Statistica Sinica **30** (2020), 1905-1924
doi:https://doi.org/10.5705/ss.202018.0321

JOINT MODELS FOR GRID POINT AND RESPONSE PROCESSES IN LONGITUDINAL AND FUNCTIONAL DATA

Daniel Gervini and Tyler J Baur

University of Wisconsin–Milwaukee

Abstract: The distribution of the grid points at which a response function is observed in longitudinal or functional data applications is often informative and not independent of the response process. Here, we propose a covariation model for estimating and making inferences about this interrelation, where we treat the data as replicated realizations of a marked point process. We derive the maximum likelihood estimators and the asymptotic distribution of the estimators. The behavior of the estimators is examined using simulations. Lastly, we apply the model to an online auction data set, and show that there is a strong correlation between bidding patterns and price trajectories.

Key words and phrases: Doubly-stochastic process, Karhunen–Loève decomposition, latent-variable model, Poisson process.

1. Introduction

In many statistical applications, the objects of analysis are samples of functions, $\{g_i(x) : i = 1, \ldots, n\}$. In general, these functions are measured at discrete points $\{x_{ij} : j = 1, \ldots, m_i\}$; thus, the observed data are actually given by $\{(x_{ij}, y_{ij}) : j = 1, \ldots, m_i,\ i = 1, \ldots, n\}$, with

$$y_{ij} = g_i(x_{ij}) + \eta_{ij}, \tag{1.1}$$

where η_{ij} is random noise. Longitudinal data often fit this framework (Rice (2004); Müller (2008)).

Functional data analyses focus on samples of functions $g_i(x)$, which are usually recovered from raw data using some form of smoothing (James, Hastie and Sugar (2000); Ramsay and Silverman (2005, Chap. 3); Yao, Müller and Wang (2005)). In general, the distribution of the grid points $\{x_{ij}\}$ is considered noninformative. However, there are situations in which the distribution of x_{ij} may be informative in its own right.

Consider, for example, the bid price trajectories shown in Figure 1. These

are bid prices for Palm M515 personal digital assistants (PDAs) on week-long eBay auctions that took place between March and May of 2003. Bidding activity tends to concentrate at the beginning and at the end of the auctions, in patterns that have been called "early bidding" and "bid sniping," respectively. Earlier analyses of these data (Shmueli and Jank (2005); Jank and Shmueli (2006, 2010)) studied the dynamics of the process using derivatives of the bid price curves. More recently, Wu, Müller and Zhang (2013) and Arribas-Gil and Müller (2014) investigated the bid time process itself. However, a joint modeling of the bid time process and the bid price curves has not yet been attempted, and there are reasons to believe these processes are not independent. For example, it is suspected that items with prices below the mean are more likely to experience bid sniping. To answer such questions, it is necessary to jointly model the bid time process $\{x_{ij}\}$ and the bid price process $\{y_{ij}\}$.

The approach we present in this paper considers the data $\{(x_{ij}, y_{ij})\}$ as n independent realizations of a marked point process. Using common point-process terminology, for each subject i, x_{ij} is viewed as an observation of a point process, and y_{ij} as a corresponding "mark" (Cox and Isham (1980); Møller and Waagepetersen (2004); Baddeley (2007); Streit (2010)). Note, however, that not all marked point processes arise as discretizations of smooth functions, as that in model (1.1) does. Therefore, the methods we propose here are specifically intended for functional and longitudinal data applications. To avoid confusion with terminology, we do not refer to the m_i observations for each subject i as "replications", as is often done in the point-process literature. Instead, we consider the whole set $\{(x_{ij}, y_{ij}) : j = 1, \ldots, m_i\}$ for each i as a single realization of the process, and the n sets as the n replications.

As pointed out by Guan and Afshartous (2007); Møller, Ghorbani and Rubak (2016), the literature on modeling marked point processes is limited and restricted to the single-replication scenario, focusing on simple summary statistics of the processes and on testing broad, generic hypotheses, such as independent marking (Guan and Afshartous (2007); Myllymäki et al. (2017); see also Baddeley (2010, Sec. 21.7)). However, the availability of replications allows us to estimate the correlations between the intensity functions of the point process $\{x_{ij}\}$ and the Karhunen–Loève components of the response process $\{y_{ij}\}$, which is not possible in a single-replication scenario. Regression models in point-process contexts have been proposed (Barrett et al. (2015); Rathbun and Shiffman (2016)); however, their goal is to incorporate covariates into intensity function models. Similarly, Scheike (1997) related longitudinal data to marked point processes, but his goal

was to model the conditional distribution of the time points given past observations. None of the aforementioned works jointly model the time points and the response processes, which is the goal of this study.

2. Latent Variable Model

A point process X is a random countable set in a space $\mathscr{S}$, where $\mathscr{S}$ is usually $\mathbb{R}$ for temporal processes, or $\mathbb{R}^2$ for spatial processes (Møller and Waagepetersen (2004, Chap. 2); Streit (2010, Chap. 2)). When each point $x \in X$ is accompanied by a random feature Y_x in some space $\mathscr{M}$, $Z = \{(x, Y_x) : x \in X\}$ is called a marked point process. As mentioned in Section 1, we are interested in the specific situation where Y_x follows the model

$$Y_x = g(x) + \eta_x, \tag{2.1}$$

with $g : \mathscr{S} \to \mathscr{M}$ the function of interest, and η_x denoting random noise. We consider only $\mathscr{M} = \mathbb{R}$ in this paper, but extensions to the multivariate case $\mathscr{M} = \mathbb{R}^k$ are straightforward.

A point process X is locally finite if $\#(X \cap B) < \infty$ with probability one, for any bounded $B \subseteq \mathscr{S}$. For a locally finite process the count function $N(B) = \#(X \cap B)$ can be defined, and $Z_B := \{(x, Y_x) : x \in X \cap B\}$ is a finite set, $Z_B = \{(x_1, y_1), \ldots, (x_{N(B)}, y_{N(B)})\}$. A Poisson process is a locally finite process, for which there exists a locally integrable function $\lambda : \mathscr{S} \to [0, \infty)$, called the intensity function, such that (i) $N(B)$ has a Poisson distribution with rate $\int_B \lambda(t)dt$, and (ii) for disjoint sets $B_1, \ldots, B_k$, the random variables $N(B_1), \ldots, N(B_k)$ are independent. A consequence of (i) and (ii) is that the conditional distribution of the points in $X \cap B$ given $N(B) = m$, is the distribution of m independent and identically distributed (i.i.d.) observations with density $\lambda(t)/\int_B \lambda$.

For replicated point processes, a single intensity function λ rarely provides an adequate fit for all replications. It is more reasonable to assume that λ is subject-specific, and then to treat it as a random effect. Such processes are called doubly stochastic processes, or Cox processes (Møller and Waagepetersen (2004, Chap. 5); Streit (2010, Chap. 8)). A doubly stochastic process is a pair (X, Λ), where $X|\Lambda = \lambda$ is a Poisson process with intensity function λ, and Λ is a random function that takes values on the space $\mathscr{F}$ of nonnegative locally integrable functions on $\mathscr{S}$. Then, the n replications of the point process can be seen as i.i.d. realizations of a doubly stochastic process (X, Λ), where X is

observable, but Λ is not. Similarly, for g in (2.1), we assume there is a process G, such that $Y \mid (X, G = g)$ follows model (2.1). Then, the n replications of the marked point process can be seen as i.i.d. realizations of (X, Y, Λ, G), where X and Y are observable, but Λ and G are not.

