}\Theta_{44} - \Theta_{26}\Theta_{42}\Theta_{43} + \Theta_{25}\Theta_{42}\Theta_{44} - \Theta_{23}\Theta_{44}^2) + \Theta_{12}(-\Theta_{25}\Theta_{41}\Theta_{44} + \Theta_{26}\Theta_{41}\Theta_{43}))$$

$$\times\gamma_1\gamma_2^2+\frac{1}{2}\Theta_{31}(\Theta_{11}\Theta_{24}\Theta_{25}+\Theta_{11}\Theta_{23}\Theta_{26}+\Theta_{12}\Theta_{21}\Theta_{26}+\Theta_{12}\Theta_{22}\Theta_{25})\frac{1}{h_{G,1}^2}\gamma_1^2\gamma_2,$$

$$H_4=\frac{1}{36}(\Theta_{24}\Theta_{26}\Theta_{42}^2+2\Theta_{23}\Theta_{25}\Theta_{42}\Theta_{44}+2\Theta_{23}\Theta_{24}\Theta_{43}\Theta_{44}-\Theta_{24}^2\Theta_{43}^2-\Theta_{23}^2\Theta_{44}^2-\Theta_{25}^2\Theta_{42}^2$$
$$-\Theta_{24}^2\Theta_{42}\Theta_{44}+2\Theta_{24}\Theta_{25}\Theta_{42}\Theta_{43}-2\Theta_{23}\Theta_{26}\Theta_{42}\Theta_{43})h_{G,1}^2\gamma_1^2\gamma_2+\frac{1}{36}\Theta_{12}(\Theta_{26}\Theta_{43}^2+\Theta_{24}\Theta_{44}^2$$
$$-2\Theta_{25}\Theta_{43}\Theta_{44})\gamma_1\gamma_2^2-\Theta_{23}\Theta_{24}\Theta_{25}\Theta_{31}\gamma_1^3\gamma_2+\frac{1}{2}\Theta_{12}(\Theta_{25}^2\Theta_{31}-\Theta_{24}\Theta_{26}\Theta_{31})\frac{1}{h_{G,1}^2}\gamma_1^2\gamma_2.$$

By Slutsky's theorem,

$$\begin{aligned}\sup_{t,s\in\mathcal{T}}|H_1-G^{(1,0)}(t,s)| &= O_p(\frac{1}{\sqrt{n}h_{G,1}^3}+h_{G,1}^4),\\ \sup_{t,s\in\mathcal{T}}|H_2-f_2^{(0,2)}(t,s)f_2^3(t,s)| &= O_p(\frac{1}{\sqrt{n}h_{G,1}^5}+h_{G,1}^2),\\ \sup_{t,s\in\mathcal{T}}|H_p| &= O_p(\frac{1}{\sqrt{n}h_{G,1}^5}+h_{G,1}^2),\end{aligned}$$

for $p=3,4$. Since $f_2^{(0,2)}(t,s)f_2^3(t,s)\neq 0$ from (B1),

$$\sup_{t,s\in\mathcal{T}}\left|\frac{2H_3}{H_2}\right|=O_p(\frac{1}{\sqrt{n}h_{G,1}^5}+h_{G,1}^2)=o_p(1),$$

as $nh_{G,1}^{10}\to\infty$, when $n\to\infty$, and the result (12) follows.

Proof of theorem 2 The properties of A_G in (B4) imply that all λ_k are positive. Define $\hat{G}^*(t,s)=\sum_{\{k:\hat{\lambda}_k>0\}}^K\hat{\lambda}_k\hat{\phi}_k(t)\hat{\phi}_k(s)$ and $\hat{G}(t,s)=\sum_{k=1}^K\hat{\lambda}_k\hat{\phi}_k(t)\hat{\phi}_k(s)$. Because $|\hat{\lambda}_k-\lambda_k|=O_p(\frac{1}{\sqrt{n}h_{G,0}^2})$ from (15) in Yao et al. (2005), $\hat{G}^*(t,s)$ and $\hat{G}(t,s)$ are asymptotically equivalent. Observing (B4) and (12), we find

