Joint models for grid point and response processes in longitudinal and functional data

Daniel Gervini and Tyler J Baur
Department of Mathematical Sciences
University of Wisconsin–Milwaukee

August 13, 2018

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. In this paper we introduce a covariation model to estimate and make inferences about this interrelation, by treating the data as replicated realizations of a marked point process. We derive maximum likelihood estimators, the asymptotic distribution of the estimators, and study the estimators' behavior by simulation. 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: Doubly-stochastic process; Karhunen–Loève decomposition; latent-variable model; Poisson process.

1 Introduction

In many statistical applications the object of analysis are samples of functions, $\{g_i(x) : i = 1, \ldots, n\}$. These functions are generally measured at discrete points $\{x_{ij} : j = 1, \ldots, m_i\}$, so the data actually observed is $\{(x_{ij}, y_{ij}) : j = 1, \ldots, m_i,\ i = 1, \ldots, n\}$ with

$$y_{ij} = g_i(x_{ij}) + \eta_{ij}, \tag{1}$$

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

Functional data analysis has focused on the analysis of the functions $g_i(x)$s, which are usually recovered from the raw data by some form of smoothing (James et al., 2000; Ramsay and Silverman, 2005, ch. 3; Yao et al., 2005). The distribution of the grid points $\{x_{ij}\}$ is generally considered noninformative. However, there are situations where the distribution of the x_{ij}s may be informative in its own right.

Consider, for example, the bid price trajectories shown in Figure 1. They are bid prices of Palm M515 Personal Digital Assistants (PDA) 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 via derivatives of the bid price curves. More recently, Wu et al. (2013) and Arribas-Gil and Müller (2014) investigated the bid time process itself. But a joint modeling of the bid time process and the bid price curves has not 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 study 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. For each subject i, the x_{ij}s are seen as a realization of a point process and the y_{ij}s as the corresponding 'marks', to use common point-process terminology (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 in model (1); the methods we propose here are specifically intended for functional and longitudinal data

applications. To avoid confusions with terminology, we will not call the m_i observations for each subject i 'replications', as is often done in the point process literature; 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 replications are the sets for $i = 1, \ldots, n$.

As pointed out by Guan and Afshartous (2007) and Møller et al. (2016), the literature on modeling marked point processes is limited, and restricted to the single replication scenario; it has focused 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). But 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 recently (Barret et al., 2015; Rathbun and Shiffman, 2016), but they aim at incorporating 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 the past observations. None of those papers aim at jointly modeling the time points and the response processes, which is the goal of this paper.

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, ch. 2; Streit, 2010, ch. 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

$$Y_x = g(x) + \eta_x, \tag{2}$$

with $g : \mathscr{S} \to \mathscr{M}$ the function of interest and η_x random noise. We will 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 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 the λs are subject-specific and treat them as random effects. Such processes are called doubly stochastic processes or Cox processes (Møller and Waagepetersen, 2004, ch. 5; Streit, 2010, ch. 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 non-negative locally integrable functions on $\mathscr{S}$. Then the n replications of the point process can be seen as independent identically distributed realizations of a doubly stochastic process (X, Λ), where X is observable but Λ is not. Similarly, for g in (2) we will assume there is a process G such that $Y \mid (X, G = g)$ follows model (2). Then the n replications of the marked point process can be seen as independent identically distributed 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 the xs and the response process G that generates the ys. To this end we will assume that G follows a finite Karhunen–Loève decomposition

$$G(x) = \nu(x) + \sum_{k=1}^{p_2} v_k \psi_k(x), \tag{3}$$

where the ψ_ks are orthonormal functions in $L^2(\mathscr{S})$ and the v_ks are uncorrelated zero-mean random variables. Any stochastic process in $L^2(\mathscr{S})$ with finite variance can be decomposed as in (3) with a possibly infinite p_2 (Ash and Gardner, 1975, ch. 1.4), but since we are interested in smooth processes in this paper, for practical purposes it is sufficient to consider only finite p_2s.

