# An Analysis of Various Spline Smoothing Techniques for Online Auctions
# AMSC 689

Brian Alford and Lakshmi Urimi

December 10, 2004

# 1 Introduction

Looking for the missing trinket for your collection? Want to see if you can get a good deal on the latest mp3 player? Chances are, someone out there has what you want and is willing to part with it. With the advent of the Internet, consumer to consumer transactions have prospered through online auctions. Internet auction sites such as eBay, UBid, Overstock, and Amazon have become increasingly popular as internet access has become commonplace in the home.

Internet auctions have grown to become a multi-billion dollar business over the past decade. They have inspired the creation of numerous businesses devoted strictly to the service and support of buying, selling, and transaction processes. In addition, the academic world has begun to devote classes, research time, and conferences just to study the impact internet auctions play on the economy.

Over the past several years, many papers investigating the qualitative aspects of different bidding strategies, the economics of bidder behavior, and techniques of auction fraud have appeared in the literature [1, 2, 3, 4]. CITE PAPERS WE READ. Aside from standard regressions on the bids in auctions, there has been little to no work devoted to studying the dynamics of online auctions in a more in-depth, mathematical fashion.

This paper aims to summarize our attempts this semester to try to analyze these dynamics further. We will provide the details behind our explorations into applying standard and monotone spline smoothing techniques to fit functions to the bid history of an auction. Representing the data as functions, rather than sequences of $(t_i, y_i)$ pairs, and analyzing these the properties of these single units is called functional data analysis. We illustrate the basics of these techniques, demonstrate applying them to several auction datasets, and give our interpretation of the various results that we have obtained throughout this semester. In addition, we analyze the properties possessed by some of these smoothed curves and investigate the impact several auction factors play in determining auction prices.

# 2 Online Auctions

Online auctions are similar in nature to traditional auctions. Items are placed up for auctions, and people that enter the auction web site can bid against others for an item. At the end of the auction, a winner is declared and that person buys the item for the amount agreed upon during the process. Of course, the online process differs from the in-person experience. To begin, a person is required to register for an account at a site so it can keep track of items he bid on and items he sell. When someone enters an auction to place a bid, instead of a real, live auctioneer, the site's computers determine the amount he needs to bid in order to become the highest bidder. He can either bid this amount or enter his own maximum bid. For every additional bid that is placed, the computers orchestrating the auction compare his bid against the current bids and determine if his bid is the winner.

When someone wants to sell an item, it is also easy. eBay charges a small fee to display the item on its web site depending upon the value of the item and how much it sells for.

Each additional comment, description, or picture on the item on that page will cost a little. If a buyer wants to obtain more information about the item, he or she can contact the seller by sending a message through the site. In eBay, sellers can maintain a reserve price. A reserve price is the minimum price a seller is willing to accept for the item, which is often kept by him secretly. The seller is not obligated to sell the item if the reserve price is not met. The winning bidder must meet or exceed the reserve price and have the highest bid in order to be guaranteed the right to purchase the item.

At eBay, a registered user is able to leave feedback for the buyer of his item or a seller from whom he purchased an item. This information is very helpful to the later buyers or sellers. Generally, the rating of a user is seen on the web site with a parenthesis beside the user ID. Negative ratings correspond to bad feedback for the user, and a feedback rating of zero corresponds to either no feedback or where negative feedback is neutralized by the positive feedback. Positive ratings obviously therefore correspond to good feedback

Generally, the length of the auction and the bid increment differs from site to site. For example, eBay has 1, 3, 7, and 10 day auctions that strictly close after the pre-determined length of the auction, whereas Amazon allows the auctions to extend as long as bidders are still bidding (a going, going, gone style more like the traditional auctions). The bid increment is the amount by which a bid will be raised each time the current bid is outbid by someone else's maximum bid. It's is determined by the current high bid.

## 2.1 Bidding

e-Bidding differs from the traditional auction bidding. In traditional bidding, the bidding will be an increasing, monotone sequence since it is face to face, and every person knows the bids of his competitors. In e-Bidding, when someone places a bid, he or she enters the maximum amount he would be willing to pay for the item. This maximum amount is kept confidential from the other bidders and the seller. The eBay system compares each incoming bid to those of the other bidders. The system places bids on the behalf of the users, using only as much of the bid as is necessary to maintain the high bid position (or to meet the reserve price) or until a bidder's maximum amount is reach. If another bidder has a higher maximum, the first person is outbid by this new bidder. But, if no other bidder has a higher maximum, the first person maintains his hold on the top bid. The bidder could end up paying a significantly smaller amount less than his maximum price! This means he does not have to keep coming back to re-bid every time another bid is placed. This service, called proxy bidding, allows the computer to bid on a person's behalf to his maximum bid. Proxy bidding makes life very much easier for a busy person who wants to buy an item being auctioned but does not have enough time to closely monitor the acution. Proxy Bidding also make the auction web sites very popular among all age groups.

To give an example of the proxy bidding, we provide the following scenario. Suppose that the current high bid is $50 and bidder A wants to bid $110. He is not obligated to pay $110 right away. The system records bidder A's maximum amount ($110) but displays only the previous high bid plus the bid increment at $50 (i.e. $51, assuming $1 is the bid increment) as the current bid price. Suppose bidder B comes along and bids $100. Then bidder B is outbid by bidder A, and the current high bid price will increase ($100 + bid increment at $100), with bidder A still holding the high bid. At this instance, suppose that bidder B revises his bid to $115. Then A is outbid by B, and the new high bid that appears on the web page will be $110 + the bid increment. An illustration of this for a real auction can be seen in Figure 1.

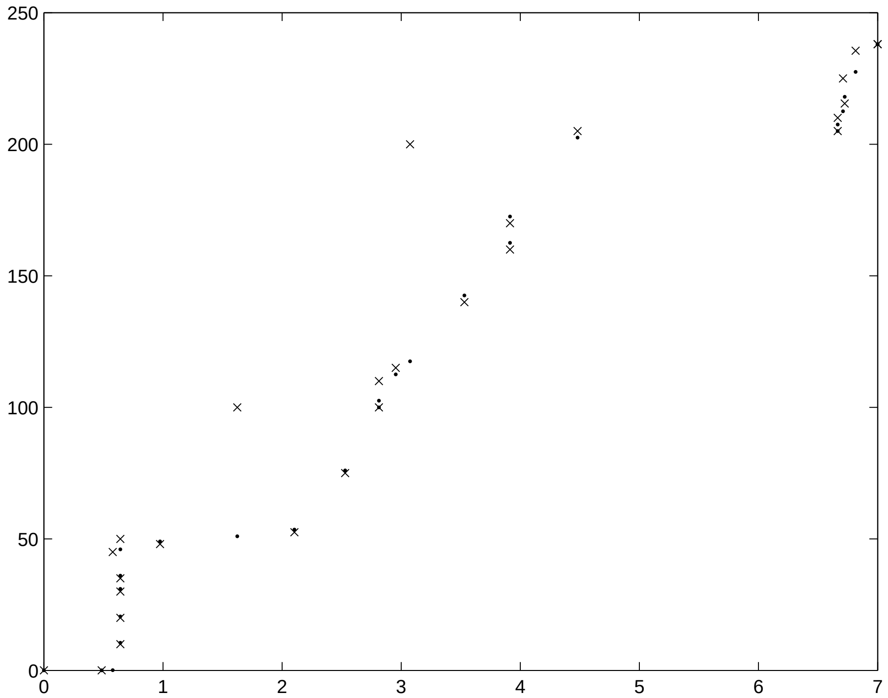



Figure 1: Sample plot of the actual bid ($\times$) with the live bid ($\cdot$) for an auction

