Stat
The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.100X/sta.0000

Support information for "Multilevel sparse functional principal component analysis"

Chongzhi Di[a]*, Ciprian M. Crainiceanu[b], Wolfgang S. Jank[c]

Received 3 March 2014; Accepted 24 March 2014

1. Details for the eBay auction application

Our second application originates from online auctions, which are challenging because they involve *user-generated* data: sellers decide when to post an auction, and bidders decide when to place bids. This can result in individual auctions that have extremely sparse observations, especially during the early parts of the auction. In fact, well-documented bidding strategies such as early bidding (Bapna et al., 2004) or last-minute bidding (Roth & Ockenfels, 2002; Shmueli et al., 2007) cause "bidding-draughts" (Jank & Shmueli, 2007) during the middle, leaving the auction with barely any observations at all. Peng & Müller (2008) and Liu & Müller (2009) studied clustering and dynamics of such sparse auction data while Jank et al. (2010) and Reithinger et al. (2008) proposed new methods for smoothing sparse auction data. Here, we study bidding records of 843 *digital camera auction* that were listed on eBay between April, 2007, and January, 2008. These auctions were on 515 types of digital cameras, from 233 distinct sellers. On average, there were 11 bids per auction. The timing of the bids was irregular and often sparse: some auctions contained as many as 56 bids, while others included as few as 1-2 bids. In this application we are particularly interested in investigating the pattern of variation of an auction's bidding path and decomposing it into components that are attributable to the product and components attributable to the bidding process.

We analyzed a subset of our online auction data that consist of 40 pairs of auctions of digital cameras. Every pair contains two auctions of exactly the same camera, so any observed differences in the auction outcome must be due to differences in the seller or the bidding process. Figure 1 displays the raw auction data for the first three cameras. Since different auctions have different durations (e.g. 1-day auctions vs. 7-day auctions), we normalize auction-time to the interval $[0, 1]$, where $t = 0$ and $t = 1$ correspond to the beginning and end of an auction, respectively. We also re-scale the y-axis ("price") to the average final price so that all auctions are comparable with respect to an item's value. Figure 1 shows that some auctions include as few as 3 bids, while others have as many as $20 - 30$ bids. The bid-timing

[a]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M2-B500, Seattle, WA 98115, USA
[b]Department of Biostatistics, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205, USA
[c]Department of Information Systems and Decision Sciences, University of South Florida, Tampa, FL 33620, USA
*Email: cdi@fhcrc.org

is irregular and unbalanced – often it is quite sparse at the beginning and middle phases of the auction but rather dense towards the end; the reason is that bidders decide when to bid and their decision is driven by individually-varying bidding strategies such as "last-minute" bidding and it may also reflect their reaction to competitor's bids.

We point out that the bids in Figure 1 are not monotonically increasing as would be expected from an ascending auction. The reason lies in eBay's *proxy bidding system*. In that system, an individual submits a "proxy bid," which is the maximum value he/she is willing to pay for the item (see Appendix for detailed explanations). Because proxy bids measure bidders' privately held willingness to pay (at the time of the bid), we refer to the trajectories in Figure 1 as the *current maximum willingness to pay* trajectories. Thus the goal of our analysis will be to study the *evolution* of willingness to pay over the course of an auction and how its variation can be broken up into components that are attributable to differences in the product and components that are attributable to the bidding process itself.