$$\sup_{t\in\mathcal{T}}|(\frac{1}{\hat{\lambda}_k}-\frac{1}{\lambda_k})\int G^{(1,0)}(t,s)\phi_k(s)ds|=O_p(\frac{1}{\sqrt{n}h_{G,0}^2})$$

and

$$\begin{aligned}&\left|\int(\hat{G}^{(1,0)}(t,s)\hat{\phi}_k(s)-G^{(1,0)}(t,s)\phi_k(s))ds\right|\\ \leq\ &\left|\int\hat{G}^{(1,0)}(t,s)(\hat{\phi}_k(s)-\phi_k(s))ds\right|+\left|\int(\hat{G}^{(1,0)}(t,s)-G^{(1,0)}(t,s))\phi_k(s)ds\right|\\ =\ &O_p(\frac{1}{\sqrt{n}h_{G,1}^3}+h_{G,1}^4+\frac{1}{\sqrt{n}h_{G,0}^2}),\end{aligned}$$

which implies (13).

Proof of theorem 3 Analogous to the proof of (21) in Yao et al. (2005).

REFERENCES

Abramson, I. and Müller, H.-G. (1994), "Estimating Direction Fields in Autonomous Equation Models, with an Application to System Identification from Cross-sectional Data," *Biometrika*, 81, 663–672.

Ahnert, K. and Abel, M. (2007), "Numerical Differentiation of Experimental Data: Local Versus Global Methods," *Computer Physics Communications*, 764–774.

Ash, R. B. and Gardner, M. F. (1975), *Topics in Stochastic Processes*, New York: Academic Press.

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

Besse, P. and Ramsay, J. O. (1986), "Principal Components Analysis of Sampled Functions," *Psychometrika*, 51, 285–311.

Chambers, J. M. (1991), *Statistical Models in S*, Duxbury Press.

de Boor, C. (1972), "On Calculating with B-splines," *Journal of Approximation Theory*, 6, 50–62.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*, London: Chapman & Hall.

Gasser, T. and Müller, H.-G. (1984), "Estimating Regression Functions and their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics*, 11, 171–185.

Gasser, T., Müller, H.-G., Köhler, W., Molinari, L., and Prader, A. (1984), "Nonparametric Regression Analysis of Growth Curves (Corr: V12 P1588)," *The Annals of Statistics*, 12, 210–229.

Hall, P. and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society B*, 68, 109–126.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006), "Properties of Principal Component

Methods for Functional and Longitudinal Data Analysis," *The Annals of Statistics*, 34, 1493–1517.

Jank, W. and Shmueli, G. (2005a), "Profiling Price Dynamics in Online Auctions Using Curve Clustering," *SSRN eLibrary*, http://ssrn.com/paper=902893.

— (2005b), "Visualizing Online Auctions," *Journal of Computational and Graphical Statistics*, 14, 299–319.

Ling, L. (2006), "Finding Numerical Derivatives for Unstructured and Noisy Data by Multiscale Kernels," *SIAM Journal on Numerical Analysis*, 1780–1800.

Liu, B. and Müller, H.-G. (2008), "Functional Data Analysis for Sparse Auction Data," in *Statistical Methods in eCommerce Research*, eds. Jank, W. and Shmueli, G., New York: Wiley & Sons, Inc, in Press.

Müller, H.-G., Stadtmüller, U., and Schmitt, T. (1987), "Bandwidth Choice and Confidence Intervals for Derivatives of Noisy Data," *Biometrika*, 74, 743–749.

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007), "Parameter Estimation for Differential Equations: A Generalized Smoothing Approach (with discussion)," *Journal of the Royal Statistical Society B*, 69, 741–796.

Ramsay, J. O. and Silverman, B. W. (2002), *Applied Functional Data Analysis*, Springer Series in Statistics, New York: Springer-Verlag.

— (2005), *Functional Data Analysis*, Springer Series in Statistics, New York: Springer, 2nd ed.

Rao, C. R. (1958), "Some Statistical Methods for Comparison of Growth Curves," *Biometrics*, 14, 1–17.

Reddy, S. K. and Dass, M. (2006), "Modeling On-Line Art Auction Dynamics Using Functional Data Analysis," *Statistical Science*, 21, 179–193.