## 2.2 Auction Data

We primarily worked with two data sets for all of our analyses. One is a collection of over 150 auctions for Palm handhelds. For those auctions, the variables we used in our analyses were auction ID, bid amount, bidder ID, bid time. Auction ID and bidder ID are self-descriptive. The other two variables are described as follows:

- bid time - the time when each bid is placed, and
- bid amount - the dollar amount placed as a maximum bid

The second data set we analyzed was data from two types of auctions on watches. This data set has over 470 auctions and contains more descriptive information than the Palm data set. The variables we used in our analyses are auction ID, auction category, high bid, start bid, seller rating, bidder ID, bidder rating, bids, and bid time. The following is a brief description of the data items:

- auction category - Rolex or Cartier (the two differen types of watches),
- high bid - highest bid in that auction,
- start bid - the opening price of the auction,
- seller rating and bidder ratings - ratings given to users according the their feedback,
- bids - the number of bids in the auction
- bid time - time when each bid is placed.

From these we computed the opening bid, final bid, and the average bidder rating for further study..

## 2.3 Live bids

The data we started with is the raw data of the exact bid history of the auction (exactly as it is displayed on eBay). We first transformed this data into a monotonic, increasing sequence to recreate the actual price see by potential bidders. i.e. we made the bids into live bids by following the rules of proxy bidding and reconstructing the path the actual price followed throughout the auction.

Reconstruction of the live bids is important because the live bids are the exact price information known by the bidder at the time his or her bid is placed. The bidder has no idea with total certainty what the maximum bid of the current high bidder is. Understanding the actual live bids is important in trying to determine the motivations behind the actions of bidders.

One downfall of the use of livebids is that the velocity of the price is not necessarily representative of the velocity of bids. If some bidder A comes along and places a bid much higher than the current livebid and some other person, bidder B, comes along and places his maximum bid just below the maximum bid of bidder A, the livebid jumps to be one bid increment over bidder B's maximum bid. If during the same amount of time in another auction 10 bids are placed during the same time between the bids of bidder A and bidder B, the live bid would increase by the same amount, but for drastically different reasons.

Another disadvantage is that the livebids do not allow us to determine when the winning bid is placed. Often, bid sniping causes the winning bid to occur in the closing moments of the auction, but occasionally, it can occur several days before the auction is over. To the livebids, there is no discrimination between these two cases. If we were to instead fit our curves to the actual maximum bids found in the bid history of each auction, we might be able to capture

# 3 Functional Data Analysis

A lot of observable quantities have an underlying functional relationship governing them. Fitting functions, be them linear or nonlinear, parametric or non-parametric, to try to represent this relationship has been the cornerstone of statistical analysis for decades. Determining the basic functional relationships between different observables is used to validate models, to forecast future outcomes, and so forth. However, often these studies only focus on the acutal observable data when the actual functions themselves can be studied in a useful manner.

The functional form of raw data can be estimated and depicted using a variety of different techniques, yet the result of fitting a function to raw data is a function that takes an independent input and creates an estimate of the observed data. Functional data analysis (FDA) is a technique that treats each function fitted to a set of data as a single observation. The goal of functional data analysis is to analyze these functions, characterize their behavior, and to see if any relationships are present between them.

A key ingredient to functional data analysis is finding ways to best represent the functional form of the observed data. In this section, we provide some of the main details of three techniques that can be used to extract the functional nature of the auction data in which we are interested: penalized spline smoothing, monotone penalized spline smoothing, and a combination of kernel smoothing with another smoothing technique. For an in-depth look at these techniques and their role in FDA, consult the text by Ramsay and Silverman [8] and the paper describing monotone smoothing by Ramsay [5]

## 3.1 Penalized Spline Smoothing

The spline smoothing method estimates a curve $x$ from observations $y_j = f(x_j) + \epsilon_j$ with two aims:

- The estimated curve gives a good fit for data, and
- The fit should not be too good; we do not want a curve that is excessively wiggly or locally variable.

In spline smoothing, the mean square error is a measure of poorness of estimation. This error can be reduced by reducing the sampling variance. This is the main reason why we impose smoothness restrictions on the estimated curve. I e. we vary the data gently so that we find a smooth curve to be our function estimate. To find the smooth function, we should know how rough the data is. One roughness measure of function is by its integrated squared second derivative.

$$PEN_2(f) = \int D^2 f(x)\, dx = ||D^2 f||^2 \tag{1}$$

This measure quantifies the total curvature of $f$ and gives the degree to which $f$ departs from a straight line. Tha is, the higher the value of $PEN$, the wigglier the function will be. This is because the second derivative of a wiggly function is large over much of the range.

In order to decrease the chance of the function being too wiggly, we should penalize this roughness(wiggliness) by adding a penalizing parameter $\lambda$ and attaching this term to the residual sum of squares. The corresponding functional is given by

$$PENSSE_\lambda(f|y) = \sum_j \{y_j - f(x_j)\}^2 + \lambda \times PEN_2(f)\,. \tag{2}$$

Our estimate of the function is obtained by finding the function $f$ that minimizes this error over all functions $f$ for which $PEN_2(f)$ is defined. The parameter $\lambda$ is a smoothing parameter that measures the rate of exchange between the fit to the dat , as measured by the residual sum of squares and the variability of function $f$ as quantified by $PEN_2(f)$. If we used a large $\lambda$, that means the function is very wiggly and we penalized it with a large value to make it look smooth. From this we can say that as $\lambda \to \infty$ the function becomes a straight line. On the other hand, as lambda decreases, the curve tends to become more and more variable since there is less and less penalty placed on its roughness. As lambda tends to zero, the curve approaches an interpolant to the data, satisfying $f(x_j) = y_j$ for all $j$. This curve does not vary much and it is the cubic spline with knots at the data points $x_j$ that is a result of the use of roughness penalty $PEN_2$.

Penalized spline smoothing is nice because is gives a quick and easy way for us to fit data in ways other than interpolation. It often provides good fits in a short amount of computational time, and it can be utilized as an inexpensive comparison for other methods. Much of the data contained in this section was from the FDA book by Ramsay and Silverman [8].

