Functional Data Analysis for Sparse Auction Data

Bitao Liu & Hans-Georg Müller
Department of Statistics
University of California, Davis

Abstract

Bid arrivals of eBay auctions often exhibit "bid sniping", a phenomenon where "snipers" place their bids at the last moments of an auction. This is one reason why bid histories for eBay auctions tend to have sparse data in the middle and denser data both in the beginning and at the end of the auction. Time spacing of the bids is thus irregular and sparse. For nearly identical products that are auctioned repeatedly, one may view the price history of each of these auctions as realization of an underlying smooth stochastic process, the *price process*. While the traditional Functional Data Analysis (FDA) approach requires that entire trajectories of the underlying process are observed without noise, this assumption is not satisfied for typical auction data. We provide a review of a recently developed version of functional principal component analysis (Yao *et al.*, 2005), which is geared towards sparse, irregularly observed and noisy data, the principal analysis through conditional expectation (PACE) method. The PACE method borrows and pools information from the sparse data in all auctions. This allows the recovery of the price process even in situations where only few bids are observed. In a modified approach, we adapt PACE to summarize the bid history for varying current times during an ongoing auction through time-varying principal component scores. These scores then serve as time-varying predictors for the closing price. We study the resulting time-varying predictions using both linear regression and generalized additive modelling, with current scores as predictors. These methods will be illustrated with a case study for 157 Palm M515 PDA auctions from e-Bay, and the proposed methods are seen to work reasonably well. Other related issues will also be discussed.

1 Introduction

eBay.com is today's biggest global online auction market place. It provides a convenient environment for millions of sellers and buyers to carry out real-time transactions on the internet. The fact that eBay makes complete auction information available to the public provides opportunities for researchers to analyze online bidding behavior. This may lead to improved transaction strategies, benefiting both sellers and bidders. Although there are millions of different auctioned products available on eBay at any moment, we only focus on online transactions of a similar or same type of product, auctioned under the same modalities (same duration setting and same billing currency (US dollars)) and during the same time period.

We refer to Jank & Shmueli (2005a) and Reddy & Dass (2006) for some prior functional data analysis approaches for online auction data; this previous work is largely based on approaches described in Ramsay & Silverman (2002, 2005). The time spacing of the bids from each auction is usually sparse for long periods during the auction and often becomes very dense as the auction is approaching its end. This phenomenon, caused by bidders who place their bids at the last moment, when an auction is near its close, is known as sniping. Since the auctioned products that we are interested in are very much alike, the corresponding realizations of the price process for these products at different auctions will be considered to be i.i.d., although at a more detailed level, some dependencies between auctions that are close in calendar time might be present. We ignore such possible dependencies in the following and make the assumption that the observed bids from each auction can be viewed as measurements of the realization of an underlying random smooth stochastic process, the *price process.*

Functional data analysis, and especially functional principal component modelling, provides useful tools for analyzing such data (see also Jank & Shmueli, 2006). Traditional functional principal component analysis (FPCA) requires dense and regular design data. Due to the high degree of data sparsity and time irregularity present in the auction data, it is therefore necessary to adjust this methodology. For this purpose, we adopt the principal analysis through conditional expectation (PACE) method that was developed in Yao *et al.*

(2005). This will allow us to recover price trajectories even if only one or two bids are available for a given auction, provided that the pooled timings of bids from all auctions included in the data form a dense grid. An alternative is to fit a functional random effects models (James *et al.*, 2000; Rice & Wu, 2001), which handles sparse functional data through a pre-specified function basis such as B-Splines.

The paper is organized as follows. In Section 2, we provide an overview of the auction system on eBay.com. In Section 3, we review the methodology to recover individual trajectories by the PACE method, which will be applied to the estimation of both log price process and the log price ratio process. In a modified approach, we adapt PACE to summarize the bid history at varying current times during an ongoing auction, providing time-varying principal component scores. These can be used to predict the closing price while an auction is still ongoing, providing instant updates when a new bid is registered during an ongoing auction. In Section 4, we illustrate the methods with data for 158 Palm M515 PDA auctions. We also present the prediction of the closing price for each auction using linear regression and generalized additive models, with the time-varying functional principal component scores as predictors. Other issues such as monotonicity of the fitted price curves will be briefly discussed.

2 Online auctions at eBay.com

The most common auctions on eBay are single-item auctions, which are organized as second-price auctions, in which bidders submit confidential bids and the bidder who first offers the highest bid wins the auction. However, the winning bidder is only obliged to pay the second-highest bid. Dutch auctions are used by eBay in the multiple item case, where the price starts out high as set by the seller and drops until someone wins the item. In this paper, we only consider the single-item auction case.

The major source of eBay's revenue is generated through auction listings and selling fees. All sellers are charged a nonrefundable Insertion Fee to enter the auction listing, with the amount depending on the starting price or reserve price (fully refundable if an item is sold) of an item. If an item is sold successfully, the seller is also charged a Final Value Fee,

the amount depending on a certain percentage of the closing price. Moreover, eBay offers a list of optional features, such as *Buy It Now*, *eBay Picture Services Fees* etc., for sellers to enhance their listings for a fee. A seller can choose the listing duration to be either 1, 3, 5, 7 or 10 days. Here, we consider 7-day auctions only, the most popular auction for single-item listings.

eBay uses an automatic bidding system for the single-item auctions that we consider here. Bidders enter the maximum amount they are willing to pay for the item, the so-called WTP (willing-to-pay) values or proxy bids. Current price, the real-time price displayed in eBay's web page is referred to as *live bid* by Jank & Shmueli (2005b). This price plays a key role in the decision-making for any bidder participating in an auction, since any new WTP values are required to be no less than the current price plus the pre-set bid increment corresponding to the current price. During the auction, eBay places new bids on behalf of a bidder according to the pre-set bid increment table, so that the bidder can bid against other bidder's maximum bid, until his or her WTP value is reached. The WTP values form the bid history and they are viewable to other bidders in an ongoing auction. However, the real time highest WTP value is always displayed the same as the current price, while the actual current highest WTP value can be higher but it is not disclosed to other bidders until the initiation of another higher WTP value from a different bidder, with the exception that the winner's WTP value is always kept confidential to other parties. eBay stores the completed listings for all auctions that ended over the past 15 days and they are available to any registered eBay user. The completed listings contain useful reference information for sellers and bidders. The listing for each auction contains a detailed bid history which includes: description of the item, item number, closing price, quantity of items auctioned, number of proxy bids received during the auction, and the WTP values (except for the winner's proxy bid), displayed in descending order, along with the corresponding bidder information, date and time.

It is important to observe that because the actual highest WTP value at any given time is hidden from other bidders during an auction, it is common that some of the earlier WTP values are higher than some later WTP values. For instance, assume the live bid for an

auction is currently \$30. Suppose **Bidder A** is currently the highest WTP value holder, say at \$50. Because the amount of WTP \$50 is invisible to other bidders, **Bidder B** joins the auction and places \$35 as his or her WTP value, causes the current price to jump to \$35 (the second highest WTP) + 1 (the bid increment corresponding to \$35) = \$36, that is, the system would now show the current price as \$36 and **Bidder B** has been outbid. Therefore, although the live bids themselves are monotone increasing over an auction, the corresponding WTP values often are not monotone.

From the data collection point of view, the WTP values are much easier to obtain than the real-time live bids, since the complete bid history of any eBay auction that ended in any prior 15 day period can be searched and retrieved from the publicly available eBay web pages. During an ongoing auction, the live bid (or current price) is displayed dynamically as a single number in eBays's auction web page. This number increases whenever there is an update on the second highest WTP value for the auction. Information from eBay is needed to obtain the entire live bid history for any completed auction. However, even without access to this information, one can recover almost all live bids from the WTP information available in the bid history, by making use of the bid increment table which is defined at eBay[1].