Reithinger, F., Jank, W., Tutz, G., and Shmueli, G. (2008), "Smoothing Sparse and Unevenly Sampled Curves using Semiparametric Mixed Models: An Application to

Online Auctions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57, 127–148.

Rice, J. A. and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves," *Journal of the Royal Statistical Society B*, 53, 233–243.

Rice, J. A. and Wu, C. O. (2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253–259.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," *The Annals of Statistics*, 45, 151–163.

Shibata, R. (1981), "An Optimal Selection of Regression Variables," *Biometrika*, 68, 45–54.

Silverman, B. W. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," *The Annals of Statistics*, 24, 1–24.

Wang, S., Jank, W., and Shmueli, G. (2008a), "Explaining and Forecasting Online Auction Prices and their Dynamics using Functional Data Analysis," *Journal of Business and Economic Statistics*, in Press.

Wang, S., Jank, W., Shmueli, G., and Smith, P. (2008b), "Modeling Price Dynamics in eBay Auctions Using Principal Differential Analysis," *Journal of the American Statistical Association*, in Press.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590.

Zhou, S. and Wolfe, D. A. (2000), "On Derivative Estimation in Spline Regression," *Statistica Sinica*, 10, 93–108.

Table 1: Simulation results for SIMULATION A, irregular and sparse data with error variance $\sigma^2 = 0.5$. Methods PACE-EIG, PACE-QUO and ME-SPLINE are described in the beginning of Section 4. and returned results for all 500 simulations.

$\sigma^2 = 0.5$	Mean of RMISE			Median of RMISE		
	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$
PACE-EIG	.067	.362	.721	.066	.359	.719
PACE-QUO	.067	.360	.821	.066	.355	.704
ME-SPLINE	.090	.565	.920	.091	.586	.942

Table 2: Simulation results for SIMULATION B, irregular but denser data, where for each method RMISE is computed only from those simulations for which the respective method returned a result. The number of these simulations out of 500 is in the last column.

$\sigma^2 = 0.1$	Mean of RMISE			Median of RMISE			Results
	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$	count
PACE-EIG	.045	.302	.633	.044	.299	.623	500
PACE-QUO	.045	.315	1.06	.044	.310	.919	500
ME-SPLINE	.058	.409	.780	.048	.316	.699	500
KERNEL	.138	1.36	22.9	.133	.891	2.10	496
CUBIC SPLINE	683	1.54×10^5	5.18×10^{14}	1.18	37.7	442	311
P-SPLINE	1.69	1.04	.984	.145	.961	.984	477

Table 3: Simulation results for SIMULATION C, regular and dense data with error variance $\sigma^2 = 0.1$. All methods returned results for all 500 simulations.

$\sigma^2 = 0.1$	Mean of RMISE			Median of RMISE		
	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$
PACE-EIG	1.20×10^{-3}	.033	.311	1.20×10^{-3}	.032	.291
PACE-QUO	1.20×10^{-3}	.023	.276	1.20×10^{-3}	.023	.274
ME-SPLINE	3.07×10^{-2}	.568	1.04	3.07×10^{-2}	.569	1.03
KERNEL	2.00×10^{-3}	.051	.308	2.00×10^{-3}	.051	.308
CUBIC SPLINE	1.90×10^{-3}	.126	5.55	1.90×10^{-3}	.124	5.31
P-SPLINE	1.70×10^{-3}	.993	1.01	1.70×10^{-3}	.993	1.01

Table 4: Simulation results for SIMULATION B, irregular but denser data, calculating the error for a method based only on those simulations for which the respective method returned an estimate. The number of these simulations out of 500 for each method is in the last column of Table 2.