## 3.2 Monotone Data Smoothing

Monotone data smoothing is another type of smoothing that estimates a non-decreasing function curve $f$ from observations $y_i = f(x_i) + \epsilon_i$ where the values $\epsilon_i$ are assumed to be independent and identically distributed with mean 0 and variance $\sigma^2$, and the argument values are within the interval [0,T]. Monotone smoothing also utilizes a penalizing parameter $\lambda$, but the functional to be minimized is slightly different than the normal smoothing spline. The fitting criterion considered here is

$$F_\lambda(y|w) = N^{-1} \sum_i \{y_i - \beta_0 - \beta_1 m(t_i)\}^2 + \lambda \int_0^T w^2(t)\, dt \tag{3}$$

where

$$m(t) = \{D^{-1} exp(D^{-1} w)\}(t) \tag{4}$$

$$D^{-1} f(t) = \int_0^t f(s)\, ds \tag{5}$$

The first term in (3) is the least squares fitting criterion that is in the usual spline smoothing, except for the linear regression parameters $\beta_0$ and $\beta_1$. These two parameters are essential beacuase $m(0) = 0$ and $Dm(0) = 1$.

The second term, the penalty or the regularization term, has some of the characteristics of the norm of the second derivative used in cubic spline smoothing, but the role played by denominator in $w = D^2 f / Df$ is important since it keeps the fitted function from becoming too flat; that is, it keeps it away from the boundary condition $Df = 0$. The limiting case $\lambda \to 0$ is a straight line. The smoothing parameter may be chosen by cross-validation, but in all our studies we chose values for the parameter that fit the data best. A small discussion on the choice of this parameter is included later in the paper, but more can be found in [5].

Monotone data smoothing is of interest to this study because the auctions that we study consist of an increasing sequence of discrete auction prices that are monotonic. However, monotone smoothing splines are harder to implement than normal smoothing splines, are a more difficult to use in functional data objects, and require a considerable amount of computation time relative to normal smoothing splines.

## 3.3 Pre-smoothing the Data

Another technique we can employ in functional data analysis is a combination of different smoothing techniques, which could help to improve our fit of the observed data when the data is not well-behaved. By employing a two-step process where the data is first smoothed by one technique and then passed along to another technique for further smoothing, it is sometimes possible to obtain better functional forms than those obtained just a single smoothing.

One common method used to pre-process the data is kernel smoothing. Kernel smoothing is a simple, straightforward method that produces a continuous, functional estimate of the data using a linear combination of the local data. I.e., for any value of time in the interval in question, the estimates are given by

$$\hat{x}(t) = \sum_{j}^{N} S_j(t) y_j \,, \tag{6}$$

where $y_j$ is an observed data point and $S_j$ is the weight assigned to the data point $y_j$. These weights can be calculated using a variety of different user-chosen kernel functions depending on whether the user needs very accurate fits or fast computation. In either case, the end result is a functional object that provides a reasonable fit to the observed data.

After sending the data to a kernel smoother, we can then use another smoothing technique, such as the normal penalized smoothing splines or monotone smoothing splines described in the previous subsections, to try to further produce a good fit of the original data. Instead of sending the original observables to the smoother, though, some of the kernel-smoothed values are passed along to be smoothed by this other technique.

We mention the possibilty of using kernel smoothing because auction live bids often are not as well-behaved as we would like. In some auctions, bids are placed very close in time to one another but for large increases in price over the preceeding bid, pushing the live price of the item upward in a very short time short amount of time. Thus, there can be large jumps between successive points. In addition, the bids are not scattered evenly throughout the auction duration. As is the case with many auctions, several bids are placed near the beginning, relatively few during the middle, and several more close to the end of the auction. These two factors contribute to the presence of erratic data for each auction path.

## 3.4 Computational Implementation of Smoothing Methods

Luckily, two other researchers, J.O. Ramsay and B.W. Silverman, have done much to advance the field of functional data analysis. In addition to their recent book on the subject[8], they have written code for use in a variety of mathematical software programs (e.g., MATLAB and R) and have it available for public download on Ramsay's web site [7]. This website has many tools and gave us all the necessary backbone tools for performing our own functional data analyses. The sections that follow use this software extensively.

# 4 Smoothing Methods on Auction Data

The goal of our studies were to apply some of the basics of data smoothing to model bid paths of online auctions and to apply some of the ideas of functional data analysis to study to the dynamics of auction prices. As mentioned earlier, online auctions have only recently begun to receive very much attention from statisticians. Applying FDA to study auctions is a very new area of research. Thus, our study was aimed at attempting to analyze the basics of applying FDA principles to online auctions. Specifically, one of our main tasks was to look at the various aspects of the three smoothing techniques introduced in the previous section and to examine. In this section, we demonstrate two of these studies and comment on our observations.

For all of our analyses, we fit curves to the livebids of each auction in the data sets for the Palm handhelds and the Cartier and Rolex watches. We wrote programs in MATLAB to read and format the data, convert the bid history to live bids, and use the FDA package created and supplied by Ramsay [6]. Our original intention was to only study and compare normal penalized smoothing splines and monotone smoothing splines. However, for some auctions we noticed that the live bids were especially spread out through the auction duration. In several cases, the normal smoothing curve behaved in a less-than-desirable manner. Thus, we decided to also add the third type of data smoothing, kernel smoothing, into some of the analyses. After first applying kernel smoothing to the live bids, we fit these new values with a normal smoothing spline.

We begin by noting several options that we did not study in detail or that we chose based on our observations: placement of knots, number of basis functions, and the smoothing penalty. We experiemented with different values of the three options but did not find that choosing them cleverly gained any benefits. For example, we attempted to place more knots and basis functions at the ends of the auctions, but the curves did not give a noticeable improvement in the fitting. Placing more knots only caused the running time of our programs to increase. For all computations throughout the remainder of this paper, we used five equispaced knots, seven sixth-order basis functions, and smoothing on the second derivative for monotone smoothing and the fourth derivative for normal smoothing.

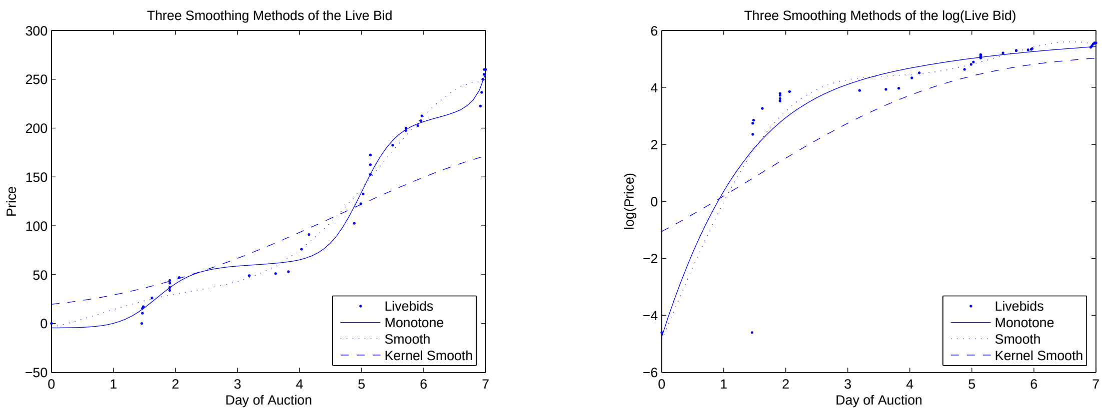
Much of the work that is shown in this section illustrates our attempt to look some of the other factors that appeared to be more significant when analyzing online auctions. In doing so, we fuel our overall attempt to determine which of the three smoothing techniques provides the best means for fitting live bids while still providing a reasonable representation of the bidding path throughout the entire auction.

