Empirical Dynamics for Longitudinal Data

December 2009

Short title: Empirical Dynamics

Hans-Georg Müller[1]
Department of Statistics
University of California, Davis
Davis, CA 95616 U.S.A.
Email: mueller@wald.ucdavis.edu

Fang Yao[2]
Department of Statistics
University of Toronto
100 St. George Street
Toronto, ON M5S 3G3
Canada
Email: fyao@utstat.toronto.edu

[1]Corresponding author. Research supported in part by NSF grant DMS08-06199.
[2]Research supported in part by a NSERC Discovery grant.

ABSTRACT

We demonstrate that the processes underlying on-line auction price bids and many other longitudinal data can be represented by an empirical first order stochastic ordinary differential equation with time-varying coefficients and a smooth drift process. This equation may be empirically obtained from longitudinal observations for a sample of subjects and does not presuppose specific knowledge of the underlying processes. For the nonparametric estimation of the components of the differential equation, it suffices to have available sparsely observed longitudinal measurements which may be noisy, and are generated by underlying smooth random trajectories for each subject or experimental unit in the sample. The drift process that drives the equation determines how closely individual process trajectories follow a deterministic approximation of the differential equation. We provide estimates for trajectories and especially the variance function of the drift process. At each fixed time point, the proposed empirical dynamic model implies a decomposition of the derivative of the process underlying the longitudinal data into a component explained by a linear component determined by a varying coefficient function dynamic equation and an orthogonal complement that corresponds to the drift process. An enhanced perturbation result enables us to obtain improved asymptotic convergence rates for eigenfunction derivative estimation and consistency for the varying coefficient function and the components of the drift process. We illustrate the differential equation with an application to the dynamics of on-line auction data.

Key words and phrases: functional data analysis, longitudinal data, stochastic differential equation, Gaussian process.

AMS Subject Classification: 62G05, 62G20

1 Introduction

Recently, there has been increasing interest in analyzing on-line auction data and to infer the underlying dynamics that drive the bidding process. Each series of price bids for a given auction corresponds to pairs of random bidding times and corresponding bid prices, generated whenever a bidder places a bid (Jank and Shmueli 2005, 2006; Bapna et al. 2008; Reddy and Dass 2006). Related longitudinal data where similar sparsely and irregularly sampled noisy measurements are obtained are abundant in the social and life sciences, for example they arise in longitudinal growth studies. While more traditional approaches of functional data analysis require fully or at least densely observed trajectories (Kirkpatrick and Heckman 1989; Ramsay and Silverman 2005; Gervini and Gasser 2005), more recent extensions cover the case of sparsely observed and noise-contaminated longitudinal data (Yao et al. 2005; Wang et al. 2005).

A common assumption of approaches for longitudinal data grounded in functional data analysis is that such data are generated by an underlying smooth and square integrable stochastic process (Sy et al. 1997; Staniswalis and Lee 1998; Rice 2004; Zhao et al. 2004; Hall et al. 2006). The derivatives of the trajectories of such processes are central for assessing the dynamics of the underlying processes (Ramsay 2000; Mas and Pumo 2007). Although this is difficult for sparsely recorded data, various approaches for estimating derivatives of individual trajectories nonparametrically by pooling data from samples of curves and using these derivatives for quantifying the underlying dynamics have been developed (Gasser et al. 1984; Reithinger et al. 2008; Wang et al. 2008a,b). Related work on nonparametric methods for derivative estimation can be found in Gasser and Müller (1984); Härdle and Gasser (1985) and on the role of derivatives for the functional linear model in Mas and Pumo (2009).

We expand here on some of these approaches and investigate an empirical dynamic equation. This equation is distinguished from previous models that involve differential equations in that it is empirically determined from a sample of trajectories, and does not presuppose knowledge of a specific parametric form of a differential equation which generates the data, except that we choose it to be a first order equation. This stands in contrast to current approaches of modeling dynamic systems, which are "parametric" in the sense that a prespecified differential equation is assumed. A typical example for such an approach has been

developed by Ramsay et al. (2007), where a prior specification of a differential equation is used to guide the modeling of the data, which is done primarily for just one observed trajectory. A problem with parametric approaches is that diagnostic tools to determine whether these equations fit the data either do not exist, or where they do, are not widely used, especially as nonparametric alternatives to derive differential equations have not been available. This applies especially to the case where one has data on many time courses available, providing strong motivation to explore nonparametric approaches to quantify dynamics. Our starting point is a nonparametric approach to derivative estimation by local polynomial fitting of the derivative of the mean function and of partial derivatives of the covariance function of the process by pooling data across all subjects (Liu and Müller 2009).

We show that each trajectory satisfies a first order stochastic differential equation where the random part of the equation resides in an additive smooth drift process, which drives the equation; the size of the variance of this process determines to what extent the time evolution of a specific trajectory is determined by the non-random part of the equation over various time subdomains and therefore is of tantamount interest. We quantify the size of the drift process by its variance as a function of time. Whenever the variance of the drift process Z is small relative to the variance of the process X, a deterministic version of the differential equation is particularly useful as it then explains a large fraction of the variance of the process.

The empirical stochastic differential equation can be easily obtained for various types of longitudinal data. This approach thus provides a novel perspective to assess the dynamics of longitudinal data and permits insights about the underlying forces that shape the processes generating the observations, which would be hard to obtain with other methods. We illustrate these empirical dynamics by constructing the stochastic differential equations that govern online auctions with sporadic bidding patterns.

We now describe a data model for longitudinally collected observations, which reflects that the data consist of sparse, irregular and noise corrupted measurements of an underlying smooth random trajectory for each subject or experimental unit (Yao et al. 2005), the dynamics of which is of interest. Given n realizations X_i of the underlying process X on a domain $\mathcal{T}$ 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

$$Y_{ij} = Y_i(T_{ij}) = X_i(T_{ij}) + \varepsilon_{ij} \quad T_{ij} \in \mathcal{T}, \tag{1}$$

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

The paper is organized as follows. In Section 2, we review expansions in eigenfunctions and functional principal components, which we use directly as the basic tool for dimension reduction – alternative implementations with B-splines or P-splines could also be considered (Shi et al. 1996; Rice and Wu 2001; Yao and Lee 2006). We also introduce the empirical stochastic differential equation and discuss the decomposition of variance it entails. Asymptotic properties of estimates for the components of the differential equation, including variance function of the drift process, coefficient of determination associated with the dynamic system and auxiliary results on improved rates of convergence for eigenfunction derivatives are the theme of Section 3. Background on related perturbation results can be found in Dauxois et al. (1982); Fine (1987); Kato (1995); Mas and Menneteau (2003). Section 4 contains the illustration of the differential equation with auction data, followed by a brief discussion of some salient features of the proposed approach in Section 5. Additional discussion of some preliminary formulas is provided in Appendix A.1, estimation procedures are described in Appendix A.2, assumptions and auxiliary results are in Appendix A.3 and proofs in Appendix A.4.

2 Empirical Dynamics

2.1 Functional Principal Components and Eigenfunction Derivatives

A key methodology for dimension reduction and modeling of the underlying stochastic processes X that generate the longitudinal data, which usually are sparse, irregular and noisy as in (1), is Functional Principal Component Analysis (FPCA). Processes are assumed to be square integrable with mean function $E\{X(t)\} = \mu(t)$ and auto-covariance function $\text{cov}\{X(t), X(s)\} = G(t,s)$, $s,t \in \mathcal{T}$, which is smooth, symmetric and non-negative definite. Using G as kernel in a linear operator leads to the Hilbert-Schmidt operator $(\boldsymbol{G}f)(t) = \int_{\mathcal{T}} G(t,s)f(s)ds$. We denote the ordered eigenvalues (in declining order) of this

operator by $\lambda_1 > \lambda_2 > \ldots \geq 0$ and the corresponding orthonormal eigenfunctions by $\phi_k(t)$. We assume that all eigenvalues are of multiplicity 1 in the sequel.

It is well known that the kernel G has the representation $G(t,s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t)\phi_k(s)$ and the trajectories generated by the process satisfy the Karhunen-Loève representation (Grenander 1950) $X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t)$. Here the $\xi_{ik} = \int_{\mathcal{T}} \{X_i(t) - \mu(t)\}\phi_k(t)dt$, $k = 1, 2, \ldots$, $i = 1, \ldots, n$, are the functional principal components (FPCs) of the random trajectories X_i. The ξ_k are uncorrelated random variables with $\mathrm{E}(\xi_k) = 0$ and $E\xi_k^2 = \lambda_k$, with $\sum_k \lambda_k < \infty$. Upon differentiating both sides, one obtains

$$X_i^{(1)}(t) = \mu^{(1)}(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(1)}(t), \tag{2}$$

where $\mu^{(1)}(t)$ and $\phi_k^{(1)}(t)$ are the derivatives of mean and eigenfunctions.

The eigenfunctions ϕ_k are the solutions of the eigen-equations $\int G(t,s)\phi_k(s)\,ds = \lambda_k\phi_k(t)$, under the constraint of orthonormality. Under suitable regularity conditions, one observes

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

which motivates corresponding eigenfunction derivative estimates. A useful representation is

$$\mathrm{cov}\{X^{(\nu_1)}(t), X^{(\nu_2)}(s)\} = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu_1)}(t)\phi_k^{(\nu_2)}(s), \quad \nu_1, \nu_2 \in \{0,1\}, \quad s,t \in \mathcal{T}, \tag{4}$$

which is an immediate consequence of the basic properties of the functional principal components ξ_k. For more details and discussion, we refer to Appendix A.1.

It is worthwhile to note that the representation (2) does not correspond to the Karhunen-Loève representation of the derivatives, which would be based on orthonormal eigenfunctions of a linear Hilbert-Schmidt operator defined by the covariance kernel $\mathrm{cov}\{X^{(1)}(t), X^{(1)}(s)\}$. A method to obtain this representation might proceed by first estimating $\mathrm{cov}\{X^{(1)}(t), X^{(1)}(s)\}$ using (4) for $\nu_1 = \nu_2 = 1$ and suitable estimates $\phi_k^{(1)}$ for eigenfunction derivatives, then directly decomposing $\mathrm{cov}\{X^{(1)}(t), X^{(1)}(s)\}$ into eigenfunctions and eigenvalues. This leads to $\mathrm{cov}\{X^{(1)}(t), X^{(1)}(s)\} = \sum_{k=1}^{\infty} \lambda_{k,1}\phi_{k,1}(t)\phi_{k,1}(s)$ and the Karhunen-Loève representation $X_i^{(1)}(t) = \mu^{(1)}(t) + \sum_{k=1}^{\infty} \xi_{ik,1}\phi_{k,1}(t)$, with orthonormal eigenfunctions $\phi_{k,1}$ (Liu and Müller 2009).

2.2 Empirical Stochastic Differential Equation

In the following we consider differentiable Gaussian processes, for which the differential equation introduced below automatically applies. In the absence of the Gaussian assumption, one may invoke an alternative least squares-type interpretation. Gaussianity of the processes implies the joint normality of centered processes $\{X(t)-\mu(t), X^{(1)}(t)-\mu^{(1)}(t)\}$ at all points $t \in \mathcal{T}$, so that

$$\begin{pmatrix} X^{(1)}(t)-\mu^{(1)}(t) \\ X(t)-\mu(t) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{\infty} \xi_k \phi_k^{(1)}(t) \\ \sum_{k=1}^{\infty} \xi_k \phi_k(t) \end{pmatrix}$$

$$\sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)^2 & \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t) \\ \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t) & \sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2 \end{pmatrix} \right). \tag{5}$$

This joint normality immediately implies a "population" differential equation of the form $E\{X^{(1)}(t)-\mu^{(1)}(t) \mid X(t)\} = \beta(t)\{X(t)-\mu(t)\}$, as has been observed in Liu and Müller (2009); for additional details see Appendix A.1. However, it is considerably more interesting to find a dynamic equation which applies to the individual trajectories of processes X. This goal necessitates inclusion of a stochastic term which leads to an empirical stochastic differential equation that governs the dynamics of individual trajectories X_i.