The one exception of an auction characteristic about which information cannot be recovered, to the best of our knowledge, is the presence of a secret reserve price for a given auction; it appears that the completed listings from eBay do not contain information about the existence of this feature of an auction and it is only viewable by the bidders in an on-going auction listing. If an auction indeed features a secret reserve price, the converted live bids will differ from the actual live bids by at most one record, namely a live bid where the secret reserve price has come into effect. As an example, assume an auction contains a secret reserve price of \$100. Suppose the current live bid is \$80 with the highest WTP value being \$90 by **Bidder A**. Now, **Bidder B** comes in with a WTP of \$120. eBay's automatic bidding system will institute a special price jump due to the secret reserve price, that is, the live bid will jump to \$100, while without the presence of special reserve, it would have

[1]http://pages.ebay.com/help/buy/bid-increments.html

been \$90 (the second highest WTP) + \$1 (the bid increment corresponding to \$90) = \$91. However, the next converted bid will no longer be influenced by the secret reserve price, so only this one-time increment of the live bid will differ from the value it would have had in the absence of a secret reserve price. Continuing this example, an incoming **Bidder C** will be prompted by the eBay system to enter a WTP value that is \$100 or more. Suppose **Bidder C** enters \$105, then the converted bid will be \$105 (the second highest WTP) + \$2.50 (the bid increment corresponding to \$105) = \$107.50, which is the correct converted live bid even if during the conversion of the data from WTP values to live bids it is not known that a secret reserve price was present.

Since the secret reserve price feature is a costly option for sellers and the unknown size of the special price jump when the reserve is met is potentially discouraging for bidders participating in such auctions, this feature is used more often in auctions of relatively expensive products. The auction data in this paper was collected from WTP values, and in the absence of knowledge about the presence of a secret reserve, the converted live bids can be assumed to reflect the price process reasonably well. We will further discuss details of the conversion of auction bids from WTP to live bids in the case study section. From now on, "bids" is used as a synonym for "live bids".

3 Functional Methods for Sparse Auction Data

3.1 Recovering longitudinal trajectories through Functional Principal Component Analysis

The observed sequence of live bids is assumed to be generated by realizations of an underlying smooth random function, denoted as price process $X(\cdot)$, where X is a square integrable function, defined on a domain $\mathcal{T}$, which is closed and bounded. The price process X is assumed to have unknown smooth mean function $\mathrm{E}X(t) = \mu(t) \in L^2$, and unknown smooth covariance function $\mathrm{cov}(X(s), X(t)) = G(s,t) \in L^2, s, t \in \mathcal{T}$. The covariance function $G(s,t)$ is assumed to have an orthogonal expansion with non-increasing eigenvalues λ_k and

eigenfunctions ϕ_k of the autocovariance operator A_G, defined by

$$(A_G f)(t) = \int_{\mathcal{T}} G(s,t) f(s) ds \tag{1}$$

for any f in L^2. Since G is symmetric and non-negative definite, this linear operator in the Hilbert space $L^2(\mathcal{T})$ is a Hilbert-Schmidt operator (see Courant & Hilbert, 1953). The orthogonal expansion of G is

$$G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t), \tag{2}$$

and the Karhunen-Loève representation (Ash & Gardner, 1975; Rice & Silverman, 1991) of a random trajectory X_i is then given by:

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t). \tag{3}$$

Here, the $\xi_{ik} = \int_{\mathcal{T}} (X_i(t) - \mu(t)) \phi_k(t) dt$ are the functional principal component score (FPCs) of X_i, for $k = 1, 2, \ldots$. These are random variables with $\mathrm{E}(\xi_{ik}) = 0$, $\mathrm{E}(\xi_{ik}\xi_{ik'}) = 0$ for $k \neq k'$, and $\mathrm{E}\xi_{ik}^2 = \lambda_k$, where $\sum_k \lambda_k < \infty$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. In other words, the FPCs form a sequence of uncorrelated random variables with decreasing variances which are given by the eigenvalues.

In practice, the sequence of observed bids differs from the values of the smooth underlying trajectory, and the differences correspond to a "measurement error". In the case of auction data, these differences are best interpreted as random aberrations of prices around the smooth underlying price trajectory, rather than as physical measurement errors. Let T_{ij} be time points at which bids are placed for price trajectory X_i, where $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ and $T_{ij} \in \mathcal{T}$. The number of bids n_i observed for trajectory X_i (i.e., the i-th auction) is assumed to be an i.i.d. random variable, independent of other random variables. With an additive measurement error assumption (see also Rice & Wu, 2001), equation (3) can be connected with actual bids Y_{ij} observed at times T_{ij} through

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

where the ε_{ij} denote the aberrations of the bids from the underlying smooth trajectories, assumed to be i.i.d. with mean 0 and constant variance σ^2, and such that ε_{ij} is independent of ξ_{ik} for all i, k.

Identifying the components of (4) starts with the nonparametric estimation of mean and covariance function, following the proposal in Yao *et al.* (2005). We use local linear scatterplot smoothers for the mean function and local linear surface smoothers for the covariance function (see Fan & Gijbels, 1996). By pooling all observed bids together, across all auctions, we obtain the estimated mean function $\hat{\mu}(t)$ through minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_1\left(\frac{T_{ij}-t}{h_\mu}\right)\{Y_{ij}-\beta_0-\beta_1(t-T_{ij})\}^2 \tag{5}$$

with respect to β_0 and β_1, where $\hat{\mu}(t)=\hat{\beta}_0(t)$, $t\in\mathcal{T}$. Here $K_1(\cdot)$ is a univariate kernel function, usually chosen as a symmetric (around 0) density function. The necessary bandwidth h_μ can be selected through leave-one-curve-out cross-validation (CV) (Rice & Silverman, 1991) or variants, such as generalized cross-validation (GCV). Often, a subjective choice through visualization is more adequate.

The assumptions about the bid aberrations ε_{ij} present in model (4) imply that only the diagonal elements of the covariance are affected, as

$$\operatorname{cov}(Y_{ij},Y_{il}|T_{ij},T_{il})=\operatorname{cov}(X(T_{ij}),X(T_{il}))+\sigma^2\delta_{jl},$$

where δ_{jl} is the Kronecker delta. For "raw covariances" $G_i(T_{ij},T_{il})=(Y_{ij}-\hat{\mu}(T_{ij}))(Y_{ij}-\hat{\mu}(T_{il}))$, $\hat{\mu}(\cdot)$ being the estimated mean function, one finds
$\mathrm{E}[G_i(T_{ij},T_{il})|T_{ij},T_{il}]\approx\operatorname{cov}(Y_{ij},Y_{il}|T_{ij},T_{il})$ and this motivates smoothing the raw covariances while omitting the diagonal terms. Let $\hat{G}(s,t)$ be a smooth estimate of the covariance surface, for example obtained by minimizing

$$\sum_{i=1}^{n}\sum_{1\le j\neq l\le n_i} K_2\left(\frac{T_{ij}-s}{h_{G_1}},\frac{T_{il}-t}{h_{G_2}}\right)\{G_i(T_{ij},T_{il})-(\beta_0+\beta_{11}(s-T_{ij})+\beta_{12}(t-T_{ij}))\}^2 \tag{6}$$

with respect to β_0, β_{11} and β_{12}, where the surface estimate is $\hat{G}(s,t)=\hat{\beta}_0(s,t), s,t\in\mathcal{T}$, according to the local least squares method. Here $K_2(\cdot,\cdot)$ is a bivariate density, and usually

one chooses the smoothing parameters such that $h_{G_1} = h_{G_2} = h_G$. Similarly to the situation for the mean, bandwidths h_G can be selected through CV, GCV or visually.

Estimating eigenfunctions and eigenvalues in model (4) corresponds to finding the solutions $\hat{\phi}_k$ and $\hat{\lambda}_k$ of the eigen-equations,

$$\int_{\mathcal{T}} \hat{G}(s,t)\hat{\phi}_k(s)ds = \hat{\lambda}_k\hat{\phi}_k(t), \tag{7}$$

where the $\hat{\phi}_k$ are subject to $\int_{\mathcal{T}} \hat{\phi}_k(t)^2dt = 1$ and $\int_{\mathcal{T}} \hat{\phi}_k(t)\hat{\phi}_m(t)dt = 0$ for $m \neq k$. The estimated eigenvalues and eigenfunctions are then obtained by spectral decomposition of the discretized smoothed covariance (Rice & Silverman, 1991; Capra & Müller, 1997).