Our main goal is to study the relationship between the intensity process Λ that generates x and the response process G that generates y. To this end, we assume that G follows a finite Karhunen–Loève decomposition

$$G(x) = \nu(x) + \sum_{k=1}^{p_2} v_k \psi_k(x), \tag{2.2}$$

where $\{\psi_k\}$ are orthonormal functions in $L^2(\mathscr{S})$, and $\{v_k\}$ are uncorrelated zero-mean random variables. Any stochastic process in $L^2(\mathscr{S})$ with finite variance can be decomposed as in (2.2), with a possibly infinite p_2 (Ash and Gardner (1975, Chap. 1.4)). However, because we are interested in smooth processes, for practical purposes, it is sufficient to consider only finite p_2.

A similar decomposition for Λ would be problematic, owing to the nonnegativity constraint. A nonnegative decomposition was proposed by Gervini (2016). However, for simplicity, we use an alternative approach. We decompose the logarithm of Λ, which is unconstrained,

$$\log \Lambda(x) = \mu(x) + \sum_{k=1}^{p_1} u_k \phi_k(x), \tag{2.3}$$

where $\{\phi_k\}$ are orthonormal functions in $L^2(\mathscr{S})$, and $\{u_k\}$ are uncorrelated zero-mean random variables.

The association between Λ and G is then determined by the association between the component scores $\mathbf{u} = (u_1, \ldots, u_{p_1})$ and $\mathbf{v} = (v_1, \ldots, v_{p_2})$ in (2.3) and (2.2), respectively. As a working model, we assume that $(\mathbf{u}, \mathbf{v})$ follows a joint multivariate normal distribution with mean zero and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \operatorname{diag}(\sigma_u^2) & \mathbf{\Sigma}_{uv} \\ \mathbf{\Sigma}_{uv}^T & \operatorname{diag}(\sigma_v^2) \end{pmatrix},$$

where σ_u^2 and σ_v^2 are the variances of the elements of $\mathbf{u}$ and $\mathbf{v}$, respectively. The error term η in (2.1) is assumed $N(0, \sigma_\eta^2)$ and independent of $\mathbf{u}$ and $\mathbf{v}$. The parameter of interest is the cross-covariance matrix $\mathbf{\Sigma}_{uv}$; the others are mostly nuisance parameters.

The signs of the component scores are not identifiable, because $-u_k$ and $-\phi_k(x)$ satisfy the same model as u_k and $\phi_k(x)$, respectively; similar reasoning holds for v_k and ψ_k. Consequently, the signs of $\Sigma_{uv,kl} = \mathrm{cov}(u_k, v_l)$ are not identifiable either, and can be chosen for convenience of interpretation for any given application.

To facilitate the estimation of the functional parameters μ, ϕ_k, ν, and ψ_k, we use semiparametric basis-function expansions. As basis functions, we can use, for instance, B-splines if $\mathscr{S} = \mathbb{R}$, or normalized Gaussian radial kernels if $\mathscr{S} = \mathbb{R}^2$; other families are possible, and perhaps better in some cases, such as simplicial bases for bivariate functions on irregular domains. We call this family $\mathscr{B}$. Let $\gamma(x)$ be the vector of basis functions $\{\gamma_1, \ldots, \gamma_q\}$ of $\mathscr{B}$, with $\gamma_j : \mathscr{S} \to \mathbb{R}$. We assume, then, that $\mu(x) = \mathbf{c}_0^T\gamma(x)$, $\phi_k(x) = \mathbf{c}_k^T\gamma(x)$, $\nu(x) = \mathbf{d}_0^T\gamma(x)$, and $\psi_k(x) = \mathbf{d}_k^T\gamma(x)$.

The model parameters are collected, for simplicity, in a single vector

$$\theta = (\mathrm{vec}\boldsymbol{\Sigma}_{uv}, \mathbf{c}_0, \ldots, \mathbf{c}_{p_1}, \mathbf{d}_0, \ldots, \mathbf{d}_{p_2}, \sigma_u^2, \sigma_v^2, \sigma_\eta^2). \tag{2.4}$$

The orthonormality constraints on ϕ_k and ψ_k can be expressed as $\mathbf{c}_k^T\mathbf{J}\mathbf{c}_l = \mathbf{d}_k^T\mathbf{J}\mathbf{d}_l = \delta_{kl}$, where δ_{kl} is Kronecker's delta and $\mathbf{J} = \int \gamma(x)\gamma(x)^T dx$.

3. Penalized Maximum Likelihood Estimation

With a slight abuse of notation, we write $\{(x_{ij}, y_{ij}) : j = 1, \ldots, m_i\}$ in vector form as $(\mathbf{x}_i, m_i, \mathbf{y}_i)$. Then, the joint density of the observations and the latent variables can be factorized as

$$f_\theta(\mathbf{x}, m, \mathbf{y}, \mathbf{u}, \mathbf{v}) = f_\theta(\mathbf{y} \mid \mathbf{x}, m, \mathbf{u}, \mathbf{v}) f_\theta(\mathbf{x}, m \mid \mathbf{u}, \mathbf{v}) f_\theta(\mathbf{u}, \mathbf{v}).$$

Because $f_\theta(\mathbf{y} \mid \mathbf{x}, m, \mathbf{u}, \mathbf{v})$ does not explicitly depend on $\mathbf{u}$, and $f_\theta(\mathbf{x}, m \mid \mathbf{u}, \mathbf{v})$ does not explicitly depend on $\mathbf{v}$, we can write

$$f_\theta(\mathbf{x}, m, \mathbf{y}, \mathbf{u}, \mathbf{v}) = f_\theta(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v}) f_\theta(\mathbf{x}, m \mid \mathbf{u}) f_\theta(\mathbf{u}, \mathbf{v}).$$

From (2.1), (2.2), (2.3), and the distributional assumptions in Section 2, we have

$$f_\theta(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v}) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left\{-\frac{1}{2\sigma_\eta^2}\|\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\|^2\right\}, \tag{3.1}$$

with $\nu(\mathbf{x}) = (\nu(x_1), \ldots, \nu(x_m))^T$ and $\mathbf{\Psi}(\mathbf{x}) = [\psi_1(\mathbf{x}), \ldots, \psi_{p_2}(\mathbf{x})]$;

$$f_\theta(\mathbf{x}, m \mid \mathbf{u}) = \exp\left\{-\int \lambda_{\mathbf{u}}(t)dt\right\} \frac{1}{m!} \prod_{j=1}^{m} \lambda_{\mathbf{u}}(x_j),$$

with $\lambda_{\mathbf{u}}(x) = \exp\{\mu(x) + \mathbf{u}^T\phi(x)\}$; and

$$f_\theta(\mathbf{u}, \mathbf{v}) = \frac{1}{(2\pi)^{(p_1+p_2)/2}(\det \mathbf{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{u}^T, \mathbf{v}^T)\mathbf{\Sigma}^{-1}(\mathbf{u}^T, \mathbf{v}^T)^T\right\}.$$

The marginal density of the observations,

$$f_\theta(\mathbf{x}, m, \mathbf{y}) = \iint f_\theta(\mathbf{x}, m, \mathbf{y}, \mathbf{u}, \mathbf{v})\ d\mathbf{u}\ d\mathbf{v},$$

has no closed form, and requires numerical integration for its evaluation, for which we use the Laplace approximation. This and other details of the implementation are discussed in the Supplementary Material.

