Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions

BY JIE PENG AND HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, CA 95616, USA

Abstract

We propose a distance between two realizations of a random process where for each realization only sparse and irregularly spaced measurements with additional measurement errors are available. Such data occur commonly in longitudinal studies and online trading data. A distance measure then makes it possible to apply distance-based analysis such as classification, clustering and multi-dimensional scaling for irregularly sampled longitudinal data. Once a suitable distance measure for sparsely sampled longitudinal trajectories has been found, we apply distance-based clustering methods to online auction data. We identify five distinct clusters of bidding patterns. Each of these bidding patterns is found to be associated with a distinct chance to obtain the auctioned item at a reasonable price.

Key words: Bidding; Clustering of trajectories; Functional data analysis; Metric in function space; Multidimensional scaling.
1 Introduction

The goal of cluster analysis is to group a collection of subjects into distinct clusters, such that those falling into the same cluster are more similar to each other than those in different clusters. Therefore a measure of similarity or dissimilarity between subjects is a necessary ingredient for clustering. A metric defined on the subject space is one way to obtain dissimilarities, simply using the distance between two subjects as a measure of dissimilarity. While one can readily choose from a variety of well-known metrics for the case of classical multivariate data, or for functional data that are in the form of continuously observed trajectories, finding a suitable distance measure for irregularly observed data can be a challenge. One such situation which we study here occurs in the commonly encountered case of irregularly and sparsely observed longitudinal data, with online auction data a prominent example [Shmueli and Jank (2005), Jank and Shmueli (2006), Shmueli, Russo and Jank (2007)].

For such online auction data, observations in the form of recorded willing-to-pay amounts for each bidder are sparse as many bidders submit only few bids during a given auction, and the timing of their bids is random. It is well known that early and late phases in an auction attract more bidding activity than the middle phase. The often frenzied bidding activity at the end of an auction is referred to as “bid sniping”, due to bidders who are determined to win a given auction. A main objective for most bidders is to win the item being auctioned at the lowest possible price.

Our study is motivated by the task to classify bidders according to their bidding activity patterns. Assuming that bidding activity reflects a bidder-specific stochastic bid price trajectory, we are then interested to define a distance between the various observed bidding behaviors, in order to derive a dissimilarity measure between bidders. The distance is to be based on the observed bids, where all bidders considered place bids for the same item at different auctions. Typical bid histories may include only two or very few bids observed for a particular bidder, and this leads to the challenge to define a distance based on sparse and irregularly timed data. Similar problems arise in many other types of longitudinal data where one also observes noisy measurements at random times. This motivates the development of a metric on the sample space of a stochastic process, where elements of this space are independent realizations of the underlying process and consist of noisy measurements made
at sparse and irregular time points. Once we have found a reasonable distance, we may base clustering methods on the resulting distance matrix, e.g., one may apply multidimensional scaling. Implementing such a procedure, we find five distinct clusters of bidders. Interestingly, the chance of obtaining the auctioned item at a low price is closely associated with the bidding pattern: If the goal of the bidder is to win the auctioned item at a reasonable price, some bidding strategies are better than others.

If the entire trajectory of each sample curve were observed, then the $L_2$ norm in the space of square integrable functions would provide a natural starting point for defining a metric. However, the $L_2$ distance is not readily calculable from the actually available noisy, sparse and irregularly sampled measurements of the bid price process. Suppose one observes a square integrable stochastic process $\{X(t) : t \in T\}$ at a random number of randomly located points in $T$, with measurements corrupted by additive i.i.d. random noise. The observations available from $n$ independent realizations of the process are $\{Y_{il} : 1 \leq l \leq n_i; 1 \leq i \leq n\}$ with

$$Y_{il} = X_i(T_{il}) + \varepsilon_{il},$$  \hspace{1cm} (1)

where $\{\varepsilon_{il}\}$ are i.i.d. with mean 0 and variance $\sigma^2$. Since $X$ is a square integrable stochastic process, by Mercer’s theorem (cf. Ash (1972)) there exists a positive semidefinite kernel $C(\cdot, \cdot)$ such that $\text{cov}(X(s), X(t)) = C(s, t)$ and we have the following a.s. expansion of the process $X_i(t)$ in terms of the eigenfunctions of the kernel $C(\cdot, \cdot)$

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

where $\mu(\cdot) = E(X(\cdot))$ is the mean function; the random variables $\{\xi_{ik} : k \geq 1\}$ for each $i$ are uncorrelated with zero mean and variance $\lambda_k$; and $\sum_{k=1}^{\infty} \lambda_k < \infty$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ are the eigenvalues of $C(\cdot, \cdot)$; $\phi_k(\cdot)$ are the corresponding orthonormal eigenfunctions.