Theorem 1. *For a differentiable Gaussian process, it holds that*

$$X^{(1)}(t) - \mu^{(1)}(t) = \beta(t)\{X(t)-\mu(t)\} + Z(t), \ t \in \mathcal{T}, \tag{6}$$

where

$$\beta(t) = \frac{\mathrm{cov}\{X^{(1)}(t), X(t)\}}{\mathrm{var}\{X(t)\}} = \frac{\sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t)}{\sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2} = \frac{1}{2}\frac{d}{dt}\log[\mathrm{var}\{X(t)\}], \ \ t \in \mathcal{T}, \tag{7}$$

and Z is a Gaussian process such that $Z(t)$, $X(t)$ are independent at each $t \in \mathcal{T}$ and where Z is characterized by $E\{Z(t)\} = 0$ and $\mathrm{cov}\{Z(t), Z(s)\} = G_z(t,s)$, with

$$\begin{aligned} G_z(t,s) \ &= \ \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k^{(1)}(s) \ - \ \beta(t)\sum_{k=1}^{\infty} \lambda_k \phi_k(t)\phi_k^{(1)}(s) \\ &\quad - \beta(s)\sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(s) \ + \ \beta(t)\beta(s)\sum_{k=1}^{\infty} \lambda_k \phi_k(t)\phi_k(s). \end{aligned} \tag{8}$$

Equation (6) provides a first order linear differential equation which includes a time-varying linear coefficient function $\beta(t)$ and a random drift process $Z(t)$. The process Z "drives" the equation at each time t. It is square integrable and possesses a smooth covariance function and smooth trajectories. It also provides an alternative characterization of the individual trajectories of the process. The size of its variance function $\mathrm{var}(Z(t))$ determines the importance of the role of the stochastic drift component.

We note that the assumption of differentiability of the process X in Theorem 1 can be relaxed. It is sufficient to require weak differentiability, assuming that $X \in W^{1,2}$, where $H^1 = W^{1,2}$ denotes the Sobolev space of square integrable functions with square integrable weak derivative (Ziemer 1989). Along these lines, eq. (6) may be interpreted as a stochastic Sobolev embedding. Observe also that the drift term Z can be represented as an integrated diffusion process. Upon combining (2) and (6), and observing that functional principal components can be represented as $\xi_k = \sqrt{\lambda_k/\gamma_k}\int_{\mathcal{T}} h_k(u)\,dW(u)$, where h_k is the k-th eigenfunction of the Wiener process W on domain $\mathcal{T} = [0,T]$ and γ_k the associated eigenvalue, such a representation is given by

$$Z(t) = \sum_{k=1}^{\infty} \sqrt{\frac{\lambda_k}{2T^3}}(2k-1)\pi \int_0^T \sin\{\frac{(2k-1)\pi}{2T}\,u\}\,\{\phi_k^{(1)}(t) - \beta(t)\phi(t)\}\,dW(u).$$

Another observation is that the joint normality in (5) can be extended to joint normality for any finite number of derivatives, assuming these are well-defined. Therefore, higher order stochastic differential equations can be derived analogously to eq. (6). However, these higher order analogues are likely to be much less relevant practically, as higher order derivatives of mean and eigenfunctions cannot be well estimated for the case of sparse noisy data or even denser noisy data.

Finally, it is easy to see that the differential equation (6) is equivalent to the following stochastic integral equation,

$$\begin{aligned} X(t) &= X(s) + \{\mu(t) - \mu(s)\} \\ &+ \int_s^t \beta(u)\{X(u) - \mu(u)\}\,du + \int_s^t Z(u)\,du, \quad \text{for any} \quad s,t \in \mathcal{T},\ 0 \le s < t, \quad (9) \end{aligned}$$

in the sense that X is the solution of both equations. For a domain with left endpoint at time 0, setting $s = 0$ in (9) then defines a classical initial value problem. Given a trajectory of the drift process Z and a varying coefficient function β, one may obtain a solution for X

numerically by Euler or Runga-Kutta integration or directly by applying the known solution formula for the initial value problem of an inhomogeneous linear differential equation.

2.3 Interpretations and Decomposition of Variance

We note that equations (6) and (9) are of particular interest on domains $\mathcal{T}$ or subdomains defined by those times t for which the variance function $\text{var}\{Z(t)\}$ is "small". From (7) and (8) one finds

$$\begin{aligned} V(t) &= \text{var}\{Z(t)\} = (\text{var}\{X^{(1)}(t)\}\text{var}\{X(t)\} - [\text{cov}\{X^{(1)}(t)), X(t)\}]^2)/\text{var}\{X(t)\} \\ &= \left(\sum_{k=1}^{\infty} \lambda_k (\phi_k^{(1)}(t))^2 \sum_{k=1}^{\infty} \lambda_k \phi_k^2(t) - \left\{ \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t) \phi_k(t) \right\}^2 \right) / \sum_{k=1}^{\infty} \lambda_k \phi_k^2(t). \end{aligned} \quad (10)$$

On subdomains with small variance function, the solutions of (6) will not deviate too much from the solutions of the approximating equation

$$X^{(1)}(t) - \mu^{(1)}(t) = \beta(t)\{X(t) - \mu(t)\},\ t \in \mathcal{T}. \quad (11)$$

In this situation, the future changes in value of individual trajectories are highly predictable and the interpretations of the dynamic behavior of processes X obtained from the shape of the varying coefficient function $\beta(t)$ apply at the individual level.

If $\beta(t) < 0$, the dynamic behavior can be characterized as "dynamic regression to the mean"; a trajectory which is away from (above or below) the mean function μ at time t is bound to move closer towards the mean function μ as time progresses beyond t. Similarly, if $\beta(t) > 0$, trajectories will exhibit "explosive" behavior, since deviations from the mean (above or below) at time t will be reinforced further as time progresses, so that trajectories are bound to move further and further away from the population mean trajectory. Intermediate cases arise when the function β changes sign, in which case the behavior will switch between explosive and regression to the mean, depending on the time subdomain. Another situation occurs on subdomains where both β and $\text{var}(Z(t))$ are very small, in which case the deviation of the derivative of an individual trajectory from the population mean derivative will also be small which means that trajectory derivatives will closely track the population mean derivative on such subdomains.

The independence of $Z(t)$ and $X(t)$ means that the right hand side of (6) provides an

orthogonal decomposition of $X^{(1)}(t)$ into the two components $\beta(t)X(t)$ and $Z(t)$ such that

$$\text{var}\{X^{(1)}(t)\} = \beta(t)^2\text{var}\{X(t)\} + \text{var}\{Z(t)\}.$$

It is therefore of interest to determine the fraction of the variance of $X^{(1)}(t)$ that is explained by the differential equation itself, i.e., the "coefficient of determination"

$$R^2(t) = \frac{\text{var}\{\beta(t)X(t)\}}{\text{var}\{X^{(1)}(t)\}} = 1 - \frac{\text{var}\{Z(t)\}}{\text{var}\{X^{(1)}(t)\}}, \tag{12}$$

which is seen to be equivalent to the squared correlation between $X(t), X^{(1)}(t)$,

$$R^2(t) = \frac{[\text{cov}\{X(t), X^{(1)}(t)\}]^2}{\text{var}\{X(t)\}\text{var}\{X^{(1)}(t)\}} = \frac{\{\sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t)\}^2}{\sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2 \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)^2}. \tag{13}$$

We are then particularly interested in subdomains of $\mathcal{T}$ where $R^2(t)$ is large, say, exceeds a prespecified threshold of 0.8 or 0.9. On such subdomains the drift process Z is relatively small compared to $X^{(1)}(t)$ so that the approximating deterministic first order linear differential equation (11) can substitute for the stochastic dynamic equation (6). In this case, short-term prediction of $X(t+\Delta)$ may be possible for small Δ, by directly perusing the approximating differential equation (11).

It is instructive to visualize an example of the function $R^2(t)$ for the case of fully specified eigenfunctions and eigenvalues. Assuming that the eigenfunctions correspond to the trigonometric orthonormal system $\{\sqrt{2}\cos(2k\pi t),\, k = 1, 2, \ldots\}$ on $[0, 1]$, we find from (13)

$$R^2(t) = \left[\sum \lambda_k k \cos(2k\pi t)\sin(2k\pi t)\right]^2 / \left[\sum \lambda_k (\cos(2k\pi t))^2 \sum \lambda_k k (\sin(2k\pi t))^2\right], \quad t \in [0, 1].$$

Choosing $\lambda_k = k^{-4}$, $\lambda_k = 2^{-k}$ and numerically approximating these sums, one obtains the functions $R^2(t)$ as depicted in Figure 1. This illustration shows that the behavior of this function often will fluctuate between small and large values and also depends critically on both the eigenvalues and the shape of the eigenfunctions.

3 Asymptotic Properties

We obtain asymptotic consistency results for estimators of the varying coefficient functions β, for the variance function $\text{var}\{Z(t)\}$ of the drift process and for the variance explained at time t by the deterministic part (11) of the stochastic equation (6), quantified by $R^2(t)$.

Corresponding estimators result from plugging in estimators for the eigenvalues λ_k, eigenfunctions ϕ_k and eigenfunction derivatives $\phi_k^{(1)}$ into the representations (7) for the function $\beta(t)$, (10) for the variance function of Z and (13) for $R^2(t)$. Here one needs to truncate the expansions at a finite number $K = K(n)$ of included eigen-components.

Details about the estimation procedures, which are based on local linear smoothing of one- and two-dimensional functions, are deferred to Appendix A.2. Our asymptotic consistency results focus on L^2 convergence rates. They peruse auxiliary results on the convergence of estimates of eigenvalues, eigenfunctions and eigenfunction derivatives, complementing and improving upon related results of Liu and Müller (2009), which were derived for convergence in the sup norm. Improved rates of convergence in the L^2 distance are the consequence of a special decomposition that we employ in the proofs to overcome the difficulty caused by the dependence of the repeated measurements.

Required regularity conditions include assumptions for the distribution of the design points, behavior of eigenfunctions ϕ_k and eigenvalues λ_k as their order k increases 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 $\mu^{(1)}(t)$, and $h_{G,0}, h_{G,1}$ for the estimation of the covariance surface and its partial derivative. We note that extremely sparse designs are covered, with only two measurements per trajectory; besides being bounded, the number of measurements N_i for the i-th trajectory is required to satisfy $P(N_i \geq 2) > 0$.

Specifically, for the observations (T_{ij}, Y_{ij}), $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, made for the i-th trajectory, we require that

(A1) N_i are random variables with $N_i \overset{\text{i.i.d.}}{\sim} N$, where N is a bounded positive discrete random variable with and $P\{N \geq 2\} > 0$, and $(\{T_{ij}, j \in J_i\}, \{Y_{ij}, j \in J_i\})$ are independent of N_i, for $J_i \subseteq \{1, \ldots, N_i\}$.

