EM I ICAL D AMIC F L GIT DI AL DATA

By Hans-Georg Müller[1] and Fang Yao[2]

University of California, Davis and University of Toronto

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.

1. I . Recently, there has been increasing interest in analyzing on-line auction data and in inferring 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, Jank and Shmueli (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, Müller and Wang (2005), Wang, Carroll and Lin (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, Taylor and Cumberland (1997), Staniswalis and Lee (1998), Rice (2004), Zhao, Marron and Wells (2004), Hall, Müller and Wang (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, Li and Huang (2008), Wang et al. (2008)]. 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 nonrandom 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   is small relative to the variance of the process   , 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, Müller and Wang (2005)], the dynamics of which is of interest. Given realizations   of the underlying process   on a domain $\mathcal{T}$ and   of an integer-valued bounded random variable   , we assume that   measurements   ,   = 1 ...   = 1 ...   , are obtained at random times   , according to

$$(1) \qquad\qquad\qquad = \quad (\quad) + \varepsilon \qquad\qquad \in \mathcal{T}$$

where ε are zero mean i.i.d. measurement errors, with var(ε) = ^{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, Pousse and Romain (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. Em ___ ___ m_ _.

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 $\mathrm{cov}\{X(s), X(t)\} = G(s,t)$, $s,t \in \mathcal{T}$, which is smooth, symmetric and nonnegative definite. Using G as kernel in a linear operator leads to the Hilbert–Schmidt operator $(\mathbf{G}f)(t) = \int_{\mathcal{T}} G(s,t) f(s)\,ds$. We denote the ordered eigenvalues (in declining order) of this operator by $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ 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 $G(s,t) = \sum_{k=1}^{\infty} \lambda_k \times \phi_k(s)\phi_k(t)$ and the trajectories generated by the process satisfy the Karhunen–Loève representation [Grenander (1950)] $X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t)$. Here the $\xi_k = \int_{\mathcal{T}} \{X(t) - \mu(t)\}\phi_k(t)\,dt$, $k = 1, 2, \ldots$, $j = 1, \ldots$, are the functional principal components (FPCs) of the random trajectories X. The ξ_k are uncorrelated random variables with $E(\xi_k) = 0$ and $E\xi_k^2 = \lambda_k$, with $\sum \lambda_k < \infty$. Upon differentiating both sides, one obtains

$$(2) \qquad X^{(1)}(t) = \mu^{(1)}(t) + \sum_{k=1}^{\infty} \xi_k \phi_k^{(1)}(t)$$

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 $\int G(s,t) \times \phi_k(s)\,ds = \lambda_k \phi_k(t)$, under the constraint of orthonormality. Under suitable regularity conditions, one observes

$$(3) \qquad \begin{aligned} -\int_{\mathcal{T}} G(s,t)\,\phi_k(s)\,ds &= -\lambda_k \phi_k(t) \\ \phi_k^{(1)}(t) &= \frac{1}{\lambda_k}\int_{\mathcal{T}} \frac{\partial}{\partial t} G(s,t)\,\phi_k(s)\,ds \end{aligned}$$

and suitable estimates $^{(1)}$ for eigenfunction derivatives, then directly decomposing $\text{cov}\{\,^{(1)}(\ )\quad^{(1)}(\ )\}$ into eigenfunctions and eigenvalues. This leads to $\text{cov}\{\,^{(1)}(\ )\quad^{(1)}(\ )\} = \sum_{=1}^{\infty}\,_{1}\quad_{1}(\ )\quad_{1}(\ )$ and the Karhunen–Loève representation $^{(1)}(\ ) = \,^{(1)}(\ ) + \sum_{=1}^{\infty}\,_{1}\quad_{1}(\ )$, with orthonormal eigenfunctions $_{1}$ [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 $\{\ (\ ) - \ (\ )\quad^{(1)}(\ ) - \,^{(1)}(\ )\}$ at all points $\in \mathcal{T}$, so that

$$
(5) \qquad
\begin{pmatrix} ^{(1)}(\ ) - \,^{(1)}(\ ) \\ (\ ) - \ (\ ) \end{pmatrix}
= \begin{pmatrix} \sum_{=1}^{\infty} \quad^{(1)}(\ ) \\ \sum_{=1}^{\infty} \quad (\ ) \end{pmatrix}
$$

$$
\sim \;_{2}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} \sum_{=1}^{\infty} \quad^{(1)}(\ )^2 & \sum_{=1}^{\infty} \quad^{(1)}(\ )\quad(\ ) \\ \sum_{=1}^{\infty} \quad^{(1)}(\ )\quad(\ ) & \sum_{=1}^{\infty} \quad(\ )^2 \end{pmatrix} \right).
$$

This joint normality immediately implies a "population" differential equation of the form $\{\,^{(1)}(\ ) - \,^{(1)}(\ ) |\ (\ )\} = \ (\ )\{\ (\ ) - \ (\ )\}$, 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 . This goal necessitates inclusion of a stochastic term which leads to an empirical stochastic differential equation that governs the dynamics of individual trajectories .

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

$$
(6) \qquad ^{(1)}(\ ) - \,^{(1)}(\ ) = \ (\ )\{\ (\ ) - \ (\ )\} + \ (\ ) \qquad \in \mathcal{T}
$$

where

$$
(7) \qquad
\begin{aligned}
(\ ) &= \frac{\text{cov}\{\,^{(1)}(\ )\quad(\ )\}}{\text{var}\{\ (\ )\}} = \frac{\sum_{=1}^{\infty} \quad^{(1)}(\ )\quad(\ )}{\sum_{=1}^{\infty} \quad(\ )^2} \\
&= \frac{1}{2} — \log[\text{var}\{\ (\ )\}] \qquad \in \mathcal{T}
\end{aligned}
$$

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

$$
(\ ) = \sum_{=1}^{\infty} \ ^{(1)}( ) \ ^{(1)}( ) - \ ( )\sum_{=1}^{\infty} \ ( ) \ ^{(1)}( )
$$

(8)

$$
- \ ( )\sum_{=1}^{\infty} \ ^{(1)}( ) \ ( ) + \ ( ) \ ( )\sum_{=1}^{\infty} \ ( ) \ ( ).
$$

Equation (6) provides a first order linear differential equation which includes a time-varying linear coefficient function   ( ) and a random drift process   ( ). The process   "drives" the equation at each time . 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 var(( )) determines the importance of the role of the stochastic drift component.