The maximum likelihood estimator of θ is the maximizer of $\sum_{i=1}^{n} \log f_\theta(\mathbf{x}_i, m_i, \mathbf{y}_i)$. However, when a large family of basis functions $\mathscr{B}$ is used, it is advisable to regularize the functional estimators by adding roughness penalties to the objective function. Therefore we define the penalized log-likelihood

$$\ell_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f_\theta(\mathbf{x}_i, m_i, \mathbf{y}_i) - \xi_1 P(\mu) - \xi_2 \sum_{k=1}^{p_1} P(\phi_k) - \xi_3 P(\nu) - \xi_4 \sum_{k=1}^{p_2} P(\psi_k), \tag{3.2}$$

where $\xi_1, \ldots, \xi_4$ are nonnegative smoothing parameters, and $P(f)$ is a roughness penalty function, such as $P(f) = \int (f'')^2$ if f is univariate, or $P(f) = \iint \{(\partial^2 f/\partial t_1^2)^2 + 2(\partial^2 f/\partial t_1/\partial t_2)^2 + (\partial^2 f/\partial t_2^2)^2\}$ if f is bivariate. The estimator of θ is then defined as

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ell_n(\theta),$$

where Θ is the parameter space

$$\begin{aligned} \Theta = \{\theta \in \mathbb{R}^d : & h_{kl}^C(\theta) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_1, \\ & h_{kl}^D(\theta) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_2, \\ & \sigma_\eta^2 > 0,\ \mathbf{\Sigma} > 0\}, \end{aligned} \tag{3.3}$$

with d the dimension of θ, $h_{kl}^C(\theta) = \mathbf{c}_k^T \mathbf{J} \mathbf{c}_l - \delta_{kl}$, $h_{kl}^D(\theta) = \mathbf{d}_k^T \mathbf{J} \mathbf{d}_l - \delta_{kl}$, and $\mathbf{\Sigma} > 0$ denoting that $\mathbf{\Sigma}$ is symmetric and positive definite. The estimating equations

for $\hat{\theta}$ and an expectation–maximization (EM) algorithm (Dempster, Laird and Rubin (1977)) for its computation are derived in the Supplementary Material. The programs implementing these algorithms are available on the first author's website.

Once $\hat{\theta}$ has been obtained, individual predictors of the latent component scores, whether for the sample units or for new data, can be obtained as $\hat{\mathbf{u}}_i = E_{\hat{\theta}}(\mathbf{u} \mid \mathbf{x}_i, m_i, \mathbf{y}_i)$ and $\hat{\mathbf{v}}_i = E_{\hat{\theta}}(\mathbf{v} \mid \mathbf{x}_i, m_i, \mathbf{y}_i)$. These integrals can also be evaluated numerically using a Laplace approximation.

This model has a number of tuning parameters that have to be chosen by the user: the numbers of functional components p_1 and p_1, the type of basis family $\mathscr{B}$ and its dimension q, and the smoothing parameters ξ in the penalized likelihood. The specific type of basis family will not have much of an impact for most applications, provided that the dimension q is sufficiently large. In this study, we use cubic B-splines with equally spaced knots for our simulations and data analyses; higher-order splines should be used if an estimation of derivatives is of interest. The dimension q is more relevant and should be relatively large to avoid bias; the variability of the estimators will be taken care of by ξ. As noted by Ruppert (2002, Sec. 3), although q can be chosen systematically using cross-validation, there is little change in the goodness of fit after a minimum dimension q has been reached; for a larger q, the fit is essentially determined by the smoothing parameters.

The choice of ξ, then, is more important, and can be done objectively using cross-validation (Hastie, Tibshirani and Friedman (2009, Chap. 7)). Leave-one-out cross-validation finds ξ that maximizes

$$\mathrm{CV}(\xi_1, \xi_2, \xi_3, \xi_4) = \sum_{i=1}^{n} \log f_{\hat{\theta}^{[-i]}}(\mathbf{x}_i, m_i, \mathbf{y}_i), \tag{3.4}$$

where $\hat{\theta}^{[-i]}$ denotes the estimator obtained without observation i. A faster alternative is to use k-fold cross-validation, where the data are split into k subsets that are used as test data; $k = 5$ is a common choice. A full four-dimensional optimization of (3.4) would be too time consuming, even with five-fold cross-validation. As a workable alternative, we suggest a sequential optimization, where each ξ_j is optimized in turn on a grid, and the others are kept fixed at an initial value chosen by the user.

A more practical alternative is to choose the parameters subjectively by visual inspection. Plots of the means and components for different ξ on a grid

can be inspected to determine how new features of the curves appear or disappear as ξ varies. Then, we can choose ξ that produces curves with features that are well defined, but not too irregular. In general, because curve shapes change smoothly with ξ, there is a relatively broad range of ξ that will produce reasonable results; thus, it is not necessary to specify a precise optimal. We use this method in our simulations and data analysis.

The choice of the numbers of components p_1 and p_2 can also be done either objectively, using cross-validation, or subjectively, by taking into account the accumulated proportions of variability $\sigma^2_{u1} + \cdots + \sigma^2_{up_1}$ and $\sigma^2_{v1} + \cdots + \sigma^2_{vp_2}$. From a practical perspective, however, the goal of this model is not so much to find the largest possible p that will best approximate the data, but to capture the most salient modes of variability of the X and Y processes, and then to estimate and interpret their correlations. From this perspective, having a few well-estimated components with significant correlations is preferable to having a higher-dimensional model without many (or any) significant correlations, even if some residual systematic variability remains unaccounted for.

4. Asymptotics and Inference

The asymptotic behavior of $\hat{\theta}$ as $n \to \infty$ can be studied using standard empirical-process techniques (Pollard (1984); Van der Vaar (2000)), because (3.2) is the average of i.i.d. functions and a nonrandom roughness penalty; for example, see Knight and Fu (2000).

"Nonparametric" asymptotics, where no assumptions about the functional parameters (other than degrees of smoothness) are made, and the dimension q of the basis family $\mathscr{B}$ is allowed to grow with n, is perhaps the most theoretically satisfying, but it is too difficult. A simpler approach is that of "parametric" asymptotics, where q is held fixed, and the functional parameters are assumed to belong to $\mathscr{B}$. This approach, in effect, ignores the smoothing bias. However, in practice, this is not a serious problem, as long as q is reasonably large. We follow this approach, which others have followed in similar semiparametric contexts (e.g., Yu and Ruppert (2002); Xun et al. (2013)), and show later by simulation that the asymptotic variance estimates provide very accurate approximations to the actual finite-sample variance of the estimators.