In the observed data model we assume that $\{T_{il} : l = 1, \ldots, n_i; 1 \leq i \leq n\}$ are randomly sampled from a (possibly unknown) distribution with a density $g$ on $T$. In the problems we shall study, $n_i$ would typically be small, reflecting that the observed data consists of sparse and noisy realizations of a stochastic process. We shall then define a distance between two such realizations $X_i$ and $X_j$ based on the observed data $Y_i$ and $Y_j$, as described in the next section. This approach is inspired by recent developments of functional data analysis methodology for longitudinal data, notably the work of Yao, Müller and Wang (2005) where
 trajectories are predicted from sparse and noisy observations, which recently was adapted to online auctions [Liu and Müller (2007)]. Approaches based on B-spline fitting with random coefficients which are suitable to fit similar data with random coefficient models have been proposed by Shi, Taylor and Weiss (1996), Rice and Wu (2000), and recently in the context of online auctions by Reithinger et al. (2007).

For an up-to-date introduction to functional data analysis, we refer to the excellent book by Ramsay and Silverman (2005). Descriptions of the rapidly evolving interface between longitudinal and functional methodology and functional models for sparse longitudinal data can be found in Rice and Wu (2000), James, Hastie and Sugar (2001), James and Sugar (2003), and the overviews provided in Rice (2004), Zhao, Marron and Wells (2004) and Müller (2005). In the following section, we discuss the proposed distance. An application to the clustering of bidding patterns in ebay online auctions is the topic of section 3, followed by concluding remarks. Proofs and auxiliary remarks can be found in an Appendix.

2 A Distance for Sparse Data

We propose a distance between the random curves \( X_i \) and \( X_j \) based on the observed data \( Y_i = (Y_{i1}, \ldots, Y_{in_i})^T \) and \( Y_j = (Y_{i1}, \ldots, Y_{jn_j})^T \), respectively. The idea is to use the conditional expectation of the \( L_2 \) distance between these two curves, given the data. Our analysis is conditional on the times of the measurements \( \{T_{il} : l = 1, \ldots, n_i; 1 \leq i \leq n \} \) and their numbers \( \{n_i : 1 \leq i \leq n \} \).

2.1 Definition

The \( L_2 \) distance between two curves \( X_i \) and \( X_j \) is defined as

\[
D(i, j) = \left\{ \int_T (X_i(t) - X_j(t))^2 \, dt \right\}^{1/2},
\]

and is not calculable as only the sparse data \( Y_i \) and \( Y_j \) are observed. Therefore, we propose to use the conditional expectation of \( D^2(i, j) \) given \( Y_i \) and \( Y_j \) as the squared distance between \( X_i \) and \( X_j \),

\[
\hat{D}(i, j) = \left\{ \mathbb{E}(D^2(i, j)|Y_i, Y_j) \right\}^{1/2}, \quad 1 \leq i, j \leq n.
\] 

(3)
Note that as a function of \( Y_i, Y_j \), the \( \tilde{D}(i, j) \)s are random variables and have the following properties, the proof of which is given in the Appendix.

**Proposition 1.** \( \tilde{D} \) satisfies the following properties:

1. \( \tilde{D}(i, j) \geq 0; \tilde{D}(i, i) = 0 \) and for \( i \neq j, \) \( P(\tilde{D}(i, j) > 0) = 1; \)

2. \( \tilde{D}(i, j) = \tilde{D}(j, i); \)

3. For \( 1 \leq i, j, k \leq n \), \( \tilde{D}(i, j) \leq \tilde{D}(i, k) + \tilde{D}(k, j). \)

Therefore \( \tilde{D} \) can be viewed as a metric on the subject space consisting of random realizations \( \{X_i(\cdot)\} \) of the underlying stochastic process \( X(\cdot) \). Since under model (2), Parzeval’s identity implies that the \( L^2 \) distance between \( X_i \) and \( X_j \) can be written as

\[
D(i, j) = ||X_i - X_j||_2 = \left\{ \sum_{k=1}^{\infty} (\xi_{ik} - \xi_{jk})^2 \right\}^{1/2},
\]

we get

\[
\tilde{D}^2(i, j) = E\left( \sum_{k=1}^{\infty} (\xi_{ik} - \xi_{jk})^2 | Y_i, Y_j \right).
\]

For an integer \( K \geq 1 \), we then define truncated versions of \( \tilde{D} \) as

\[
\tilde{D}^{(K)}(i, j) = \left\{ E\left( \sum_{k=1}^{K} (\xi_{ik} - \xi_{jk})^2 | Y_i, Y_j \right) \right\}^{1/2}
= \left\{ \sum_{k=1}^{K} \text{var}(\xi_{ik}|Y_i) + \text{var}(\xi_{jk}|Y_j) + (E(\xi_{ik}|Y_i) - E(\xi_{jk}|Y_i))^2 \right\}^{1/2}. \quad (4)
\]

Note that it follows from these definitions that \( E(\tilde{D}^2(i, j)) = E(D^2(i, j)) \) and also for the truncated versions \( E(\tilde{D}^{(K)}(i, j))^2 \) = \( \sum_{k=1}^{K} 2\lambda_k = E(D^{(K)}(i, j))^2 \), so that these conditional expectations are unbiased predictors of the corresponding squared \( L^2 \) distances.