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

$$
( ) = \sum_{=1}^{\infty} \sqrt{\frac{ }{2 \ ^3}} (2 \ -1) \ \int_0 \sin\left\{\frac{(2 \ -1) }{2}\right\}\{ \ ^{(1)}( ) - \ ( ) \ ( )\} \ ( ).
$$

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 (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:

$$
( ) = \ ( ) + \{ \ ( ) - \ ( )\}
$$

(9)

$$
+ \int \ ( )\{ \ ( ) - \ ( )\} \ + \int \ ( )
$$

for any $\in \mathcal{T} \ 0 \le$

in the sense that   is the solution of both equations. For a domain with left end-point at time 0, setting $= 0$ in (9) then defines a classical initial value problem. Given a trajectory of the drift process   and a varying coefficient function   , one may obtain a solution for   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   for which the variance function var{ ( )} is "small." From (7) and (8) one finds

$$( ) = \mathrm{var}\{\ (\ )\}$$

$$= \left(\mathrm{var}\{\ ^{(1)}(\ )\}\,\mathrm{var}\{\ (\ )\} - [\mathrm{cov}\{\ ^{(1)}(\ )\quad (\ )\}]^2\right)\ \mathrm{var}\{\ (\ )\}$$

(10)

$$= \left(\sum_{=1}^{\infty}\ (\ ^{(1)}(\ ))^2 \sum_{=1}^{\infty}\quad \text{sc -289738 -1.061 Td(k)Tj-0.0002 Tc 10.9589 0 00 9.}\right.$$

and $\beta(t)$ 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, that is, the "coefficient of determination"

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

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

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

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)

$$R^2(t) = \Big[\sum_k \lambda_k k \cos(2k\pi t)\sin(2k\pi t)\Big]^2$$

$$\Big/\Big[\sum_k \lambda_k(\cos(2k\pi t))^2\sum_k\lambda_k k^2(\sin(2k\pi t))^2\Big] \qquad t \in [0, 1].$$

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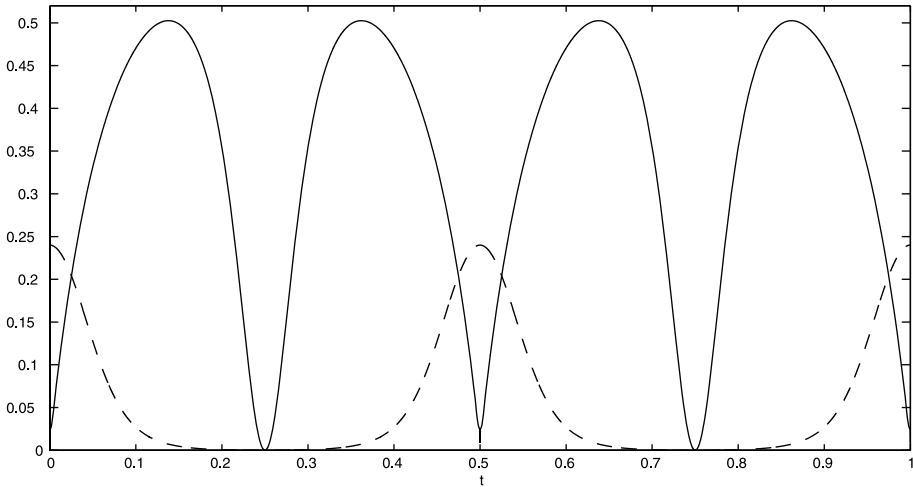
FIG. 1. *"Coefficient of determination" functions* $\rho^2(\ )$ (12), (13), *quantifying the fraction of variance explained by the deterministic part of the dynamic equation* (6), *illustrated for the trigonometric basis* $\{\sqrt{2}\cos(2\pi\ )\ =1\ 2\ \ldots\}$ *on* [0 1] *and eigenvalue sequences*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

asymptotic consistency results focus on $\ ^2$ 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 $\ ^2$ 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 and eigenvalues as their order increases and the large sample behavior of the bandwidths $\ _0\ _1$ for the estimation of the mean function and its first derivative $\ ^{(1)}(\ )$, and $\ _0\ _1$ 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 for the th trajectory is required to satisfy $(\ \geq 2)\ \ 0$.

Specifically, for the observations $(\ \ \ )$, $\ =1\ \ldots\ $, $\ =1\ \ldots\ $, made for the th trajectory, we require that:

(A1)   are random variables with $\overset{i.i.d.}{\sim}$, where   is a bounded positive discrete random variable with and $\{\ \geq 2\}\ \ 0$, and $(\{\ \in\ \}\ \{\ \ \ \in\ \})$ are independent of , for $\ \subseteq \{1\ \ldots\ \}$.

Writing $\mathbf{T}\ =(\ _1\ \ldots\ \ )$ and $\ =(\ _1\ \ldots\ \ )$, the triples $\{\mathbf{T}\ \ \}$ are assumed to be i.i.d. For the bandwidths used in the smoothing steps for $\ (\ )$

and $^{(1)}(\ )$ in (21), $(\ )$ and $^{(1\ 0)}(\ )$ in (22), we require that, as $\to \infty$,

(A2) $\max(\ _0\ _1\quad _0\quad _1) \to 0, \qquad _0 \to \infty, \quad ^3\ _1 \to \infty, \quad ^2\ _0 \to \infty,$
$\qquad ^4\ _1 \to \infty.$

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

(14) $\qquad\qquad _1 = \ _1 - \ _2 \quad = \min_{\leq}(\ _{-1} - \quad - \ _{+1}) \qquad \geq 2.$

For the kernels used in the local linear smoothing steps and underlying density and moment functions, we require assumptions (B1) and (B2) in the Appendix. Denote the ^{2} norm by $\|\ \| = \{\int_{\mathcal{T}} \ ^2(\ )\quad\}^{1\ 2}$, the Hilbert–Schmidt norm by $\|\Phi\| = \{\int_{\mathcal{T}}\int_{\mathcal{T}}\{\Phi^2(\ )\quad\quad\}^{1\ 2}$ and also define $\|\Phi\|^2 = \{\int_{\mathcal{T}} \Phi^2(\ )\quad\}^{1\ 2}$.