A similar decomposition for Λ would be problematic due to the non-negativity constraint. A nonnegative decomposition was proposed by Gervini (2016). However,

for simplicity we will use an alternative approach in this paper, and decompose instead the logarithm of Λ, which is unconstrained:

$$\log \Lambda(x) = \mu(x) + \sum_{k=1}^{p_1} u_k \phi_k(x), \tag{4}$$

where the ϕ_ks are orthonormal functions in $L^2(\mathscr{S})$ and the u_ks 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 (3) and (4). As a working model, we will assume that $(\mathbf{u}, \mathbf{v})$ follows a joint multivariate normal distribution with mean zero and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \operatorname{diag}(\boldsymbol{\sigma}_u^2) & \mathbf{\Sigma}_{uv} \\ \mathbf{\Sigma}_{uv}^T & \operatorname{diag}(\boldsymbol{\sigma}_v^2) \end{pmatrix},$$

where $\boldsymbol{\sigma}_u^2$ and $\boldsymbol{\sigma}_v^2$ are the variances of the u_ks and the v_ks, respectively. The error term η in (2) is assumed $N(0, \sigma_\eta^2)$ and independent of the u_ks and the v_ks. The parameter of interest here is the cross-covariance matrix $\mathbf{\Sigma}_{uv}$; the others are mostly nuisance parameters.

The signs of the component scores are not identifiable, since $-u_k$ and $-\phi_k(x)$ satisfy the same model as u_k and $\phi_k(x)$, and similarly with the v_ks and ψ_ks. Consequently, the signs of $\Sigma_{uv,kl} = \operatorname{cov}(u_k, v_l)$ are not identifiable either and can be chosen for convenience of interpretation for any given application.

To facilitate estimation of the functional parameters μ, ϕ_ks, ν and ψ_ks, we will use semiparametric basis-function expansions. As basis functions one can take, 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 will call this family $\mathscr{B}$. Let $\boldsymbol{\gamma}(x)$ 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 $\mu(x) = \mathbf{c}_0^T \boldsymbol{\gamma}(x)$, $\phi_k(x) = \mathbf{c}_k^T \boldsymbol{\gamma}(x)$, $\nu(x) = \mathbf{d}_0^T \boldsymbol{\gamma}(x)$ and $\psi_k(x) = \mathbf{d}_k^T \boldsymbol{\gamma}(x)$.

The model parameters will be collected, for simplicity, in a single vector

$$\boldsymbol{\theta} = (\operatorname{vec} \mathbf{\Sigma}_{uv}, \mathbf{c}_0, \ldots, \mathbf{c}_{p_1}, \mathbf{d}_0, \ldots, \mathbf{d}_{p_2}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_v^2, \sigma_\eta^2). \tag{5}$$

The orthonormality constraints on the ϕ_ks and the ψ_ks 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 $\mathbf{J} = \int \boldsymbol{\gamma}(x)\boldsymbol{\gamma}(x)^T dx$.

3 Penalized maximum likelihood estimation

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

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

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

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

From (2), (3), (4) and the distributional assumptions made in Section 2, we have:

$$f_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v}) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left\{ -\frac{1}{2\sigma_\eta^2} \|\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\|^2 \right\}, \tag{6}$$

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

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

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

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

The marginal density of the observations,

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

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

The maximum likelihood estimator of $\boldsymbol{\theta}$ would be the maximizer of $\sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i)$. 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. So we define the penalized log-likelihood

$$\ell_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i) - \xi_1 P(\mu) - \xi_2 \sum_{k=1}^{p_1} P(\phi_k) - \xi_3 P(\nu) - \xi_4 \sum_{k=1}^{p_2} P(\psi_k), \quad (7)$$

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

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

where Θ is the parameter space

$$\begin{aligned} \Theta \;=\; & \{\boldsymbol{\theta} \in \mathbb{R}^d : h_{kl}^C(\boldsymbol{\theta}) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_1, \\ & h_{kl}^D(\boldsymbol{\theta}) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_2, \\ & \sigma_\eta^2 > 0, \; \boldsymbol{\Sigma} > 0\}, \end{aligned} \quad (8)$$

with d the dimension of $\boldsymbol{\theta}$, $h_{kl}^C(\boldsymbol{\theta}) = \mathbf{c}_k^T \mathbf{J} \mathbf{c}_l - \delta_{kl}$, $h_{kl}^D(\boldsymbol{\theta}) = \mathbf{d}_k^T \mathbf{J} \mathbf{d}_l - \delta_{kl}$, and $\boldsymbol{\Sigma} > 0$ denoting that $\boldsymbol{\Sigma}$ is symmetric and positive definite. The estimating equations for $\hat{\boldsymbol{\theta}}$ and an EM algorithm (Dempster et al., 1977) for its computation are derived in the Supplementary Material. The programs implementing these algorithms are available on the first author's website.