## 4.1 Live Bids or Log(Live Bids)?

One topic that we investigated was whether the data smoothing techniques were affected by working with the live bids directly or whether they were more effective in smoothing the natural logarithm of the live bids. The entire analysis in this subsection provides some of our observations from fitting the data contained in the Palm handheld dataset.

In Figure 2(a) and Figure 2(b), we have displayed one auction from this dataset. We notice in the first figure that both the normal smoothing spline and the monotone spline perform fairly well at fitting the live bids while the kernel smoothing just captures the upward trend in the data. Throughout the auction, the normal and monotone spline both have a similar shape and path everywhere except between days two and three and on the last day of the auction. The monotone spline appears, though, to capture the data movement slightly better for this auction.

For the log(live bids), the curves look quite different than those in Figure 2(a), mostly because the auction used for this example's starting price was a penny, making the opening bid small on the natural log scale. However, the general comparison between the three curves in Figure 2(b) coincides with those from the figure for the normal live bids. The smoothing spline and monotone spline both fit the data well (except for the first bid around day 1.5), and the kernel smoothing curve gradually progrsses upward. The standard smoothing spline appears to adjust to the variablility in the data later on in the auction, though the monotone curve still performs well.



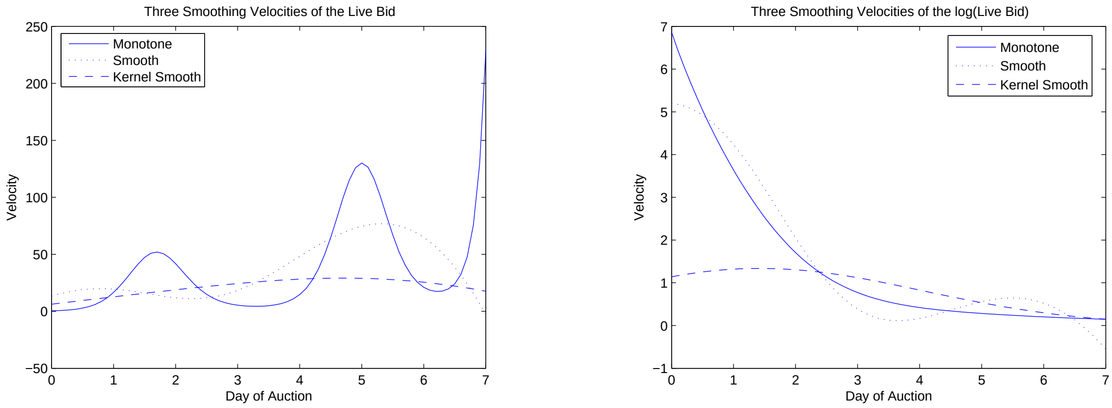

(a) Live bids

(b) log(Live bids)

Figure 2: Illustration of functional data fitting of the log(live bid) by normal and monotone smoothing splines for various values of the smoothing parameter

Running the program used to compute the three types of smoothing curves terminates in the shortest amount of time when log(live bids) are studied. Even when the stopping criterion is altered for the log(live bids) to reflect its change of scale, the program to find the curves still takes longer for the standard live bids.

Thus far, then, it appears that there is no clear indicator as to whether normal live bids or log(live bids) are to be preferred. However, looking at the first derivatives of these curves gives us a little insight as to why the normal live bids could be more important. In Figure 3(a) and Figure 3(b), the derivatives of these smoothing splines are given.




(a) Velocity for live bids

(b) Velocity for log(live bids)

Figure 3: Illustration of functional data fitting of the log(live bid) by normal and monotone smoothing splines for various values of the smoothing parameter

The velocities in Figure 3(a) can be directly interpreted. We easily can pick out the more active areas of the auction by noticing the local extrema and shape of the velocity curves. The derivatives accurately depict the flucutations in the movement of the price close to the beginning, three-quarters of the way through the auction, and during the auction's final moments. This is not so clear for the velocities for the log(live bids) in Figure 3(b). Though the velocity curve for the normal spline has some oscillation, representing some sort of bidding activity, the decreasing nature of the derivative of the monotone curve makes it appear that the auction is rather boring. The velocity curve makes it appear like there is not much interest in the auction.

It is not totally clear as to whether live bids or log(live bids) is a clear winner when it comes to performing the data smoothing. The log(live bids) run a little quicker than the non-transformed live bids, but the interpretation of the curves and derivative curves is a little decieving. However, we note that in datasets like the one for the Palm handhelds used for the analyses in this section, the items up for auction are all of a similar quality and, for the most part, have similar starting and ending prices. Thus, the curves for all the auctions had relatively similar live bid patterns and therefore, had similar bid paths fitted by the smoothing techniques. The normal live bids work well because the curves for all auctions are relatively similar. Whenever more heterogeneous data are used (e.g., many different items with widely ranging opening and final prices), the curves will be closer to one another and less variable if logarithms are taken first.

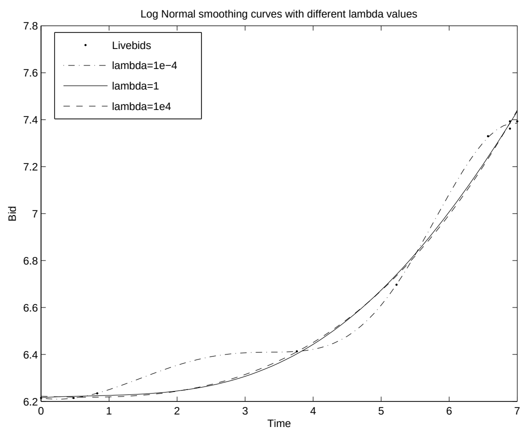
## 4.2 Choosing an Appropriate Smoothing Parameter

Another study we performed was to observe the effect of the smoothing parameter $\lambda$ on the normal smoothing, monotone smoothing and kernel smoothing when both live bids and log(live bids) are used. For this analysis, we looked at a single auction from the watches data set and displayed the smoothing splines that arose when we employed three different values of the smoothing paramter $\lambda$.

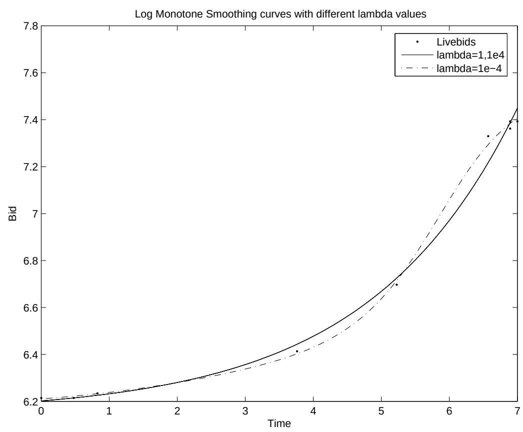
### 4.2.1 Log(Live Bids)

- Smoothing Spline: For very small values of $\lambda$, the curve fits through the data but it is wiggly. For a zero $\lambda$, the smoothing curve will have very wild oscillations which means that it is not penalized at all to fit smoothly to data. For $\lambda$ value one, we got an impressive smooth curve. For any value greater than one, the smoothness is not improved. So, from our observation, we felt that$\lambda = 1$ is a good penalizing parameter for normal smoothing curves. These three curves can be seen in Figure 4(a).