The following result provides asymptotic rates of convergence in the ^{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 $^{(1)}$, with constants and rates that hold uniformly in the order ≥ 1.

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

$$\|\ ^{(\ )} - \ ^{(\ )}\| = \left(\frac{1}{\sqrt{\ ^{2\ +1}}} + \ ^2\ \right)$$

(15)

$$\|\ ^{(\ 0)} - \ ^{(\ 0)}\| = \left(\frac{1}{\sqrt{\ ^{\ +1}}} + \ ^2\ \right).$$

For $^{(1)}(\ )$ *corresponding to* $\quad$ *of multiplicity* 1,

(16)
$$\|\ ^{(1)}(\ ) - \ ^{(1)}(\ )\|$$
$$= \left(\frac{1}{\ }\left\{\frac{1}{\sqrt{\ ^2}\ _1} + \ ^2\ _1 + \frac{1}{\ }\left(\frac{1}{\sqrt{\ }\ _0} + \ ^2\ _0\right)\right\}\right)$$

where the $(\cdot)$ *term in* (16) *is uniform in* ≥ 1.

An additional requirement is that variances of processes $\quad$ and $^{(1)}$ are bounded above and below, since these appear in the denominators of various representations, for example, in (10) and (13),

(A3) $\inf_{\in \mathcal{T}}\ ^{(\ )}(\ ) \geq \quad 0$ and $\|\ ^{(\ )}\| \quad \infty$ for $= 0\ 1$,

implying that $\|\ ^{(\ )}\| \quad \infty$ by the Cauchy–Schwarz inequality. Define remainder terms

(17) $\qquad (\ ) = \sum_{= \ +1}^{\infty}\ \{\ ^{(\ )}(\ )\}^2 \qquad * \ (\ ) = \sum_{= \ +1}^{\infty}\ \ ^{(\ )}(\ )\ ^{(\ )}(\ );$

by the Cauchy–Schwarz inequality, $\| \quad ^* \quad \| \leq \| \qquad \|$.

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

$$\text{(A4)} \qquad = (\min\{\sqrt{\quad}^2{}_1 \quad {}^{-2}{}_1\})$$

$$\sum_{=1} {}^{-1} = (\min(\sqrt{\quad} \quad {}_0 \quad {}^{-2}{}_0\})$$

$$\max_{=0\,1} \| \qquad \| \to 0 \qquad \text{as} \quad \to \infty.$$

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

$$\text{(18)} \qquad = \{(\sqrt{\quad}^2{}_1)^{-1} + {}^2{}_1\} + \left(\sum_{=1} {}^{-1}\right)\{(\sqrt{\quad} {}_0)^{-1} + {}^2{}_0\}.$$

Note that $\operatorname{cov}\{ {}^{(1)}( ) \quad {}^{(1)}( )\} = {}^{(1\,1)}(\quad) = \sum_{=1}^{\infty} {}^{(1)}( ) {}^{(1)}( )$ with corresponding plug-in estimate $\hat{}^{(1\,1)}(\quad) = \sum_{=1} \hat{} \hat{}^{(1)}( ) \hat{}^{(1)}( )$, where $=$ $( )$ is the included number of eigenfunctions. The plug-in estimate for $( )$ is based on (7) and given by $\hat{} ( ) = \sum_{=1} \hat{} \hat{}^{(1)}( ) \hat{} ( ) \sum_{=1} \hat{} \hat{} ( )^2$ and analogously the plug-in estimate $\hat{}$ of is based on representation (8), using the estimate $\hat{}$. In a completely analogous fashion one obtains the estimates $\hat{}^2( )$ of $^2( )$ from (13) and $\hat{} ( )$ of the variance function $( ) = \operatorname{var}(( ))$ of the drift process from (10). The ^{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),

$$\text{(19)} \qquad \begin{aligned} \| \hat{}^{(1\,1)} - {}^{(1\,1)} \| &= (+ \| {}^*{}_1 \|) \\ \| \hat{}^{(1\,1)} - \end{aligned}$$

this result. To see this, one may use $\sup_{\geq 1} |\hat{} - | = (\| \hat{} - \|)$ and $\| \hat{} - \| = {}^{*-1} (\| \hat{} - \|)$ where $\hat{}$ is any estimate of [Bosq (2000)]. Here the $(\cdot)$ terms are uniform in and $_1^* = {}_1 - {}_2$ $^* = \min_{\leq} ({}_{-1} - - {}_{+1})$ for ≥ 2.