The first result in this section, Theorem 1, establishes the consistency of the estimator $\hat{\theta}$. The proof, given in the Supplementary Material, essentially follows along the lines of the classical consistency proof of maximum likelihood

estimators, with the caveat that the indeterminate sign of the functional components requires special handling. We further assume that the components have multiplicity one; thus, we define

$$\begin{aligned}\Theta = \{\theta \in \mathbb{R}^s : {} & h^C_{kl}(\theta) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_1, \\ & h^D_{kl}(\theta) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_2, \\ & \sigma^2_\eta > 0,\ \mathbf{\Sigma} > 0,\ \sigma_{u1} > \cdots > \sigma_{up_1} > 0,\ \sigma_{v1} > \cdots > \sigma_{vp_2} > 0, \\ & c_{k1} \geq 0,\ k = 1, \ldots, p_1,\ d_{k1} \geq 0,\ k = 1, \ldots, p_2\},\end{aligned} \tag{4.1}$$

and make the following assumptions:

A1 The signs of the functional components $\hat{\phi}_{k,n}$ and $\hat{\psi}_{k,n}$ are specified so that the first nonzero basis coefficient of each $\hat{\phi}_{k,n}$ and $\hat{\psi}_{k,n}$ is positive (then, $\hat{\theta}_n \in \Theta$ for Θ defined in (4.1).)

A2 The true functional parameters μ_0, ν_0, ϕ_{k0}, and ψ_{k0} of models (1.1)–(2.2)–(2.3) belong to the functional space $\mathscr{B}$ used for the estimation, and the basis coefficients $c_{k1,0}$ and $d_{k1,0}$ are not zero. The signs of ϕ_{k0} and ψ_{k0} are then specified such that $c_{k1,0} > 0$ and $d_{k1,0} > 0$; therefore, there is a unique θ_0 in Θ, such that $f_{\theta_0}(\mathbf{x}, m, \mathbf{y})$ is the true density of the data.

A3 $\xi_n \to \mathbf{0}$ as $n \to \infty$, where $\xi_n = (\xi_{1n}, \xi_{2n}, \xi_{3n}, \xi_{4n})^T$ is the vector of smoothing parameters in (3.2).

The requirement in assumption A2 that the first basis coefficients $c_{k1,0}$ and $d_{k1,0}$ of each ϕ_{k0} and ψ_{k0} be nonzero and, therefore, can be taken as strictly positive, is somewhat artificial. Clearly, ϕ_{k0} and ψ_{k0} must have at least one nonzero basis coefficient; however, it need not be the first, nor any other, in particular. However, a condition such as this is necessary to uniquely identify a "true" parameter θ_0, which would otherwise be unidentifiable, owing to sign ambiguity. This condition has to be consistent with the sign-specification rule for the estimators in assumption A1.

Theorem 1. *Under assumptions* A1–A3, $\hat{\theta}_n \xrightarrow{P} \theta_0$ *as* $n \to \infty$.

To establish the asymptotic normality of the estimators, we follow the approach of Geyer (1994), and use the tangent cone of the parameter space. The definition and properties of tangent cones can be found in Rockafellar and Wets (1998, Chap. 6). Using Theorem 6.31 of Rockafellar and Wets (1998), the tangent

cone of Θ at θ_0 is

$$\mathscr{T}_0 = \{\delta \in \mathbb{R}^s : \nabla h_{kl}^C(\theta_0)^T \delta = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_1, \\ \nabla h_{kl}^D(\theta_0)^T \delta = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_2\}.$$

The explicit forms of $\nabla h_{kl}^C(\theta)$ and $\nabla h_{kl}^D(\theta)$ are derived in the Supplementary Material. Let $\mathbf{A}$ be the $s_1 \times s$ matrix with rows $\nabla h_{kl}^C(\theta_0)^T$ and $\nabla h_{kl}^D(\theta_0)^T$, where $s_1 = \{p_1(p_1+1)/2 + p_2(p_2+1)/2\}$, and let $\mathbf{B}$ be an orthogonal complement of $\mathbf{A}$, that is, an orthogonal $(s - s_1) \times s$ matrix, such that $\mathbf{A}\mathbf{B}^T = \mathbf{O}$.

The next theorem gives the asymptotic distribution of $\hat{\theta}_n$. In addition to $\mathbf{B}$ defined above, it uses Fisher's information matrix,

$$\begin{aligned} \mathbf{F}_0 &= E_{\theta_0}\{\nabla \log f_{\theta_0}(\mathbf{x}, m, \mathbf{y}) \nabla \log f_{\theta_0}(\mathbf{x}, m, \mathbf{y})^T\} \\ &= -E_{\theta_0}\{\nabla^2 \log f_{\theta_0}(\mathbf{x}, m, \mathbf{y})\}, \end{aligned}$$

where ∇ and ∇^2 are taken with respect to the parameter θ, and $\mathsf{D}\mathbf{P}(\theta)$, the Jacobian matrix of the smoothness penalty vector $\mathbf{P}(\theta) = (P(\mu), \sum_{k=1}^{p_1} P(\phi_k), P(\nu), \sum_{k=1}^{p_2} P(\psi_k))^T$ of (3.2). Explicit expressions for these derivatives are given in the Supplementary Material. We make one additional assumption:

A4 $\sqrt{n}\xi_n \to \kappa$ as $n \to \infty$, for a finite κ.

Theorem 2. *Under assumptions* A1–A4, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(-\mathbf{V}\mathsf{D}\mathbf{P}(\theta_0)^T\kappa, \mathbf{V})$ *as* $n \to \infty$, *with* $\mathbf{V} = \mathbf{B}^T(\mathbf{B}\mathbf{F}_0\mathbf{B}^T)^{-1}\mathbf{B}$.

Fisher's information matrix $\mathbf{F}_0$ can be estimated by

$$\hat{\mathbf{F}}_0 = \frac{1}{n}\sum_{i=1}^{n} \nabla \log f_{\hat{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i) \nabla \log f_{\hat{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i)^T$$

and $\mathbf{V}$ by $\hat{\mathbf{V}} = \mathbf{B}^T(\mathbf{B}\hat{\mathbf{F}}_0\mathbf{B}^T)^{-1}\mathbf{B}$. The accuracy of the approximation of $\hat{\mathbf{V}}$ to the actual finite-sample variance of the estimators depends on the ratio n/s. We found in our simulations (Section 5) that ratios of $n/s \geq 3$ offer very accurate approximations. However, this does impose some limitations on how large the basis family dimension q and the number of components p_1 and p_2 can be, for any given n.

5. Simulations

We examine the finite-sample behavior of the estimators by simulation, to assess their consistency as the sample size increases and the accuracy of the

approximation of the asymptotic variances.