Writing $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{iN_i})^T$ and $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$, the triples $\{\boldsymbol{T}_i, \boldsymbol{Y}_i, N_i\}$ are assumed to be i.i.d. For the bandwidths used in the smoothing steps for $\mu(t)$ and $\mu^{(1)}(t)$ in (21), $G(t,s)$ and $G^{(1,0)}(t,s)$ in (22), we require that, as $n \to \infty$,

(A2) $\max(h_{\mu,0}, h_{\mu,1}, h_{G,0}, h_{G,1}) \to 0$, $nh_{\mu,0} \to \infty$, $nh_{\mu,1}^3 \to \infty$, $nh_{G,0}^2 \to \infty$, $nh_{G,1}^4 \to \infty$.

To characterize the behavior of estimated eigenfunction derivatives $\hat{\phi}^{(1)}(t)$, define

$$\delta_1 = \lambda_1 - \lambda_2, \quad \delta_k = \min_{j \leq k}(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}), \quad k \geq 2. \tag{14}$$

For the kernels used in the local linear smoothing steps and underlying density and moment functions, we require assumptions (B1)-(B2) in the Appendix. Denote the L^2 norm by $\|f\| = \{\int_{\mathcal{T}} f^2(t)dt\}^{1/2}$, the Hilbert-Schmidt norm by $\|\Phi\|_s = \{\int_{\mathcal{T}}\int_{\mathcal{T}}\{\Phi^2(t,s)dtds\}^{1/2}$, and also define $\|\Phi\|_u^2 = \{\int_{\mathcal{T}} \Phi^2(t,t)dt\}^{1/2}$.

The following result provides asymptotic rates of convergence in the L^2 norm for the auxiliary estimates of mean functions and their derivatives as well as covariance functions and their partial derivatives, which are briefly discussed in Appendix A.2. A consequence is a convergence result for the eigenfunction derivative estimates $\hat{\phi}_k^{(1)}$, with constants and rates that hold uniformly in the order $k \geq 1$.

Theorem 2. *Under (A1)-(A2) and (B1)-(B3), for* $\nu \in \{0,1\}$,

$$\|\hat{\mu}^{(\nu)} - \mu^{(\nu)}\| = O_p\left(\frac{1}{\sqrt{nh_{\mu,\nu}^{2\nu+1}}} + h_{\mu,\nu}^2\right), \quad \|\hat{G}^{(\nu,0)} - G^{(\nu,0)}\|_s = O_p\left(\frac{1}{\sqrt{n}h_{G,\nu}^{\nu+1}} + h_{G,\nu}^2\right). \tag{15}$$

For $\phi_k^{(1)}(t)$ *corresponding to* λ_k *of multiplicity* 1,

$$\|\hat{\phi}_k^{(1)}(t) - \phi_k^{(1)}(t)\| = O_p\left(\frac{1}{\lambda_k}\Big\{\frac{1}{\sqrt{n}h_{G,1}^2} + h_{G,1}^2 + \frac{1}{\delta_k}\big(\frac{1}{\sqrt{n}h_{G,0}} + h_{G,0}^2\big)\Big\}\right), \tag{16}$$

where the $O_p(\cdot)$ *term in (16) is uniform in* $k \geq 1$.

An additional requirement is that variances of processes X and $X^{(1)}$ are bounded above and below, since these appear in the denominators of various representations, e.g., in (10) and (13),

(A3) $\inf_{t\in\mathcal{T}} G^{(\nu,\nu)}(s,s) \geq c > 0$ and $\|G^{(\nu,\nu)}\|_u < \infty$ for $\nu = 0, 1$,

implying that $\|G^{(\nu,\nu)}\|_s < \infty$ by the Cauchy-Schwarz inequality. Define remainder terms

$$R_{K,\nu}(t) = \sum_{k=K+1}^{\infty} \lambda_k\{\phi^{(\nu)}(t)\}^2, \quad R_{K,\nu}^*(s,t) = \sum_{k=K+1}^{\infty} \lambda_k\phi^{(\nu)}(s)\phi^{(\nu)}(t); \tag{17}$$

by the Cauchy-Schwarz inequality, $\|R_{K,\nu}^*\|_s \leq \|R_{K,\nu}\|_u$.

In order to obtain consistent estimates of various quantities, a necessary requirement is that the first K eigen-terms approximate the infinite-dimensional process sufficiently well. The increase in the sequence $K = K(n)$ as $n \to \infty$ therefore needs to be tied to the spacing and decay of eigenvalues,

(A4) $K = o(\min\{\sqrt{n}h_{G,1}^2, h_{G,1}^{-2}\})$, $\sum_{k=1}^{K}\delta_k^{-1} = o(\min(\sqrt{n}h_{G,0}, h_{G,0}^{-2}\})$, $\max_{\nu=0,1}\|R_{K,\nu}\| \to 0$, as $n \to \infty$.

If the eigenvalues decrease rapidly and merely a few leading terms are needed, condition (A4) is easily satisfied. We use "$\overset{p}{\asymp}$" to connect two terms which are asymptotically of the same order in probability, i.e., the terms are O_p of each other. Define the sequence

$$\alpha_n = K\{(\sqrt{n}h_{G,1}^2)^{-1} + h_{G,1}^2\} + (\sum_{k=1}^{K}\delta_k^{-1})\{(\sqrt{n}h_{G,0})^{-1} + h_{G,0}^2\}. \tag{18}$$

Note that $\mathrm{cov}\{X^{(1)}(s), X^{(1)}(t)\} = G^{(1,1)}(s,t) = \sum_{k=1}^{\infty}\lambda_k\phi_k^{(1)}(s)\phi_k^{(1)}(t)$ with corresponding plug-in estimate $\hat{G}_K^{(1,1)}(s,t) = \sum_{k=1}^{K}\hat{\lambda}_k\hat{\phi}_k^{(1)}(s)\hat{\phi}_k^{(1)}(t)$, where $K = K(n)$ is the included number of eigenfunctions. The plug-in estimate for $\beta(t)$ is based on (7) and given by $\hat{\beta}_K(t) = \sum_{k=1}^{K}\hat{\lambda}_k\hat{\phi}_k^{(1)}(t)\hat{\phi}_k(t)/\sum_{k=1}^{K}\hat{\lambda}_k\hat{\phi}_k(t)^2$ and analogously the plug-in estimate $\hat{G}_{z,K}$ of G_z is based on representation (8), using the estimate $\hat{\beta}_K$. In a completely analogous fashion one obtains the estimates $\hat{R}_K^2(t)$ of $R^2(t)$ from (13) and $\hat{V}_K(t)$ of the variance function $V(t) = \mathrm{var}(Z(t))$ of the drift process from (10). The L^2 convergence rates of these estimators of various components of the dynamic model (6) are given in the following result.

Theorem 3. *Under (A1)-(A4) and (B1)-(B3),*

$$\|\hat{G}_K^{(1,1)} - G^{(1,1)}\|_s = O_p(\alpha_n + \|R_{K,1}^*\|_s), \quad \|\hat{G}_K^{(1,1)} - G^{(1,1)}\|_u = O_p(\alpha_n + \|R_{K,1}\|), \tag{19}$$

$$\begin{aligned}\|\hat{\beta}_K - \beta\| &\overset{p}{\asymp} \|\hat{G}_{z,K} - G_z\|_s \overset{p}{\asymp} \|\hat{G}_{z,K} - G_z\|_u \overset{p}{\asymp} \|\hat{R}_K^2 - R^2\| \overset{p}{\asymp} \|\hat{V}_K - V\| \\ &= O_p(\alpha_n + \|R_{K,0}\| + \|R_{K,1}\|).\end{aligned} \tag{20}$$

The weak convergence and L^2 consistency for the estimated eigenvalues $\{\rho_k\}$ and eigenfunctions $\{\psi_k\}$ of the drift process Z is an immediate consequence of this result. To see this, one may use $\sup_{k\geq 1}|\hat{\rho}_k - \rho_k| = O_p(\|\hat{G}_z - G_z\|_s)$ and $\|\hat{\psi}_k - \psi_k\| = {\delta_k^*}^{-1}O_p(\|\hat{G}_z - G_z\|_s)$ where $\hat{G}_z$ is any estimate of G_z (Bosq 2000). Here the $O_p(\cdot)$ terms are uniform in k and $\delta_1^* = \rho_1 - \rho_2$, $\quad \delta_k^* = \min_{j\leq k}(\rho_{j-1} - \rho_j, \rho_j - \rho_{j+1})$ for $k \geq 2$.

4 Application to Online Auction Data

4.1 Data and population level analysis

To illustrate our methods, we analyze the dynamic system corresponding to online auction data, specifically using eBay bidding data for 156 online auctions of Palm Personal Digital

Assistants in 2003 (courtesy of Wolfgang Jank). The data are "live bids" that are entered by bidders at irregular times and correspond to the actual price a winning bidder would pay for the item. This price is usually lower than the "willingness-to-pay" price, which is the value a bidder enters. Further details regarding the proxy bidding mechanism for the 7-day second-price auction design that applies to these data can be found in Jank and Shmueli (2005, 2006); Liu and Müller (2008, 2009).

The time unit of these 7-day auctions is hours and the domain is the interval $[0, 168]$. Adopting the customary approach, the bid prices are log-transformed prior to the analysis. The values of the live bids Y_{ij} are sampled at bid arrival times T_{ij}, where $i = 1, \ldots, 156$ refers to the auction index and $j = 1, \ldots, N_i$ to the total number of bids submitted during the i-th auction; the number of bids per auction is found to be between 6 and 49 for these data. We adopt the point of view that the observed bid prices result from an underlying price process which is smooth, where the bids themselves are subject to small random aberrations around underlying continuous trajectories. Since there is substantial variability of little interest in both bids and price curves during the first three days of an auction, when bid prices start to increase rapidly from a very low starting point to more realistic levels, we restrict our analysis to the interval [96,168] (in hours), thus omitting the first three days of bidding. This allows us to focus on the more interesting dynamics in the price curves taking place during the last four days of these auctions.

Our aim is to explore the price dynamics through the empirical stochastic differential equation (6). Our study emphasizes description of the dynamics over prediction of future auction prices and consists of two parts: A description of the dynamics of the price process at the "population level", which focuses on patterns and trends in the population average and is reflected by dynamic equations for conditional expectations. The second and major results concern the quantification of the dynamics of auctions at the individual or "auction-specific level", where one studies the dynamic behavior for each auction separately, but using the information gained across the entire sample of auctions. Only the latter analysis involves the stochastic drift term Z in the stochastic differential equation (6). We begin by reviewing the population level analysis, which is characterized by the deterministic part of eq. (6), corresponding to the equation $E(X^{(1)}(t) - \mu^{(1)}(t)|X(t) - \mu(t)) = \beta(t)\{X(t) - \mu(t)\}$. This equation describes a relationship that holds for conditional means but not necessarily for

individual trajectories.

For the population level analysis, we require estimates of the mean price curve μ and its first derivative $\mu^{(1)}$, and these are obtained by applying linear smoothers to (21) to the pooled scatterplots that are displayed in Figure 2 (for more details, see Appendix A.2). One finds that both log prices and log price derivatives are increasing throughout, so that at the log-scale the price increases are accelerating in the mean, as the auctions proceed.

A second ingredient for our analysis are estimates for the eigenfunctions and eigenvalues (details in Appendix A.2). Since the first three eigenfunctions were found to explain 84.3%, 14.6% and 1.1% of the total variance, three components were selected. The eigenfunction estimates are shown in the left panel of Figure 3, along with the estimates of the corresponding eigenfunction derivatives in the right panel. For the interpretation of the eigenfunctions it is helpful to note that the sign of the eigenfunctions is arbitrary. We also note that variation in the direction of the first eigenfunction ϕ_1 corresponds to the major part of the variance. The variances $\lambda_1\phi_1^2(t)$ that are attributable to this eigenfunction are seen to steadily decrease as t is increasing, so that this eigenfunction represents a strong trend of higher earlier and smaller later variance in the log price trajectories.