### 2.2 Estimation

In the following, we discuss the estimation of \( \tilde{D}^{(K)}(i, j) \) (4). Given an integer \( K \geq 1 \), let \( \Lambda^{(K)} = \text{diag}\{\lambda_1, \cdots, \lambda_K\} \) be the \( K \times K \) diagonal matrix with diagonal elements \( \{\lambda_1, \cdots, \lambda_K\} \).

For \( 1 \leq i \leq n, \ 1 \leq k \leq K \), let \( \mu_i = (\mu(T_{i1}), \cdots, \mu(T_{im_i}))^T, \ \xi_{i}^{(K)} = (\xi_{i1}, \cdots, \xi_{ik})^T, \ \phi_{ik} = (\phi_k(T_{i1}), \cdots, \phi_k(T_{im_i}))^T \) and \( \Phi^{(K)}_i = (\phi_{i1}, \cdots, \phi_{iK}). \) Define

\[
\tilde{\xi}_{i}^{(K)} = \Lambda^{(K)}(\Phi^{(K)}_i)^T \Sigma_{Y_i}^{-1}(Y_i - \mu_i), \quad (5)
\]
where \( \Sigma_i = \text{cov}(Y_i, Y_i) = (C(T_{iI}, T_{iI}')) + \sigma^2 I_{n_i} \).

Note that, \( \tilde{\xi}^{(K)} \) is the best linear unbiased predictor (BLUP) of \( \xi^{(K)} \), since \( \text{cov}(\xi^{(K)}, Y_i) = \Lambda^{(K)}(\Phi_i^{(K)})^T \). Moreover, if we have a finite dimensional process, such that for some integer \( K > 0, \lambda_k = 0 \) for \( k > K \) in model (2).

Then in this case, \( \Sigma_i = \Phi_i^{(K)} \Lambda^{(K)}(\Phi_i^{(K)})^T + \sigma^2 I_{n_i} \), and \( \Lambda \Phi_i^T (\Phi_i \Lambda \Phi_i^T + \sigma^2 I_{n_i})^{-1} = (\Phi_i^T \Phi_i + \sigma^2 \Lambda^{-1})^{-1} \Phi_i^T \) (upper subscript \( K \) omitted for simplicity), so that

\[
\tilde{\xi}_i = (\Phi_i^T \Phi_i + \sigma^2 \Lambda^{-1})^{-1} \Phi_i^T (Y_i - \mu_i),
\]

which also is the solution of the penalized least-squares problem

\[
\min_{\xi} (Y_i - \mu_i - \Phi_i \xi)^T (Y_i - \mu_i - \Phi_i \xi) + \sigma^2 \sum_{k=1}^K \xi_k^2 / \lambda_k.
\]

If one assumes normality of the processes in models (1) and (2), i.e., \( \xi_{ik} \sim N(0, \lambda_k) \) and \( \varepsilon_{it} \sim N(0, \sigma^2) \) and they are independent, the joint distribution of \( \{Y_i, \xi_i^{(K)}\} \) is multivariate normal with

\[
\begin{pmatrix}
Y_i \\
\xi_i^{(K)}
\end{pmatrix}
\sim \text{Normal}
\left(
\begin{pmatrix}
\mu_i \\
0
\end{pmatrix},
\begin{pmatrix}
\Sigma_i & \Phi_i^{(K)} \Lambda^{(K)} \\
\Lambda^{(K)}(\Phi_i^{(K)})^T & \Lambda^{(K)}
\end{pmatrix}
\right).
\]

Therefore the conditional distribution of \( \xi_i^{(K)} \) given \( Y_i \) is normal with mean

\[
E(\xi_i^{(K)} | Y_i) = \Lambda^{(K)}(\Phi_i^{(K)})^T \Sigma_i^{-1} (Y_i - \mu_i) = \tilde{\xi}_i^{(K)}
\]

and variance

\[
\text{var}(\xi_i^{(K)} | Y_i) = \Lambda^{(K)} - \Lambda^{(K)}(\Phi_i^{(K)})^T \Sigma_i^{-1} \Phi_i^{(K)} \Lambda^{(K)}.
\]