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

This model has a number of tuning parameters that have to be chosen by the user: the number of functional components p_1 and p_1, the type of basis family $\mathscr{B}$ and its dimension q, and the smoothing parameters ξs in the penalized likelihood. The specific type of basis family will not have much of an impact for most applications, provided the dimension q is large enough. In this paper we use cubic B-splines with equally spaced knots for our simulations and data analyses; higher-order splines

should be used if 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 the ξs. As noted by Ruppert (2002, sec. 3), although q can be chosen systematically by cross-validation, there is little change in goodness of fit after a minimum dimension q has been reached, and the fit will essentially be determined by the smoothing parameters thereafter.

The choice of ξs, then, is more important, and can be done objectively by cross-validation (Hastie *et al.*, 2009, ch. 7). Leave-one-out cross-validation finds ξs that maximize

$$\mathrm{CV}(\xi_1, \xi_2, \xi_3, \xi_4) = \sum_{i=1}^{n} \log f_{\hat{\boldsymbol{\theta}}^{[-i]}}(\mathbf{x}_i, m_i, \mathbf{y}_i), \tag{9}$$

where $\hat{\boldsymbol{\theta}}^{[-i]}$ denotes the estimator obtained without observation i. A faster alternative is to use k-fold cross-validation, where the data is split into k subsets that are alternatively used as test data; $k = 5$ is a common choice. Full four-dimensional optimization of (9) would be too time consuming even with five-fold cross-validation, so as a workable alternative we suggest a sequential optimization, where each ξ_j in turn is optimized on a grid while 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 ξs on a grid can be inspected to see how new features of the curves appear or disappear as ξ varies, and choose ξs that produce curves with well-defined but not too irregular features. In general, since curve shapes change smoothly with ξ, there is a relatively broad range of ξs that will produce reasonable results; it is not necessary to specify a precise optimal. We use this method in our simulations and data analysis in this paper.

The choice of the number of components p_1 and p_2 can also be done either objectively by 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 ps that will best approximate the data, but to capture the most salient modes of variability of the X and Y processes and estimate and interpret their correlations; from this perspective, a few well-estimated components with significant correlations will be preferable to 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 $\hat{\boldsymbol{\theta}}$ as $n \to \infty$ can be studied via standard empirical-process techniques (Pollard, 1984; Van der Vaart, 2000), since (7) is the average of independent identically distributed functions plus a non-random roughness penalty, as in e.g. Knight and Fu (2000).

A '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 the 'parametric' asymptotics, where q is held fixed and the functional parameters are assumed to belong to $\mathscr{B}$. This approach, in effect, ignores smoothing bias, but in practice this is not a serious problem as long as q is reasonably large. We will then follow this approach, which others have followed in similar semiparametric contexts (e.g. Yu and Ruppert, 2002, and 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 consistency of the estimator $\hat{\boldsymbol{\theta}}$. 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 will also assume that the components have multiplicity one, so we define

$$\begin{aligned}\Theta \;=\; & \{\boldsymbol{\theta} \in \mathbb{R}^s : h^C_{kl}(\boldsymbol{\theta}) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_1, \\ & h^D_{kl}(\boldsymbol{\theta}) = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_2, \\ & \sigma^2_\eta > 0,\ \boldsymbol{\Sigma} > 0,\ \sigma_{u1} > \cdots > \sigma_{up_1} > 0,\ \sigma_{v1} > \cdots > \sigma_{vp_2} > 0, \\ & c_{k1} \geq 0,\ k = 1, \ldots, p_1,\ d_{k1} \geq 0,\ k = 1, \ldots, p_2\},\end{aligned} \tag{10}$$

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 non-zero basis coefficient of each $\hat{\phi}_{k,n}$ and $\hat{\psi}_{k,n}$ is positive (then $\hat{\boldsymbol{\theta}}_n \in \Theta$ for Θ defined in (10).)

A2 The true functional parameters μ_0, ν_0, ϕ_{k0}s and ψ_{k0}s of model (1)-(3)-(4) belong to the functional space $\mathscr{B}$ used for 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 so that $c_{k1,0} > 0$ and $d_{k1,0} > 0$; therefore there is a unique $\boldsymbol{\theta}_0$ in Θ such that $f_{\boldsymbol{\theta}_0}(\mathbf{x}, m, \mathbf{y})$ is the true density of the data.

A3 $\boldsymbol{\xi}_n \to \mathbf{0}$ as $n \to \infty$, where $\boldsymbol{\xi}_n = (\xi_{1n}, \xi_{2n}, \xi_{3n}, \xi_{4n})^T$ is the vector of smoothing parameters in (7).