The contrast between large variance of the trajectories at earlier times and smaller variances later reflects the fact that auction price trajectories are less determined early on when both relatively high as well as low prices are observed, while at later stages prices differ less as the end of the auction is approached and prices are constrained into a narrower range. Correspondingly, the first eigenfunction derivative is steadily increasing (decreasing if the sign is switched), with notably larger increases (decreases) both at the beginning and at the end and a relatively flat positive plateau in the middle part.

The second eigenfunction corresponds to a contrast between trajectory levels during the earlier and the later part of the domain, as is indicated by its steady increase and the sign change, followed by a slight decrease at the very end. This component thus reflects a negative correlation between early and late log price levels. The corresponding derivative is positive and flat, with a decline and negativity towards the right endpoint. The third eigenfunction, explaining only a small fraction of the overall variance, reflects a more complex contrast between early and late phases on one hand and a middle period on the other, with equally more complex behavior reflected in the first derivative.

The eigenfunctions and their derivatives in conjunction with the eigenvalues determine the varying coefficient function β, according to eq. (7). The estimate of this function is obtained by plugging in the estimates for these quantities and is visualized in the left panel of Figure 5, demonstrating small negative values for the function β throughout most of the domain, with a sharp dip of the function into the negative realm near the right end of the auctions.

For subdomains of functional data, where the varying coefficient or "dynamic transfer" function β is negative, as is the case for the auction data throughout the entire time domain, one may interpret the population equation $E(X^{(1)}(t)-\mu^{(1)}(t)|X(t)-\mu(t)) = \beta(t)\{X(t)-\mu(t)\}$ as indicating "dynamic regression to the mean". By this we mean the following: When a trajectory value at a current time t falls above (resp., below) the population mean trajectory value at t, then the conditional mean derivative of the trajectory at t falls below (resp., above) the mean. The overall effect of this negative association is that the direction of the derivative is such that trajectories tend to move towards the overall population mean trajectory as time progresses.

Thus, our findings for the auction data indicate that "dynamic regression to the mean" takes place to a small extent throughout the auction period and to a larger extent near the right tail, at the time when the final auction price is determined (see also Liu and Müller 2009). One interpretation is that at the population level, prices are self-stabilizing, which tends to prevent price trajectories running away towards levels way above or below the mean trajectory. This self-stabilization feature gets stronger towards the end of the auction, where the actual "value" of the item that is being auctioned serves as a strong homogenizing influence. This means that in a situation where the current price level appears particularly attractive, the expectation is that the current price derivative is much higher than for an auction with an unattractive current price, for which then the corresponding current price derivative is likely lower. The net effect is a trend for log price trajectories to regress to the mean trajectory as time progresses.

4.2 Auction-specific dynamics

We illustrate here the proposed stochastic differential equation (6). First estimating the function β, we obtain the trajectories Z_i of the drift process. These trajectories are presented

in Figure 4 for the entire sample of auctions. These trajectories quantify the component of the derivative process $X^{(1)}$ that is left unexplained by the varying coefficient function and linear part of the dynamic model (6). The trajectories Z_i exhibit fluctuating variances across various subdomains. The subdomains for which these variances are small are those where the deterministic approximation (11) to the stochastic differential equation works best. It is noteworthy that the variance is particularly small on the subdomain starting at around 158 hours towards the endpoint of the auction at 168 hours, since auction dynamics are of most interest during these last hours. It is well known that towards the end of the auctions, intensive bidding takes place, in some cases referred to as "bid sniping", where bidders work each other into a frenzy to outbid each other in order to secure the item that is auctioned.

The right panel of Figure 5 shows the first three eigenfunctions of Z, which are derived from the eigenequations derived from estimates $\hat{G}_{z,K}$ of covariance kernels $G_{z,K}$ (8) that are obtained as described after eq. (22). In accordance with the visual impression of the trajectories of Z in Figure 4, the first eigenfunction reflects maximum variance in the middle portion of the domain and very low variance at both ends. Interestingly, the second eigenfunction reflects high variance at the left end of the domain where prices are still moving upwards quite rapidly, and very low variance near the end of the auction. This confirms that overall variation is large in the middle portion of the auctions, so that the drift process in eq. (6) plays an important role in that period.

Further explorations of the modes of variation of the drift process Z can be based on the functional principal component scores of Z_i. Following Jones and Rice (1992), we identify the three auctions with the largest absolute values of the scores. A scatterplot of second and first principal component scores with these auctions highlighted can be seen in the left upper panel of Figure 6. The corresponding individual (centered) trajectories of the drift process Z are in the right upper panel, and the corresponding trajectories of centered processes X and $X^{(1)}$ in the left and right lower panels. The highlighted trajectories of Z are indeed similar to the corresponding eigenfunctions (up to sign changes) and we find that they all exhibit the typical features of small variance near the end of the auction for Z and X and of large variance for $X^{(1)}$.

For the two trajectories corresponding to maximal scores for first and second eigenfunction of Z we find that near the end of the auctions their centered derivatives turn negative. This

is in line with dynamic regression to the mean, or equivalently, negative varying coefficient function β, as described in Section 4.1. Here the trajectories for X at a current time t are above the mean trajectory, which means the item is pricier than the average price at t. As predicted by dynamic regression to the mean, log price derivative trajectories at t are indeed seen to be below the mean derivative trajectories at t. The trajectory corresponding to maximal score for the third eigenfunction also follows dynamic regression to the mean: Here the trajectory for X is below the overall mean trajectory, so that the negative varying coefficient function β predicts that the derivative trajectory $X^{(1)}$ should be above the mean, which indeed is the case.

That the variance of the drift process Z is small near the endpoint of the auction is also evident from the estimated variance function $V(t) = \mathrm{var}(Z(t))$ in the left panel of Figure 7, overlaid with the estimated variance function $\mathrm{var}(X^{(1)}(t))$ of $X^{(1)}$. The latter is rapidly increasing towards the end of the auction, indicating that the variance of the derivative process is very large near the auction's end. This means that price increases vary substantially near the end across auctions. The large variances of derivatives coupled with the fact that $\mathrm{var}(Z(t))$ is small near the end of the auction implies that the deterministic part (11) of the empirical differential equation (6) explains a very high fraction of the variance in the data. This corresponds to a very high, indeed close to the upper bound 1, value of the coefficient of determination $R^2(t)$ (12), (13) in an interval of about 10 hours before the endpoint of an auction, as seen in the right panel of Figure 7. We therefore find that the dynamics during the endrun of an auction can be nicely modeled by the simple deterministic approximation (11) to the stochastic dynamic equation (6), which always applies.

This finding is corroborated by visualizing the regressions of $X^{(1)}(t)$ versus $X(t)$ at various fixed times t. These regressions are linear in the Gaussian case and may be approximated by a linear regression in the least squares sense in the non-Gaussian case. The scatterplots of $\hat{X}_i^{(1)}(t) - \hat{\mu}^{(1)}(t)$ versus $\hat{X}_i(t) - \hat{\mu}(t)$ for times $t = 125$ hours and $t = 161$ hours (where the time domain of the auctions is between 0 and 168 hours) are displayed in Figure 8. This reveals the relationships to be indeed very close to linear. These are regressions though the origin. The regression slope parameters are not estimated from these scatterplot data which are contaminated by noise, but rather are obtained directly from (6), as they correspond to $\beta(t)$. Thus one simply may use the already available slope estimates, $\hat{\beta}(125) = -0.015$

and $\hat{\beta}(1) = -0.072$. The associated coefficients of determination, also directly estimated via (13) and the corresponding estimation procedure, are found to be $\hat{R}^2(125) = 0.28$ and $\hat{R}^2(161) = 0.99$.

As the regression line fitted near the end of the auction at $t = 161$ hours explains almost all the variance, the approximating deterministic differential equation (11) can be assumed to hold at that time (and at later times as well, all the way to the end of the auction). At $t = 125$ the regression line explains only a fraction of the variance, while a sizable portion of variance resides in the drift process Z, so that the stochastic part in the dynamic system (6) cannot be comfortably ignored in this time range. These relationships can be used to predict derivatives of trajectories and thus price changes at time t for individual auctions, given their log price trajectory values at t. We note that such predictions apply to fitted trajectories, not for the actually observed prices which contain an additional random component that is unpredictable, according to model (1). We find that at time $t = 161$, regression to the mean is observed at the level of individual auctions: An above (below) average log price level is closely associated with a below (above) average log price derivative. This implies that a seemingly very good (or bad) deal will not be quite so good (or bad) when the auction ends.

5 Discussion

The main motivation of using the dynamic system approach based on eq. (6) is that ir provides a better description of the mechanisms that drive longitudinal data but are not directly observable. The empirical dynamic equation may also suggest constraints on the form of parametric differential equations that are compatible with the data. In the auction example, the dynamic equation quantifies both the nature and extent of how expected price increases depend on auction stage and current price level. This approach is primarily phenomenological and does not directly lend itself to the task of predicting future values of individual trajectories.

That expected conditional trajectory derivatives satisfy a first order differential equation model (which we refer to as the "population level" since this statement is about conditional expectations) simply follows from Gaussianity and in particular does not require additional assumptions. This suffices to infer the stochastic differential equation described in eq. (5),

which we term "empirical differential equation", as it is determined by the data. Then the function R^2, quantifying the relative contribution of the drift process Z to the variance of $X^{(1)}$, determines how closely individual trajectories follow the deterministic part of the equation. We could equally consider stochastic differential equations of other orders, but practical considerations favor the modeling with first order equations.

We find in the application example that online auctions follow a dynamic regression to the mean regime for the entire time domain, which becomes more acute near the end of the auction. This allows to construct predictions of log price trajectory derivatives from trajectory levels at the same t. These predictions get better towards the right endpoint of the auctions. This provides a cautionary message to bidders, since an auction that looks particularly promising since it has a current low log price trajectory is likely not to stay that way and larger than average price increases are expected down the line. Conversely, an auction with a seemingly above average log price trajectory is likely found to have smaller than average price increases down the line.

This suggests that bidders take a somewhat detached stance, watching auctions patiently as they evolve. In particular, discarding auctions that appear overpriced is likely not a good strategy as further price increases are going to be smaller than the average for such auctions. It also implies that bid snipers are ill advised: A seemingly good deal is not likely to stay that way, suggesting a more relaxed stance. Conversely, a seller who anxiously follows the price development of an item, need not despair if the price seems too low at a time before closing, as it is likely to increase rapidly towards the end of the auction.

For prediction purposes, drift processes Z_i for individual auctions are of great interest. In time domains where their variance is large, any log price development is possible. Interestingly, the variance of drift processes is very small towards the right tail of the auctions, which means that the deterministic part of the differential equation (6) is relatively more important, and log price derivatives during the final period of an auction become nearly deterministic and thus predictable.

Other current approaches of statistical modeling of differential equations for time course data (e.g., Ramsay et al. 2007) share the idea of modeling with a first order equation. In all other regards these approaches are quite different, as they are based on the prior notion that a differential equation of a particular and known form pertains to the observed time

courses and moreover usually have been developed for the modeling of single time courses. This established methodology does not take into account the covariance structure of the underlying stochastic process. In contrast, this covariance structure is a central object in our approach and is estimated nonparametrically from the entire ensemble of available data, across all subjects or experiments.