Furthermore, \( \tilde{\xi}^{(K)} \) becomes the best predictor of \( \xi^{(K)} \), and with (7) and (8),

\[
(\tilde{D}^{(K)}(i, j))^2 = \text{tr}(\Lambda^{(K)} - \Lambda^{(K)}(\Phi_i^{(K)})^T \Sigma_i^{-1} \Phi_i^{(K)} \Lambda^{(K)}) + \text{tr}(\Lambda^{(K)} - \Lambda^{(K)}(\Phi_j^{(K)})^T \Sigma_j^{-1} \Phi_j^{(K)} \Lambda^{(K)}) + \|\Lambda^{(K)}(\Phi_i^{(K)})^T \Sigma_i^{-1} (Y_i - \mu_i) - \Lambda^{(K)}(\Phi_j^{(K)})^T \Sigma_j^{-1} (Y_j - \mu_j)\|_2^2.
\]

Therefore, \( \tilde{D}^{(K)}(i, j) \) can then be estimated by plugging in estimates for the model components, i.e., for mean curve \( \mu(\cdot) \), covariance kernel \( C(\cdot, \cdot) \), first \( K \) eigenvalues \( \{\lambda_k : k = 1, \cdots, K\} \) and corresponding eigenfunctions \( \{\phi_k : k = 1, \cdots, K\} \), and the error variance \( \sigma^2 \).

Although (9) is derived under the normality assumption, its expectation always equals to the expectation of \( D^{(K)}(i, j)^2 \) (which is \( \sum_{k=1}^K 2\lambda_k \)), regardless of distributional assumptions.
Assuming that mean, covariance and eigenfunctions are all smooth, following Yao et al. (2005), we may apply local linear smoothers [Fan and Gijbels (1996)] based on the pooled data for function and surface estimation, fitting local lines in one dimension for the mean function and local planes in two dimensions for the covariance kernel. Denoting the resulting estimates of $\mu(\cdot), C(\cdot, \cdot)$ by $\hat{\mu}(\cdot), \hat{C}(\cdot, \cdot)$, the estimates of eigenfunctions and eigenvalues are given by the solutions $\hat{\phi}_k$ and $\hat{\lambda}_k$ of the eigen-equations based on $\hat{C}$,

$$\int_T \hat{C}(s, t) \hat{\phi}_k(s) ds = \hat{\lambda}_k \hat{\phi}_k(t),$$

where $\{\hat{\phi}_k : 1 \leq k \leq K\}$ are orthonormal and this system of equations is solved by discretizing the smoothed covariance [Rice and Silverman (1991)]. The estimate $\hat{\sigma}^2$ of $\sigma^2$ is obtained by first subtracting $\hat{C}(t, t)$ from a local linear smoother of $C(t, t) + \sigma^2$, denoted by $\hat{V}(t)$, then averaging over a subset of $T$ [Yao et al. (2005)]. Further details can be found in the Appendix.

The estimate of $\tilde{D}^{(K)}(i, j)$ is then given by

$$\hat{D}^{(K)}(i, j) = \left\{ \begin{array}{ll} \text{tr}(\hat{\Lambda}^{(K)} - \hat{\Lambda}^{(K)}(\hat{\Phi}_i^{(K)})^T \hat{\Sigma}_Y^{-1} \hat{\Phi}_i^{(K)} \hat{\Lambda}^{(K)}) \\
+ \text{tr}(\hat{\Lambda}^{(K)} - \hat{\Lambda}^{(K)}(\hat{\Phi}_j^{(K)})^T \hat{\Sigma}_Y^{-1} \hat{\Phi}_j^{(K)} \hat{\Lambda}^{(K)}) \\
+ \| (\hat{\Lambda}^{(K)}(\hat{\Phi}_i^{(K)})^T \hat{\Sigma}_Y^{-1}(Y_i - \hat{\mu}_i) - \hat{\Lambda}^{(K)}(\hat{\Phi}_j^{(K)})^T \hat{\Sigma}_Y^{-1}(Y_j - \hat{\mu}_j) \|_2^2 \right\}^{1/2} \tag{10}$$

where $\hat{\Lambda}^{(K)} = \text{diag}\{\hat{\lambda}_1, \cdots, \hat{\lambda}_K\}; \hat{\Phi}_i^{(K)} = (\hat{\phi}_{i1}, \cdots, \hat{\phi}_{iK});$ and the $(l, l')$ entry of $\hat{\Sigma}_Y$ is $(\hat{\Sigma}_Y)_{ll'} = \hat{C}(T_{il}, T_{il'}) + \hat{\sigma}^2 \delta_{ll'}$. The following result shows the consistency of these estimates for the target distance $\tilde{D}(i, j)$, providing some assurance that the estimated distance is close to the targeted one if enough components are included and the number of observed random curves is large enough.

**Theorem 1.** Under the same assumptions as Lemma 2 in the Appendix,

$$\lim_{K \to +\infty} \lim_{n \to +\infty} \hat{D}^{(K)}(i, j) = \tilde{D}(i, j) \text{ in probability.}$$

The proof is in the Appendix.