- Monotone Smoothing: For very small values of $\lambda$, the curve is a little wiggly but fits the actual data points well. When $\lambda$ increases to one, the curve fits to data very nicely and becomes much smoother. Like the normal smoothing spline, there is no improvement in smoothness for $\lambda$ values a few magnitudes above $\lambda$ equal to one. In Figure 4(b), these curves are displayed.




(a) Normal Smoothing

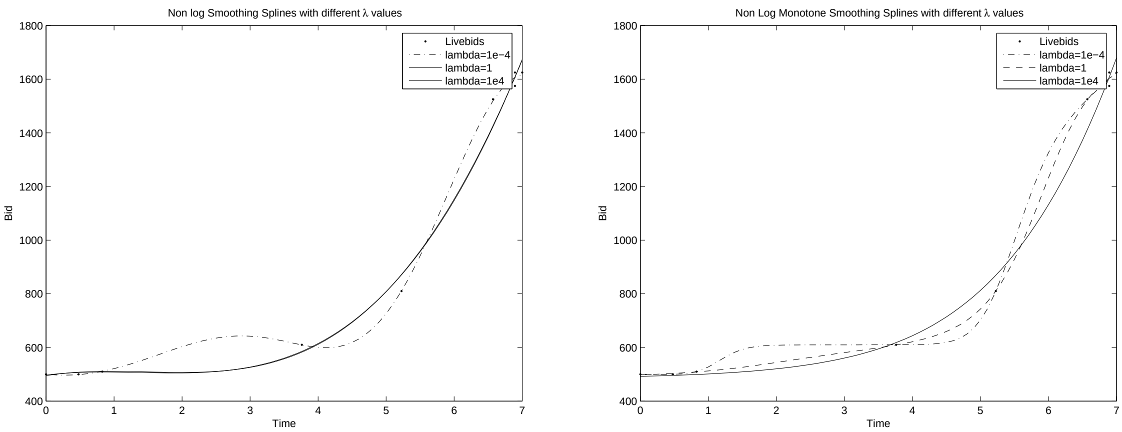



(b) Monotone Smoothing

Figure 4: Illustration of functional data fitting of the log(live bid) by normal and monotone smoothing splines for various values of the smoothing parameter

### 4.2.2 Live Bids:

- Smoothing Spline: This behaves almost similar to smoothing splines whenever log(live bids) are used except that the non-positive $\lambda$ values yield wigglier curves than those for log(live bids). The case when $\lambda = 1$ seems to give more well-fitting curves than the others. These curves are given in Figure 5(a).
- Monotone Smoothing: These splines also behave similar to their log(live bids) counterparts but with one exception. In Figure 5(b), we can observe that the smaller values of the smoothing parameter ($\lambda < 1$), the curve just fits the data but is not very smooth at all. For $\lambda = 1$ the curve is much better and improves very gradually as $\lambda$ increases.




(a) Normal Smoothing (b) Monotone Smoothing

Figure 5: Illustration of functional data fitting of the live bid by normal and monotone smoothing splines for various values of the smoothing parameter

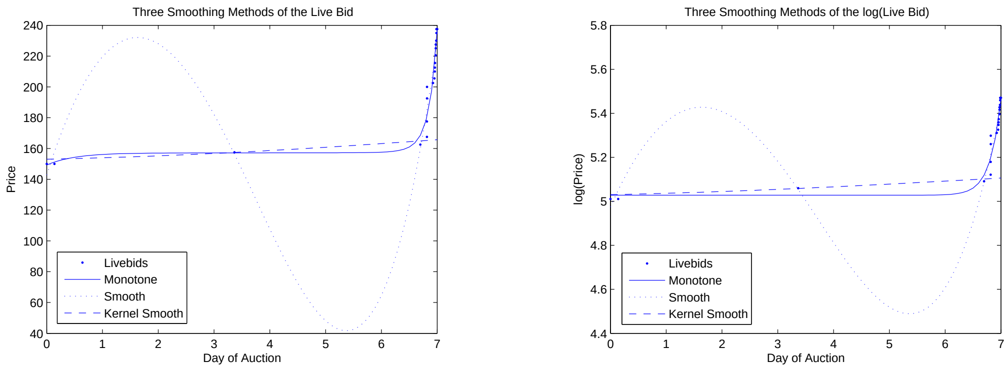
## 4.3 Normal vs. Monotone Smoothing Splines

One of the main goals of our study was to determine whether we should focus solely on monotone smoothing splines since our live bid data is monotone, or whether the smoothing splines would suffice in most cases. Throughout our experimentation with the normal and monotone smoothing splines, we observed many different auctions and viewed both the smoothed curves and velocities for both types of splines.

Some of the differences between normal and monotone smooting splines have already been observed. First, the monotone splines seem to more accurately represent the velocity of the auction price, particularly when the normal live bids are used as the observed data. In Figure 2(a), we can examine the end of the auction and see that both the normal and monotone smoothing splines are close passing through to the live bid observed data points.

However, the concavity of the two curves at the end of the auction is quite different. This is evident in Figure 4(a). The monotone velocity accurately depicts the fact that the auction is speeding up towards the finish, while velocity of the normal spline makes it seem as if little activity is occurring at the end of the auction. The smoothing spline does not always perform this poorly with regards to the velocity, but in general we saw the monotone spline to be especially better at detecting large velocity increases at the end of the auction.

Another downfall of the normal smoothing spline is that it sometimes obtains an excellent fit for the data, but the resulting curve is not a realistic auction-price path. An example of this activity is given in Figures 6(a) and 6(b). In both figures, the smoothing spline nearly interpolates the live bid data points exactly (even with a smoothing parameter value of 0.1). However, the large periodic oscillation that it uses to obtain the fit looks nothing like an auction path. The monotone spline, however, spikes only at the end of the auction, when the bid snipers attempt to steal the auction.




(a) Live bids (b) log(live bids)

Figure 6: When normal smoothing splines go bad

The biggest burden of using monotone splines is the computational burden required to obtain good, meaningful curves. The monotone splines are found by iteratively solving a difficult minimization problem, where the optimizer attempts to decrease the sum of the squared error until the improvement is less than a given tolerance. These iterations can take a considerable amount of time relative to both standard spline smoothing, even kernel smoothing. Whereas the normal spline smoothing can cycle through the dataset of over 470 auctions on watches, the monotone smoothing takes several hours to complete.