Appendix

A.1 Additional details and discussion for preliminary formulas

Formula (4) is an extension of the covariance kernel representation in terms of eigenfunctions, given by $\mathrm{cov}(X(t), X(s)) = \sum \lambda_k \phi_k(t)\phi_k(s)$ (Ash and Gardner 1975), which itself is a stochastic process version of the classical multivariate representation of a covariance matrix C in terms of its eigenvectors e_k and eigenvalues λ_k, $C = \sum \lambda_k e_k e_k'$. Specifically, using representation (2), one finds $\mathrm{cov}(X^{(\nu_1)}(t), X^{(\nu_2)}(s)) = \sum_{k,l} \mathrm{cov}(\xi_k, \xi_l)\phi_k^{(\nu_1)}(t)\phi_k^{(\nu_2)}(s)$, and (4) follows upon observing that $\mathrm{cov}(\xi_k, \xi_l) = \lambda_k$ for $k = l$ and $= 0$ for $k \neq l$.

Regarding the "population differential equation" $E\{X^{(1)}(t) - \mu^{(1)}(t) \mid X(t)\} = \beta(t)\{X(t) - \mu(t)\}$, observe that for any jointly normal random vectors (U_1, U_2) with mean 0 and covariance matrix C with elements $c_{11}, c_{12}, c_{21} = c_{12}, c_{22}$, it holds that $E(U_2|U_1) = (c_{21}/c_{11})U_1$. Applying this to the jointly normal random vectors in eq. (5) then implies this population equation. The specific form for the function β in eq. (7) is obtained by plugging in the specific terms of the covariance matrix given on the right hand side of eq. (5).

Applying (4), observing $\mathrm{var}(X(t)) = \sum_k \lambda_k \phi_k(t)^2$, and then taking the log-derivative leads to $\frac{d}{dt}\log(\mathrm{var}(X(t))) = 2\,[\sum_k \lambda_k \phi_k(t)\phi_k^{(1)}(t)]/[\sum_k \lambda_k \phi_k^2(t)]$, establishing the last equality in representation (7).

A.2 Estimation procedures

Turning to estimation, in a first step we aggregate 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

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i}\kappa_1\left(\frac{T_{ij}-t}{h_{\mu,\nu}}\right)\left\{Y_{ij}-\sum_{m=0}^{\nu+1}\alpha_m(T_{ij}-t)^m\right\}^2 \tag{21}$$

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 (3), 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, we minimize the pooled scatterplot of pairwise raw covariances

$$\sum_{i=1}^{n}\sum_{1\le j\neq l\le N_i}\kappa_2\left(\frac{T_{ij}-t}{h_{G,\nu}},\frac{T_{il}-s}{h_{G,\nu}}\right)\left\{G_i(T_{ij},T_{il})-(\sum_{m=0}^{\nu+1}\alpha_{1m}(T_{ij}-t)^m+\alpha_{21}(T_{il}-s))\right\}^2 \tag{22}$$

for fixed (t,s) with respect to α_{1m} and α_{21} for $m = 1, \ldots, \nu+1$, where $G_i(T_{ij},T_{il}) = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{il} - \hat{\mu}(T_{il}))$, $j \neq l$, 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 essential to overcome the limitations of the sparse sampling designs. We note that the case of no derivative $\nu = 0$ is always included, and solving the eigenequations on the left hand side of eq. (3) numerically for that case leads to the required estimates $\hat{\lambda}_1, \hat{\lambda}_2, \ldots$ of the eigenvalues and $\hat{\phi}_1, \hat{\phi}_2, \ldots$ of the eigenfunctions. The estimates $\hat{\phi}_1^{(1)}, \hat{\phi}_2^{(1)}, \ldots$ of the eigenfunction derivatives are then obtained from the right hand side of eq. (3), plugging in the estimates for eigenfunctions and eigenvalues, followed by a numerical integration step.

The plug-in estimates $\hat{\beta}_K, \hat{G}_{z,K}, \hat{V}_K, \hat{R}^2_K$ are then obtained from the corresponding representations (7), (8), (10), (13), by including K leading components in the respective sums. While for theoretical analysis and asymptotic consistency one requires $K = K(n) \to \infty$, the number of included eigen-terms K in practical data analysis can be chosen by various criteria, e.g., AIC/BIC based on marginal/conditional pseudo-likelihood or thresholding of the total variation explained by the included components (Liu and Müller 2009). One key feature of the covariance surface smoothing step in (22) 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 (22) when no derivatives are involved can then be used to find consistent estimates for the error variance σ^2 (Yao et al. 2005).

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 expectation can be consistently estimated in the case of Gaussian processes by applying principal analysis by conditional expectation (PACE) (Yao et al. 2005). For $\boldsymbol{X}_i = (X_i(T_{i1}),\ldots,X_i(T_{iN_i}))^T$, $\boldsymbol{Y}_i = (Y_{i1},\ldots,Y_{iN_i})^T$, $\boldsymbol{\mu}_i = (\mu(T_{i1}),\ldots,\mu(T_{iN_i}))^T$, $\boldsymbol{\phi}_{ik} = (\phi_k(T_{i1}),\ldots,\phi_k(T_{iN_i}))^T$, if ξ_{ik} and ε_{ij} in (1) are jointly Gaussian, then by standard properties of the Gaussian distribution,

$$E(\xi_{ik} \mid \boldsymbol{Y}_i) = \lambda_k \boldsymbol{\phi}_{ik}^T \Sigma_{Y_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i), \tag{23}$$

where $\Sigma_{Y_i} = \mathrm{cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_i) = \mathrm{cov}(\boldsymbol{X}_i, \boldsymbol{X}_i) + \sigma^2 I_{N_i}$. This implies $E(X_i^{(\nu)}(t)|Y_{i1},\ldots,Y_{iN_i}) = \sum_{k=1}^{\infty} E(\xi_{ik} \mid \boldsymbol{Y}_i)\phi_k^{(\nu)}(t) = \{\sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu)}(t)\boldsymbol{\phi}_{ik}^T\}\, \Sigma_{Y_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$, $\nu = 0, 1$. The unknown quantities can be estimated by simply plugging in the variance, eigenvalue, eigenfunction and eigenfunction derivative estimates discussed above, again coupled with truncating the number of included components at K.

A.3 Additional assumptions and auxiliary results

Denote the densities of T and (T_1, T_2) by $f_1(t)$, $f_2(t,s)$ and define an interior domain by $\mathcal{T} = [a,\, b]$ with $\mathcal{T}_\delta = [a-\delta,\, b+\delta]$ for some $\delta > 0$. Regularity conditions for the densities and the targeted moment functions as well as their derivatives are as follows, where ℓ_1, ℓ_2 are non-negative integers,

(B1) $f_1^{(5)}(t)$ exists and is continuous on $\mathcal{T}_\delta$ with $f(t) > 0$, $\frac{\partial^5}{\partial t^{\ell_1} \partial s^{\ell_2}} f_2(t,s)$ exists and is continuous on $\mathcal{T}_\delta^2$ for $\ell_1 + \ell_2 = 5$.

(B2) $\mu^{(5)}(t)$ exists and is continuous on $\mathcal{T}_\delta$, $\frac{\partial^5}{\partial t^{\ell_1} \partial s^{\ell_2}} G(t,s)$ exists and is continuous on $\mathcal{T}_\delta^2$ for $\ell_1 + \ell_2 = 5$.

We say that a bivariate kernel function κ_2 is of order $(\boldsymbol{\nu}, \ell)$, where $\boldsymbol{\nu}$ is a multi-index

$\boldsymbol{\nu} = (\nu_1, \nu_2)$, if

$$\int\int u^{\ell_1} v^{\ell_2} K_2(u,v)dudv = \begin{cases} 0 & 0 \le \ell_1 + \ell_2 < \ell, \ell_1 \neq \nu_1, \ell_2 \neq \nu_2, \\ (-1)^{|\nu|}\nu_1!\nu_2! & \ell_1 = \nu_1, \ell_2 = \nu_2, \\ \neq 0 & \ell_1 + \ell_2 = \ell, \end{cases} \tag{24}$$

where $|\boldsymbol{\nu}| = \nu_1 + \nu_2 < \ell$. The univariate kernel κ_1 is said to be of order (ν, ℓ) for a univariate $\nu = \nu_1$, if (24) holds with $\ell_2 = 0$ on the right hand side, integrating only over the argument u on the left hand side. For the kernel functions κ_1, κ_2 used in the smoothing steps to obtain estimates for $\mu(t)$ and $\mu^{(1)}(t)$ in (21) and for $G(t,s)$ and $G^{(1,0)}(t,s)$ in (22) we assume

(B3) Kernel functions κ_1 and κ_2 are non-negative with compact supports, bounded and of order $(0,2)$ and $((0,0),2)$, respectively.

The following lemma provides the weak L^2 convergence rate for univariate and bivariate weighted averages defined below. For arbitrary real functions $\theta : \Re^2 \to \Re$ and $\theta^* : \Re^4 \to \Re$, define $\tilde{\theta}(t) = E\{\theta(t_{ij}, Y_{ij})|T_{ij} = t\}$ and $\tilde{\theta}^*(t) = E\{\theta^*(t_{ij}, t_{il}, Y_{ij}, Y_{il})|T_{ij} = t, T_{il} = s\}$, let $\theta_\nu(t) = \tilde{\theta}^{(\nu)}(t)f_1(t)$ for a single index ν and $\theta^*_{\boldsymbol{\nu}}(t) = f_2(t,s)\frac{\partial^{|\nu|}}{\partial^{\nu_1}\partial^{\nu_2}}\tilde{\theta}^*(t,s)$ for a multi-index $\boldsymbol{\nu} = (\nu_1, \nu_2)$ and define the weighted kernel averages, employing bandwidths h_1, h_2,

$$\hat{\theta}(t) = \frac{1}{E(N)nh_1^{\nu+1}} \sum_{i=1}^{n}\sum_{j=1}^{N_i} \theta(T_{ij}, Y_{ij})\kappa_1\left(\frac{t - T_{ij}}{h_1}\right), \tag{25}$$

$$\hat{\theta}^*(t,s) = \frac{1}{E\{N(N-1)\}nh_2^{|\boldsymbol{\nu}|+2}} \sum_{i=1}^{n} \sum_{1\le j\neq l\le N_i} \theta^*(T_{ij}, T_{il}, Y_{ij}, Y_{il})\kappa_2\left(\frac{t - T_{ij}}{h_2}, \frac{s - T_{il}}{h_2}\right). \tag{26}$$

For establishing convergence results for the general weighted averages (25), assume that

(B2†) Derivatives $\tilde{\theta}^{(\ell)}(t)$ exist and are continuous on $\mathcal{T}_\delta$, $\frac{\partial^5}{\partial t^{\ell_1}\partial s^{\ell_2}}\tilde{\theta}^*(t,s)$ exists and is continuous on $\mathcal{T}_\delta^2$ for $\ell_1 + \ell_2 = \ell$.

(B3†) The univariate kernel κ_1 is of order (ν, ℓ) and the bivariate kernel κ_2 is of order $(\boldsymbol{\nu}, \ell)$.