### 2.3 Distance based scaling

Multidimensional scaling (MDS) aims to find a projection of given original objects for which one has a distance matrix into $p$ dimensional (Euclidean) space for any $p \geq 1$, often chosen
as $p = 2$ or 3 which provides best visualization. The projected points in $p$-space represent the original objects (e.g., random curves) in such a way that their distances match as well as possible with the original distances or dissimilarities $\{\delta_{ij}\}$, according to some target criterion. In our setting, these original distances will be the estimated conditional $L_2$ distances (10) between the sparsely observed random trajectories. Various techniques exist for implementing the MDS projection, including metric and nonmetric scaling.

In classical metric scaling one treats dissimilarities $\{\delta_{ij}\}$ directly as Euclidean distances and then uses the spectral decomposition of a doubly centered matrix of dissimilarities [Cox and Cox (2001)]. It is well known that there is an equivalence between principal components analysis and classical scaling when dissimilarities are truly Euclidean distances (if the subjects are points in an Euclidean space). Metric least squares scaling finds configuration points $\{x_i\}$ in a $p$ dimensional space with distances $\{d_{ij}\}$ matching $\{\delta_{ij}\}$ as closely as possible, by minimizing a loss function $S$, e.g., $S = \sum_{i<j} \frac{\delta_{ij}^2}{\sum_{i<j} \delta_{ij}} [\text{Sammon (1969)}]$. Two other popular optimality criteria for metric MDS are metric stress and s-stress, which are special cases of the criterion

$$\text{minimize } \sum_{i<j} w_{ij} [(d_{ij}^2)^r - (\delta_{ij}^2)^r]^2,$$

usually implemented with $w_{ij} = 1$. The stress criterion corresponds to the case $r = 1/2$ and was originally proposed by Kruskal (1964) for nonmetric MDS; while the s-stress criterion corresponds to $r = 1$ and was popularized by Takane, Young and DeLeeuw (1977). The s-stress criterion leads to a smooth minimization problem in contrast to stress. Kearsley, Tapia and Trosset (1998) applied Newton’s method to find solutions using these criteria. In practice, the stress criterion is often normalized by the sum of squares of the dissimilarities, thus becoming scale free. Similarly, the s-stress criterion is normalized with the sum of the 4th powers of the dissimilarities.

3 Clustering Bidders in Online Auctions

Online auctions are generating increasingly large amounts of data for which analysis tools are still scarce. We illustrate the usefulness of the proposed distance measure with ebay online auction data [Shmueli and Jank (2005), Jank and Shmueli (2006)] for 158 seven day
online auctions of Palm M515 Personal Digital Assistants (PDA) that took place between March and May, 2003. We are aiming at classifying the bidders according to their bidding behavior.

The recorded bid values correspond to willingness-to-pay (WTP) amounts, entered by a bidder at a random bidding time during the auction period. We removed an anomalous auction that contained identical bids of the same value placed by the same and only bidder for this auction, and the analysis reported here is therefore based on 157 auctions. The WTP values are converted into so-called 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 will not exceed the current highest WTP value; (iv) the closing price is the same as the winner’s bid (the converted live bid corresponding to the second last WTP value). For further details and assumptions see Liu and Müller (2007).

Among these 157 auctions, we found 113 bids in 44 of the auctions that had WTP values that were no higher than the previous live bids. In other words, these bidders placed WTP values that had no effect on the current live bids. Since these bids did not affect current price, we removed these ineffective bids from the data, which resulted in data reflecting 3687 bids in 157 auctions. The data consists of the random times when bidding took place, the bid prices and bidder identification. To study bidding patterns, we randomly selected one bidder for each auction such that the bidding patterns for each of 157 different bidders are included. For the \( i \)-th bidder, the series of bid times is \( T_{il}^i \), \( l = 1, \ldots, n_i \), \( i = 1, \ldots, n \) (relative to the auction starting time in hours), and the corresponding live bids are \( Y_{il}^i \) (after conversion from the recorded WTP values). These data form the sparsely observed functional data. Bidder behavior is assumed to be manifested in an unobserved random trajectory of which the observed actual bids are just a snapshot.

Our goal is to determine whether there are distinct types of bidding behavior. The mean trajectory for all 157 bidders is fitted by pooling all data as described in the Appendix. In a preprocessing step, the pooled residuals toward the overall mean curve are calculated, and the standard deviation calculated from the pooled residuals is used to remove those
bidders with outlying bids, which we define to be bidders whose bids fall outside of three standard deviations of the mean curve; after this outlier removal $n = 151$ bidders remained in the analysis. Bid times ranged from 0.24 hours to 168 hours, with an average of 4.61 bids per bidder. The data is modelled as being generated from 151 independent realizations of an underlying but unobserved stochastic bid process, as described in Section 1. The mean trajectory (Figure 1) and covariance surface of this bid process are estimated as described in the Appendix, with bandwidths selected by leave-one-curve-out cross validation. We chose $K = 5$ components when estimating the conditional $L_2$ distance $\tilde{D} = (\tilde{D}(i, j))_{1\leq i, j \leq 151}$ (see Equations (3) and (9)), based on the fact that these accounted for about 90% of total variation of the random trajectories. Multidimensional scaling (MDS) was then applied to the distance matrix $\tilde{D}$, projecting into a space of dimension $p = 2$, using the matlab function mdscale.