We believe that monotone smoothing is very advantageous for the analysis of auction data. First, the prices we input to the smoother were the livebid prices. They are the exact monotonic sequence of prices seen by anyone examining an auction while the auction is ongoing. Fitting monotonic data with a function that is also monotonic preserves the increasing nature of the price. In addition, they represent the data well and give good representations of the path of the auction price and the velocity of price that might be present in an auction. However, we feel that the complexity of the calculation of the monotone splines is a very undesirable feature when quick analyses need to be performed.

In studies where time is not an issue, we recommend using monotone splines citing the aforementioned notes on their robustness. However, for quick analyses, we feel that the normal splines will suffice just fine in most cases. In fact, we think that the normal spline fitting could be improved quite dramatically if we could develop some type of rejection criterion that would look not only at its ability to fit the live bid data, but also how realistic it represents an the path of live bid in an auction.

# 5 Regressions on the Auction Price

Regressions in general present an attempt to analyze at a set of data points to look for dependencies and relationships between the different factors that determine the data. Simple linear regressions test whether or not one variable influences another. Often, as is the case with auctions, there is more than one specific factor that contributes to the behavior of the observed quantity. When several different factors influence the movement of the dependent variable, a multiple regression is needed to analyze these influences. The general formula for a multiple regression of a single auction is shown in Equation (7)

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \ldots + \beta_n \mathbf{x}_n \, , \tag{7}$$

where $\mathbf{y}$ is the dependent variable, $\mathbf{x}_i$ is the $i^{th}$ regressor, and $\beta_i$ is the a parameter to be determined by the fitting performed in the regression. Each $\beta_i$ captures the relationship between the dependent variable and the $i^{th}$ independent variables when the other independent variables are held constant.

## 5.1 Regressions on Functional Data

Because online auctions are very complex, we performed multiple regressions on several independent variables that we believe play a role in the determination of the price of the item being auctioned: opening price, selling price, seller rating, current number of bids, and average bidder rating. The first three regressors are static and are constant over time for each auction. Two of the static variables are fixed at the beginning of the auction by the seller (opening price, seller rating) while the other static variable is determined at the end (final price). The current number of bids and average bidder rating are time-dependent, dynamic variables that change while the auction is ongoing as bids are placed.

Performing a regression on auction curves is different than a regression performed on a typical datset for two main reasons. First, the data from auctions is time series data. Since we wish to analyze the relationship between the price, which changes over time, with several independent variables, the regression coefficients of these variables will likely be time-dependent. Second, our data is a set of curves, not discrete values. Thus, we must first create a grid of $m$ discrete time values $t_i$ over the entire time interval $[0, T]$ and perform a multiple regression at each of these individual times $t_i$. With the $m$ regressions throughout the auction duration, we can trace the value of each individual $\beta_i$ over time.

The regressions for the dynamic variables and the static variables were a bit different from one another. At each time point the regressor variable can be found by just matching each curve with the variable associated with that auction. I.e., at each $t_i$, each static variable can just be looked up in a data structure containing various pieces of information for each auction. For the dynamic variables, we calculated the value of the variable at discrete timepoints throughout the auction duration and used spline smoothing to create smooth functions representing the value of the variable over the time interval. When the regression is performed at each discrete time, the smoothed function of these variables for each auction is evaluated, and this value is used in the regression.

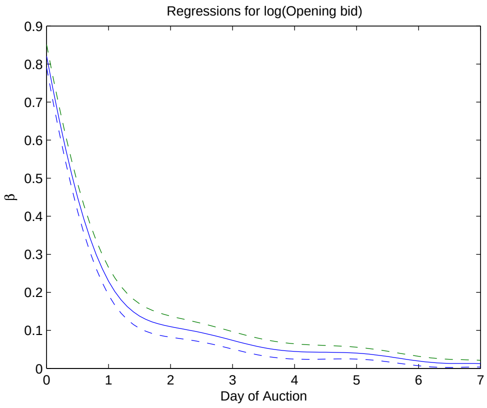
The subsections that follow illustrate these regressions for the five predictors noted above on the data set consisting of over 470 auctions of Cartier and Rolex watches. We started by using the three smoothing techniques to fit the log(live bids) data to get a family of auction curves for the entire data set. For each family, we discretized the time interval into 71 timesteps, performed a multiple regression at each of these time values, and finally traced these regression coefficients over time to determine the curves for each beta. For each regressor, we only show the multiple grid regressions for the monotone splines and the kernel- smoothed splines, though they are not always the best fits near the extremes of the auctions.

We attempted to also perform the regressions on the normal smoothing splines, but several were so erratic that the regressions were meaningless. If we could expel the auction curves with wild behavior, we would be able to compare the regressions with those in the following subsections. In addition, we performed multiple regressions on the velocities and accerelations of the of the curves but did not include them in this paper. We felt the regressions had little meaning because the confidence bounds for each $\beta$ curve crossed several times throughout the auction and sometimes diverged. We would like to study this in further detail in future work.

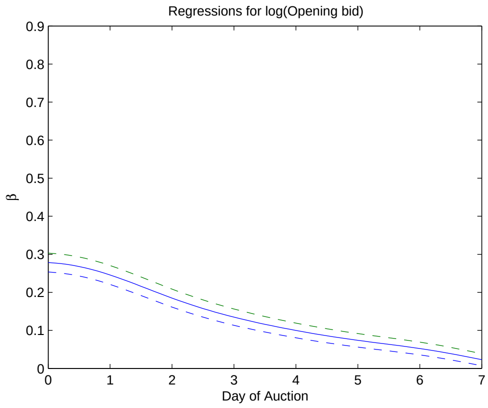
We carried out the regressions using MATLAB's `regress` function, keeping the value for $\beta_1$ through $\beta_5$, along with their confidence interval for each $t_i$. With the 71 values for each $\beta_i$, we performed a simple spline smoothing on both the value of the beta and the confidence limits to obtain a better curve than a simple trace of the 71 values of beta over time.

## 5.2 Opening Price

The opening price of the auction is the first regressor we studied. It is a static variable set by the seller when the item is placed up for auction. The auction price after the first bidder sets his maximum bid, regardless of how high this maximum is over the opening bid, is always equal to the opening price. The opening price appears to have a direct impact on the bidding during the auction. Some sellers choose to set the starting price low to attract attention and draw a lot of bidders. Other sellers choose to set the starting price high to ensure that they get a fair value for the item they're selling. Still others combine these two ideas by setting a low starting bid and placing a reserve on the item to ensure the auction price reaches a value acceptable to the seller.