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 non-zero, and therefore can be taken strictly positive, is somewhat artificial; clearly the ϕ_{k0}s and ψ_{k0}s must have at least one non-zero basis coefficient, but it need not be the first one or anyone else in particular. However, some condition like this is necessary to uniquely identify a 'true' parameter $\boldsymbol{\theta}_0$, which would otherwise be unidentifiable due 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,* $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$ *as* $n \to \infty$.

To establish the asymptotic normality of the estimators we follow the approach of Geyer (1994), which makes use of the tangent cone of the parameter space. The definition and properties of tangent cones can be found in Rockafellar and Wets (1998, ch. 6). Using Theorem 6.31 of Rockafellar and Wets (1998), the tangent cone of Θ at $\boldsymbol{\theta}_0$ is

$$\begin{aligned} \mathscr{T}_0 &= \{\boldsymbol{\delta} \in \mathbb{R}^s : \nabla h^C_{kl}(\boldsymbol{\theta}_0)^T \boldsymbol{\delta} = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_1, \\ &\quad \nabla h^D_{kl}(\boldsymbol{\theta}_0)^T \boldsymbol{\delta} = 0, \quad k = 1, \ldots, l, \quad l = 1, \ldots, p_2\}. \end{aligned}$$

The explicit forms of $\nabla h^C_{kl}(\boldsymbol{\theta})$ and $\nabla h^D_{kl}(\boldsymbol{\theta})$ are derived in the Supplementary Material. Let $\mathbf{A}$ be the $s_1 \times s$ matrix with rows $\nabla h^C_{kl}(\boldsymbol{\theta}_0)^T$ and $\nabla h^D_{kl}(\boldsymbol{\theta}_0)^T$, 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 $\hat{\boldsymbol{\theta}}_n$. In addition to $\mathbf{B}$ defined above, it uses Fisher's information matrix,

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

where ∇ and ∇^2 are taken with respect to the parameter $\boldsymbol{\theta}$, and $\mathsf{D}\mathbf{P}(\boldsymbol{\theta})$, the Jacobian matrix of the smoothness penalty vector $\mathbf{P}(\boldsymbol{\theta}) = (P(\mu), \sum_{k=1}^{p_1} P(\phi_k), P(\nu), \sum_{k=1}^{p_2} P(\psi_k))^T$ of (7). Explicit expressions for these derivatives are given in the Supplementary Material. We make an additional assumption:

A4 $\sqrt{n}\boldsymbol{\xi}_n \to \boldsymbol{\kappa}$ as $n \to \infty$, for a finite $\boldsymbol{\kappa}$.

Theorem 2 *Under assumptions A1–A4,* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} \mathrm{N}(-\mathbf{V}\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)^T\boldsymbol{\kappa}, \mathbf{V})$ *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

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

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 $n/s \geq 3$ offer very accurate approximations. This imposes 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 studied the finite sample behavior of the estimators by simulation, assessing their consistency as the sample size increases and the goodness of the approximation of the asymptotic variances.

We generated data from model (2)-(3)-(4) 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, $r = 10$ and $r = 30$, giving expected numbers of observations per curve 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 variation in overall level above or below the mean, whereas the second components ϕ_2 and ψ_2 are contrasts, where

e.g. 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 were of the form $\sigma^2_{u1} = .3^2\alpha$, $\sigma^2_{u2} = .3^2(1-\alpha)$, $\sigma^2_{v1} = .7^2\alpha$ and $\sigma^2_{v2} = .7^2(1-\alpha)$. Two choices of α were considered: $\alpha = .60$ and $\alpha = .75$. The cross-covariance matrix $\mathbf{\Sigma}_{uv}$ was diagonal with elements $\mathbf{\Sigma}_{uv,11} = .7\sigma_{u1}\sigma_{v1}$ and $\mathbf{\Sigma}_{uv,22} = .7\sigma_{u2}\sigma_{v2}$. The random-noise variance was $\sigma^2_\eta = .3^2$. We considered four sample sizes n: 50, 100, 200 and 400. The combinations of rs, αs and ns give us a total of 16 sampling models.