4. A

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 are sampled at bid arrival times , where $= 1 \ldots 156$ refers to the auction index and $= 1 \ldots$ to the total number of bids submitted during the 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 uses the information gained across the entire sample of auctions. Only the latter analysis involves the stochastic drift term in the stochastic differential equation (6). We begin by reviewing the population level analysis, which is characterized by the deterministic part of (6), corresponding to the equation $({}^{(1)}( ) - {}^{(1)}( )| ( ) - ( )) =$

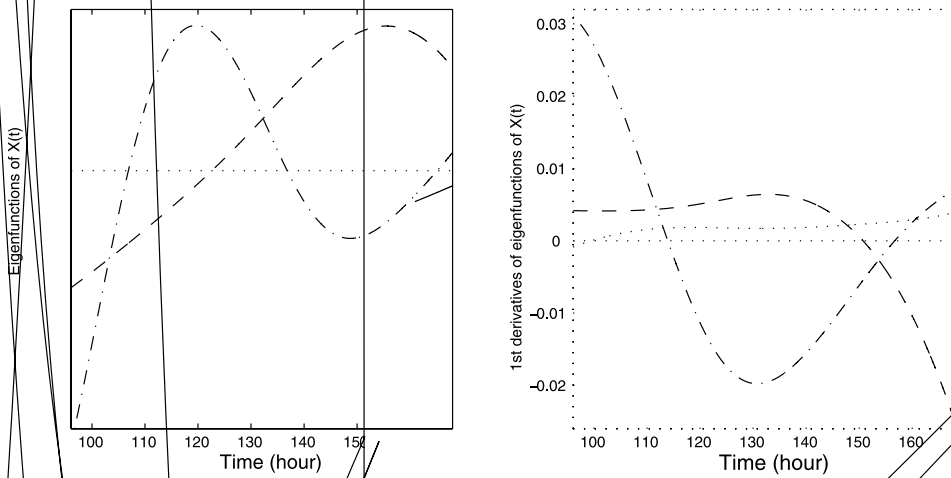110   120   130   140   150   160
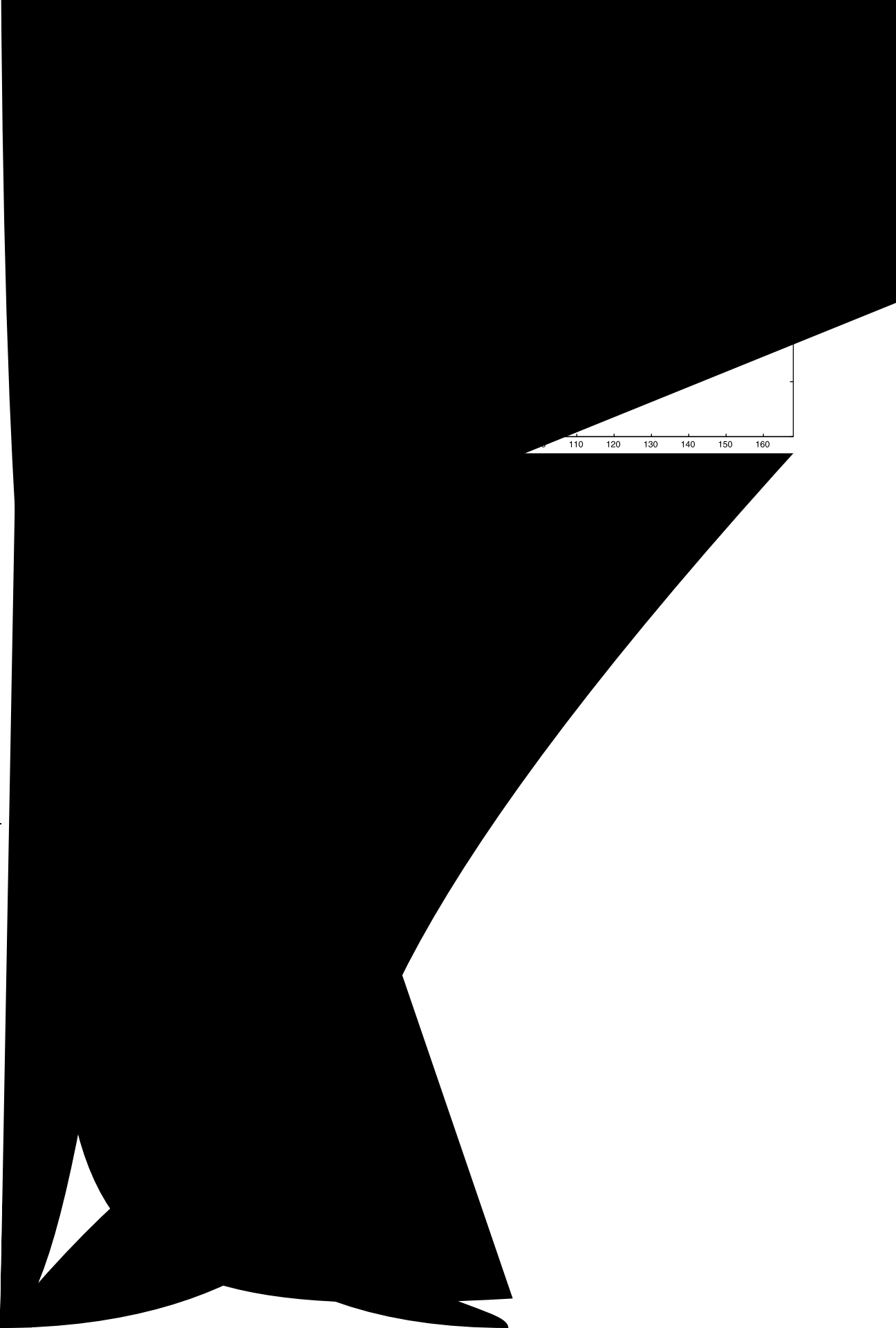
FIG. 3. *Smooth estimates of the first* (*solid*), *second* (*dashed*) *and third* (*dotted*) *eigenfunctions of process* (*left panel*) *and of their derivatives* (*right panel*), $\hat{\rho}_1^{(1)}$ (*solid*), $\hat{\rho}_2^{(1)}$ (*dashed*) *and* $\hat{\rho}_3^{(1)}$ (*dash-dotted*).

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 toward 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 , according to (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 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 is negative, as is the case for the auction data throughout the entire time domain, one may interpret the population equation ($^{(1)}($) − $^{(1)}($)| ( ) − ( )) = ( ){ ( ) − ( )} as indicating "dynamic regression to the mean." By this we mean the following: when a trajectory value at a current time falls above (resp., below) the population mean trajectory value at , then the conditional mean derivative of the trajectory at 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 toward 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 toward levels way above or below the mean trajectory. This self-stabilization feature gets stronger toward 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 (from the perspective of a buyer) 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 , we obtain the trajectories of the drift process. These trajectories are presented in Figure 4 for the entire sample of auctions. They quantify the component of the derivative process $^{(1)}$ that is left unexplained by the varying coefficient function and linear part of the dynamic model (6). The trajectories 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.