$\sigma^2 = 0.1$	Mean of RMISE [0,1/4]			Mean of RMISE [1/4,3/4]			Mean of RMISE [3/4,1]		
	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$	$\hat{X}(t)$	$\hat{X}'(t)$	$\hat{X}''(t)$
PACE-EIG	.248	.510	.897	.042	.234	.544	.027	.515	.878
PACE-QUO	.248	.538	1.55	.042	.242	.920	.027	.549	1.39
ME-SPLINE	.266	.637	1.33	.060	.343	.611	.027	.599	1.21
KERNEL	1.06	5.08	91.3	.130	.799	9.73	.046	.940	28.3
CUBIC SPLINE	1.37×10^4	7.66×10^5	1.70×10^5	289	237	294	85.3	4.42×10^5	4.07×10^{15}
P-SPLINE	1.23	.746	.923	.327	1.01	.998	5.51	1.61	1.00

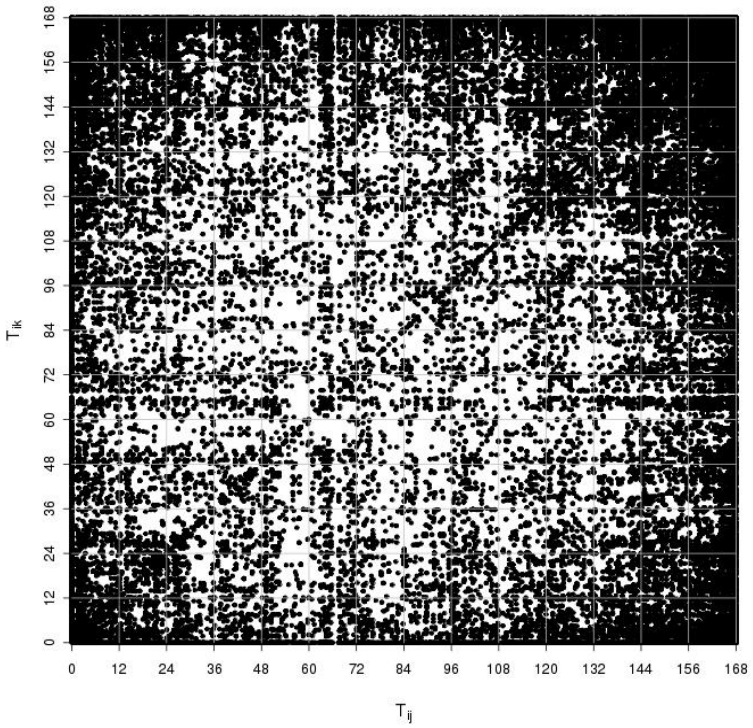


Figure 1: Design plot displaying all pairs of times where bids arrive within the same auction, for 156 online eBay auctions of Palm M515 personal digital assistants.

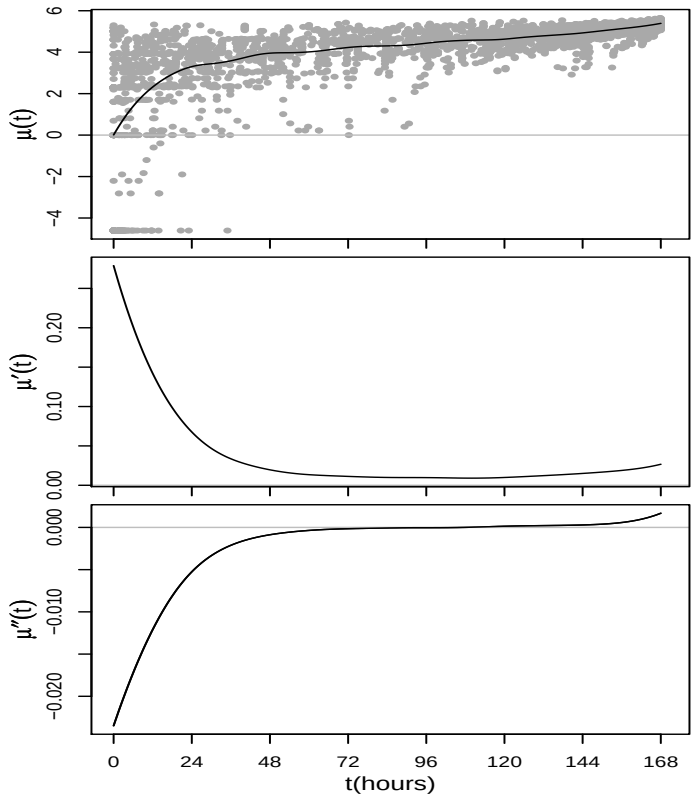


Figure 2: Estimates of the mean function and its first two derivatives.