Lemma 1. *Under (A1), (B1), (B2†), (B3†), and if* $\max\{h_1, h_2\} \to 0$, $nh_1^{2\nu+1} \to \infty$ *and* $nh_2^{2|\boldsymbol{\nu}|+2} \to \infty$, *as* $n \to \infty$,

$$\|\hat{\theta} - \theta_\nu\| = O_p(\frac{1}{\sqrt{nh_1^{2\nu+1}}} + h_1^{\ell-\nu}), \qquad \|\hat{\theta}^* - \theta^*_{\boldsymbol{\nu}}\|_s = O_p(\frac{1}{\sqrt{n}h_2^{|\boldsymbol{\nu}|+1}} + h_2^{\ell-|\boldsymbol{\nu}|}). \tag{27}$$

A.4 Technical proofs

Proof of Theorem 1. Since $X, X^{(1)}$ are jointly Gaussian processes, it is clear that Z is Gaussian. Formula (7) for $\beta(t)$ is obtained by observing that for joint Gaussian r.v.s, $E\{X^{(1)}(t) - \mu^{(1)}(t) \mid X(t) - \mu(t)\} = [\mathrm{cov}\{X^{(1)}(t), X(t)\}/\mathrm{var}\{X(t)\}]\{X(t) - \mu(t)\}$. Then the properties of the functional principal component scores lead directly to

$$\mathrm{cov}\{X^{(1)}(t), X(t)\} = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t), \quad \mathrm{var}\{X(t)\} = \sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2, \tag{28}$$

whence $\beta(t) = E\{X^{(1)}(t) - \mu^{(1)}(t) \mid X(t) - \mu(t)\}$. This implies $E\{Z(t)\} = 0$. According to (6), $\mathrm{cov}\{Z(t), X(t)\} = \mathrm{cov}\{X^{(1)}(t), X(t)\} - \beta(t)\mathrm{var}\{X(t)\} = 0$, for all $t \in \mathcal{T}$, using (28) and (7). This implies the independence of $Z(t), X(t)$, due to the Gaussianity. Next observe $\mathrm{cov}\{Z(t), Z(s)\} = \mathrm{cov}\{X^{(1)}(t) - \beta(t)X(t), X^{(1)}(s) - \beta(s)X(s)\}$, from which one obtains the result by straightforward calculation. □

Proof of Lemma 1. Since $N_i \overset{\text{i.i.d.}}{\sim} N$ and N is a bounded and integer-valued random variable. Denote the upper bound by M. To handle the one-dimensional case in (25), we observe

$$\hat{\theta}(t) = \sum_{j=1}^{M} \frac{1}{E(N)} \frac{1}{nh_1^{\nu+1}} \sum_{i=1}^{n} \theta(T_{ij}, Y_{ij})\kappa_1\left(\frac{t - T_{ij}}{h_1}\right) \mathbf{1}(N_i \geq j) \equiv \sum_{j=1}^{M} \frac{1}{E(N)} \hat{\theta}_{\nu j}(t),$$

where $\mathbf{1}(\cdot)$ is the indication function. Note that for each j, $\hat{\theta}_{\nu j}$ is obtained from an i.i.d. sample. Slightly modifying the proof of Theorem 2 in Hall (1984) for a kernel of order (ν, ℓ) provides the weak convergence rate $\|\hat{\theta}_{\nu j} - \theta_\nu P(N \geq j)\| = O_p\{(nh_1^{2\nu+1})^{-1/2} + h_1^{\ell-\nu}\}$. It is easy to check that $\sum_{j=1}^{M} P(N \geq j) = E(N)$, as N is a positive integer-valued random variable. Therefore,

$$\|\hat{\theta} - \theta_\nu\| \leq \sum_{j=1}^{M} \frac{P(N \geq j)}{E(N)} \left\| \frac{\hat{\theta}_{\nu j}}{P(N \geq j)} - \theta_\nu \right\| = O_p\left(\frac{1}{\sqrt{nh_1^{2\nu+1}}} + h_1^{\ell-\nu}\right).$$

Analogously, for the two-dimensional case in (26), let

$$\hat{\theta}^*_{\boldsymbol{\nu} j} = \frac{1}{nh_2^{|\boldsymbol{\nu}|+2}} \sum_{i=1}^{n} \theta^*(T_{ij}, T_{il}, Y_{ij}, Y_{il})\kappa_2\left(\frac{t - T_{ij}}{h_2}, \frac{s - T_{il}}{h_2}\right) \mathbf{1}\{N_i \geq \max(j, l)\},$$

and then $\hat{\theta}^*_{\boldsymbol{\nu}} = \sum_{1 \leq j \neq l \leq M} [E\{N(N-1)\}]^{-1} \hat{\theta}^*_{\boldsymbol{j}}$. Similarly to the above, one has $\|\hat{\theta}^*_{\boldsymbol{\nu} j} - \theta_{\boldsymbol{\nu}} P\{N \geq \max(j, l)\}\|_s = O_p\{nh_2^{2|\boldsymbol{\nu}|+2})^{-1/2} + h_2^{\ell-|\boldsymbol{\nu}|}\}$. Again it is easy to verify that $E\{N(N-1)\} = \sum_{1 \leq j \neq l \leq M} P\{N \geq \max(j, l)\}$. The triangle inequality for the L^2 distance entails $\|\hat{\theta}^* - \theta_{\boldsymbol{\nu}}\|_s = O_p\{(nh_2^{2|\boldsymbol{\nu}|+2})^{-1/2} + h_2^{\ell-|\boldsymbol{\nu}|}\}$. □

Proof of Theorem 2. Note that the estimators $\hat{\mu}$, $\hat{\mu}^{(1)}$, $\hat{G}$ and $\hat{G}^{(1,0)}$ all can be written as functions of the general averages defined in (25), (26). Slightly modifying the proof of Theorem 1 in Liu and Müller (2009), with sup rates replaced by the L^2 rates given in Lemma 1, then leads to the optimal L^2 weak convergence rates for $\hat{\mu}^{\nu}$ and $\hat{G}^{(\nu,0)}$ in (15).

For the convergence rate of $\hat{\phi}^{(1)}$, Lemma 4.3 in Bosq (2000) implies that

$$|\hat{\lambda}_k - \lambda_k| \leq \|\hat{G} - G\|_s, \qquad \|\hat{\phi}_k - \phi_k\| \leq 2\sqrt{2}\delta_k^{-1}\|\hat{G} - G\|_s, \tag{29}$$

where δ_k is defined in (14) and $\hat{G}$ is an arbitrary estimate (or perturbation) of G. Denote the linear operators generated from the kernels $G^{(1,0)}$ and $\hat{G}^{(1,0)}$ by $\boldsymbol{G}^{(1,0)}$, respectively, $\hat{\boldsymbol{G}}^{(1,0)}$. Noting that $\delta_k \leq \lambda_k$, one finds

$$\begin{aligned}
\|\phi_k^{(1)} - \phi_k^{(1)}\| &\leq \frac{1}{\hat{\lambda}_k}\|\hat{\boldsymbol{G}}^{(1,0)}\hat{\phi}_k - \boldsymbol{G}^{(1,0)}\phi_k\| + \|\boldsymbol{G}^{(1,0)}\phi_k\| \cdot |\frac{1}{\hat{\lambda}_k} - \frac{1}{\lambda_k}| \\
&\overset{p}{\asymp} \frac{1}{\lambda_k}\{\|\hat{G}^{(1,0)} - G^{(1,0)}\|_s + \|\hat{\phi}_k - \phi_k\|\} + \frac{|\hat{\lambda}_k - \lambda_k|}{\lambda_k^2} \\
&\overset{p}{\asymp} \frac{1}{\lambda_k}\{\|\hat{G}^{(1,0)} - G^{(1,0)}\|_s + \frac{1}{\delta_k}\|\hat{G} - G\|_s\},
\end{aligned} \tag{30}$$

which implies (16). $\square$

Proof of Theorem 3. From (29) it is easy to see that $|\hat{\lambda}_k - \lambda_k| = o_p(\|\hat{\phi}_k - \phi_k\|)$, and from both (29) and (30) that $\|\hat{\phi}_k - \phi_k\| = o_p(\|\hat{\phi}_k^{(1)} - \phi_k^{(1)}\|)$ uniformly in k. One then finds that $\|\hat{G}^{(1,1)} - G^{(1,1)}\|_s$ is bounded in probability by

$$\sum_{k=1}^{K}\|\hat{\lambda}_k\hat{\phi}_k^{(1)}\hat{\phi}_k^{(1)} - \lambda_k\phi_k^{(1)}\phi_k^{(1)}\|_s + \|\sum_{k=K+1}^{\infty}\lambda_k\phi_k^{(1)}\phi_k^{(1)}\|_s \overset{p}{\asymp} \sum_{k=1}^{K}\lambda_k\|\hat{\phi}_k^{(1)} - \phi_k^{(1)}\| + \|R_{K,1}^*\|,$$

which implies that $\|\hat{G}^{(1,1)} - G^{(1,1)}\|_s = O_p(\alpha_n + \|R_{K,1}^*\|)$, where α_n is defined in (18) and the remainder terms in (17). Similar arguments lead to $\|\hat{G}^{(1,1)} - G^{(1,1)}\|_u = O_p(\alpha_n + \|R_{K,1}\|)$, noting $\|R_{K,1}^*\| \leq \|R_{K,1}\|$ due to the Cauchy-Schwarz inequality.

Regarding $\|\hat{\beta}_K - \beta\|$, one has implies that $\|\sum_{k=K+1}^{\infty}\lambda_k\{\phi_k^{(\nu)}\}^2\| \to 0$ for $\nu = 0, 1$, as $K \to \infty$.

$$\begin{aligned}
\|\hat{\beta}_K - \beta\| &\leq \frac{1}{\inf_t G(t,t)}\Big(\sum_{k=1}^{K}\|\hat{\lambda}_k\hat{\phi}_k\hat{\phi}_k^{(1)} - \lambda_k\phi_k\phi_k^{(1)}\| + \|\sum_{k=1}^{K}\lambda_k\phi_k\phi_k^{(1)}\|\Big) \\
&\quad + \Big\|\frac{\sum_{k=1}^{K}\hat{\lambda}_k\hat{\phi}_k\hat{\phi}_k^{(1)}}{\sum_{k=1}^{K}\hat{\lambda}_k\hat{\phi}_k^2\sum_{k=1}^{\infty}\lambda_k\phi_k^2}\Big\|\Big(\sum_{k=1}^{K}\|\hat{\lambda}_k\hat{\phi}_k^2 - \lambda_k\phi_k^2\| + \|\sum_{k=K+1}^{\infty}\lambda_k\phi_k^2\|\Big)
\end{aligned}$$

$$\overset{p}{\asymp} \sum_{k=1}^{K} \lambda_k \|\hat{\phi}_k^{(1)} - \phi_k^{(1)}\| + \|R_{K,0}\| + \sqrt{\|R_{K,0}\| \cdot \|R_{K,1}\|},$$

applying the Cauchy-Schwarz inequality to $\| \sum_{k=K+1}^{\infty} \lambda_k \phi_k \phi_k^{(1)} \|$. Observing $\sqrt{\|R_{K,0}\| \cdot \|R_{K,1}\|} \leq (\|R_{K,0}\| + \|R_{K,1}\|\}/2$ yields $\|\hat{\beta}_K - \beta\| = O_p(\alpha_n + \|R_{K,0}\| + \|R_{K,1}\|)$.