(a) Monotone Smoothing




(b) Kernel Smoothing

Figure 7: Regression coefficient for the opening price predictor for both monotone and kernel smoothing

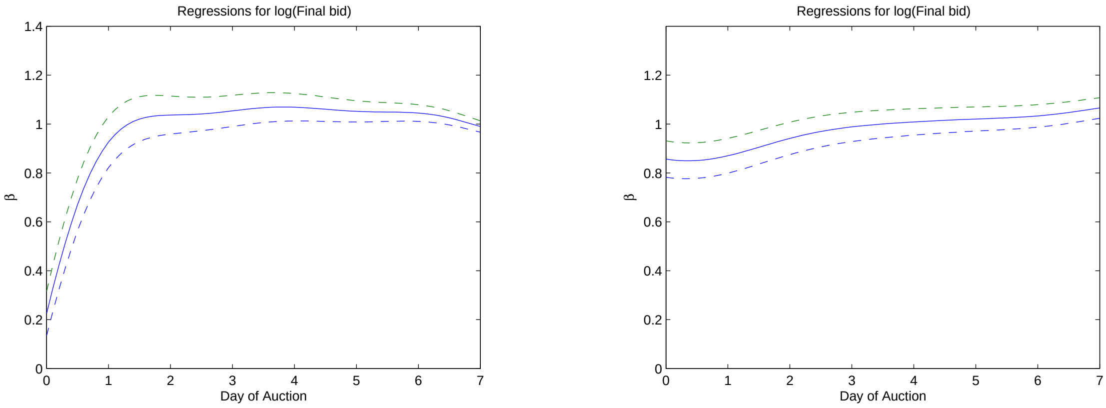
In Figure 7(a), we see the relationship between the logarithm of the auction price versus the price item opens for. The decreasing value of $\beta$ from around 0.8 to 0.1 after two days shows that the impact of the opening bid has a large role to play with the price movement at the beginning of the auction, as expected. As the auction progresses, this impact becomes less and less significant. This gradual decreasing nature can also be seen in Figure 7(b).

The result illustrated in Figure 7(a) and Figure 7(b) is as expected. Regardless of whether the auction starts out high or low, the opening bid has a direct impact on the early parts of the auction. In many auctions, a lot of action occurs both at the beginning of the auction and near the end. For auctions that have high starting bids, few bids are placed at the beginning, and thus the auction price remains high simply because the opening bid was set high. For those auctions that have a low starting price, often the price is buoyed up quickly at the beginning of the auction because a lot of attention is attracted to the auction because of the low starting price. In either case, the opening price influences the bidding path of the auction.

## 5.3 Final Price

The second variable we studied was the price at which the item was sold. The final price is also a static variable, but it is never known until the auction closes. An interesting note about final prices is that since eBay auctions are second-price auctions, the final price may not have been the same value as the highest maximum bid. However, as outsiders examining the auction, we will never know the maximum value someone was willing to pay for the item; only the winning bidder knows that information.

We obviously expect that a rather strong relationship between the final price and the auction price at the end of the auction, but we also expect that it governs the behavior of the price throughout the auction. This is particularly the case with items that are commonly available through means other than auctions and the fair value of the item being auctioned is known (e.g., consumer electronics, cars, and movie titles). Bidders are not likely to allow someone to purchase a high-demand item at a large discount to its market value; they will bid to try to obtain the item at a discount for themselves, bid on the item, and therefore, increase the item's price. This behavior we expect to see with a regression on the final price.




(a) Monotone Smoothing

(b) Kernel Smoothing

Figure 8: Regression coefficient for the final price predictor for both monotone and kernel smoothing

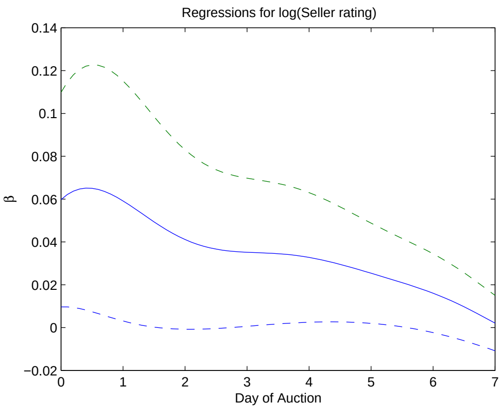
In Figure 8(a) and Figure 8(b) , the positive relationship between the final price and the auction price is evident by the increasing nature the $\beta$ curve for the monotone spline in the first part of the auction. From day two onward, this relationship levels off at about a value of one. The fit for $\beta$ for the kernel-smoothed curve also demonstrates the importance of the final price throughout. We imagine that this is indeed consistent with our belief about bidder behavior during the early days of the auction. Whether the opening price is large or small, bidders attempt to find what they consider to be a good "deal," using their knowledge and research about the going rate of the item.

We suspect that this relationship is not necessarily true for all items being auctioned. Items that are hard to find and very unique likely will not attract a lot of bidders, and the fair value might be hard to ascertain. Therefore, the final price in those auctions may not play as big a role in determining the auction price.

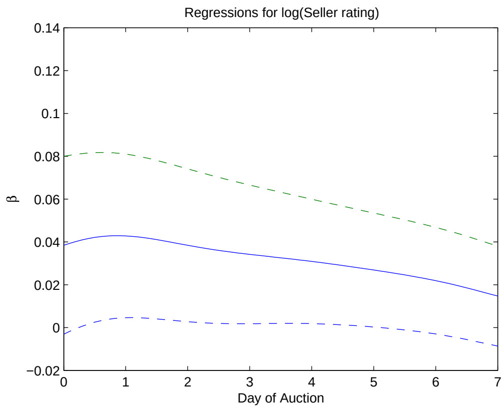
## 5.4 Seller Rating

Seller rating was the next regressor we examined. The seller rating attributed to each auction is determined by the seller's previous transactions and is set at once he or she puts the item up for auction. We imagine that auctions with higher seller ratings would fetch more attention, bidding, and a higher final price than sellers with lower ratings. This occurs because a bidder might be reluctant to take the risk of bidding on items sold by sellers with bad ratings. He or she might not be willing to pay for an item only to find that it arrives in worse condition than stated on the auction listing, to have the item arrive weeks after the auction is completed, or in the worst case, to be scammed and never receive the item.

As we can see in both Figure 9(a) and Figure 9(b), seller impact has little influence in the determination of the auction price. The figure shows that the $\beta$ for seller rating mostly remains in the insignificant range of values between 0 and 0.05. We believe this to be the case because the items being auctioned in our study were somewhat similiar in that they were either a) items rare and rather expensive or b) items wanted by many people. In the first case, the bidder might be willing to bid on an item despite the low seller rating because the item is unique and hard to find. For the second case, when an item is in demand by many people, the bidder might be willing to bid on an item with a lower seller rating just to make sure that he or she is able to obtain the item.