Traditional FPCA uses numerical integration to estimate the FPC scores $\xi_{ik} = \int_{\mathcal{T}}(X_i(t) - \mu(t))\phi_k(t)dt$. For price trajectories X_i, the estimated FPC scores via the integration method would be $\hat{\xi}^I_{ik} = \sum_{j=1}^{n_i}(Y_{ij} - \hat{\mu}(T_{ij}))\hat{\phi}_k(T_{ij})(T_{ij} - T_{i,j-1})$, with $T_{i,j-1} = 0$, assuming the T_{ij} are ordered by size. This method requires the observed bids to be dense and regularly spaced, and recorded without aberrations from the price trajectories. In this situation, $Y_{ij} = X_{ij}$. However, this estimator suffers seriously if the observed bids are noisy, sparse or irregularly spaced.

As an alternative to numerical integration, Yao *et al.* (2003) proposed a shrinkage estimator to estimate the ξ_{ik} for dense and noisy data with missing values. The PACE method developed by Yao *et al.* (2005) aims at the estimation of ξ_{ik} for sparse, irregularly observed and noisy data, and thus provides a natural approach for the auction data. Here we give a brief overview of some pertinent details. Let $\widetilde{\mathbf{X}}_i = (X_i(T_{i1}), \ldots, X_i(T_{in_i}))^T$, $\widetilde{\mathbf{Y}}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, $\widetilde{\boldsymbol{\mu}}_i = (\mu(T_{i1}), \ldots, \mu(T_{in_i}))^T$, $\widetilde{\boldsymbol{\phi}}_i = (\phi(T_{i1}), \ldots, \phi(T_{in_i}))^T$, and assume ξ_{ik} and ε_{ij} in (4) are jointly Gaussian. The best prediction of the k-th FPC score ξ_{ik} of X_i, given the observed bids for this price trajectory, is the conditional expectation, which is (Yao *et al.* (2005))

$$\xi^*_{ik} = \mathrm{E}[\xi_{ik}|\widetilde{\mathbf{Y}}_i] = \lambda_k\boldsymbol{\phi}^T_{ik}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_i}(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i), \tag{8}$$

where $\boldsymbol{\Sigma}_{\mathbf{Y}_i} = \mathrm{cov}(\widetilde{\mathbf{Y}}_i, \widetilde{\mathbf{Y}}_i) = \mathrm{cov}(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{X}}_i) + \sigma^2\mathbf{I}_{n_i}$. Substituting estimates for $\lambda_k, \boldsymbol{\phi}_{ik}, \boldsymbol{\Sigma}_{\mathbf{Y}_i}$, and $\boldsymbol{\mu}_i$, the estimates for the predicted ξ_{ik} then become

$$\hat{\xi}_{ik} = \hat{\mathrm{E}}[\xi_{ik}|\widetilde{\mathbf{Y}}_i] = \hat{\lambda}_k\hat{\boldsymbol{\phi}}^T_{ik}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathbf{Y}_i}(\widetilde{\mathbf{Y}}_i - \hat{\boldsymbol{\mu}}_i), \tag{9}$$

where the $(j,l)^{th}$ element of $\hat{\mathbf{\Sigma}}_{\mathbf{Y}_i}$ is $(\hat{\mathbf{\Sigma}}_{\mathbf{Y}_i})_{j,l} = \hat{G}(T_{ij}, T_{il}) + \hat{\sigma}^2\delta_{jl}$. Here, $\hat{G}(T_{ij}, T_{il})$ can be estimated from (6) and $\hat{\sigma}^2$ is estimated as described in equation (2) of Yao *et al.* (2005).

The idea for the estimate of $\hat{\sigma}^2$ is simply to compare a smoothed version of just the diagonal elements of the covariance matrix with the diagonal of the resulting surface estimate along the direction perpendicular to the diagonal of the covariance matrix. Under the Gaussian assumption, the estimator (9) targets the best prediction, however not the actual values of the FPC scores. Assuming the major modes of variation of the infinite-dimensional price processes X_i correspond to the first K eigenfunctions, then the estimate for the predicted price trajectory X_i is

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

To select a reasonable number of K eigenfunctions to approximate the infinite-dimensional process, one can compute the one-curve leave-out score (Rice & Silverman, 1991), aiming at minimizing the leave-one-curve-out prediction error

$$CV(K) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\{Y_{ij} - \hat{Y}_i^{(-i)}(T_{ij})\}^2, \tag{11}$$

where $\hat{Y}_i^{(-i)}(T_{ij}) = \hat{\mu}^{(-i)}(T_{ij}) + \sum_{k=1}^{K}\hat{\xi}_{ik}^{(-i)}\hat{\phi}_k^{(-i)}(T_{ij}), j = 1, \ldots, n_i$, are the predicted bids for price trajectory X_i. These values are estimated after removing the data for the trajectory X_i itself, where the estimated mean function and eigenfunctions are evaluated at the observed bid times for this trajectory.

The construction of the cross-validation score for a given number of K components involves fitting model (4) n-times, which can be computationally expensive when n is large. Moreover, in practice, the final choice of K by the CV method tends to be large and often unstable as CV does not sufficiently restrict the degrees of freedom. As an alternative to the CV method, Yao *et al.* (2005) proposed AIC (Shibata, 1981) and BIC (Schwarz, 1978) type of criteria based on a pseudo-Gaussian log-likelihood, computed conditionally on the observed $\hat{\xi}_{ik}$,

$$\hat{L} = \sum_{i=1}^{n}\left\{-\frac{n_i}{2}\log(2\pi) - \frac{n_i}{2}\log\hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2}\left(\widetilde{\mathbf{Y}}_i - \widehat{\widetilde{\boldsymbol{\mu}}}_i - \sum_{k=1}^{K}\hat{\xi}_{ik}\widehat{\widetilde{\boldsymbol{\phi}}}_{ik}\right)^T\left(\widetilde{\mathbf{Y}}_i - \widehat{\widetilde{\boldsymbol{\mu}}}_i - \sum_{k=1}^{K}\hat{\xi}_{ik}\widehat{\widetilde{\boldsymbol{\phi}}}_{ik}\right)\right\}.$$

Criteria for the selection of K are minimizing either $AIC(K) = -2\hat{L} + 2K$ or $BIC(K) = -2\hat{L} + K\log(N)$ where $N = \sum_{i=1}^{n} n_i$. Note that for each given K one may compute $AIC(K)$ or $BIC(K)$ by fitting the model (4) once only, which is computationally more efficient than the CV method.

In practice, we found the choices obtained from these criteria to be reasonable and usually better than CV. In addition to the above criteria, choice of K by means of a scree plot is relatively simple and often quite adequate. To construct a scree plot, one needs to fit a model with a relatively large number of FPCs, say M components, so that λ_M is close to zero. Next, one plots λ_k against k for $k = 1, \ldots, M$. The idea is to choose the number K such that eigenvalues beyond λ_K can be ignored in terms of the fraction of additional variance they would explain.

The proportion of variation left unexplained by the truncated expansion in (10) can be defined as

$$V(K) = 1 - \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{M} \lambda_k} \tag{12}$$

and is estimated by plugging in eigenvalue estimates. The selected K will be a point where ideally, the function $V(K)$ starts to flatten after an initial decline. The estimators in model (4) are shown to be asymptotically consistent under mild conditions in Yao *et al.* (2005), where also asymptotic point-wise and simultaneous confidence bands for the predicted individual trajectories are derived.

3.2 Analyzing evolving bid trajectories

For nearly identical products that are auctioned repeatedly over a given period of time under similar auction settings (such as same duration of an auction, same currency in use etc.), one can model the bid history for each auction as "price process" using the FDA approach as described in the previous section, under the assumption that these auctions are independent of each other.

Through the PACE method, one may recover the price trajectory for each online auction. Sometimes the trajectories underlying a bid history observed up to a current time t of an ongoing auction are of particular interest, especially for online applications where an auction

has been incompletely observed but one aims at characterizing the ongoing auction at an intermediate time. The PACE method can be readily adapted to this setting.