To study $\|\hat{G}_{z,K} - G_z\|_s$, we investigate the L^2 convergence rates of

$$I_1 = \|\hat{\beta}_K \sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k \hat{\phi}_k^{(1)} - \beta \sum_{k=1}^{\infty} \lambda_k \phi_k \phi_k^{(1)}\|_s, \quad I_2 = \|\hat{\beta}_K \hat{G}_K \hat{\beta}_K - \beta G \beta\|_s,$$

where β (resp. $\hat{\beta}_K$) and ϕ_k (resp. $\hat{\phi}_k$) share the same argument and we define $\hat{G}_K(t,s) = \sum_{k=1}^{K} \hat{\lambda} \hat{\phi}_k(t) \hat{\phi}_k(s)$. In analogy to the above arguments, $I_1 \overset{p}{\asymp} \|\hat{\beta}_K - \beta\| + \sum_{k=1}^{K} \lambda_k \|\hat{\phi}_k^{(1)} - \phi_k^{(1)}\| + \sqrt{\|R_{K,0}\| \cdot \|R_{K,1}\|}$, $I_2 = \|\hat{\beta}_K - \beta\| + \sum_{k=1}^{K} \lambda_k \|\hat{\phi}_k - \phi_k\| + \|R_{K,0}\|$. This leads to $\|\hat{G}_{z,K} - G_z\|_s = O_p(\alpha_n + \|R_{K,0}\| + \|R_{K,1}\|)$. The same argument also applies to $\|\hat{G}_{z,K} - G_z\|_u$. Next we study $\|\hat{R}_K^2 - R^2\|$ and find that $\| \sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k^{(\nu_1)} \hat{\phi}_k^{(\nu_2)} - \sum_{k=1}^{\infty} \lambda_k \phi^{(\nu_1)} \phi^{(\nu_2)} \| \overset{p}{\asymp} \sum_{k=1}^{K} \lambda_k (\|\hat{\phi}_k^{(\nu_1)} - \phi_k^{(\nu_1)}\| + \|\hat{\phi}_k^{(\nu_2)} - \phi_k^{(\nu_2)}\|) + \|R_{K,\nu_1}\| + \|R_{K,\nu_2}\| = O_p(\alpha_n + \|R_{K,\nu_1}\| + \|R_{K,\nu_2}\|)$. Analogous arguments apply to $\|\hat{V}_K - V\|$, completing the proof. □

Acknowledgments

We are grateful to two referees for helpful comments that led to an improved version of the paper.

References

Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes.* Academic Press [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics, Vol. 27.

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

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

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate Analysis* **12** 136–154.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.

Fine, J. (1987). On the validity of the perturbation method in asymptotic theory. *Statistics* **18** 401–414.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics. Theory and Applications* **11** 171–185.

Gasser, T., Müller, H.-G., Köhler, W., Molinari, L. and Prader, A. (1984). Nonparametric regression analysis of growth curves. *The Annals of Statistics* **12** 210–229.

Gervini, D. and 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. (1984). Integrated square error properties of kernel estimators of regression functions. *The Annals of Statistics* **12** 241–260.

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.

Härdle, W. and Gasser, T. (1985). On robust kernel estimation of derivatives of regression functions. *Scandinavian Journal of Statistics. Theory and Applications* **12** 233–240.

Jank, W. and Shmueli, G. (2005). Profiling price dynamics in online auctions using curve clustering. *SSRN eLibrary* .

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

Jones, M. C. and Rice, J. A. (1992). Displaying the important features of large collections of similar curves. *The American Statistician* **46** 140–145.

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

Kirkpatrick, M. and Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of Mathematical Biology* **27** 429–450.

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

Liu, B. and Müller, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics. *Journal of the American Statistical Association* **104** 704–714.

Mas, A. and Menneteau, L. (2003). Perturbation approach applied to the asymptotic study of random operators. In *High dimensional probability, III (Sandjberg, 2002)*, vol. 55 of *Progr. Probab.* Birkhäuser, Basel, 127–134.

Mas, A. and Pumo, B. (2007). The ARHD model. *Journal of Statistical Planning and Inference* **137** 538–553.

Mas, A. and Pumo, B. (2009). Functional linear regression with derivatives. *Journal of Nonparametric Statistics* **21** 19–40.

Ramsay, J. (2000). Differential equation models for statistical functions. *Cnadian Journal of Statistics* **28** 225–240.

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: Series B (Statistical Methodology)* **69** 741–796.

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

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). Modelling price paths in on-line auctions: smoothing sparse and unevenly sampled curves by using semiparametric mixed models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **57** 127–148.

Rice, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica* 631–647.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259.

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. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **45** 151–163.

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **93** 1403–1418.

Sy, J. P., Taylor, J. M. G. and Cumberland, W. G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53** 542–555.

Wang, L., Li, H. and Huang, J. Z. (2008a). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association* **103** 1556–1569.

Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association* **100** 147–157.

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* **103** 1100–1118.

Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 3–25.

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

Zhao, X., Marron, J. S. and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica* **14** 789–808.

Ziemer, W. (1989). *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*. Springer, New York.

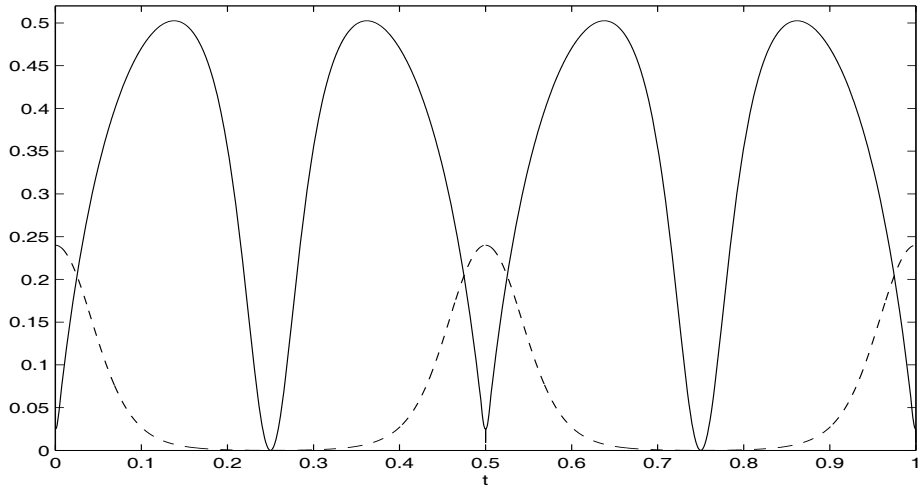

Figure 1: "Coefficient of determination" functions $R^2(t)$ (12), (13), quantifying the fraction of variance explained by the deterministic part of the dynamic equation (6), illustrated for the trigonometric basis $\{\sqrt{2}\cos(2k\pi t),\ k = 1, 2, \ldots\}$ on $[0, 1]$ and eigenvalue sequences $\lambda_k = k^{-4}$ (solid) and $\lambda_k = 2^{-k}$ (dashed).

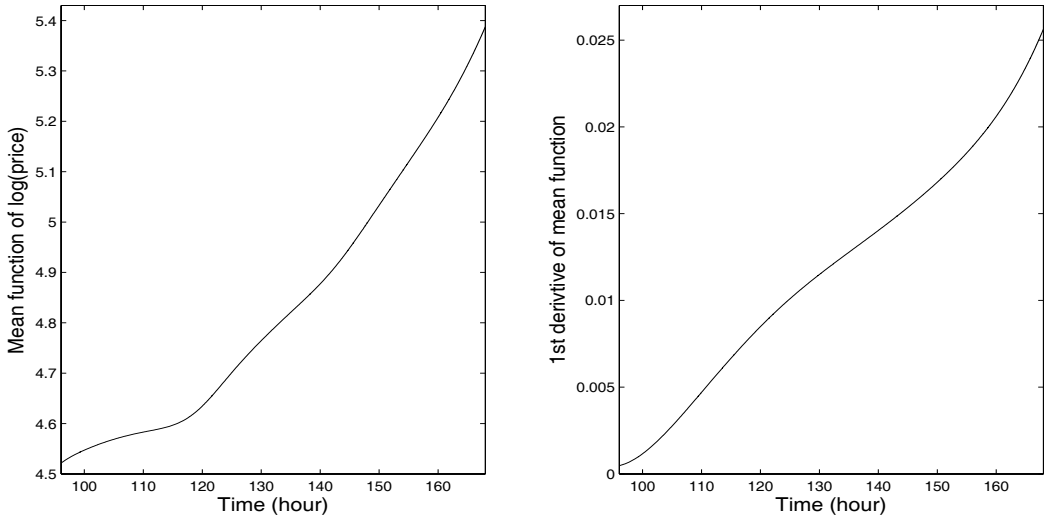

Figure 2: Smooth estimate of the mean function of log(Price) in the left panel and of its first derivative in the right panel.

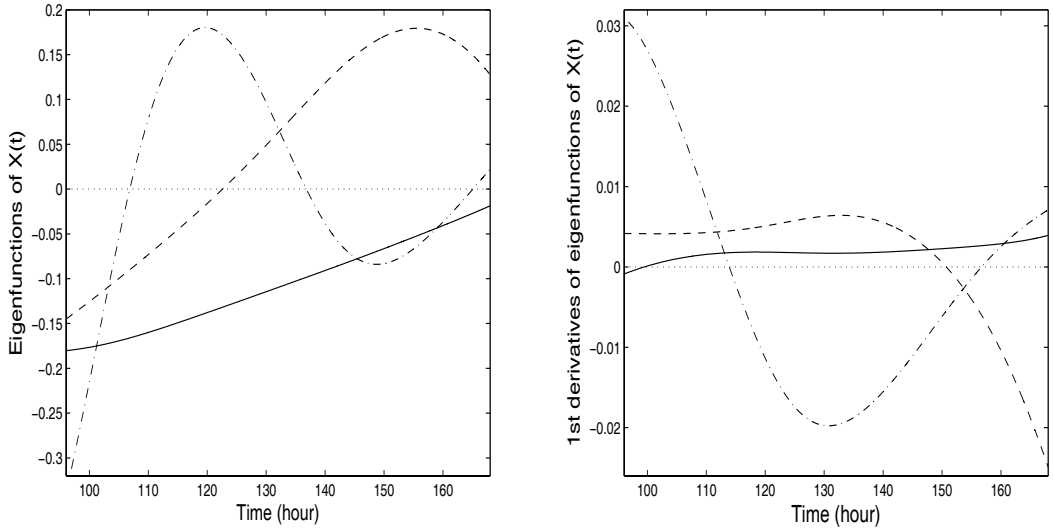

Figure 3: Smooth estimates of the first (solid), second (dashed) and third (dotted) eigenfunctions of process X (left panel) and of their derivatives (right panel), $\hat{\phi}_1^{(1)}$ (solid), $\hat{\phi}_2^{(1)}$ (dashed) and $\hat{\phi}_3^{(1)}$ (dash-dotted).

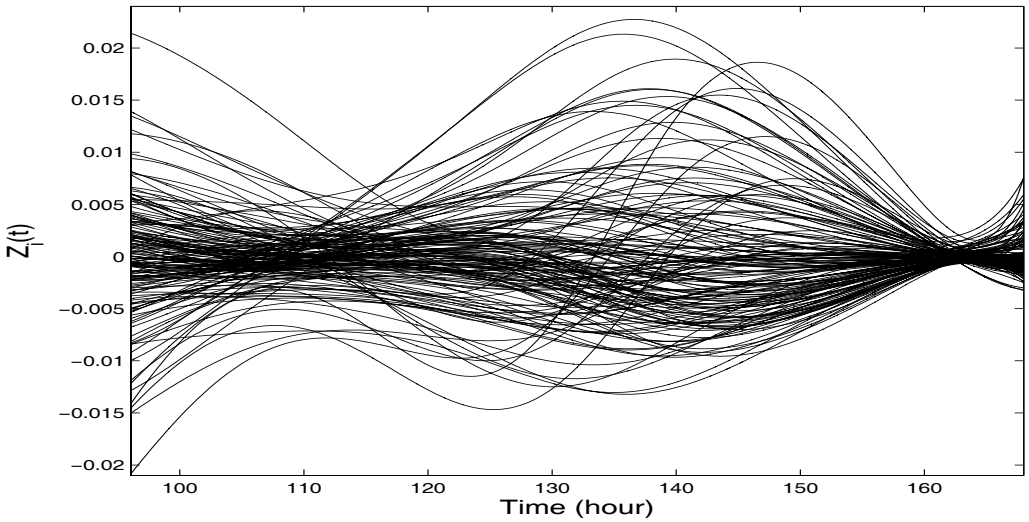

Figure 4: Smooth estimates for the trajectories of the drift process Z.