For estimation, we considered cubic B-spline families with five and ten equally spaced knots. The smoothing parameters were visually chosen, 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 $r = 30$. They 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 measures of estimation error we considered the root mean squared errors. For scalar parameters, e.g. σ_η, they are defined as usual: $E^{1/2}\{(\hat{\sigma}_\eta - \sigma_\eta)^2\}$. For functional parameters, e.g. $\mu(x)$, they are defined in terms of the L^2-norm: $E^{1/2}(\|\hat{\mu} - \mu\|^2)\}^{1/2}$. For the random-effect predictors, e.g. the $\hat{u}_{i1}$s, they are defined as $E^{1/2}\{\sum_{i=1}^n(\hat{u}_{i1} - u_{i1})^2/n\}$. The sign of the $\hat{\phi}_k(x)$s and the $\hat{\psi}_k(x)$s, 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 the $\hat{u}_{ik}$s, $\hat{v}_{ik}$s and the elements of $\hat{\mathbf{\Sigma}}_{uv}$ were changed accordingly. For reasons of space we only report here the results for the six sampling models with $\alpha = .75$, $n \leq 200$ and estimators obtained using five-knot splines (Table 1). The rest of the results can be found in the Supplementary Material and are largely in line with the ones reported here. Also given in the Supplementary Material are plots of the functional estimators, which help assess the relative weight 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 big impact on the accuracy of the estimators, particularly of the components ϕ_1 and ϕ_2. A look at the plots in the Supplementary Material reveals that most of the error of $\hat{\phi}_1$ and $\hat{\phi}_2$ comes from the bias rather than the variance, and, for a given n, the bias decreases fast as r increases. Part of the bias of $\hat{\phi}_1$ and $\hat{\phi}_2$ can be attributed to component reversal,

	$r = 10$			$r = 30$		
Parameter	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\Sigma_{uv,11}$	.054	.031	.025	.038	.026	.019
$\Sigma_{uv,21}$	.057	.038	.024	.028	.017	.011
$\Sigma_{uv,12}$	.036	.023	.015	.021	.014	.010
$\Sigma_{uv,22}$	.023	.017	.012	.014	.009	.006
μ	.121	.102	.090	.096	.082	.072
ν	.124	.099	.087	.163	.144	.136
ϕ_1	.738	.515	.376	.436	.261	.188
ϕ_2	.882	.726	.558	.588	.389	.290
ψ_1	.243	.249	.206	.138	.090	.061
ψ_2	.216	.216	.176	.145	.097	.068
σ_{u1}	.065	.057	.029	.039	.027	.020
σ_{u2}	.065	.069	.038	.033	.024	.018
σ_{v1}	.070	.058	.096	.062	.047	.036
σ_{v2}	.071	.082	.065	.037	.027	.018
σ_η	.067	.081	.062	.012	.011	.010
u_{i1}	.217	.184	.170	.154	.140	.134
u_{i2}	.163	.141	.121	.118	.104	.097
v_{i1}	.167	.159	.162	.168	.151	.143
v_{i2}	.153	.148	.138	.105	.083	.072

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, for model with variance proportion $\alpha = .75$.

				$r = 10$					
	$n = 100$			$n = 200$			$n = 400$		
Parameter	True	Med	MAE	True	Med	MAE	True	Med	MAE
$\Sigma_{uv,11}$	.31	.63	.32	.24	.28	.04	.16	.16	.01
$\Sigma_{uv,21}$	.38	.73	.35	.24	.36	.12	.15	.21	.06
$\Sigma_{uv,12}$	.23	.42	.19	.15	.21	.06	.11	.12	.02
$\Sigma_{uv,22}$	.17	.30	.13	.12	.14	.02	.13	.09	.05
				$r = 30$					
$\Sigma_{uv,11}$	.25	.45	.20	.18	.21	.03	.12	.13	.01
$\Sigma_{uv,21}$	.17	.32	.15	.11	.16	.04	.08	.10	.02
$\Sigma_{uv,12}$	.14	.24	.10	.10	.12	.02	.06	.07	.01
$\Sigma_{uv,22}$	.09	.18	.09	.06	.09	.02	.04	.05	.01

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 $\alpha = .75$

which is more frequent for the models with $\alpha = .60$ than for $\alpha = .75$. 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 $\hat{\phi}_1$ and $\hat{\phi}_2$.

Table 2 compares the true finite-sample standard deviations of the elements of $\hat{\boldsymbol{\Sigma}}_{uv}$ with their median asymptotic approximations, and also provides median absolute errors of these approximations, for the estimators based on five-knot splines and models with variance proportion $\alpha = .75$; for $\alpha = .60$ and for ten-knot splines the results are given in the Supplementary Material. The dimension of $\boldsymbol{\theta}$ for five-knot splines is $s = 63$, so Fisher's information matrix estimator $\hat{\mathbf{F}}_0$ is singular for $n = 50$; thus 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 ten-knot splines, where the dimension of $\boldsymbol{\theta}$ is $s = 93$, the tables in the Supplementary Material show that the approximation is accurate for $n \geq 400$. This suggests ratios $n/s \geq 3$ as sufficient for accurate asymptotic approximations of the variances.