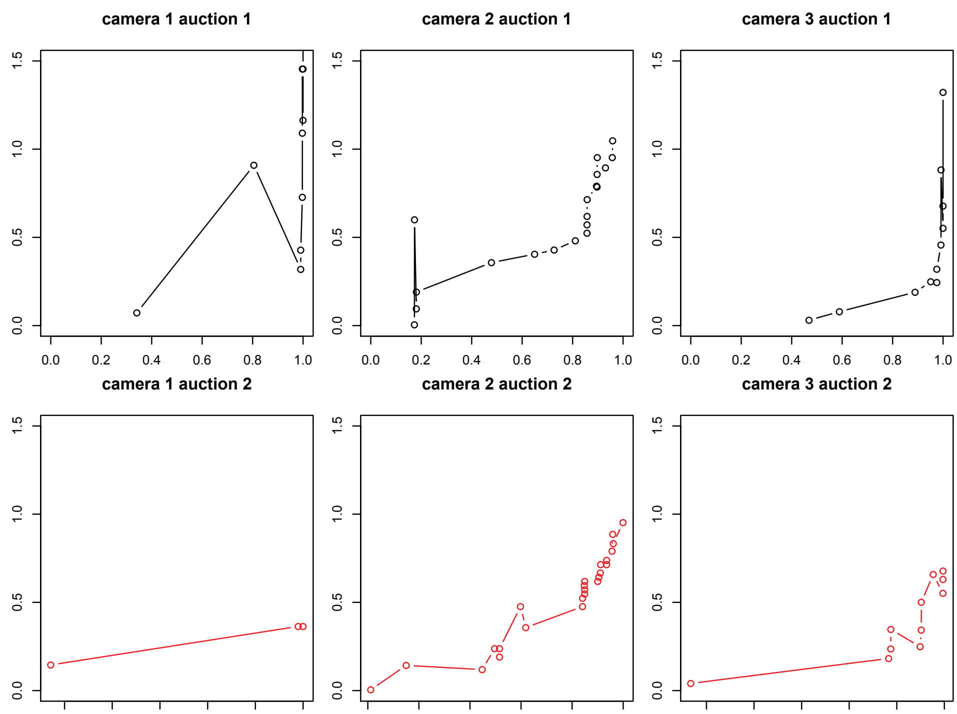

Figure 1. Auction trajectories for three digital cameras with two auctions per camera. The "x" axis represents normalized bidding time with range $t \in [0, 1]$, and the "Y" axis corresponds to re-scaled bidding prices. The two subfigures in each column show two auctions from a specific type of digital camera.

To accomplish that goal, we view bidding trajectories as noisy realizations of underlying smooth processes, and fit MFPCA models with one way ANOVA,

$$Y_{ij}(t) = \mu(t) + Z_i(t) + W_{ij}(t) + \epsilon_{ij}(t) = \mu(t) + \sum_{k=1}^{N_1} \xi_{ik}\, \phi_k^{(1)}(t) + \sum_{l=1}^{N_2} \zeta_{ijl}\, \phi_l^{(2)}(t) + \epsilon_{ij}(t).$$

Here, $\mu(t)$ is the overall mean function, $Z_i(t)$ is camera specific deviation, $W_{ij}(t)$ is within camera auction-specific deviation, and $\epsilon_{ij}(t)$ is residual noise or measurement error. Figure 2 shows observed data from all auctions and estimated mean function $\hat{\mu}(t)$. As expected, bidding prices increase with time on average, more rapidly towards the

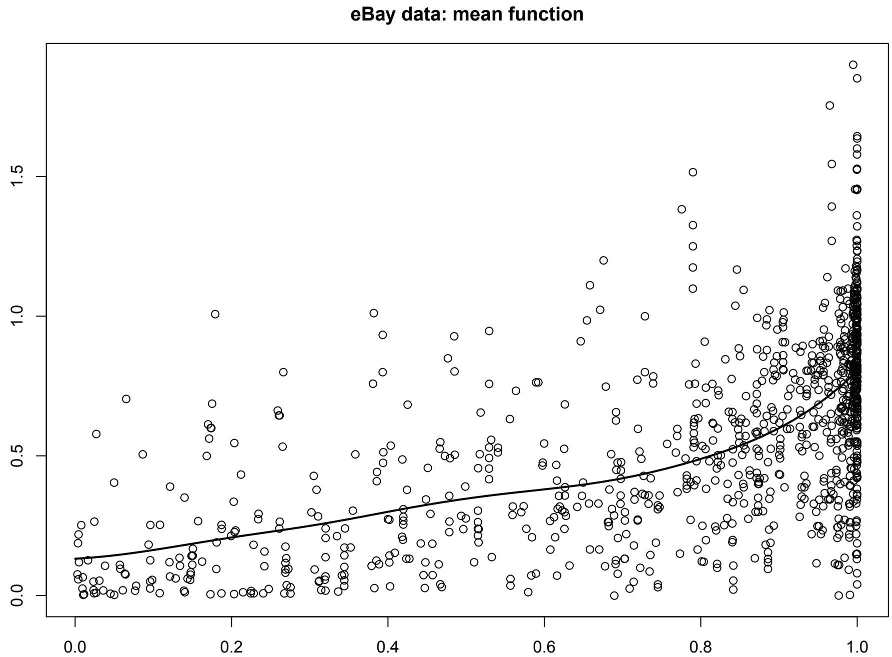

Figure 2. The raw auction data and estimated mean function $\hat{\mu}(t)$. The black curve corresponds to $\hat{\mu}(t)$, which demonstrates the average shape of auction trajectories across all auctions.