Let t, where $0 < t < T$, be the current length of time an auction has been running, where $\mathcal{T} = [0, T]$ is the domain of the price process and the auction is assumed to begin at time 0. Then the price process for an auction observed up to time t is denoted by $X(s, t)$, $0 \leq s \leq t$ (see also Müller & Zhang, 2005). The corresponding mean and covariance function are (i) $\mathrm{E}X(s, t) = \mu(s, t)$, with $\mu(s, t) = \mu(s)$ for $s \leq t$, and (ii) $\mathrm{cov}(X(s_1, t), X(s_2, t)) = G_t(s_1, s_2)$, with $G_t(s_1, s_2) = G(s_1, s_2)$ for $0 \leq s_1, s_2 \leq t$, respectively. The orthogonal expansion of the covariance function is $G_t(s_1, s_2) = \sum_{k=1}^{\infty} \lambda_{kt}\phi_{kt}(s_1)\phi_{kt}(s_2)$, where $\lambda_{1t} \geq \lambda_{2t} \geq \ldots \geq 0$ are eigenvalues and $\phi_{1t}(\cdot), \phi_{2t}(\cdot), \ldots$, are orthonormal eigenfunctions of the autocovariance operator A_{G_t}. The eigenvalues and eigenfunctions correspond to the solution of the eigen-equations $\int_0^t G_t(s, u)\phi_{kt}(u)du = \lambda_{kt}\phi_{kt}(s)$. Then the Karhunen-Loève representation of price trajectory of X_i up to time t corresponds to

$$X_i(s, t) = \mu(s, t) + \sum_{k=1}^{\infty} \xi_{ikt}\phi_{kt}(s), \quad 0 \leq s \leq t, \quad t \geq 0, \tag{13}$$

where $\xi_{ikt} = \int_0^t \{X_i(s, t) - \mu(s, t)\}\phi_{kt}(s)ds$, the FPC for the trajectory X_i in $[0, t]$. As usual, we have $\mathrm{E}(\xi_{ikt}) = 0$ and $\mathrm{E}(\xi_{ikt}\xi_{ik't}) = 0$ for $k \neq k'$ and $\mathrm{E}(\xi_{ikt}^2) = \lambda_{kt}$ with $\sum_k \lambda_{kt} < \infty$.

The bids again may be viewed as contaminated values of the price trajectory as above, and estimation follows the same principles as described in section 3.1, with the modification that the input data (or observed bids) Y_{ijt} are restricted to $T_{ij} \in [0, t]$ only. As t varies, one thus obtains a time-varying version of the PACE method, which provides continuous updates for the characteristics of the price process as it evolves over time.

3.3 Predicting closing price through bid histories

In addition to modeling the evolution of price trajectories for each auction, we are also interested in predicting the final price using currently available bid information at each time t, $0 < t < T$. More specifically, instead of using only a current bid at or near time t as a predictor, we aim at relating the closing price of an auction, i.e., the value of the price process at time T, to the entire price process from 0 to current time t. The underlying

rationale for this approach is that the price history contains much richer information than current price alone.

If price processes are well approximated by projecting on the function space spanned by the first K eigenfunctions, it suffices to use the first K time-varying functional principal component scores ξ_{kt} to represent trajectories (price processes) X_i up to current time t. We may assume various regression models with the final price Y^* as response and ξ_{kt} as predictors. For example, Hastie & Tibshirani (1993) proposed a series of varying coefficient generalized linear models for cross-sectional data, while Hoover *et al.* (1998) and Fan & Zhang (1999, 2000) considered modelling varying-coefficient functions for longitudinal data.

When $g_t(\cdot)$, the time-varying link function, is chosen as the identity function, the varying coefficient linear model becomes

$$\mathrm{E}(Y^*|\xi_{1t},\ldots,\xi_{Kt}) = \beta_{0t} + \sum_{k=1}^{K} \xi_{kt}\beta_{kt}, \tag{14}$$

where Y^* is assumed to be an independent Gaussian random variable, the varying intercept function β_{0t} is the mean of the price function at time t, the ξ_{kt} are the uncorrelated individual random components with zero mean, and β_{kt} are the corresponding varying coefficients for ξ_{kt}. The least squares estimates of the varying coefficients of $\widetilde{\beta}_t = (\beta_{0t},\ldots,\beta_{Kt})^T$ will be usually obtained for each t, $\hat{\widetilde{\beta}}_t = \mathrm{argmin}_{\widetilde{\beta}_t} \sum_{i=1}^{n}\{Y_i^* - (\beta_{0t} + \sum_{k=1}^{K} \hat{\xi}_{ikt}\beta_{kt})\}^2$. The predicted final price for an auction with price process $X_i(s), s \le t$ is then

$$\hat{Y}_{it}^* = \hat{\beta}_{0t} + \sum_{k=1}^{K} \hat{\xi}_{ikt}\hat{\beta}_{kt}. \tag{15}$$

A varying coefficient generalized additive model (Hastie & Tibshirani, 1990) provides a flexible alternative to predict the final price. Under the identity link, model (14) becomes

$$E(Y^*|\xi_{1t},\ldots,\xi_{Kt}) = \sum_{k=1}^{K} f_{kt}(\xi_{kt}) \tag{16}$$

where the $f_{kt}(\cdot)$ are unknown smooth functions with $\mathrm{E}f_{kt}(\xi_{kt}) = 0$ and the predicted final price is

$$\hat{Y}_{it}^* = \hat{\beta}_{0t} + \sum_{k=1}^{K} \hat{f}_{kt}(\hat{\xi}_{ikt}). \tag{17}$$

There are currently two available R packages which provide estimation for model (16) with various link functions. One is the **gam** package (author: Trevor Hastie) which uses the local scoring algorithm and iteratively fits the weighted additive models by backfitting. An alternative version is the **mgcv** package (author: Simon Woods) which estimates the unknown functions through penalized iteratively reweighted least squares (IRLS) (see Wood, 2000), with a focus on automatic smoothing parameter choice via GCV. For any fixed t, model (17) involves the estimation of $\hat{\beta}_{0t}$ and of the unknown functions $\hat{f}_{kt}$. Since we are interested in fitting many models with varying values of t, the automatic bandwidth choice of **mgcv** is an attractive feature, and this version was selected for our applications.

To assess goodness-of-fit for models (14) and (16), one can use the mean square prediction error function,

$$\text{MSPE}(t) = \frac{1}{n}\sum_{i=1}^{n}(Y_i^* - \hat{Y}_{it}^*)^2. \tag{18}$$

As t moves closer to the auction end time T, the $\hat{\xi}_{ikt}$ contain more and more information about the auction; thus, MSPE is expected to be monotone falling as t increases, apart from random fluctuations. We will illustrate this feature for the Palm M515 PDA auction example in the case study section below.

4 Case study

4.1 Preprocessing of data on Personal Digital Assistant auctions

The data used in this paper was collected by Wolfgang Jank[2]. It contains 158 auctions of Palm M515 Personal Digital Assistants (PDA) that took place between March and May, 2003. The bid values correspond to WTP values. The WTP values are converted into live bids according to the following conversion rules: (i) the first bid (opening bid) in each of the auction records is set by the seller at the start of each auction and is considered as the first live bid; (ii) the WTP value that is placed by the first bidder is considered the opening bid; (iii) any other current live bid is equal to current second highest WTP value plus the bid increment corresponding to this price, as discussed above, with the constraint that the sum

[2]http://www.smith.umd.edu/ceme/statistics/data.html

will not exceed the current highest WTP value; (iv) any non-increasing live bids causing by new update(s) from the current highest WTP value holder will be excluded; (v) the closing price is the same as the winner's bid (the converted live bid corresponding to the second last WTP value). The conversion relies on the assumption that no secret reserve price is present in any of the auctions. We will discuss this assumption below. One auction that did not contain any information about the bidder ID was excluded, since rule (iv) can't be applied in this case. We further removed another auction that contained identical bids of the same value ($199) placed by the same and only bidder for the auction, and the analysis reported here is therefore based on 156 auctions.