−0.05   −0.02   0   0.02   0.04   0.06
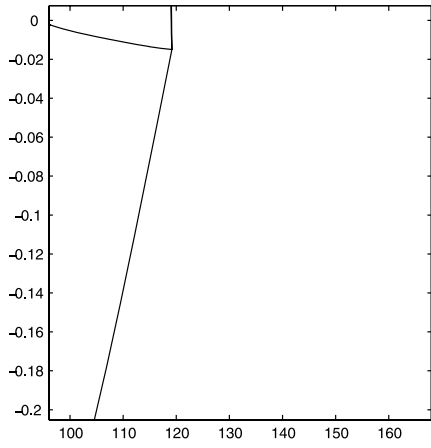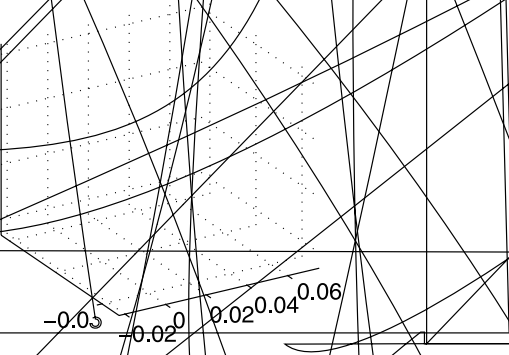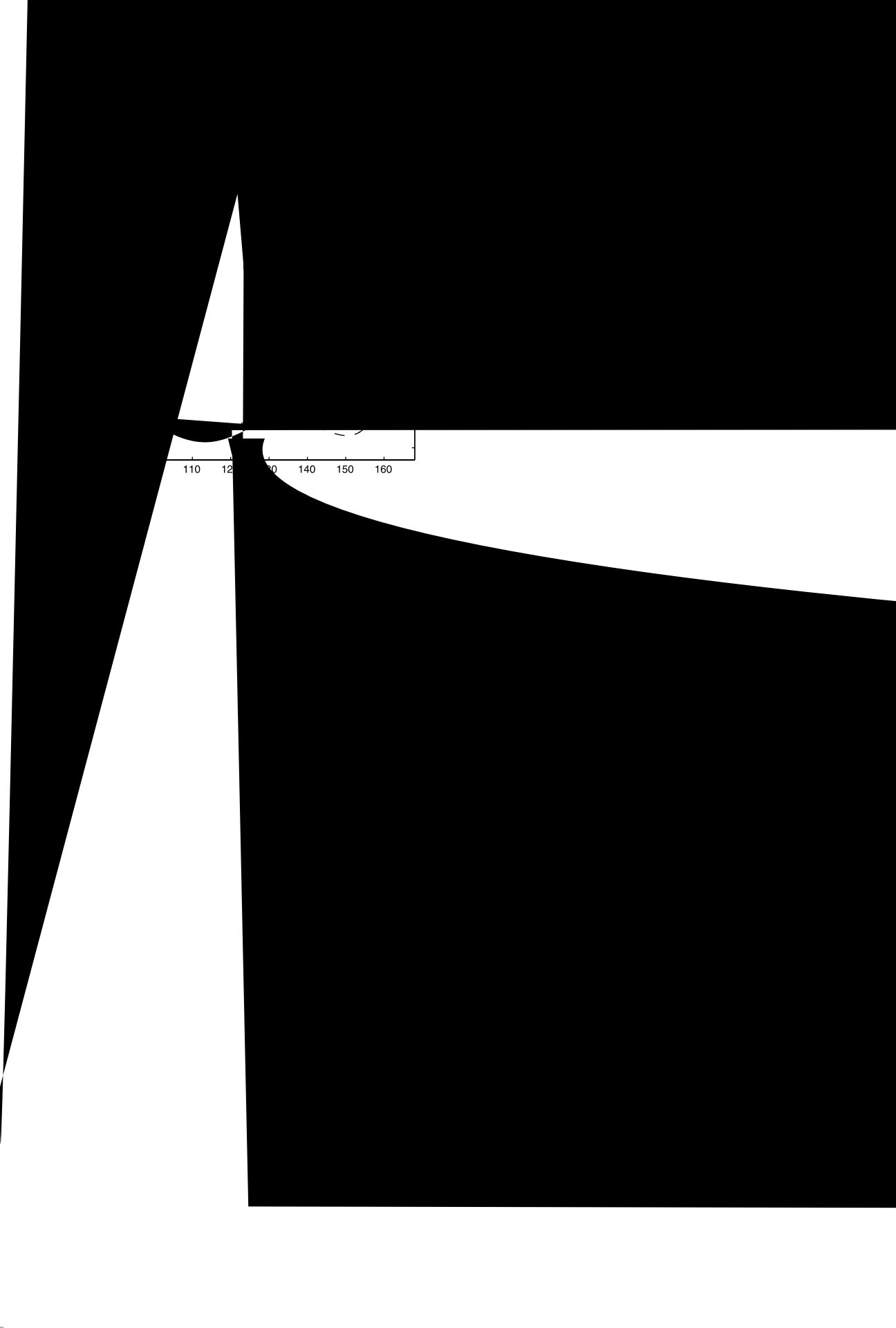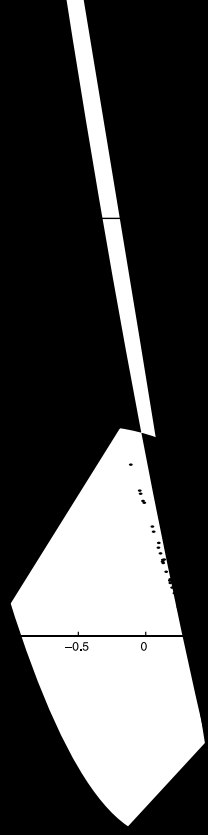
average log price level is closely associated with a below (above) average log price derivative. This implies that a seemingly very good (or bad) deal tends to be not quite so good (or bad) when the auction ends.