end of auctions. Next, we look at level 1 and level 2 principal components, which extract dominating modes of variations at the between and with camera level, respectively. Eigenvalues and Eigenfunctions are shown in Table 1 and Figure 3, respectively. Based on Table 1, level 1 and 2 explain 35.9% and 64.1% of the total variation, respectively. Thus, there is substantial amount of variation at both levels.

	Level 1 (Proportion explained: 35.9%)			Level 2 (Proportion explained: 64.1%)	
Component	1	2	3	1	2
eigenvalue ($\times 10^{-3}$)	6.2	3.6	2.3	19.4	3.4
% var	48.2	27.8	18.1	84.6	14.6
cum. % var	48.2	76.0	94.1	84.6	99.2

Table 1. Estimated eigenvalues from eBay auction data. Level 1 and 2 correspond to between camera and within camera variations, respectively. “% var" means percentage of variance explained by corresponding components, while “cum. % var" means cumulative percentage of variance explained by current and prior components.

We first look at level 1 eigenvalues and eigenfunctions, which capture modes of variations across different cameras. In Figure 3, the first row show the shapes of three leading eigenfunctions $\{\phi_1^{(1)}(t), \phi_2^{(1)}(t), \phi_3^{(1)}(t)\}$, while the second row display the types of variations resulting from them. The first eigenfunction (PC1) is mostly positive, indicating that auctions loading positively (negatively) on this component will always have higher (lower) values than average. Its magnitude is small at the beginning, increases and reaches maximum around t=0.2, then gradually decreases to 0