We generated data from models (2.1)–(2.3) with $p_1 = p_2 = 2$. We considered a temporal process on $\mathscr{S} = [0, 1]$, with $\mu(x) \equiv \sin \pi x - \log 1.98 + \log r$, $\nu(x) = 5x$, $\phi_1(x) = \sqrt{2} \sin \pi x$, $\phi_2(x) = \sqrt{2} \sin 2\pi x$, $\psi_1(x) = \phi_1(x)$, and $\psi_2(x) = \phi_2(x)$. The baseline intensity function $\lambda_0(x) = \exp \mu(x)$ integrates to r; we chose two different values, namely, $r = 10$ and $r = 30$, giving expected numbers of observations per curve of 10.5 and 31.3, respectively. The lower rate $r = 10$ corresponds to the sparse situation where most individual trajectories cannot be recovered by smoothing. The first components ϕ_1 and ψ_1 are essentially size components, explaining the variation in the overall level above or below the mean. The second components ϕ_2 and ψ_2 are contrasts, where, for example, a positive score corresponds to curves that are above the mean on the first half of $\mathscr{S}$, and below the mean on the second half.

The component variances are of the form $\sigma_{u1}^2 = 0.3^2\alpha$, $\sigma_{u2}^2 = 0.3^2(1 - \alpha)$, $\sigma_{v1}^2 = 0.7^2\alpha$, and $\sigma_{v2}^2 = 0.7^2(1 - \alpha)$. Two choices of α were considered: $\alpha = 0.60$ and $\alpha = 0.75$. The cross-covariance matrix $\mathbf{\Sigma}_{uv}$ was diagonal with elements $\mathbf{\Sigma}_{uv,11} = 0.7\sigma_{u1}\sigma_{v1}$ and $\mathbf{\Sigma}_{uv,22} = 0.7\sigma_{u2}\sigma_{v2}$. The random-noise variance was $\sigma_\eta^2 = 0.3^2$. We considered four sample sizes n: 50, 100, 200, and 400. The combinations of r, α, and n, yield 16 sampling models.

For the estimation, we considered cubic B-spline families with five and 10 equally spaced knots. The smoothing parameters were chosen visually, as explained in Section 3, from a few trial samples from each of the six models, with $r = 10$, and each of the two knot sequences; the same smoothing parameters were used for the respective models with $r = 30$. These are listed in the Supplementary Material. The Monte Carlo study, then, considered a total of 32 scenarios, with two families of estimators per sampling model. Each scenario was replicated 300 times.

As a measure of the estimation error, we considered the root mean squared error. For scalar parameters, for example, σ_η, these errors are defined in the usual way: $E^{1/2}\{(\hat{\sigma}_\eta - \sigma_\eta)^2\}$. For functional parameters, for example, $\mu(x)$, these errors are defined in terms of the L^2-norm: $E^{1/2}(\|\hat{\mu} - \mu\|^2)^{1/2}$. For the random-effect predictors, for example, $\{\hat{u}_{i1}\}$, the errors are defined as $E^{1/2}\{\sum_{i=1}^n (\hat{u}_{i1} - u_{i1})^2/n\}$. The signs of $\hat{\phi}_k(x)$ and $\hat{\psi}_k(x)$, which in principle are indeterminate, were chosen as the signs of the inner products $\langle \hat{\phi}_k, \phi_k \rangle$ and $\langle \hat{\psi}_k, \psi_k \rangle$; the signs of $\hat{u}_{ik}$, $\hat{v}_{ik}$, and the elements of $\hat{\mathbf{\Sigma}}_{uv}$ were changed accordingly. For brevity, we only report here the results for the six sampling models with $\alpha = 0.75$, $n \leq 200$, and estimators obtained using five-knot splines (Table 1). The rest of the results can be found

Table 1. Simulation Results. Root mean squared errors of estimators based on five-knot B-splines under different baseline rates r and sample sizes n.

	$r = 10$			$r = 30$		
Parameter	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\Sigma_{uv,11}$	0.054	0.031	0.025	0.038	0.026	0.019
$\Sigma_{uv,21}$	0.057	0.038	0.024	0.028	0.017	0.011
$\Sigma_{uv,12}$	0.036	0.023	0.015	0.021	0.014	0.010
$\Sigma_{uv,22}$	0.023	0.017	0.012	0.014	0.009	0.006
μ	0.121	0.102	0.090	0.096	0.082	0.072
ν	0.124	0.099	0.087	0.163	0.144	0.136
ϕ_1	0.738	0.515	0.376	0.436	0.261	0.188
ϕ_2	0.882	0.726	0.558	0.588	0.389	0.290
ψ_1	0.243	0.249	0.206	0.138	0.090	0.061
ψ_2	0.216	0.216	0.176	0.145	0.097	0.068
σ_{u1}	0.065	0.057	0.029	0.039	0.027	0.020
σ_{u2}	0.065	0.069	0.038	0.033	0.024	0.018
σ_{v1}	0.070	0.058	0.096	0.062	0.047	0.036
σ_{v2}	0.071	0.082	0.065	0.037	0.027	0.018
σ_{η}	0.067	0.081	0.062	0.012	0.011	0.010
u_{i1}	0.217	0.184	0.170	0.154	0.140	0.134
u_{i2}	0.163	0.141	0.121	0.118	0.104	0.097
v_{i1}	0.167	0.159	0.162	0.168	0.151	0.143
v_{i2}	0.153	0.148	0.138	0.105	0.083	0.072

in the Supplementary Material, and are largely in line with those reported here. Also given in the Supplementary Material are plots of the functional estimators, which help assess the relative weights of the bias and variance in the overall mean squared error.

We see in Table 1 that the estimation errors decrease as n increases, as expected, for both baseline rates r. However, the latter has a significant impact on the accuracy of the estimators, particularly for the components ϕ_1 and ϕ_2. The plots in the Supplementary Material show that most of the errors of $\hat{\phi}_1$ and $\hat{\phi}_2$ come from the bias, rather than the variance and, for a given n, the bias decreases rapidly as r increases. Part of the bias of $\hat{\phi}_1$ and $\hat{\phi}_2$ can be attributed to a component reversal, which is more frequent for the models with $\alpha = 0.60$ than it is for $\alpha = 0.75$. This is also the case, but to a lesser degree, for $\hat{\psi}_1$ and $\hat{\psi}_2$, which, for each (n, r) combination, are more accurate estimators of their respective parameters than are $\hat{\phi}_1$ and $\hat{\phi}_2$.

Table 2 compares the true finite-sample standard deviations of the elements of $\hat{\mathbf{\Sigma}}_{uv}$ with their median asymptotic approximations, as well as median absolute errors of these approximations, for the estimators based on five-knot splines and

Table 2. Simulation Results. True standard deviations and median and median absolute errors of estimated asymptotic standard deviations ($\times 10$) of estimators under different baseline rates r and sample sizes n, for estimators based on five-knot B-splines and variance proportion $\alpha = 0.75$.