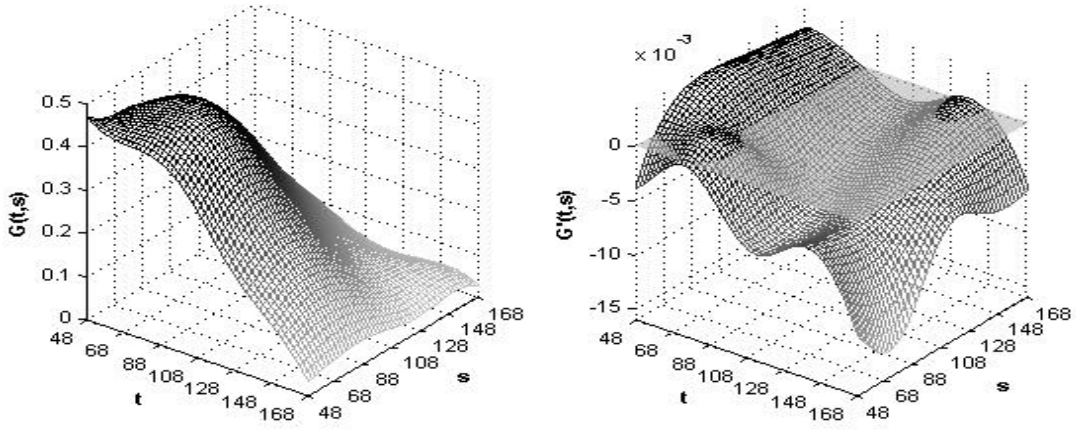


Figure 3: Estimated smooth covariance surface $G(t, s)$ (left panel) and its first partial derivative $G^{(1,0)}(t, s)$ in direction t (right panel).

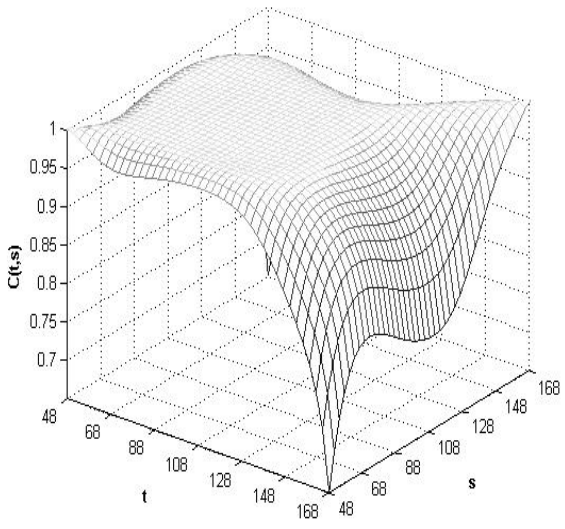


Figure 4: Fitted correlation surface $\hat{C}$ obtained from eq. (15) for the auction log price trajectories.