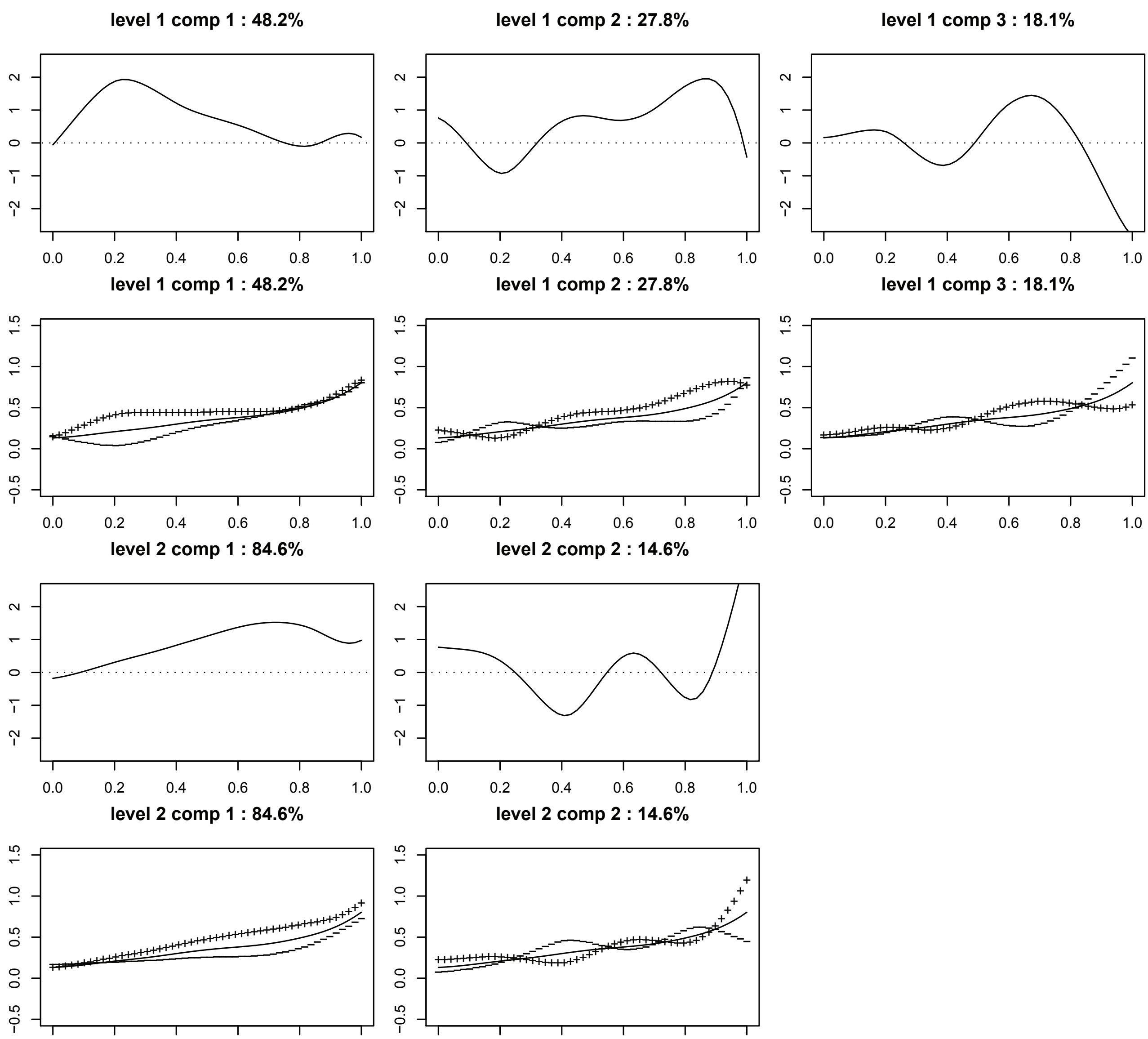

Figure 3. Estimated eigenfunctions at level 1 (Rows 1-2) and level 2 (Rows 3-4). In Row 1 and 3, the solid lines represent estimated eigenfunctions. In Row 2 and 4, solid lines represent the overall mean function $\mu(t)$, and the lines indicated by "+" and "-" are mean function plus or minus a multiple of $\phi_1^{(1)}(t)$, respectively.

around t=0.8, and is close to 0 after t=0.8. The subfigure in Row 2 Column 1 demonstrates the type of variation induced by this component. The solid line represents the overall mean function, $\mu(t)$, and the lines indicated by "+" and "-" are overall mean function plus or minus a multiple of $\phi_1^{(1)}(t)$, respectively. Namely, "+" represents $\mu(t) + c\,\phi_1^{(1)}(t)$, and "-" represents $\mu(t) + c\,\phi_1^{(1)}(t)$ for some constant c. In terms of the bidding context, the "+" line represents an auction in which the price increases rapidly at the beginning, almost flattens in the middle phase and increases somewhat towards the end. The "-" line corresponds to an auction in which the price flattens at the beginning, but keeps increasing after t=0.2. This component explains 48.2% of variation at the between camera level. One can interpret other eigenfunctions in a similar manner. For example, the second component that explains 27.8% of variation characterizes mostly the variation at the later parts ($t = 0.4$ to $t = 0.95$) of auctions. The third principal component explains 18.1% of the variation. These three components together explain more than 90% of level 1 variation.

The importance of these three principal components in the auction context is as follows. The first PC suggests that while price (or more precisely, willingness to pay) increases over the course of auctions, there is significant variation in bidder's valuations during the early auction stages. This is in line with earlier research that has documented that the early auction phase is often swamped with “bargain hunters," i.e. bidders that are not very serious about the product itself but would like to “steal" it at a bargain price. Over time, the effect of bargain hunting diminishes and only serious bidders (with reasonable valuations) remain. The second PC shows a distinct difference in variation between the early stages (until to 0.3 or so) and the later stages. This is in line with auctions that experience different early- and late-stage dynamics (Jank & Shmueli, 2007). For instance, auctions that start out “fast" early, often see slow-downs in price during the later stages; conversely, auctions that procrastinate during the initial phases see tremendous price-acceleration towards the end. The second PC captures these differences in price dynamics. The third PC is noteworthy because it captures primarily differences in final prices. (Note that PC3 is largest in magnitude at t=1.0.) The reason is that although we are considering auctions for the same type of product (i.e. digital cameras), one camera (e.g. a *Nikon Coolpix*) has a different value compared to another camera (e.g. a *Canon Powershoot*) – PC3 captures differences in prices. It is also noteworthy that price differences only surface in the third principal component – the reason is that, as pointed out earlier, our trajectories are already scaled to the average price and hence control some of the variation in price.