For the goodness-of-fit criterion that the MDS algorithm minimizes, we considered the criteria strain, sammon, metricstress and metricsstress. Among these, sammon and metricstress failed to converge for the data at hand; we display the results of the MDS projection for strain and metricsstress in Figure 2. This figure reveals that the best separation of buyers into distinct subgroups is obtained from the application of metricsstress. Applying K-means cluster analysis to the metricsstress MDS results, 5 clusters of bidders were identified (Figure 3). Based on the associated bidding patterns, the groups of bidders falling into the respective clusters are labelled “A” (Aggressive), “S” (Slow start), “L” (Low end), “H” (High end) and “F” (Fast start, slow increase). In Figure 3, “O” (Outlier) is used to denote “unclassifiable” bidders whose bidding behavior does not fit naturally into any of the five clusters.

Figures 4 and 5 show the individual trajectories and the fitted mean trajectories for the 5 groups of bidders, respectively. Two groups of bidders stand out: the A ($n_A = 24$ “aggressive”) and S ($n_S = 33$ “slow start”) bidders. Bidders in either of these two groups usually bid over a sustained period of time (with interquartile ranges 50 to 137 hours and 63 to 134 hours, respectively, and average number of bids 6.16 and 5.75, respectively). While “A” bidders increase their bids fast initially, followed by moderate increases toward the end of an auction, with an overall concave mean bid trajectory, “S” bidders tend to slowly raise their bids at the beginning and then increase them faster near the end of the auction, giving
rise to a convex mean curve. The “L” \((n_L = 38)\) and “H” \((n_H = 43)\) bidders tend to place bids over a short period of time near the end of an auction (with interquartile ranges 159 to 167 hours and 116 to 167 hours, respectively; average number of bids 3.44 and 3.67, respectively), and participate in “bid sniping”. Near the end of an auction, “L” bidders place relatively low bids, so these bidders aim at a good bargain, while “H” bidders place relatively high bids, so the primary interest of these bidders seems to be to secure the item, while price plays a secondary role. Finally, “F” \((n_F = 11)\) bidders place bids an average number of times throughout the auction (interquartile range 27 to 163 hours; average number of bids 5.18), with relatively high bids at the beginning (mostly starting around 50 dollars), and from then on only cautiously increase their bids, ending their bidding at a fairly modest price level. Their mean bid trajectory is almost linear.

Regarding the auction outcome, “H” bidders have the highest winning rate, at around 34.9%; this is not surprising, as these bidders tend to place high bids near the end of an auction. The success rates for the other four bidder groups are: For “A” bidders, 12.5%; for “S” bidders, 3%; for “L” bidders, 16.2%; and for “F” bidders, 9%. By examining the prices paid when winning the bid, “H” bidders also pay the most: 205$ on average, while the corresponding figures are 188$ for group A, 155$ for group S, 150$ for group L, and 132$ for group F bidders. It emerges that “L” bidders have the best overall chance to win the item at a reasonable price. These bidders tend to place their bids only when the auction is in the final stages, and place modestly priced bids at this time; their winning rate is much lower than that for “H” bidders, but the price they pay is also markedly lower. Overall, these results confirm the effectiveness of the “bid sniping” strategy in ebay online auctions, and the best overall strategy as adopted by the “L” bidders is to seek out auctions that are at a relatively low price near their ending phase and then to place modestly priced bids right before the auction ends.

4 Discussion

The proposed functional distance is defined conditionally on observed “snapshots” of the underlying trajectories. It provides a useful tool that allows to apply distance-based methods for sparsely and irregularly observed functional and longitudinal data. This includes multidi-
mensional scaling for dimension reduction and various clustering methods. We demonstrate that the estimated distances as estimated from such data converge to the target values.

Insights on bidding patterns can be gained by applying these methods to online auction data. Bidding behaviors fall into five distinct clusters with specific characteristics of the corresponding bid trajectories. These distinct bidding strategies are clearly associated with the chance to win the item that is auctioned, and also with the final price a winning bidder pays for the item. It turns out that the five bidding strategies can be clearly distinguished in terms of how well they achieve these goals. The strategy of placing bids near the end of an auction at moderate bid levels emerges as a winning strategy. The proposed methodology is more generally useful for all longitudinal studies where clustering of subjects is of interest.