5. D   . The main motivation of using the dynamic system approach based on (6) is that it 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 (5) which we term "empirical differential equation" as it is determined by the data. Then the function ^{2}, quantifying the relative contribution of the drift process to the variance of $^{(1)}$, 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 us to construct predictions of log price trajectory derivatives from trajectory levels at the same . These predictions get better toward 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 toward the end of the auction.

For prediction purposes, drift processes 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 toward 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

for each with respect to $_{rr'}$ for $_{rr'} = 0 \ldots + 1$, from which one obtains $\hat{}^{(\ )}(\ ) = \hat{}\ (\ )\ !$ [Fan and Gijbels (1996)].

According to (3), we will also need estimates of $— (\ ) = {}^{(\ 0)}$. 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 mini-mize the pooled scatterplot of pairwise raw covariances

$$
\text{(22)} \quad \sum_{=1} \sum_{1 \le\ \ne\ \le}\ {}_2\!\left(\frac{-}{}\ \frac{-}{}\right)
$$

$$
\times \left\{\ (\qquad) - \left(\sum_{rr'=0}^{+1} {}_{1rr'}(\ -\ )^{rr'} + {}_{21}(\ -\ )\right)\right\}^2
$$

for fixed $(\ )$ with respect to $_{1rr'}$ and $_{21}$ for $_{rr'} = 1 \ldots + 1$, where $(\ ) = (\ - \hat{}(\ ))(\ - \hat{}(\ ))$, $\ne\ $, $_2$ is a kernel chosen as a bivariate density function, and is a bandwidth. This leads to $\hat{}^{(\ 0)}(\ ) = \hat{}_1(\ )\ !$.

The pooling that takes place in the scatterplots for estimating the derivatives of and of is the means to accomplish the borrowing of information across the sample, which is essential to overcome the limitations of the sparse sampling de-signs. We note that the case of no derivative $= 0$ is always included, and solving the eigenequations on the left-hand side of (3) numerically for that case leads to the required estimates $\hat{}_1\ \hat{}_2 \ldots$ of the eigenvalues and $\hat{}_1\ \hat{}_2 \ldots$ of the eigen-functions. The estimates $\hat{}_1^{(1)}\ \hat{}_2^{(1)} \ldots$ of the eigenfunction derivatives are then obtained from the right-hand side of (3), plugging in the estimates for eigenfunc-tions and eigenvalues, followed by a numerical integration step.

The plug-in estimates, $\hat{}\quad \hat{}\quad \hat{}\quad \hat{}^2$, are then obtained from the corre-sponding representations, (7), (8), (10), (13), by including leading components in the respective sums. While for theoretical analysis and asymptotic consistency one requires $= (\ ) \to \infty$, the number of included eigen-terms in practi-cal data analysis can be chosen by various criteria, for example, 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 $=$); the expected value for these elements includes the measurement error variance ^{2} in addition to the variance of the process. The dif-ference 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 ^{2} [Yao, Müller and Wang (2005)].

To obtain estimates for the derivatives of the trajectories , a realistic target is the conditional expectation $\{\ {}^{(\ )}(\ )|\ _1 \ldots\ \}$. It turns out that this condi-tional expectation can be consistently estimated in the case of Gaussian processes

by applying principal analysis by conditional expectation (PACE) [Yao, Müller and Wang (2005)]. For $= ((_1) \ldots ( )) , = (_1 \ldots) ,$ $\mu = ((_1) \ldots ( )) , \phi = ((_1) \ldots ( )) ,$ if and ε in (1) are jointly Gaussian, then by standard properties of the Gaussian distribution,

$$(23) \qquad (\mid) = \phi \Sigma^{-1}(- \mu)$$

where $\Sigma = \mathrm{cov}( ) = \mathrm{cov}( ) + ^2 .$ This implies $(^{()}( )\mid _1 \ldots$ $) = \sum_{=1}^{\infty} (\mid) ^{()}( ) = \{\sum_{=1}^{\infty} ^{()}( )\phi \}\Sigma^{-1}(- \mu) = 0 1.$ 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 .

A.3. A ⌣⌣⌣⌣⌣ **m** ⌣⌣⌣⌣⌣. Denote the densities of and $(_1 _2)$ by $_1( ), _2( ),$ and define an interior domain by $\mathcal{T} = [ ]$ with $\mathcal{T} = [- +]$ for some $0.$ Regularity conditions for the densities and the targeted moment functions as well as their derivatives are as follows, where $_1 _2$ are nonnegative integers:

(B1) $_1^{(5)}( )$ exists and is continuous on $\mathcal{T}$ with $( ) 0, \frac{^5}{_1 _2} _2( )$ exists and is continuous on $\mathcal{T}^2$ for $_1 + _2 = 5;$

(B2) $^{(5)}( )$ exists and is continuous on $\mathcal{T} , \frac{^5}{_1 _2} ( )$ exists and is continuous on $\mathcal{T}^2$ for $_1 + _2 = 5.$

We say that a bivariate kernel function $_2$ is of order (v), where v is a multi-index $v = (_1 _2)$, if

$$(24) \qquad \int\int ^{_1} ^{_2} _2( )$$
$$= \begin{cases} 0 & 0 \le _1 + _2 \quad _1 \ne _1 \quad _2 \ne _2, \\ (-1)^{\mid \mid} _1! _2! & _1 = _1 \quad _2 = _2, \\ \ne 0 & _1 + _2 = , \end{cases}$$

where $\mid v\mid = _1 + _2 .$ The univariate kernel $_1$ is said to be of order $( )$ for a univariate $= _1,$ if (24) holds with $_2 = 0$ on the right-hand side, integrating only over the argument on the left-hand side. For the kernel functions $_1, _2$ used in the smoothing steps to obtain estimates for $( )$ and $^{(1)}( )$ in (21) and for $( )$ and $^{(1\,0)}( )$ in (22) we assume

(B3) Kernel functions $_1$ and $_2$ are nonnegative with compact supports, bounded and of order $(0\,2)$ and $((0\,0)\,2)$, respectively.