We next look at level 2 eigenvalues and eigenfunctions, which characterize variations between different realizations of the same camera. The third and fourth rows of Figure 3 display the leading eigenfunctions at level 2. The first principal component, which explains the majority (84.6%) of the level 2 variation, captures variations mostly at the later parts of auctions. For example, according to Row 1 Column 1 of Figure 3, an auction with positive loading demonstrates more rapid increase in bidding prices between $t = 0.2$ and $t = 0.8$, while an auction with negative loading demonstrates slower increase in bidding prices between $t = 0.2$ and $t = 0.8$ and faster increase after $t = 0.8$. The second principal component explains 14.6% of the level 2 variation.

From a domain-level point of view, level 2 eigenvalues are extremely curious: since we are now looking at the second level of the hierarchy, we are controlling for product-specific differences. In other words, all observed variation is due to the bidding process and bidders' change in valuation for a product. As pointed out earlier, most auction theory suggest that a bidder's willingness to pay remains constant across the duration of an auction. That is, while one bidder may have a different valuation compared to another bidder, each bidder's individual valuation does not change as a result of the auction process. Aggregated across all bidders, we would thus expect to see willingness to pay to be distributed evenly – and with constant variance – across the entire auction duration. However, looking at the the first principal component, we see that willingness to pay is most variable during mid- to end-auction (between t=0.5 and 0.7). That is, even after controlling for product uncertainty, bidders' valuations change, and their valuation is most volatile immediately before the final phase of an auction starts. In addition, PC2 suggests that willingness to pay becomes extremely volatile towards the end of the auction. This may be caused by heated competition and last-moment bidding frenzy that often sets in during that part of the auction.

We also show fitted auction trajectories (Figure 4) for four selected cameras with two auctions per camera. The first two columns compare two cameras that load different on the first level 1 principal component, but similarly on remaining components. The third and fourth columns compare to cameras that load different on the second level 1 principal component, but similarly on remaining components.