				$r = 10$					
	$n = 100$			$n = 200$			$n = 400$		
Parameter	True	Med	MAE	True	Med	MAE	True	Med	MAE
$\Sigma_{uv,11}$	0.31	0.63	0.32	0.24	0.28	0.04	0.16	0.16	0.01
$\Sigma_{uv,21}$	0.38	0.73	0.35	0.24	0.36	0.12	0.15	0.21	0.06
$\Sigma_{uv,12}$	0.23	0.42	0.19	0.15	0.21	0.06	0.11	0.12	0.02
$\Sigma_{uv,22}$	0.17	0.30	0.13	0.12	0.14	0.02	0.13	0.09	0.05
				$r = 30$					
$\Sigma_{uv,11}$	0.25	0.45	0.20	0.18	0.21	0.03	0.12	0.13	0.01
$\Sigma_{uv,21}$	0.17	0.32	0.15	0.11	0.16	0.04	0.08	0.10	0.02
$\Sigma_{uv,12}$	0.14	0.24	0.10	0.10	0.12	0.02	0.06	0.07	0.01
$\Sigma_{uv,22}$	0.09	0.18	0.09	0.06	0.09	0.02	0.04	0.05	0.01

the models with variance proportion $\alpha = 0.75$; for $\alpha = 0.60$ and for 10-knot splines, the results are given in the Supplementary Material. The dimension of θ for five-knot splines is $s = 63$; thus, Fisher's information matrix estimator $\hat{\mathbf{F}}_0$ is singular for $n = 50$; therefore, we only report the results for $n \geq 100$. Overall, we see that the asymptotic standard deviations are very accurate estimators of the true standard deviations for $n \geq 200$. For 10-knot splines, where the dimension of θ is $s = 93$, the tables in the Supplementary Material show that the approximation is accurate for $n \geq 400$. This suggests ratios of $n/s \geq 3$ are sufficient for accurate asymptotic approximations of the variances.

6. Application: Online Auction Data

The eBay auction data mentioned in Section 1 were downloaded from the companion website of Jank and Shmueli (2010). In this sample, there were 194 items sold at auction, and each auction lasted seven days. A subsample of 20 bid-price trajectories is shown in Figure 1. The dots are the actual bids; the solid lines were drawn for better visualization. Figure 1 shows that bidding activity tends to concentrate at the beginning and at the end of the auctions, in patterns known as early bidding and bid sniping, respectively. The bid sniping phenomenon has been observed in dynamic studies of auction prices in the form of slopes of derivatives of bid prices (Jank and Shmueli (2005); Wang et al. (2008)). Some articles (e.g., Backus et al. (2015)) have pointed out that bid sniping is annoying for bidders, and partly as a consequence of this, the number of items auctioned

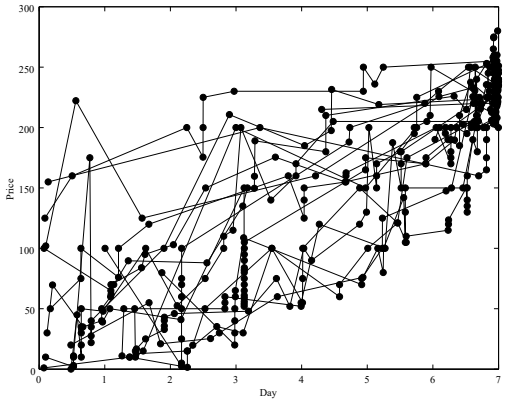

Figure 1. Online Auction Data. Price trajectories of Palm digital assistants auctioned at eBay (first 20 trajectories in a sample of 194).

at eBay has steadily decreased over the years, compared with the number of items sold at fixed prices (Einav et al. (2015)). It has been hypothesized that bid sniping is triggered by the perception that an item's current bid price is low. We do not establish causation here, because our models are not intended for that; however, the results obtained below are in line with this hypothesis.

To estimate the functional means and components, we used cubic B-splines with five equally spaced knots. We found the smoothing parameters graphically (the plots can be found in the Supplementary Material), obtaining $\xi_1 = \xi_2 = \xi_4 = 10^{-4}$ and $\xi_3 = 10^{-6}$. From preliminary trial fits with five components for each process, we found that the first two components of X explain 77% of the variability, and the first three components of Y explain essentially 100% of the variability (the other two eigenvalues are negligible); therefore, we chose $p_1 = 2$ and $p_2 = 3$. The estimated mean and components are shown in Figure 2. Figure 2(a) shows the baseline intensity function $\lambda_0(t) = \exp \mu(t)$ of the bidding process. Here, we see that most of the bidding activity tends to occur toward the end of an auction. Some items attract, overall, more bids than others, and this is explained by the first component (Fig. 2(c)): a positive score on ϕ_1 corresponds to an intensity function λ above the baseline. The second component is related to bid sniping: for items with positive scores on ϕ_2, the number of bids in the final two days of the auction will be above the mean. With regard to the bid price, Fig. 2(b) shows the mean price trajectory $\nu(t)$, and Fig. 2(d) shows the components. The first component is associated with the price level: items with positive scores on ψ_1 show prices above the mean over the whole auction period.

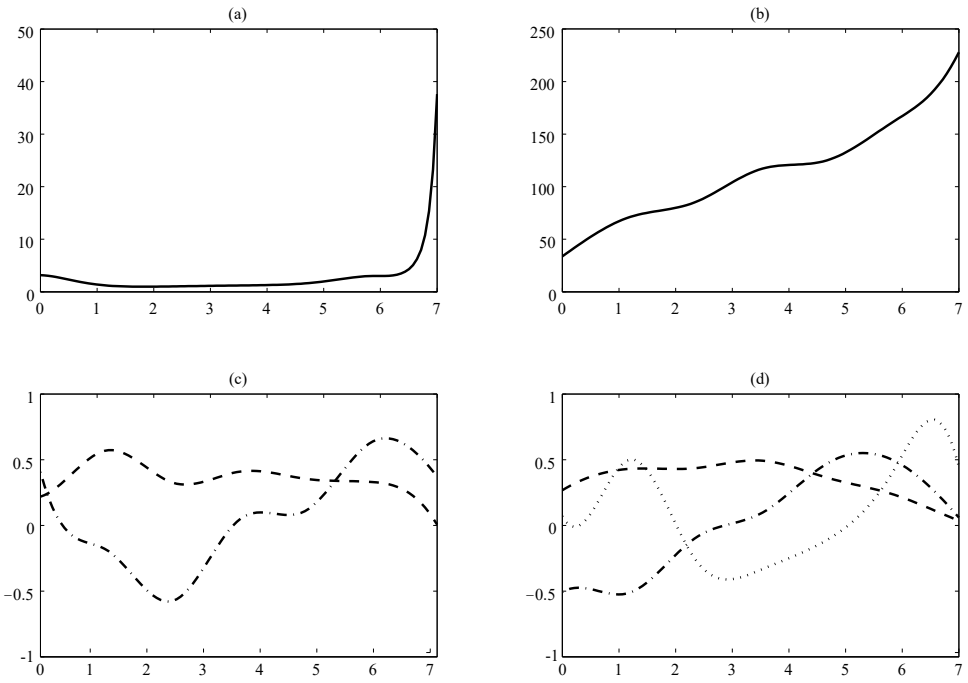

Figure 2. Online Auction Data. (a) Baseline intensity function of bidding time process. (b) Mean price trajectory. (c) Components of bidding time process, ϕ_1 (dashed line) and ϕ_2 (dash-dot line). (d) Components of price trajectories, ψ_1 (dashed line), ψ_2 (dash-dot line), and ψ_3 (dotted line).