The following lemma provides the weak ^{2} convergence rate for univariate and bivariate weighted averages defined below. For arbitrary real functions $: \Re^2 \to \Re$ and $^*: \Re^4 \to \Re$, define $\tilde{}( ) = \{ (^*) ( ) \Delta \Delta \Delta \quad \Delta \Delta$

$\{ \ ^*(\qquad)| \quad = \qquad = \}$, let $(\ ) = \ ^{\sim( )}(\ ) \ _1(\ )$ for a single index

and $\ _v^*(\ ) = \ _2(\ )\frac{|v|}{_1 \ _2} \ ^{\sim*}(\ )$ for a multi-index $v = (\ _1 \quad _2)$, and define the weighted kernel averages, employing bandwidths $_1 \quad _2$,

$$(25) \qquad \hat{}(\ ) = \frac{1}{(\ ) \ _1^{+1}} \sum_{=1} \sum_{=1} (\qquad) \ _1\left(\frac{-}{_1}\right)$$

$$\hat{}^*(\ ) = \frac{1}{\{ \ (\ -1)\} \ _2^{|v|+2}} \sum_{=1} \sum_{1\le \ \ne \ \le} \ ^*(\qquad \qquad) \ _2$$

$$(26)$$

$$\times \left(\frac{-}{_2} \quad \frac{-}{_2}\right).$$

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

(B2†) Derivatives $^{\sim( )}(\ )$

Gaussianity. Next observe $\mathrm{cov}\{\ (\ )\quad(\ )\}=\mathrm{cov}\{\ ^{(1)}(\ )-\ (\ )\ (\ )\ ^{(1)}(\ )-$ $(\ )\ (\ )\}$, from which one obtains the result by straightforward calculation. $\square$

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

$$\hat{}(\ )=\sum_{=1}\frac{1}{(\ )}\frac{1}{\ _1^{\ +1}}\sum_{=1}(\ \ \ )\ _1\left(\frac{-}{\ _1}\right)\mathbf{1}(\ \geq\ )$$

$$\equiv\sum_{=1}\frac{1}{(\ )}\hat{}\ (\ )$$

where $\mathbf{1}(\cdot)$ is the indication function. Note that for each , $\hat{}$ is obtained from an i.i.d. sample. Slightly modifying the proof of Theorem 2 in Hall (1984) for a kernel of order ( ) provides the weak convergence rate $\|\hat{}\ -\ (\ \geq\ )\|=$ $\{(\ _1^{2\ +1})^{-1}\ ^2+\ _1^{-}\}$. It is easy to check that $\sum_{=1}(\ \geq\ )=(\ )$, as is a positive integer-valued random variable. Therefore,

$$\|\hat{}-\ \|\leq\sum_{=1}\frac{(\ \geq\ )}{(\ )}\left\|\frac{\hat{}}{(\ \geq\ )}-\ \right\|=\left(\frac{1}{\sqrt{\ _1^{2\ +1}}}+\ _1^{-}\right).$$

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

$$\hat{}_{\boldsymbol{v}}^{*}=\frac{1}{\ _2^{|\boldsymbol{v}|+2}}\sum_{=1}\ ^{*}(\ \ \ \ \ \ \ )\ _2\left(\frac{-}{\ _2}\ \frac{-}{\ _2}\right)\mathbf{1}\{\ \geq\max(\ \ )\}$$

and then $\hat{}_{\boldsymbol{v}}^{*}=\sum_{1\leq\ \neq\ \leq}[\ \{\ (\ -1)\}]^{-1}\hat{}^{*}$. Similarly to the above, one has $\|\hat{}_{\boldsymbol{v}}^{*}-\ _{\boldsymbol{v}}\{\ \geq\max(\ \ )\}\|=\{(\ _2^{2|\boldsymbol{v}|+2})^{-1}\ ^2+\ _2^{-|\boldsymbol{v}|}\}$. Again it is easy to verify that $\{\ (\ -1)\}=\sum_{1\leq\ \neq\ \leq}\ \{\ \geq\max(\ \ )\}$. The triangle inequality for the $\ ^2$ distance entails $\|\hat{}^{*}-\ _{\boldsymbol{v}}\|=\{(\ _2^{2|\boldsymbol{v}|+2})^{-1}\ ^2+\ _2^{-|\boldsymbol{v}|}\}$. $\square$

PROOF OF THEOREM 2. Note that the estimators $\hat{}$, $\hat{}^{(1)}$, $\hat{}$ and $\hat{}^{(1\ 0)}$ 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 $\ ^2$ rates given in Lemma 1, then leads to the optimal $\ ^2$ weak convergence rates for $\hat{}$ and $\hat{}^{(\ 0)}$ in (15).