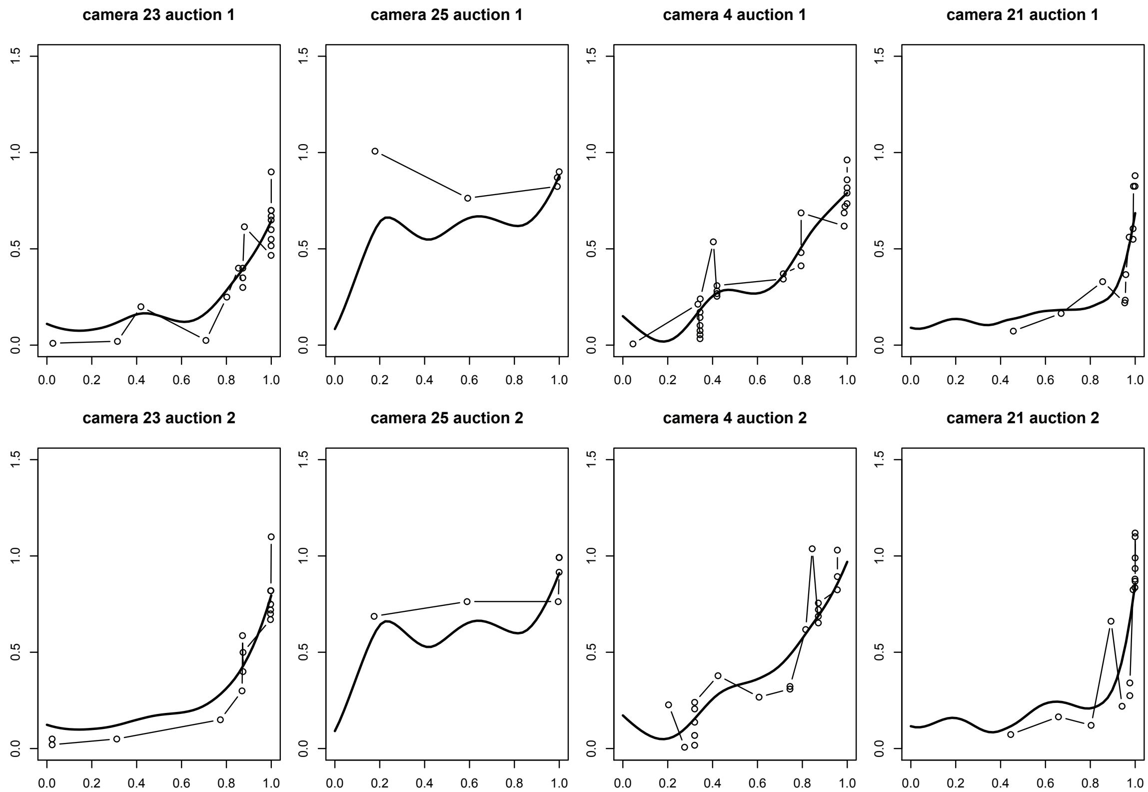

Figure 4. Fitted curves for four selected digital cameras. The dots and thin lines show observed bidding prices, while thick solid lines display fitted curves. Camera 23 and 25 load positively and negatively on the first level 1 principal component, respectively, but similarly on remaining components. Camera 4 and 21 load positively and negatively on the second level 1 principal component, respectively, but similarly on remaining components.

Overall, the significance of our analysis is that it allows, for the very first time, to partition bidders' willingness to pay into temporally different segments. In fact, level 1 analysis shows that early differences in willingness to pay can be attributed mostly to product differences; this is not surprising since different products are expected to be valuated differently. The surprising aspect is that bidders "digest" this difference (not immediately but eventually) during the early auction stage. However, differences in willingness to pay still persist, even when controlling for the product. Our level 2 analysis suggests that bidders uncertainty about the valuation is largest during mid- to late-auction. During this period of time, other bidders place bids, bidders update their own valuation and respond to the action of competitors. And finally, the second PC of our level 2 analysis suggest that bidding frenzy or last-moment bidding can result in huge variation in valuations which - so is our guess - is mostly driven by emotion rather than by rationale.

These results are interesting, especially when we compare the leading principal component from level 1 to that from level 2. These two components suggest the difference in the patterns of variation at two levels. Between different cameras, the early to mid auction behavior seems to explain much variation. Thus, auctions for some cameras start with relatively high price and bidding prices increase slowly, while auctions for others may start with low price and bidding prices increase faster later. If we consider the same camera, different realizations of auctions differ mainly in the mid to end auctions. The MFPCA methodology allows us to separate these two sources of variations, and study modes of variations at both levels.

2. Software implementation

R codes to implement the proposed multilevel sparse functional principal component analysis can be downloaded at `http://works.bepress.com/di/18`.

References

Bapna, R, Goes, P, Gupta, A & Jin, Y (2004), ‘User heterogeneity and its impact on electronic auction market design: An empirical exploration,’ *MIS Quarterly*, **28**(1), pp. 21–43.

Jank, W & Shmueli, G (2007), ‘Studying heterogeneity of price evolution in eBay auctions via functional clustering,’ *Handbook of Information Systems Series: Business Computing*.

Jank, W, Shmueli, G & Zhang, S (2010), ‘A Flexible Model for Price Dynamics in Online Auctions,’ *Journal of the Royal Statistical Society - Series C (Forthcoming)*.

Liu, B & Müller, H (2009), ‘Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics,’ *Journal of the American Statistical Association*, **104**(486), pp. 704–717.

Peng, J & Müller, H (2008), ‘Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions,’ *Annals of Applied Statitics*, **2**(3), pp. 1056–1077.

Reithinger, F, Jank, W, Tutz, G & Shmueli, G (2008), ‘Smoothing sparse and unevenly sampled curves using semiparametric mixed models: An application to online auctions,’ *Journal of the Royal Statistical Society - Series C*, **57**(2), pp. 127–148.

Roth, A & Ockenfels, A (2002), ‘Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the Internet,’ *American Economic Review*, **92**(4), pp. 1093–1103.

Shmueli, G, Russo, R & Jank, W (2007), ‘The BARISTA: A model for bid arrivals in online auctions,’ *The Annals of Applied Statistics*, **1**(2), pp. 412–441.