The start of all auctions is recalibrated to time 0, so that the first day of an auction ends after 24 hrs, and all auctions end at the end of day 7, corresponding to 168 hrs. Although the recorded bid time is accurate up to seconds, we converted the time unit into hours, that is, the time domain is between 0 and 168 hours for all auctions. Since the difference between the minimum bid ($0.01) and the maximum bid ($283.50) is large, we decided to transform the bids to the log-scale, which leads to live bid values ranging from -4.61 and 5.65 and also brings the price data closer to Gaussianity which is required for the predictions of individual trajectories in the PACE method. The analysis is therefore implemented for the log price process.

4.2 Analysis of log-bids

We begin the analysis by applying the PACE method to model the log price process for each auction on time domain $t_{ij} \in [0, 168)$ (time units in hours), for auctions indexed by $i = 1, \ldots, 156$ and bids indexed by $j = 1, \ldots, n_i$. Though the closing prices for each auction were recorded, we only used the live bids that were registered before the end of the 168^{th} hour (the end of day 7) as input data.

The number of live bids per auction ranged from 9 to 52. Figure 1 shows the boxplot of the daily number of aggregated bids from all 156 auctions, based on the 7-day scale. The median number of bids placed on the first day of an auction was 3 and then it dropped to and stayed at 1 for the second and third day. On day 4, the median number of bids

per auction fell to essentially zero, follow by a small increase on day 5, where the median number of bids per auction rose to 1. On day 6, this number climbed to 2 and it shot up rapidly to 9 on day 7, the last day of an auction. The highly irregular spacing of the times when bids were recorded reflects high variability of bidding behavior and the sniping phenomenon described earlier.

When modelling individual price trajectories, the classical FPCA approach does not work well for the estimation of the principal component scores, due to the difficulties caused by the irregular bid times for the numerical integration step. The PACE method however allows recovery of individual trajectories, as long as the pooled bid times over all auctions are dense in the domain, and also all pairs of bid times are dense on the domain squared; both assumptions are satisfied for the auction data. To model log price processes, we assume the observed bids are contaminated by aberrations which are added to the underlying smooth log price trajectories. As mentioned before, no secret reserve price is included in the modeling for any of the auctions, and should this assumption be violated for some of the auctions, the resulting bid conversion error will simply correspond to one of the bid aberrations that our model allows for.

As a first step in the analysis of the auction data, we pooled all bids together and used (5) to estimate the mean function for the log price trajectories (using the Epanechnikov kernel as weight function K_1). The bandwidth choices resulting from both one-curve-leave-out CV or GCV appeared to be undersmoothing, and we therefore augmented them through visual assessment. We found $h_\mu = 12$hr to provide a reasonable fit for the mean function, which can be seen in the top right corner of Figure 3. The bids aggregated from all auctions are displayed as grey dots and the estimated mean function is shown overlaid as a solid black curve. On average, log prices are seen to increase rapidly around the first day of an auction and then the increases taper off until the final phase of an auction.

Next, we use (6) to obtain the estimated smooth covariance function for the log price trajectories. Here the input data are the raw covariances obtained by pooling all auction data together and removing the diagonal elements which are contaminated by bid aberrations. The bivariate kernel in (6) is chosen as the product of two one-dimensional Epanechnikov

kernels. Bandwidth choices obtained from CV appeared to be undersmoothing while the GCV bandwidths (42hr, 42hr) were found adequate. We then apply spectral decomposition to the smooth covariance function at a pre-selected time grid to obtain estimated eigenvalues and eigenfunctions, the latter defined on this time grid. The estimated predictions for the FPC scores were then computed via (9).

In the spectral decomposition step, it is crucial to determine a reasonable number K of included components, representing the relevant modes of variation of the log price trajectories. Choices from CV, AIC or BIC were found to be too large. As an alternative selection tool we applied the scree plot approach based on fraction of variance explained (12) to determine K. It turned out that the choice $K = 3$ (out of $M = 20$) was satisfactory, as the first 3 components accounted for 97.65% of total variation. The resulting estimated eigenfunctions can be seen in the last column of Figure 3. The first eigenfunction (72.69% of total variation) represents an overall trend and has a similar shape as the mean function. The second eigenfunction (22.44% of total variation) increases sharply in the first 3 days and then declines slowly afterwards. This component may represent the dynamics of early price increments, up to mid-auction. The third eigenfunction (2.52% of total variation) increases in the first 2 days, followed by a decrease in the next 2 days, and then by a slow increase until the end of auction.

Figure 2 demonstrates the obtained fits for 16 estimated log price trajectories corresponding to 16 randomly selected auctions. The trajectories were fitted through (10) with $K = 3$. Overall, the price trajectories fit the data reasonably well. As the observed bids are monotone increasing within an auction, we might assume the same holds for the log price process. The PACE method does not enforce such a constraint in the estimation procedure and the fitted log price trajectories are not guaranteed to be monotone increasing. For example, the four graphs in the bottom panel display cases of non-monotone fitted log price curves. The estimated trajectories from PACE are displayed as broken lines. A simple device to monotonize the log price trajectories is to apply the "Pooled-Adjacent-Violators Algorithm" (PAVA), as introduced in Barlow *et al.* (1972), to the fitted curves obtained from the PACE method. Any non-increasing estimated bids will be successively pooled to-

gether with their adjacent values and replaced by the average of the pooled values until no more non-monotone estimated bids are encountered. The R function **isoreg()** is one of the implementations of the PAVA algorithm and is used here to obtain the monotonized curves for the four auctions shown in the bottom panel of Figure 2. The monotonized trajectories are displayed as solid black lines in the graphs. The flat line segments in each of these graphs reflect the PAVA averaged values of the estimated bids for non-monotone segments. For further discussion of PAVA in the context of nonparametric smoothing methods we refer to Friedman & Tibshirani (1984).

To further study the auction dynamics, we also applied the time-varying PACE method (13), choosing various current times t. For demonstration purposes, we illustrate the evolution of mean and the first eigenfunction for the log price processes based on bid histories observed up to current times $t = 24$hr, 72hr and 120hr, as illustrated in Figure 3. The estimated mean functions are shown in the graphs of the top panel. The bandwidth choices for the mean functions were $h_\mu = 11$hr, 9hr, 9hr, respectively for the three times, via visual choice. From these graphs, we find that the estimated mean function for log prices increases sharply within the first day ($t \leq 24$hr) of an auction and it then flattens as time progresses, as seen from the graphs for $t \leq 72$hr (day 3) and $t \leq 120$hr (day 5). The first 3 estimated eigenfunctions are displayed in the next three rows of Figure 3. They were obtained from the spectral decomposition of the three estimated covariance surfaces, choosing bandwidths $h_G = 11$hr (visually), 18hr (GCV), and 30hr (GCV), respectively. For each time t, the first three eigenfunctions explain about $96 \sim 97\%$ of the total variation, with the first eigenfunction explaining about 70%, the second about 20%, and the third about $3 \sim 5\%$ (except for $t \leq 24$hr, which is about 8%). The evolution of all eigenfunctions is smooth and gives rise to similar interpretations as for the fits over the entire time domain of an auction.

One of the advantages of the time-varying PACE method is that one can summarize the bid history up to a current time t via the time-varying FPC scores and then use these scores as predictors for the closing price. We implemented such a scheme and obtained time-varying FPCs for the bid histories observed up to $24, \ldots, 168$ hours. All bandwidths were selected through GCV and the number of included components (FPC scores) through AIC,

which led to an average of about 10 included components. To predict closing price from bid histories to current time t, we compared two regression models with closing price as response and the time-varying FPC scores as predictors. One model was a linear regression using **lm()** in R for (15). As a nonlinear alternative, we also fitted a Generalized Additive Model (Hastie & Tibshirani, 1990) (**gam()**) using the **mgcv** package in R for (17). The unknown smooth functions f_j are estimated through cubic smoothing splines and the smoothing parameters chosen by GCV. To check the goodness-of-fit of the time-varying predictions, we use (18) to calculate the MSPE for different current times t. The results are shown in Figure 4. As expected, prediction errors decline as time progresses. The time-varying GAM clearly outperforms the prediction based on multiple linear regression in this setting.