(a) Monotone Smoothing




(b) Kernel Smoothing

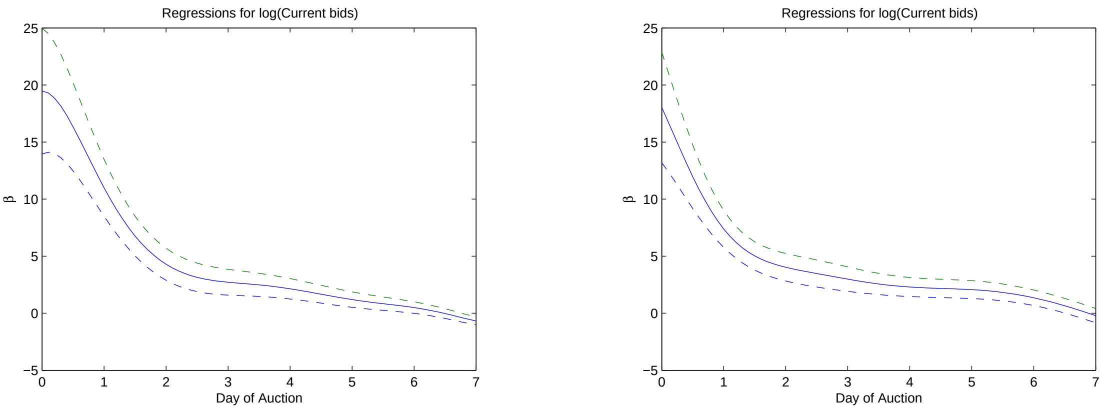
Figure 9: Regression coefficient for the seller rating predictor for both monotone and kernel smoothing

We feel that this study understates the impact that the seller rating plays in determining the auction price. To more accurately analyze the impact of the seller rating, we would need to obtain data for concurrent auctions on the same or very similar items. Doing so would allow us to directly study the premise that bidders choose to bid on items with higher seller ratings by comparing these auctions. Obtaining enough data for such a study to be significant could be rather difficult, and the study would certainly become more complex. However, this would give us a better indicator of the seller rating versus

## 5.5 Current Number of Bids

The first of our dynamic variables, this variable captures up the current number of successful bids placed on the item, either by a user or via proxy bidding. As described in Section 5, at each $t_i$ in the grid placed on the time interval, the value of the smoothing spline fitted to the current number of bids is evaluated for each auction. This value is then paired with the price of the auction to determine the values passed along to the regression.

The analysis of the current number of bids is important because it measures the impact of a large number of bids on auction price movement. Sometimes, a large number of bids implies that many bidders were battling for the item and drove the price upward. Other times, a high number of bids might only be the result of a bid sniper program attemping to probe another bidder's high bid at the end of the auction. Such actions occur when several bids with small incremental increases occur to try to steal the auction away from the current high bidder with the smallest improvement over the current price. In such cases, more and more bids may not move the price very much at all.




(a) Monotone Smoothing

(b) Kernel Smoothing

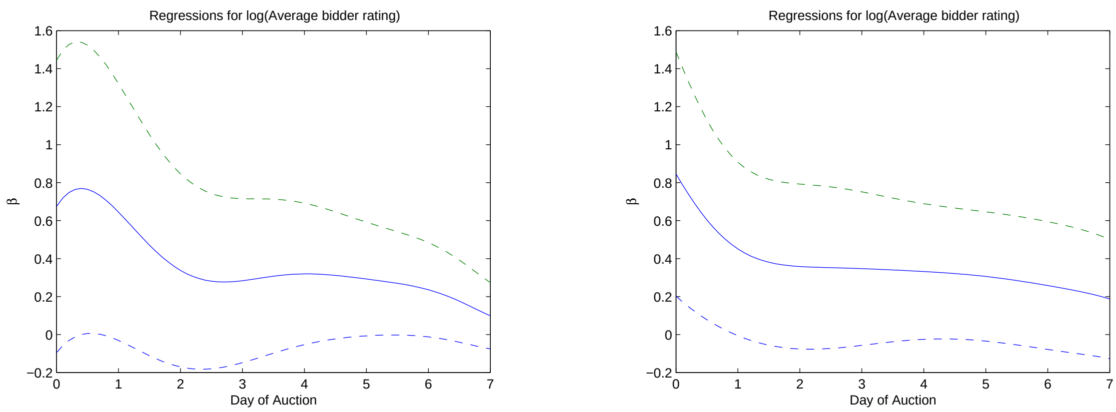
Figure 10: Regression coefficient for the current number of bids predictor for both monotone and kernel smoothing

The values for $\beta$ in both Figure 10(a) and Figure 10(a) start out rather large and indicate that the auction price at the beginning of the auction is influenced greatly by the current number of bids.

## 5.6 Average Bidder Rating

Average bidder rating is the other dynamic variable we considered in our regressions. It captures the average bidder rating of all the bidders for an auction up to a given time. Similar to the current number of bids, we smoothed the average bidder rating for each auction with a spline since the average bidder rating varies a great deal throughout an auction. At each discrete time value, this function is evaluated and paired with each auction price for the regresssion.

Economic theory suggests that a bidder should not conceal his or her maximum bid but instead release it with the first bid he or she places [1]. Often, inexperienced bidders only bid a small increment over the current price, attempting to concealing his or her maximum bid until later in the auction. By analyzing the average bidder rating of the auctions, we are looking for behavior of this sort.




(a) Monotone Smoothing

(b) Kernel Smoothing

Figure 11: Regression coefficient for the final price predictor for both monotone and kernel smoothing

In Figure 11(a) and Figure 11(b), we see almost the exact same behavior for $\beta$ in both curves. We interpret the positive value in the first two days of the auction as a sign of the following possible activity. Experienced bidders often are not afraid to reveal their maximum bid and often place a bid early on in the auction duration substantially larger than the live bid. Other bidders soon after attempt to probe his high bid by bumping up the price quickly to attempt to become the highest bidder. Without the high bid of the experienced bidder, the price of the item would likely not increase until later on in the auction.

# 6 Conclusion

In this paper, we have introduced online auctions as an interesting area where new statistical data analysis techniques can be applied. We have provided a brief overview of some of the different smoothing techniques available at our disposal to model the dynamics of these online auction price. We also illustrate some of our own observations and experiences that have arisen from our hands-on application of three smoothing techniques to fitting auction data. In the final section, we have also given a deeper look at functional regression to try to better pinpoint some of the determinants of auction prices.

Although we feel like we have made progress in our aims to apply the FDA tools to auction data, we are only at the beginning of the study. Since online auctions and FDA are rather nascent, there is a wide realm of things that we have not investigated this semester. In the future, we can hopefully obtain more data from a larger variety of auctions and can further explore the nuances of the different smoothing techniques to look for the most appropriate way to model online auctions.

# References


[1] Bajari, P. and A. Hortacsu. "Economic Insights from Internet Auctions." Working paper. http://www.econ.duke.edu/ bajari/research.html.

[2] Bapna, R., P. Goes, and A. Gupta. "User Heterogeneity and its impact on electronic auction market design: An empirical exploration." *MIS Quarterly* 28, no. 1 (2004): 21-43.

[3] Kauffman, R. and C. Wood. "Running up the bid: Detecting, predicting, and preventing reserve price shilling in online auctions." Working paper. http://misrc.umn.edu/workingpapers/abstract/0304.asp.

[4] Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves. "Pennies from eBay: the determinants of price in online auctions" Working paper. http://ideas.repec.org/p/fth/vander//00-w03.html.

[5] Ramsay, J. "Estimating smooth monotone functions." *Journal of the Royal Statistical Society B* 60, no. 2 (1998): 365-375.

[6] —. Functional Data Analysis. http://ego.psych.mcgill.ca/misc/fda/index.html.

[7] —. Personal Homepage for J.O. Ramsay. http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html.

[8] Ramsay, J. and B. Silverman. *Functional Data Analysis*. New York: Springer-Verlag, 1991.