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

$$(29)\qquad|\hat{}\ -\ |\leq\|\hat{}-\ \|\qquad\|\hat{}\ -\ \|\leq2\sqrt{2}\ ^{-1}\|\hat{}-\ \|$$

where is defined in (14) and $\hat{}$ is an arbitrary estimate (or perturbation) of . Denote the linear operators generated from the kernels $\ ^{(1\ 0)}$ and $\hat{}^{(1\ 0)}$ by $\mathbf{G}^{(1\ 0)}$,

respectively, $\hat{\mathbf{G}}^{(1\,0)}$. Noting that $\quad \leq \quad$, one finds

$$\| \quad^{(1)} - \quad^{(1)} \| \leq \frac{1}{\hat{}} \| \hat{\mathbf{G}}^{(1\,0)\,\hat{}} - \mathbf{G}^{(1\,0)} \quad \| + \| \mathbf{G}^{(1\,0)} \quad \| \cdot \left| \frac{1}{\hat{}} - \frac{1}{} \right|$$

$$(30) \qquad \asymp \frac{1}{} \{ \| \quad^{\hat{}\,(1\,0)} - \quad^{(1\,0)} \| + \| \hat{} - \quad \| \} + \frac{| \hat{} - |}{{}^2}$$

$$\asymp \frac{1}{} \left\{ \| \quad^{\hat{}\,(1\,0)} - \quad^{(1\,0)} \| + \frac{1}{} \| \hat{} - \quad \| \right\}$$

which implies (16). $\square$

PROOF OF THEOREM 3. From (29) it is easy to see that $| \hat{} - | = \quad (\| \hat{} - \|)$, and from both (29) and (30) that $\| \hat{} - \| = \quad (\| \quad^{\hat{}(1)} - \quad^{(1)} \|)$ uniformly in $\quad$. One then finds that $\| \quad^{\hat{}\,(1\,1)\Delta}$

To study $\|\hat{\ } - \ \|$, we investigate the $\ ^2$ convergence rates of

$$_1 = \left\| \hat{\ } \sum_{=1} \hat{\ } \ \hat{\ } \ \hat{\ }^{(1)} - \sum_{=1}^{\infty} \ ^{(1)} \right\| \qquad _2 = \| \hat{\ } \ \hat{\ } \ \hat{\ } - \ \|$$

where $\ $ (resp., $\hat{\ }$) and $\ $ (resp., $\hat{\ }$) share the same argument, and we define $\hat{\ }(\ ) = \sum_{=1} \hat{\ } \ \hat{\ }(\ ) \ \hat{\ }(\ )$. In analogy to the above arguments, $_1 \asymp \| \hat{\ } - \ \| + \sum_{=1} \| \hat{\ }^{(1)} - \ ^{(1)} \| + \sqrt{\| \ _0 \| \cdot \| \ _1 \|}$, $_2 = \| \hat{\ } - \ \| + \sum_{=1} \| \hat{\ } - \ \| + \| \ _0 \|$. This leads to $\| \hat{\ } - \ \| = \ (\ + \| \ _0 \| + \| \ _1 \|)$. The same argument also applies to $\| \hat{\ } - \ \|$. Next we study $\| \hat{\ }^2 - \ ^2 \|$ and find that $\| \sum_{=1} \hat{\ } \ \hat{\ }^{(_1)} \hat{\ }^{(_2)} - \sum_{=1}^{\infty} \ ^{(_1)} \ ^{(_2)} \| \asymp \sum_{=1} (\| \hat{\ }^{(_1)} - \ ^{(_1)} \| + \| \hat{\ }^{(_2)} - \ ^{(_2)} \|) + \| \ _1 \| + \| \ _2 \| = \ (\ + \| \ _1 \| + \| \ _2 \|)$. Analogous arguments apply to $\| \hat{\ } - \ \|$, completing the proof. $\square$

REFERENCES

ASH, R. B. and GARDNER, M. F. (1975). *Topics in Stochastic Processes. Probability and Mathematical Statistics* **27**. Academic Press, New York. MR0448463

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, New York. MR1783138

DAUXOIS, J., POUSSE MR1783138 1 Tf Tc -13.3051e)24(w(T36(B)T.29.061 TcR1783138)Tj3.999 -1.221 T0650

JANK, W. and SHMUELI, G. (2005). Profiling price dynamics in online auctions using curve clustering. SSRN eLibrary. Working Paper RHS-06-004, Smith School of Business, Univ. Maryland.

JANK, W. and SHMUELI, G. (2006). Functional data analysis in electronic commerce research. *Statist. Sci.* **21** 155–166. MR2324075

JONES, M. C. and RICE, J. A. (1992). Displaying the important features of large collections of similar curves. *Amer. Statist.* **46** 140–145.

KATO, T. (1995). *Perturbation Theory for Linear Operators*. Springer, Berlin. MR1335452

KIRKPATRICK, M. and HECKMAN, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.* **27** 429–450. MR1009899

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.) 269–290. Wiley, New York. MR2414047

LIU, B. and MÜLLER, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics. *J. Amer. Statist. Assoc.* **104** 704–714. MR2541589

MAS, A. and MENNETEAU, L. (2003). Perturbation approach applied to the asymptotic study of random operators. In *High Dimensional Probability, III (Sandjberg, 2002). Progress in Probability* **55** 127–134. Birkhäuser, Basel. MR2033885

MAS, A. and PUMO, B. (2007). The ARHD model. *J. Statist. Plann. Inference* **137** 538–553. MR2298956

MAS, A. and PUMO, B. (2009). Functional linear regression with derivatives. *J. Nonparametr. Stat.* **21** 19–40. MR2483857

RAMSAY, J. (2000). Differential equation models for statistical functions. *Canad. J. Statist.* **28** 225–240. MR1777224

RAMSAY, J. O., HOOKER, G., CAMPBELL, D. and CAO, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 741–796. MR2368570

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

REDDY, S. K. and DASSR 7945(n)344((fe1(Dn)3J0.0S98 Tc 7.1731 0 0 7.1731 247.6547 32 -2883)29.69S)Tj-(HMUI

WANG, S., JANK, W., SHMUELI, G. and SMITH, P. (2008). Modeling price dynamics in ebay auctions using principal differential analysis. *J. Amer. Statist. Assoc.* **103** 1100–1118.

YAO, F. and LEE, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 3–25. MR2212572

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

ZHAO, X., MARRON, J. S. and WELLS, M. T. (2004). The functional data analysis view of longitudinal data. *Statist. Sinica* **14** 789–808. MR2087973

ZIEMER, W. (1989). *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*. Springer, New York. MR1014685

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616
USA
E-MAIL: mueller@wald.ucdavis.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
100 ST. GEORGE STREET
TORONTO, ON M5S 3G3
CANADA
E-MAIL: fyao@utstat.toronto.edu