4.3 Analysis of log price increments

As an alternative to modelling log price processes, we also considered modelling the log price increment processes. Here the observed log bid ratios will be used as input data, rather than log bid values. We removed any observed bids that fell below \$1 (inclusive) or were not strictly larger than previous bids. All second live bids fall into this category, since they are the same as the opening bids. If Y_{ij} is a bid observed at time $T_{ij} \in [0, 168)$ (time in hours) and $Y_{ij} > 1$, we define the log bid ratios as

$$q_{ij} = \log \frac{Y_{ij}}{Y_{i,j-1}} \quad \text{with} \quad Y_{i0} \equiv 1, \quad i = 1, \ldots, 156,\ j = 1, \ldots, n_i.$$

Then $q_{i1} = \log Y_{i1}$ so that the first ratio is relative to \$1.

The analysis and estimation procedures follow exactly those described for the log price processes in section 4.2, substituting responses q_{ij} for Y_{ij}. The resulting estimated mean function for the log price ratio processes is shown in the top left corner of Figure 6 with $h_\mu = 12$hr (chosen visually). Unsurprisingly, the estimated mean function for the log price increment is positive throughout, since the observed bids are monotone increasing. The log bid increments are largest in the beginning and drop off sharply after the first day, after which they continue to decrease mildly. Eigenfunctions are again obtained through spectral decomposition of the smooth covariance surface with $h_G = 12$hr (chosen visually). The scree plot approach indicates that choice of $K = 2$ or 3 components is adequate.

The resulting estimated eigenfunctions are displayed in Figure 6. The fractions of total variation that is explained by the first three eigenfunctions are $89.33\%, 5.77\%$ and 2.89%, respectively. The shape of the first eigenfunction is very similar to the mean function. The second eigenfunction first decreases in the first two days and then slowly increases. The third eigenfunction declines rapidly in the first day, followed by a large increase during the second day and it then flattens. The estimated log price ratio trajectories for the same selected auctions as those shown in Figure 2 are illustrated in Figure 5, based on three FPCs. Note that the estimated trajectories fit the data reasonably well, even with the relatively small number of repeated measurements (observed log bid ratios) per auction. The log bid increments appear to be largest up to the first or second day and then become almost zero. It appears that the effect of different opening prices is attenuated after the second day of the auction. We conclude from this analysis that the log price process itself contains more information about the overall online auction dynamics than does the increment process which mainly reflects the dynamics at the beginning of an auction.

5 Conclusions and discussion

We show how a recent FPCA approach that is designed for the recovery of trajectories from sparse, irregular and noisy longitudinal repeated measurements can be easily adapted to auction data. The implementation of this approach through the PACE method is found to be useful to fit log price processes and log bid ratio processes. Especially the fits for the log price process and the associated eigenfunctions provide interesting insights into the time dynamics of online auctions. Adding a PAVA step easily leads to monotonized fitted trajectories. Alternative monotonized versions with more smoothness could be obtained by coupling PACE with other recent monotonization methods (e.g., Hall & Huang (2001)) and connecting such methods with PACE provides a topic for future research.

In addition to studying online auction dynamics, prediction of the closing price is clearly an important aspect. In section 3.3, we consider the concept of time-varying FPC scores, which summarize the bid history from the beginning to current time t of an auction. This is a particularly attractive feature for online auctions, because one usually has to make a

decision at a current time t, based on available information about the auction history to time t. As we show, prediction based on these scores works well and can be computed for an arbitrary current time t. Our results from the case study indicate that the time-varying GAM provides good predictions and performs uniformly better than prediction by multiple linear regression.

There are many possible extensions of these approaches. For example, one can study the inclusion of other variables that are involved in an online auction. These can be time-dependent, such as the intensity of bids over time and feedback scores over time, or cross-sectional, such as opening bid, seller's rating, indicator for auctions that end during a weekend, etc. Future research may explore the use and extension of existing techniques for functional regression with functional response or generalized functional linear model with scalar response, when the model contains one or more predictor functions in addition to other scalar components.

6 Acknowledgments

We are indebted to Wolfgang Jank for allowing us to use his auction data, for introducing the auction topic to us and for many valuable discussions. This research was supported by NSF grants DMS03-54448 and DMS05-05537.

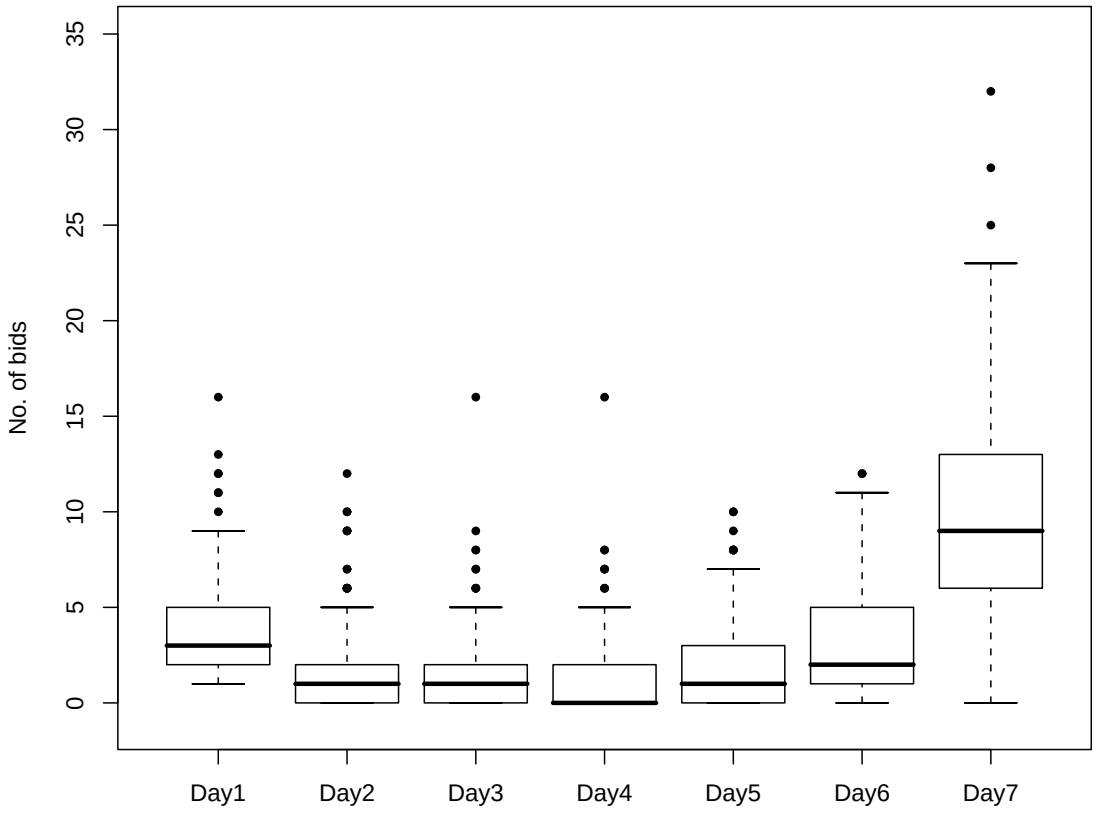

Figure 1: Boxplot of number of aggregated bids from 156 auctions placed on each day of seven day auctions. The solid black dots are potential outliers and the bold solid line represents the median daily number of aggregated bids.