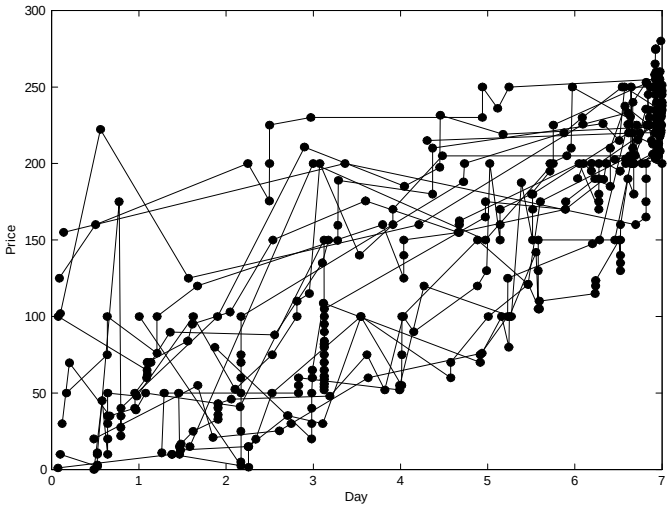

Figure 1: Online Auction Data. Price trajectories of Palm Digital Assistants auctioned at eBay (first 20 trajectories in a sample of 194).

6 Application: online auction data

The eBay auction data mentioned in Section 1 was 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 are 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 that have been called 'early bidding' and 'bid sniping', respectively. 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 at eBay has steadily decreased over the years compared to 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 will not establish causation here, since our models are not intended for that, but 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, and we see that most of the bidding activity tends to occur towards the end of the 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 last two days of the auction will be above the mean. Regarding bid price, Fig. 2(b) shows the mean price trajectory $\nu(t)$ and Fig. 2(d) the components. The first component is associated with price level: items with positive scores on ψ_1 will show prices above the mean over the whole auction period. 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 towards the end.

The estimated cross-covariance and cross-correlation matrices were

$$\hat{\mathbf{\Sigma}}_{uv} = \begin{pmatrix} -256.9 & 48.1 & 22.6 \\ -83.1 & -36.9 & -1.5 \end{pmatrix} \text{ and } \hat{\boldsymbol{\rho}}_{uv} = \begin{pmatrix} -.69 & .41 & .28 \\ -.54 & -.77 & -.05 \end{pmatrix}.$$

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

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

which are very similar to one another. We can conclude that all correlations 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 $\hat{u}_{ik}$s versus $\hat{v}_{ik}$s for the significant components. Normal probability plots of the component scores and the residuals $\hat{\eta}_{ij}$s are shown in the Supplementary Material. The component scores appear to be largely Gaussian; only the $\hat{u}_{i1}$s show a mild departure from normality. The residuals $\hat{\eta}_{ij}$s show tails somewhat heavier than Normal, but no gross outliers are evident.