The second component is a contrast: items with positive scores on ψ_2 tend to show prices below the mean at the beginning of the auction, and above the mean toward the end.

The estimated cross-covariance and cross-correlation matrices are

$$\hat{\mathbf{\Sigma}}_{uv} = \begin{pmatrix} -256.9 & 48.1 & 22.6 \\ -83.1 & -36.9 & -1.5 \end{pmatrix} \text{ and } \hat{\rho}_{uv} = \begin{pmatrix} -0.69 & 0.41 & 0.28 \\ -0.54 & -0.77 & -0.05 \end{pmatrix},$$

respectively. The asymptotic standard deviations of the elements of $\hat{\mathbf{\Sigma}}_{uv}$ obtained from Theorem 2, and the bootstrap standard deviations based on 100 wild bootstrap replications are

$$\mathrm{sd}_{\mathrm{asymp}}(\hat{\mathbf{\Sigma}}_{uv}) = \begin{pmatrix} 73.3 & 17.7 & 9.9 \\ 20.5 & 6.8 & 5.7 \end{pmatrix} \text{ and } \mathrm{sd}_{\mathrm{boot}}(\hat{\mathbf{\Sigma}}_{uv}) = \begin{pmatrix} 76.7 & 18.3 & 13.4 \\ 22.3 & 7.5 & 5.3 \end{pmatrix},$$

respectively, which are very similar to one another. We conclude that all corre-

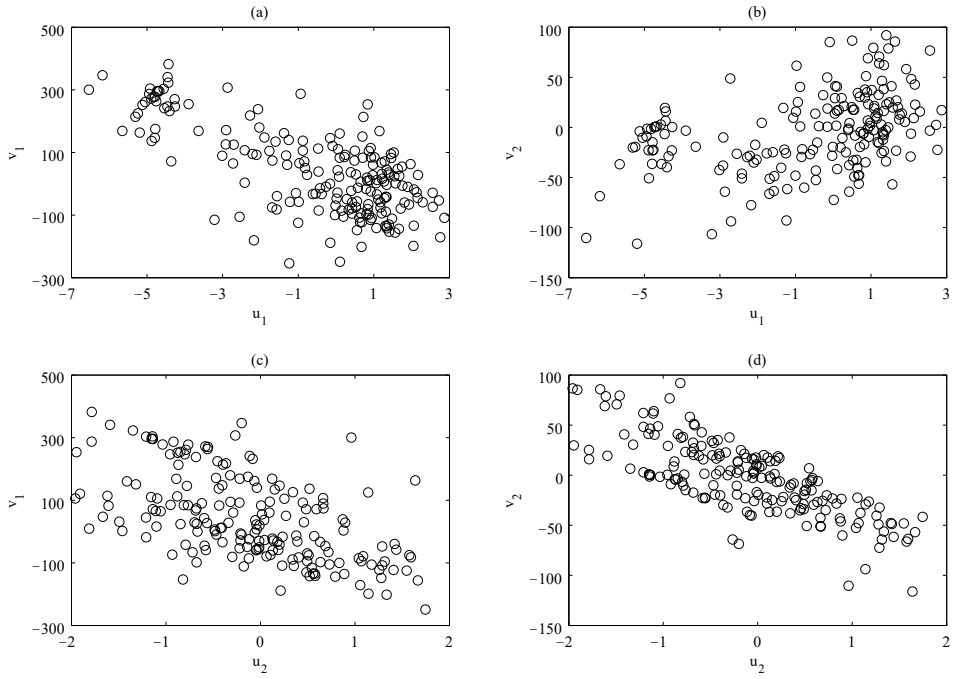

Figure 3. Online Auction Data. Scatter plots of component scores of the bidding time process versus component scores of the price trajectories.

lations involving the first two components of each process are statistically significant, but none of the correlations involving ψ_3 are.

Figure 3 shows scatter plots of the estimated random effects $\hat{u}_{ik}$ versus $\hat{v}_{ik}$ for the significant components. Normal probability plots of the component scores and the residuals $\hat{\eta}_{ij}$ are shown in the Supplementary Material. The component scores appear to be largely Gaussian; only $\hat{u}_{i1}$ shows a mild departure from normality. The residuals $\hat{\eta}_{ij}$ show tails somewhat heavier than normal, but no gross outliers are evident.

These results are in line with our intuition. The negative correlations between v_1 and both u_1 and u_2 show that items with perceived low prices tend to attract more bidders and trigger bid sniping. The strong negative correlation between u_2 and v_2 shows that bid sniping is particularly associated with price trajectories that are found to be well below the mean on the fifth day of the auction.

To illustrate with a few specific cases, Figure 4 shows the price trajectories of items with the largest and smallest scores v_1 and v_2, respectively. Figure

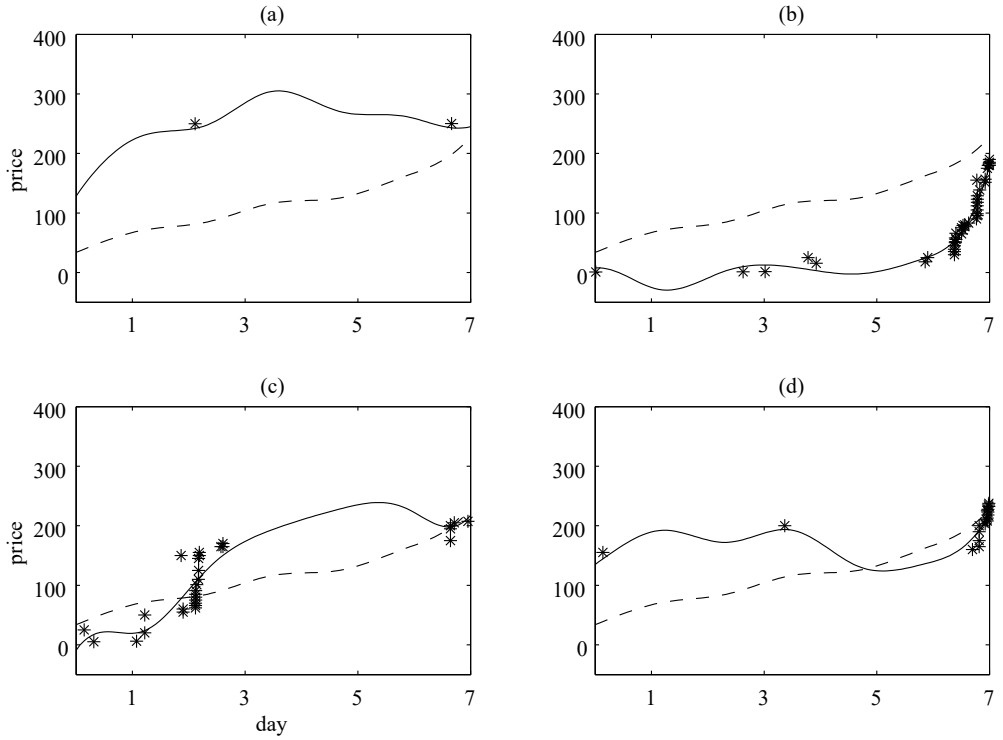

Figure 4. Online Auction Data. Estimated price trajectories (solid line) and mean price trajectory (dashed line), along with actual bids (asterisks) for items with (a) the largest score on the first Y-component, (b) the lowest score on the first Y-component, (c) the largest score on the second Y-component, and (d) the lowest score on the second Y-component.