5 Appendix

Proof of Proposition 1. The first two properties are obvious from the definition of \( \tilde{D} \), while the third one can be easily proved by applying Cauchy-Schwarz inequality. Note that \( D \) is a metric, thus satisfying the triangle inequality: \( D(i, j) \leq D(i, k) + D(k, j) \). Therefore

\[
D^2(i, j) \leq D^2(i, k) + D^2(k, j) + 2D(i, k)D(k, j).
\]

Since \( \mathbb{E}(D^2(i, j)|Y_i, Y_j, Y_k) = \mathbb{E}(D^2(i, j)|Y_i, Y_j) = \tilde{D}^2(i, j) \), we then have

\[
\tilde{D}^2(i, j) \leq \tilde{D}^2(i, k) + \tilde{D}^2(k, j) + 2\mathbb{E}(D(i, k)D(k, j)|Y_i, Y_j, Y_k)
\]

and by the Cauchy-Schwarz inequality

\[
\mathbb{E}(D(i, k)D(k, j)|Y_i, Y_j, Y_k) \leq \left\{ \mathbb{E}(D^2(i, k)|Y_i, Y_j, Y_k) \right\}^{1/2} \left\{ \mathbb{E}(D^2(k, j)|Y_i, Y_j, Y_k) \right\}^{1/2} = \tilde{D}(i, k)\tilde{D}(j, k).
\]

This concludes the proof.

Model fitting. Following Yao et al. (2005), we fit the model by local linear smoothers based on pooled data. For the mean curve \( \mu(\cdot) \), applying weighted least squares, one obtains

\[
(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_1 \left( \frac{T_{ij} - t}{h_\mu} \right) \{ Y_{ij} - \beta_0 - \beta_1(t - T_{ij}) \}^2,
\]
where \( K_1 \) is a one dimensional kernel, for example the univariate Epanechnikov kernel
\[ K_1(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}(x), \]
and \( h_\mu \) is the bandwidth which can be selected by leave-one-curve-out cross validation. The resulting estimate is \( \hat{\mu}(t) = \hat{\beta}_0(t) \). The covariance kernel \( C(\cdot, \cdot) \) is fitted by two dimensional weighted least squares,
\[
(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{1 \leq j \neq l \leq n} K_2(\frac{T_{ij} - s}{h_G}, \frac{T_{il} - t}{h_G})\{G_i(T_{ij}, T_{il}) - f(\beta, s, t, T_{ij}, T_{il})\}^2, \tag{12}
\]
where \( G_i(T_{ij}, T_{il}) = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{il} - \hat{\mu}(T_{il})), f(\beta, s, t, T_{ij}, T_{il}) = \beta_0 + \beta_1(s - T_{ij}) + \beta_2(t - T_{il}), \) and \( K_2 \) is a two dimensional kernel, for example the bivariate Epanechnikov kernel
\[ K_2(x, y) = \frac{9}{16}(1 - x^2)(1 - y^2)I_{[-1,1]}(x)I_{[-1,1]}(y). \]
Then \( \hat{C}(s, t) = \hat{\beta}_0(s, t) \). \( h_G \) is the bandwidth and can be selected by leave-one-curve-out cross validation. Note that since \( E(G_i(T_{ij}, T_{il})) \approx C(T_{ij}, T_{il}) + \sigma^2\delta_{jl} \), therefore in (12), we should only use the off diagonal entries of the empirical covariance: \( G_i(T_{ij}, T_{il}), j \neq l \). Let \( V(t) = C(t, t) + \sigma^2 \), then \( V(\cdot) \) is fitted by
\[
(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_1(\frac{T_{ij} - t}{h_V})\{G_i(T_{ij}) - \beta_0 - \beta_1(t - T_{ij})\}^2, \tag{13}
\]
where \( K_1 \) is the one dimensional kernel, \( h_V \) is the bandwidth, and \( \hat{V}(t) = \hat{\beta}_0(t) \). Then the estimate of the error variance \( \sigma^2 \) is given by
\[
\hat{\sigma}^2 = \frac{2}{T} \int_{T_1} (\hat{V}(t) - \hat{C}(t, t))dt, \tag{14}
\]
where \( T_1 \) is the middle half interval of \( T \).

We next state a series of auxiliary lemmas. The first lemma summarizes asymptotic results from Yao et al. (2005). The set of assumptions (A1.1-A4) and B(1.1)-B(2.2b) is given in Yao et al. (2005) and will not be repeated here.