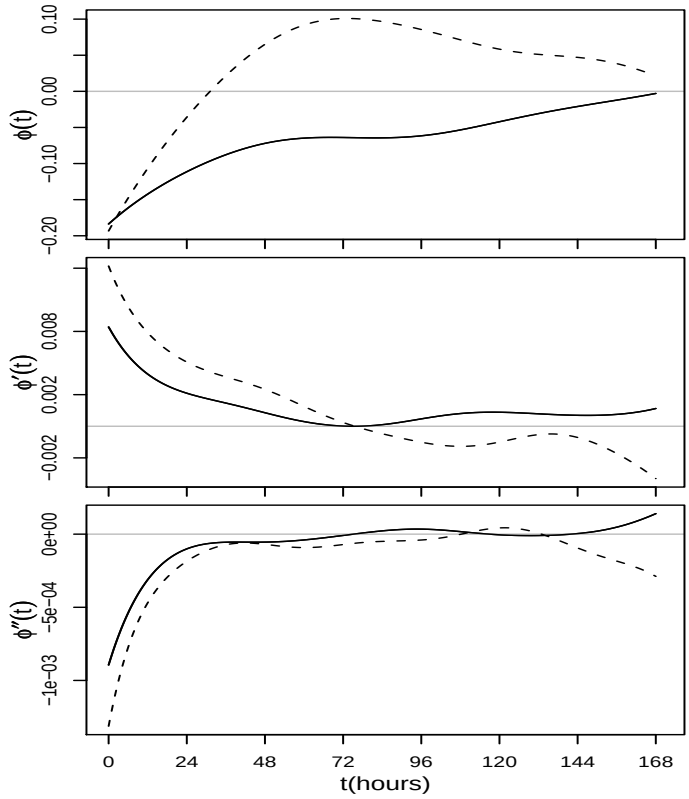


Figure 5: Estimates of first (solid) and second (dashed) eigenfunctions and their derivatives.