4(a) shows the item with largest v_1 score and, consequently, low u_1 score: an expensive item that attracted only two bids. Figure 4(b) shows the opposite, the item with the lowest v_1 score and, consequently, large u_1 and u_2 scores: an underpriced item that attracted many bids toward the end of the auction, which is a typical case of bid sniping. Figure 4(c) shows the item with the largest v_2 score and, consequently, a large u_1 score, but a low u_2 score: this was an item that started off with a low price and attracted many bids at the beginning of the auction; this sent the price above the mean early in the auction period, and then did not attract many late bidders. Figure 4(d), the item with the lowest v_2 score, shows the opposite situation: the few bids placed at the beginning of the auction period were well above the mean, but toward the end, some lower bids are placed (an unusual, but possible situation) which triggered bid sniping.

7. Discussion

We have presented a unified model for the joint statistical analysis of a functional response variable and the distribution of the grid points at which the variable is measured. Although the problems of estimating sparse functional data and intensity functions of point processes have been considered in the literature, to date, this has been done separately.

Our model allows statistical inferences for the correlations between the components of the grid-point process and the response variable. For this, we have developed a parametric asymptotic theory in Section 4, where $\sqrt{n}$-consistency is obtained, but at the price of ignoring the asymptotic bias. When the latter is negligible, for example, when the target functions are smooth and the basis family used for the estimation is sufficiently large, the asymptotic approximation is very accurate, as shown in the simulations and example in Sections 5 and 6, respectively. However, if the target functions are more irregular, and the asymptotic bias is more significant, truly nonparametric asymptotics with the dimension of the basis family growing with n would be more appropriate, although the rate of convergence would be lower than $\sqrt{n}$. This is still an open problem.

The model in Section 2 uses latent variables with distributions that are assumed normal. Of course, this is always going to be an approximation, at best. While mild departures from normality may not affect the validity of the results, more serious deviations, such as gross outliers or very heavy-tailed distributions, most likely will. For the sake of brevity, we do not include a thorough robustness analysis in this paper. However, the model and the proposed maximum likelihood estimators can be modified easily to accommodate heavier-tailed distributions, such as Student's t distributions, for the latent variables. This is left to future research.

Supplementary Material

The online supplementary material contains proofs, technical derivations, and additional simulation results.

Acknowledgments

This research was partly supported by US National Science Foundation grant DMS 1505780.

References

Arribas-Gil, A. and Müller, H.-G. (2014). Pairwise dynamic time warping for event data. *Computational Statistics and Data Analysis* **69**, 255–268.

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

Backus, M., Blake, T., Masterov, D. V. and Tadelis, S. (2015). Is sniping a problem for online auction markets? *NBER Working Paper* No. 20942.

Baddeley, A. (2007). Spatial point processes and their applications. In *Stochastic Geometry*, Lecture Notes in Mathematics 1892, 1–75.

Baddeley, A. (2010). Multivariate and marked point processes. In *Handbook of Spatial Statistis* (Edited by A. E. Gelfand, P. J. Diggle, P. Guttorp and M. Fuentes), 299–337. CRC Press, Boca Raton.

Barrett, J., Diggle, P., Henderson, R. and Taylor-Robinson, D. (2015). Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**, 131–148.

Cox, D. R. and Isham, V. (1980). *Point Processes*. Chapman and Hall/CRC, Boca Raton.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* **39**, 1–38.

Einav, L., Farronato, C., Levin, J. D. and Sundaresan, N. (2015). Sales mechanisms in online markets: What happened to Internet auctions? *NBER Working Paper* No. 19021.

Gervini, D. (2016). Independent component models for replicated point processes. *Spatial Statistics* **18**, 474–488.

Geyer, C. J. (1994). On the asymptotics of constrained M-estimation. *The Annals of Statistics* **22**, 1993–2010.

Guan, Y. and Afshartous, D. R. (2007). Test for independence between marks and points of marked point processes: a subsampling approach. *Environmental and Ecological Statistics* **14**, 101–111.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. 2nd Edition. Springer, New York.

James, G., Hastie, T. G. and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.

Jank, W. and Shmueli, G. (2005). Price dynamics in online auctions using curve clustering. *Robert H. Smith School Research Paper. RHS-06-004*.

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

Jank, W. and Shmueli, G. (2010). *Modeling Online Auctions*. Wiley & Sons, New York.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics* **28**, 1356–1378.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Møller, J., Ghorbani, M. and Rubak, E. (2016). Mechanistic spatio-temporal point process models for marked point processes, with a view to forest stand data. *Biometrics* **72**, 687–

696.

Müller, H. G. (2008). Functional modeling of longitudinal data. *Longitudinal Data Analysis* **1**, 223–252.

Myllymäki, M., Mrkvička, T., Seijo, H. and Grabarnik, P. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* **79**, 381–404.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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

Rathbun, S. L. and Shiffman, S. (2016). Mixed effects models for recurrent events data with partially observed time-varying covariates: Ecological momentary assessment of smoking. *Biometrics* **72**, 46–55.

Rice, J. A. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica* **14**, 631–647.

Rockafellar, R. T. and Wets, R. J. (1998). *Variational Analysis*. Springer, New York.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**, 735–757.

Scheike, T. H. (1997). A general framework for longitudinal data through marked point processes. *Biometrical Journal* **39**, 57–67.

Shmueli, G. and Jank, W. (2005). Visualizing online auctions. *Journal of Computational and Graphical Statistics* **14**, 299–319.

Streit, R. L. (2010). *Poisson Point Processes: Imaging, Tracking, and Sensing*. Springer, New York.

Van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.

Wang, S., Jank, W., Shmueli, G. and Smith, P. (2008). Modeling price dynamics in eBay auctions using differential equations. *Journal of the American Statistical Association* **103**, 1100–1118.

Wu, S., Müller, H.-G. and Zhang, Z. (2013). Functional data analysis for point processes with rare events. *Statistica Sinica* **23**, 1–23.

Xun, X., Cao, J., Mallick, B., Maity, A. and Carroll, R. J. (2013). Parameter estimation of partial differential equations. *Journal of the American Statistical Association* **108**, 1009–1020.

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

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association* **97**, 1042–1054.

3200 N Cramer Street, Milwaukee, IL 53211, USA.

E-mail: gervini@uwm.edu

3200 N Cramer Street, Milwaukee, IL 53211, USA.

E-mail: tjbaur@uwm.edu

(Received October 2017; accepted November 2018)