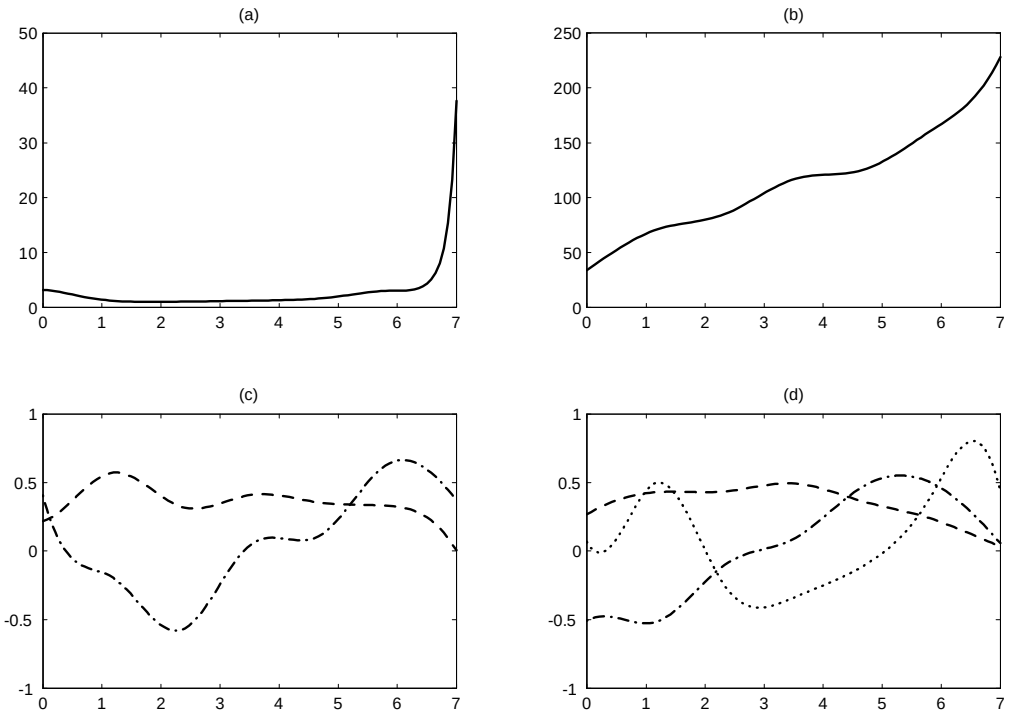

Figure 2: Online Auction Data. (a) Baseline intensity function of bidding time process. (b) Mean price trajectory. (c) Components of bidding time process, ϕ_1 (dashed line) and ϕ_2 (dash-dot line). (d) Components of price trajectories, ψ_1 (dashed line), ψ_2 (dash-dot line) and ψ_3 (dotted line).

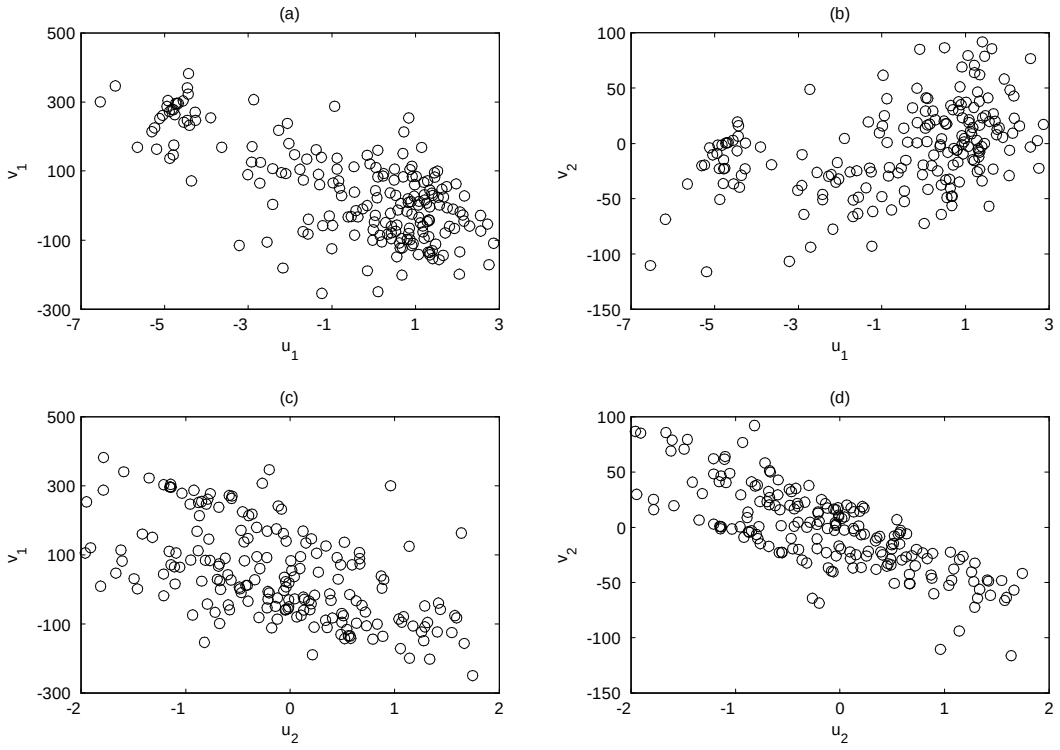

Figure 3: Online Auction Data. Scatterplots of component scores of the bidding time process versus component scores of price trajectories.