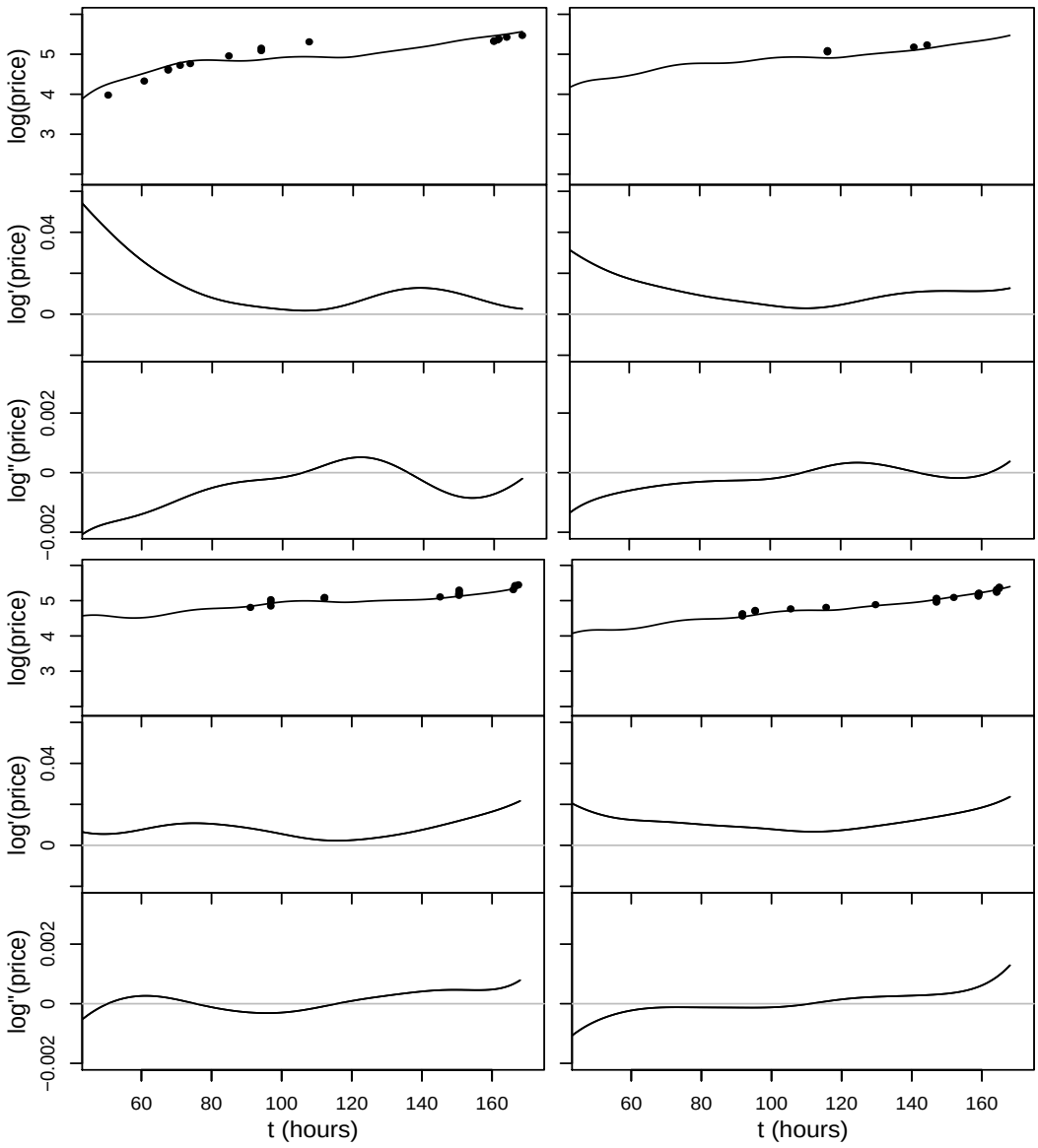


Figure 6: Fitted log price trajectories and their first and second derivatives (from top to bottom in each subpanel) for four randomly selected auctions.

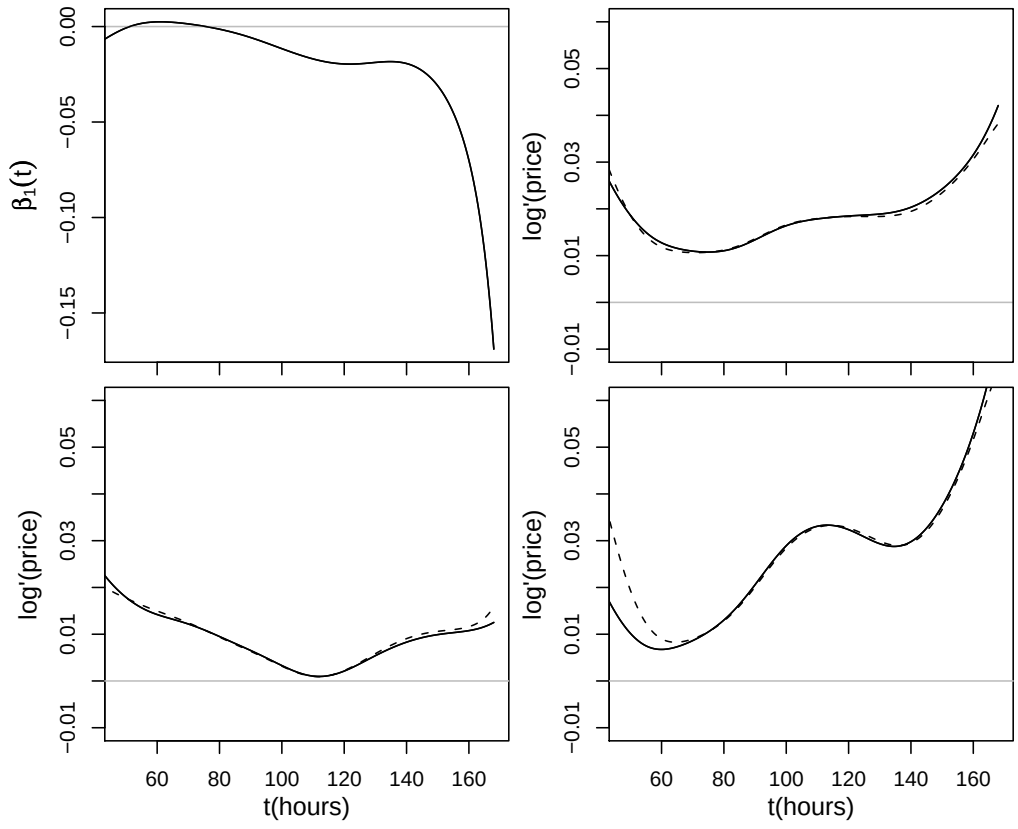


Figure 7: Estimated dynamic transfer function β_1 linking X' with X through $E(X'(t)|X(t)) = \mu'(t) + \beta_1(t)(X(t) - \mu(t))$ (top left panel), and comparisons of predicted derivative trajectories $\hat{X}^{*'} = \hat{\mu}' + \hat{\beta}_1(\hat{X} - \hat{\mu})$ using the estimated dynamic transfer function $\hat{\beta}_1$ (dashed curves) with direct derivative estimates $\hat{X}'$ (solid curves) obtained from PACE-EIG for three randomly selected auctions.