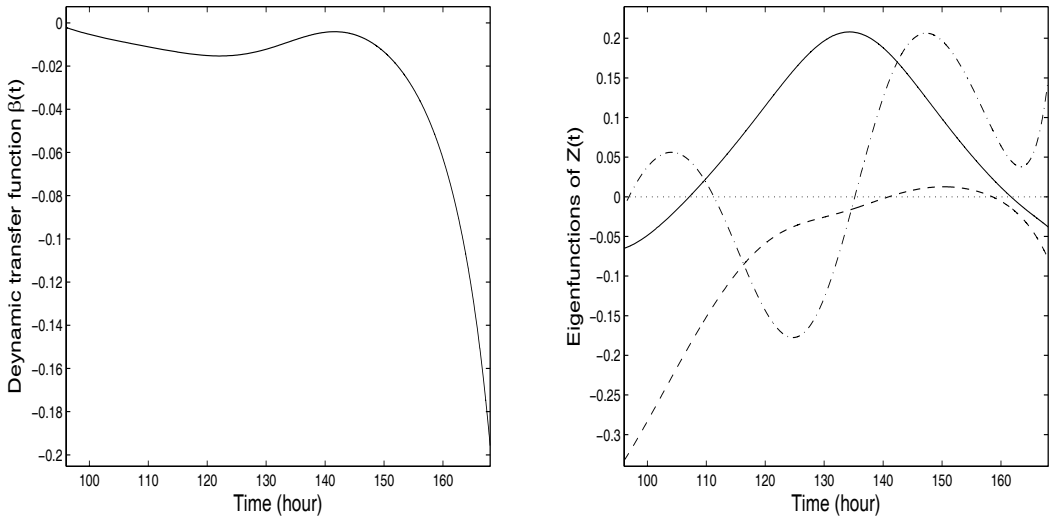

Figure 5: Left: Smooth estimate of the dynamic varying coefficient function β. Right: Smooth estimates of the first (solid), second (dashed) and third (dash-dotted) eigenfunction of Z based on (8).

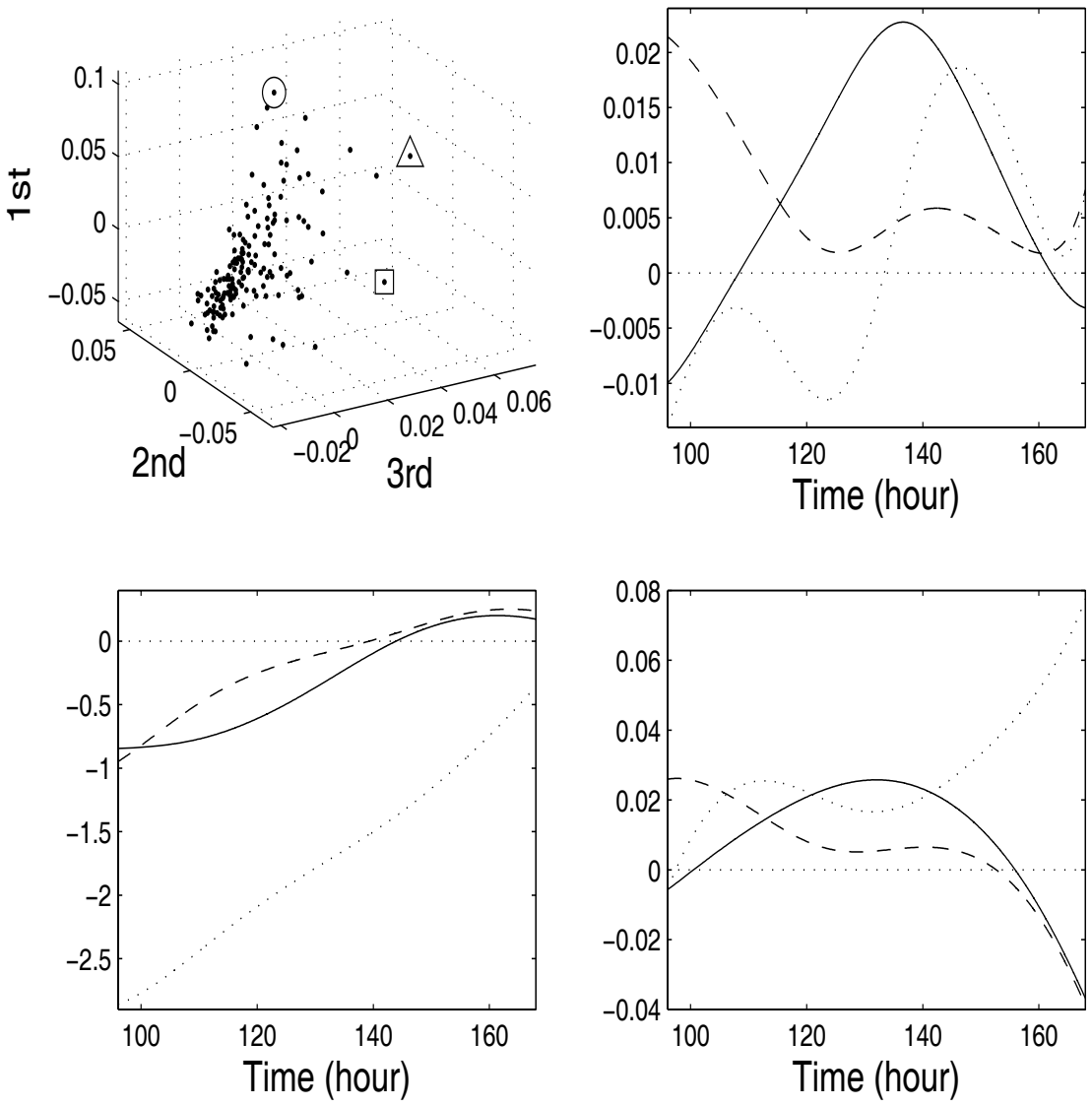

Figure 6: Top left: Point cloud corresponding to the three leading FPC scores of trajectories Z_i, where the point marked by a circle corresponds to the auction with the largest (in absolute value) first score, the point marked with a square to the auction with the largest second score and the point marked with a "triangle" to the auction with the largest third score, respectively. Top right: The trajectories Z_i of the drift process for these three auctions, where the solid curve corresponds to the trajectory of the "circle" auction, the dashed curve to the "square" auction and the dash-dotted curve to the "triangle" auction. Bottom left: Corresponding centered trajectories X_i. Bottom right: Corresponding centered trajectory derivatives $X_i^{(1)}$.

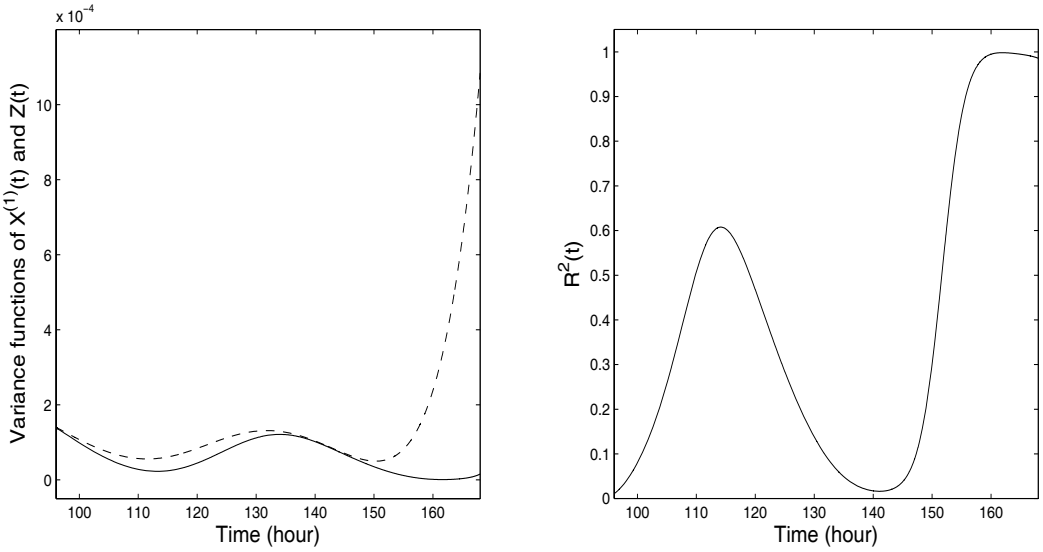

Figure 7: Left: Smooth estimates of the variance functions of $X^{(1)}(t)$ (dashed) and $Z(t)$ (solid). Right: Smooth estimate of $R^2(t)$ (12), the variance explained by the deterministic part of the dynamic equation at time t.

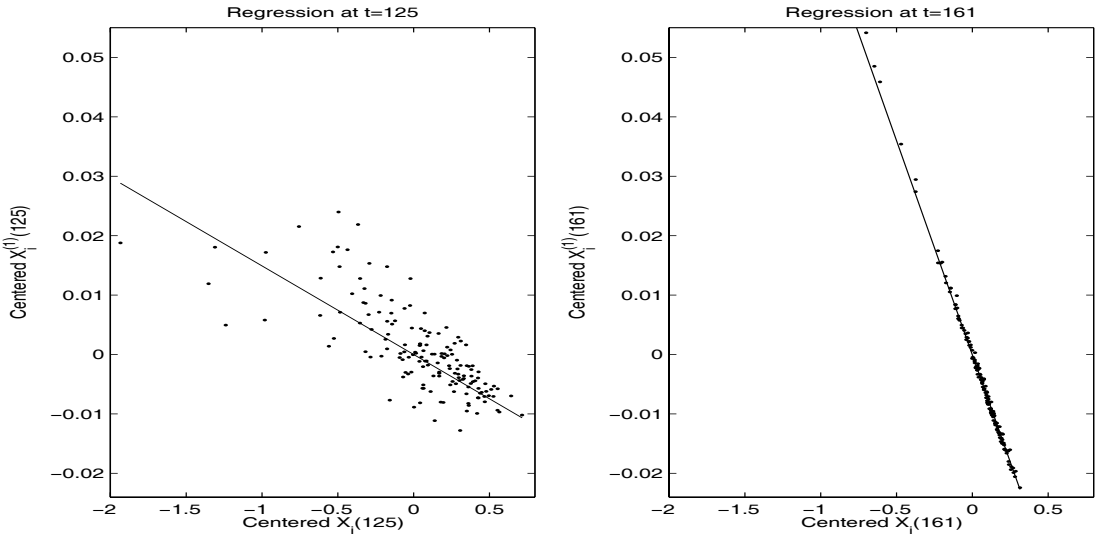

Figure 8: Regression of $X_i^{(1)}(t)$ on $X_i(t)$ (both centered) at $t = 125$ hours (left panel) and $t = 161$ hours (right panel), respectively, with regression slopes $\beta(125) = -.015$ and coefficient of determination $R^2(125) = 0.28$, respectively, $\beta(161) = -.072$ and $R^2(161) = 0.99$, demonstrating that the deterministic part (11) of the empirical differential equation (6) explains almost the entire variance of $X^{(1)}$ at $t = 161$ hours but only a fraction of variance at $t = 125$ hours.