These results are in line with 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 snipping 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 largest and smallest scores v_1 and v_2. Figure 4(a) shows the item with largest v_1 score, and consequently low u_1 score: an expensive item that attracted only two bets. Figure 4(b) shows the opposite, the item with lowest v_1 score and consequently large u_1 and u_2 scores: an underpriced item that attracted a lot of bids towards the end of the auction, a typical case of bid sniping. Figure 4(c) shows the item with largest v_2 score, and consequently large u_1 score but low u_2 score: and item that started off with a low price and attracted many bids at the beginning of the auction, which sent the price above the mean early in the auction period and

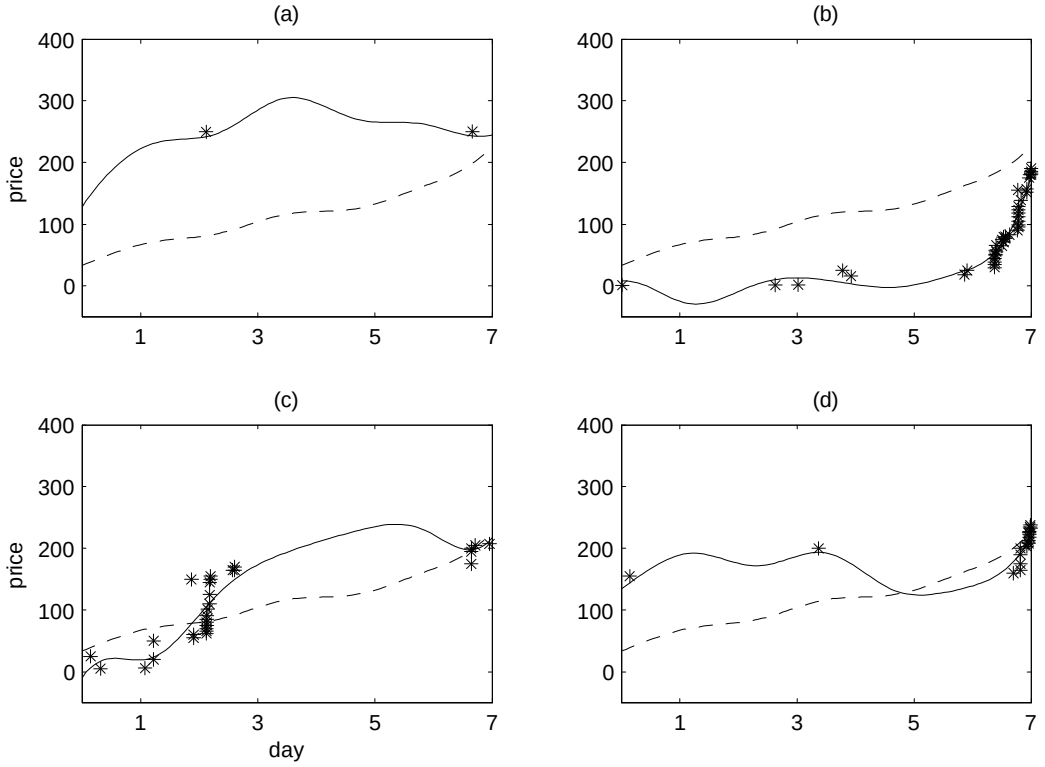

Figure 4: Online Auction Data. Estimated price trajectories (solid line) and mean price trajectory (dashed line) along with actual bets (asterisks) for items with (a) largest score on first Y-component, (b) lowest score on first Y-component, (c) largest score on second Y-component, and (d) lowest score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

then did not attract many late bidders. Figure 4(d), the item with 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 towards the end some lower bids are placed (an unusual but possible situation) which triggered bid snipping.

7 Discussion

In this paper 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 had been considered in the literature, that had been done separately up to this point. Work on canonical correlation analysis

for sparse data (Shin and Lee, 2015) is not really applicable in this setting, because instead of two sparsely observed functional variables we have a single functional variable and a random grid, which involves a completely different model and estimation process.

Our model allows statistical inference for the correlations between 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 asymptotic bias. When the latter is negligible, for example when the target functions are smooth and the basis family used for estimation is large enough, the asymptotic approximation is very accurate, as we showed by simulation and example in Sections 5 and 6. However, if the target functions were more irregular and the asymptotic bias more significant, a 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 whose distributions 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 like gross outliers or very heavy-tailed distributions most likely will. For reasons of space we could not embark on a thorough robustness analysis in this paper, but the model and maximum likelihood estimators we proposed can be easily modified to accommodate heavier-tailed distributions, like Student's t distributions, for the latent variables. This is also a matter for future research.

8 Acknowledgement

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, pp. 1–75. Springer, New York.

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

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* **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* **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.

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. (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* **79** 381–404.

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

Ramsay, J.O., and Silverman, B.W. (2005). *Functional Data Analysis (second 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.

Shin, H., and Lee, S. (2015). Canonical correlation analysis for irregularly and sparsely observed functional data. *Journal of Multivariate Analysis* **134** 1–18.

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.

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.