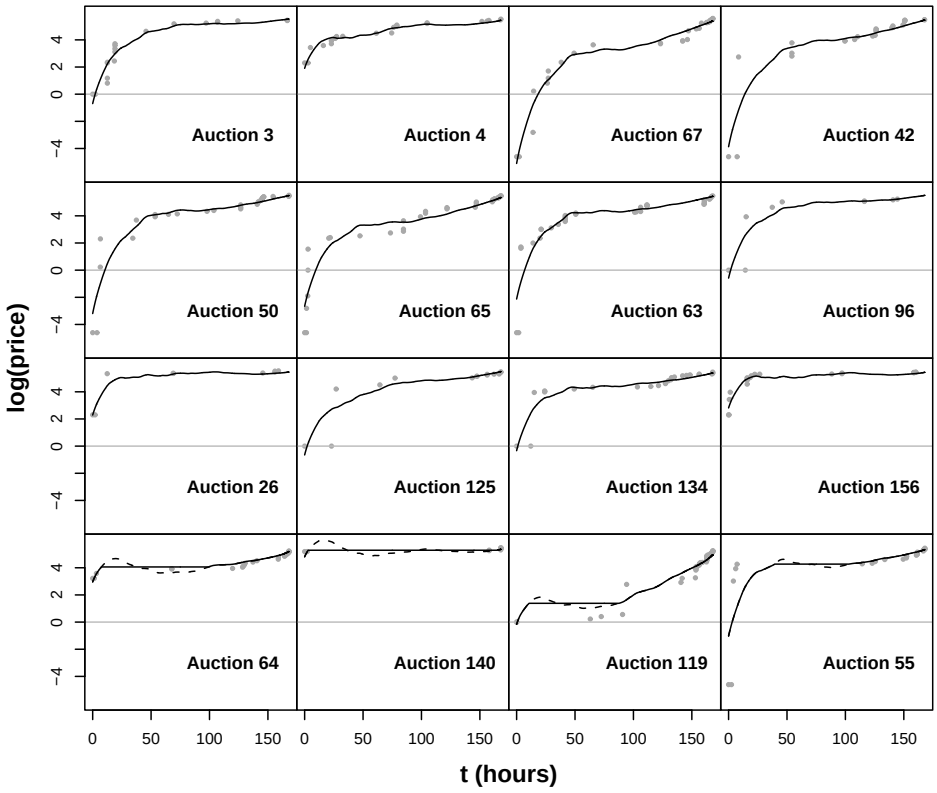

Figure 2: Estimated log(price) trajectories for 16 randomly selected Palm M515 PDA auctions. The observed log(bids) are displayed as grey dots and the black solid lines represent the fitted trajectories obtained from the PACE method. The four plots in the bottom panel illustrate the results of monotonization. The broken lines represent the estimated log price functions obtained from the PACE method, while the black solid lines correspond to monotonized functions resulting from the PAVA algorithm.

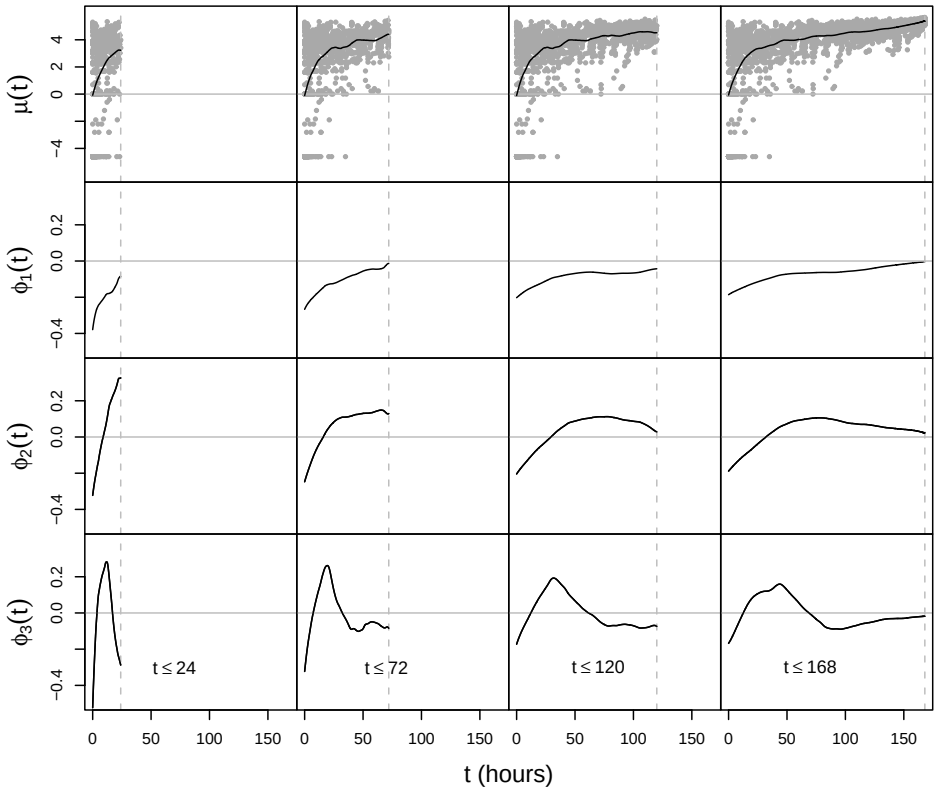

Figure 3: The top panels display the time-varying estimated mean functions $\hat{\mu}$ and the aggregated bids from all auctions for four different current times t, choosing $t = 24$, 72, 120 and 168 hours. The lower panels display the first three eigenfunctions, $\hat{\phi}_1$ (second row panels), $\hat{\phi}_2$ (third row panels) and $\hat{\phi}_3$ (bottom panels), for bid history observed up to time t.

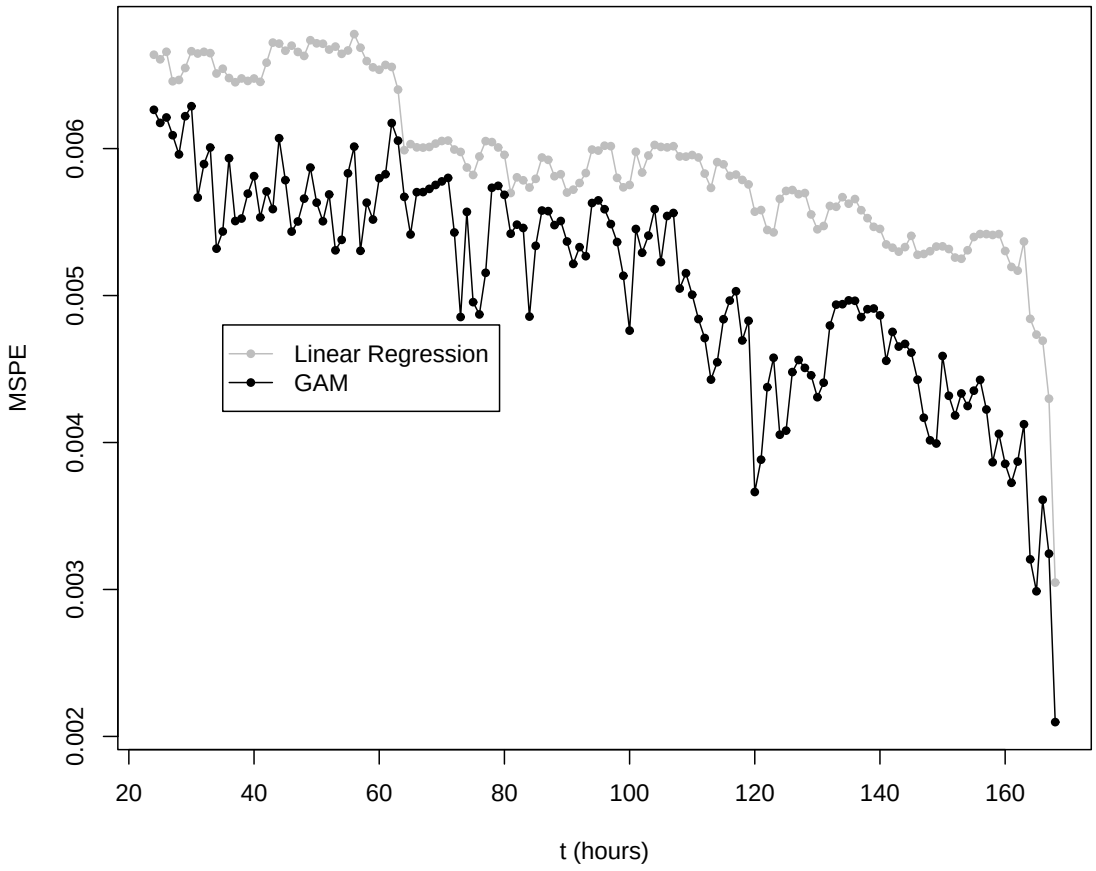

Figure 4: Mean Square Prediction Error (MSPE) versus current time t for time-varying predictions, using both linear regression and generalized additive modelling, with time-varying functional principal component scores obtained from PACE for bid histories up to current time as predictors and log(closing price) as response.