**Lemma 1. (Theorem 1, Corollary 1 and Theorem 2 in Yao et al. (2005)).** Under (A1.1-A4) and (B1.1)-B(2.2b) with \( \nu = 0, \ell = 2 \) in (B2.2a) and \( \nu = (0,0), \ell = 2 \) in (B2.2b),
\[
\sup_{t \in T} |\hat{\mu}(t) - \mu(t)| = O_p(\frac{1}{\sqrt{n}h_\mu}), \quad \sup_{t,s \in T} |\hat{C}(s, t) - C(s, t)| = O_p(\frac{1}{\sqrt{nh_G^2}}),
\]
\[
|\hat{\sigma}^2 - \sigma^2| = O_p(\frac{1}{\sqrt{n}}(\frac{1}{h_G^2} + \frac{1}{h_V})), \quad |\hat{\lambda}_k - \lambda_k| = O_p(\frac{1}{\sqrt{nh_G^2}}),
\]
\[
||\hat{\phi}_k - \phi_k||_H = O_p(\frac{1}{\sqrt{nh_G^2}}), \quad \sup_{t \in T} |\hat{\phi}_k(t) - \phi_k(t)| = O_p(\frac{1}{\sqrt{nh_G^2}}).
\]
Lemma 2. Under the normality assumption and the set of assumptions as in Lemma 1,
\[ \lim_{n \to +\infty} \hat{D}^{(K)}(i, j) = \check{D}^{(K)}(i, j) \text{ in probability.} \]

Proof. Recall that \( \hat{D}^{(K)}(i, j) \) is given by (10), which under the normality assumption equals \( \check{D}^{(K)}(i, j) \) with unknown model components replaced by their estimates. Therefore by Lemma 1 and Slutsky’s Theorem, the result follows immediately.

Lemma 3.
\[ \lim_{K \to \infty} \hat{D}^{(K)}(i, j) = \check{D}(i, j) \text{ in probability.} \]

Proof. By definition
\[ \hat{D}^2(i, j) - \check{D}^{(K)}(i, j)^2 = E\left( \sum_{k=K+1}^{\infty} (\xi_{ik} - \xi_{jk})^2 \mid Y_i, Y_j \right) \geq 0. \]

Thus
\[ E(\hat{D}^2(i, j) - \check{D}^{(K)}(i, j)^2) = E\left( \sum_{k=K+1}^{\infty} (\xi_{ik} - \xi_{jk})^2 \right). \]

Note that \( \xi_{ik} \) and \( \xi_{jk} \) are independent with mean zero and variance \( \lambda_k \), and therefore
\[ E\left( \sum_{k=K+1}^{\infty} (\xi_{ik} - \xi_{jk})^2 \right) = 2 \sum_{k=K+1}^{\infty} \lambda_k. \]

Since \( \sum_{k=1}^{\infty} \lambda_k < \infty \), then
\[ \lim_{K \to \infty} \sum_{k=K+1}^{\infty} \lambda_k = 0. \]

Therefore by Markov’s inequality and the fact that \( \hat{D}^2(i, j) - \check{D}^{(K)}(i, j)^2 \geq 0 \), for any \( \epsilon > 0 \)
\[ P(|\hat{D}^2(i, j) - \check{D}^{(K)}(i, j)^2| > \epsilon) \leq E(\hat{D}^2(i, j) - \check{D}^{(K)}(i, j)^2)/\epsilon = 2 \sum_{k=K+1}^{\infty} \lambda_k/\epsilon \to 0, \]
from which and Slutsky’s Theorem the result follows.

Proof of Theorem 1. Note that
\[ |\hat{D}^{(K)}(i, j) - \check{D}(i, j)| \leq |\hat{D}^{(K)}(i, j) - \check{D}^{(K)}(i, j)| + |\check{D}^{(K)}(i, j) - \check{D}(i, j)|. \]

By Lemma 3, for any \( \epsilon > 0 \) and any \( \delta > 0 \), there exists \( K_0 \) such that for any \( K \geq K_0 \),
\[ P(|\hat{D}^{(K)}(i, j) - \check{D}(i, j)| \geq \epsilon/2) \leq \delta/2. \]
By Lemma 2, for each $K > 0$, there exists $n_0(K) > 0$ such that for any $n \geq n_0(K)$,

$$P(|\hat{D}^{(K)}(i, j) - \tilde{D}^{(K)}(i, j)| \geq \epsilon/2) \leq \delta/2.$$ 

Therefore for $K \geq K_0$, $n \geq n_0(K)$, $P(|\hat{D}^{(K)}(i, j) - \tilde{D}(i, j)| \geq \epsilon) \leq \delta$, which concludes the proof.

Acknowledgements

This research was supported in part by National Science Foundation grants DMS03-54448 and DMS05-05537. We are grateful to Wolfgang Jank for sharing his e-bay auction data.

References


Figure 1: Mean bid process trajectory for $n = 151$ bidders participating in eBay online auctions.
Figure 2: Results for two types of multidimensional scaling (MDS) applied to proposed conditional functional distance and projecting to two-dimensional space. Top panel: MDS using the \textit{strain} criterion. Bottom panel: MDS using the \textit{metricsstress} criterion.
Figure 3: K-means clustering applied to MDS with *metricsstress*, revealing five clusters of bidding behaviors.
Figure 4: Individual fitted trajectories of the bid process for the five clusters. The five groups of bidders are labelled A (Aggressive), S (Slow start), L (Low end), H (High end) and F (Fast start, slow increase).
Figure 5: Mean bid trajectories for each of the five clusters.