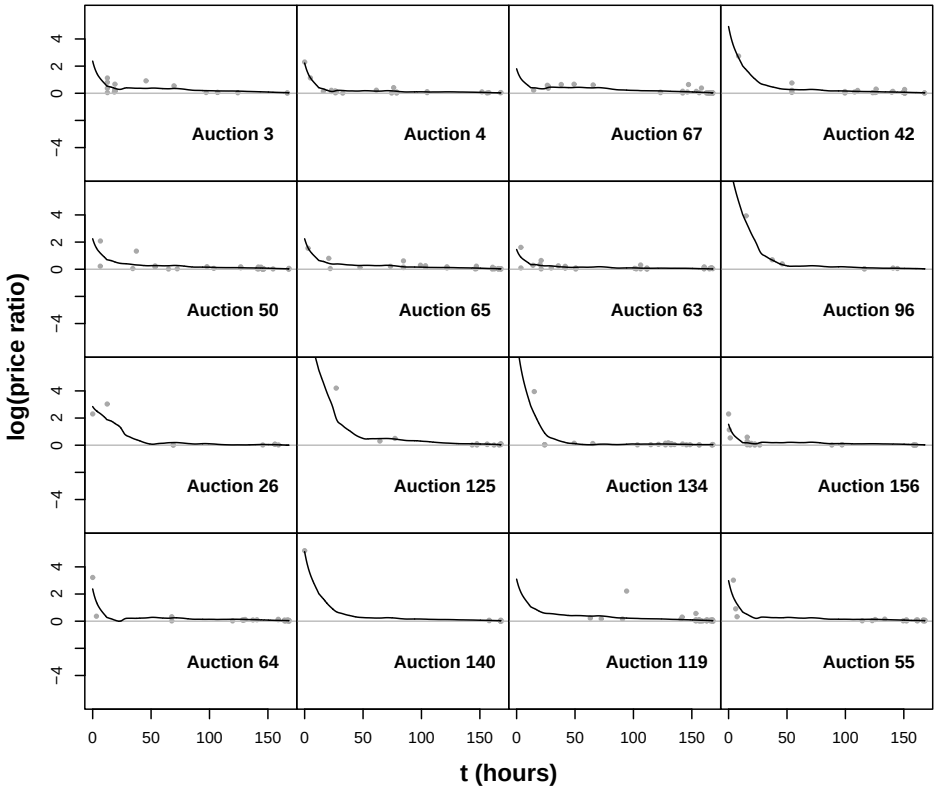

Figure 5: Estimated log(price ratio) trajectories for the same selected Palm M515 PDA auctions as displayed in Figure 2. The observed log(bid ratios) are defined as $\log \frac{Y_{ij}}{Y_{i,j-1}}$ with $Y_{i0} \equiv 1$, where Y_{ij} denotes the bid that is observed at time $T_{ij} \in [0, 168)$ (time units in hours). The aggregated log(bid ratios) are displayed as grey dots and the solid black lines represent fitted trajectories obtained from the PACE method.

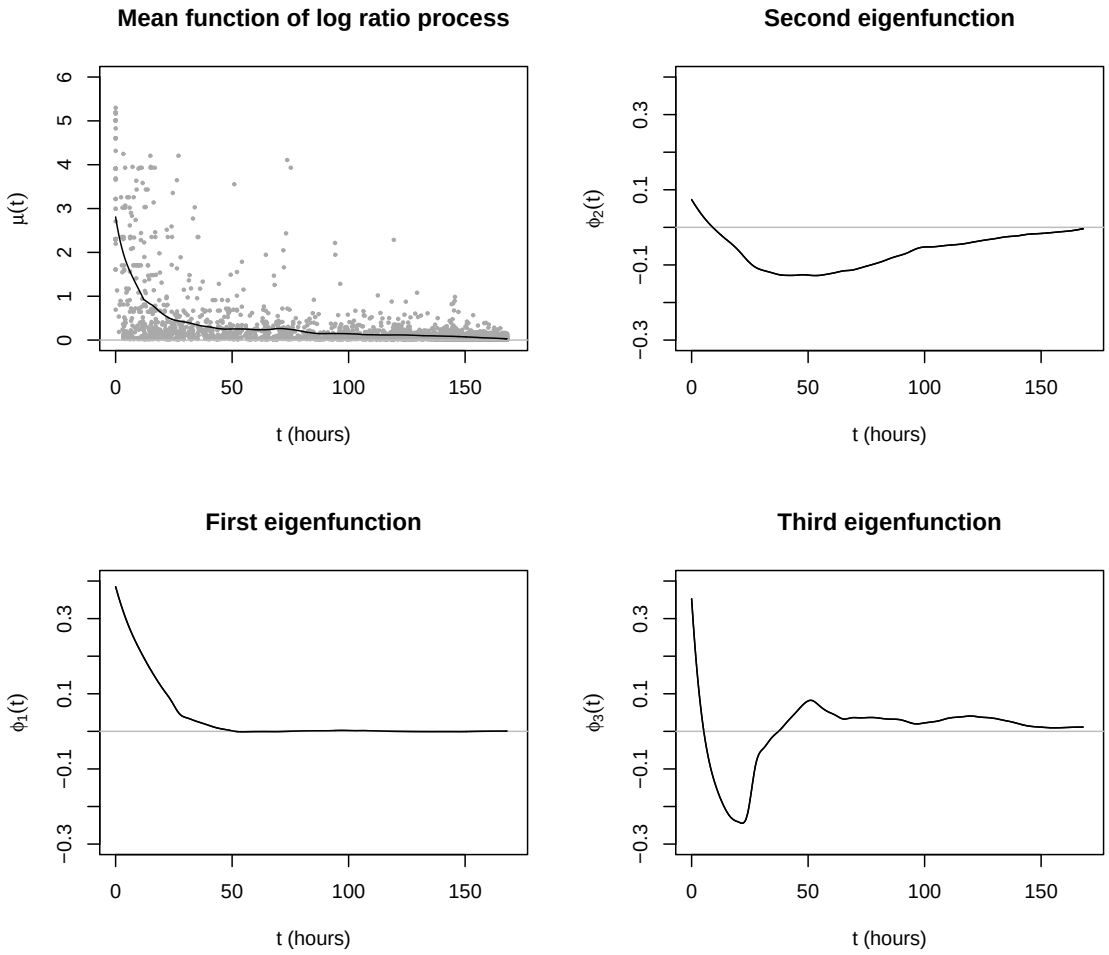

Figure 6: Estimated mean function $\hat{\mu}$ with observed log(bid ratios), aggregated over all auctions in left upper panel, and estimated first three eigenfunctions in the other panels.

References

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

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. John Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics.

Capra, W. B. & Müller, H.-G. (1997). An accelerated-time model for response curves. *Journal of American Statistical Association* **92**, 72–83.

Courant, R. & Hilbert, D. (1953). *Methods of mathematical physics. Vol. I*. John Wiley & Sons Inc., New York.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications*, vol. 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Fan, J. & Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 303–322.

Fan, J. & Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* **27**, 1491–1518.

Friedman, J. & Tibshirani, R. (1984). The monotone smoothing of scatterplots. *Technometrics* **26**, 243–250.

Hall, P. & Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics* **29**, 624–647.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*, vol. 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **55**, 757–796. With discussion and a reply by the authors.

Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–822.

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

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

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

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

Müller, H.-G. & Zhang, Y. (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* **61**, 1064–1075.

R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramsay, J. O. & Silverman, B. W. (2002). *Applied functional data analysis*. Springer Series in Statistics. Springer-Verlag, New York. Methods and case studies.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, 2nd edn.

Reddy, S. K. & Dass, M. (2006). Modeling on-line art auction dynamics using functional data analysis. *Statistical Science* **21**, 179–193.

Rice, J. A. & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**, 233–243.

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika* **68**, 45–54.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 413–428.

Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A. & Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* **59**, 676–685.

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