Forecasting eBay's Online Auction Prices using Functional Data Analysis

Shanshan Wang*, Wolfgang Jank† and Galit Shmueli‡

University of Maryland, College Park, MD 20742 §

December 10, 2004

Abstract

The goal of this work is to derive models for forecasting the final price of ongoing online auctions. This forecasting task is important not only to the participants of an auction who compete against each other for the lowest price, but also to designers of bidder-side agents. Forecasting prices in online auctions is challenging from a statistical point-of-view because traditional forecasting models do not apply. The reasons for this are three typical features of online auction data: a) unequally spaced bids; b) the limited time horizon of an auction; c) the dynamics of bidding change drastically over time. We propose a dynamic forecasting model for the auction price that can overcome these challenges. We use modern functional data analysis methods that take into account the price velocity and the price acceleration as the basis for our forecasting model. We show that our model has high forecast accuracy and it outperforms traditional methods. Our results also allow for new statistical insight into auction forecasting. We find that the forecasting accuracy increases as we predict further into the future, that is, further towards the auction end, and we tie this finding together with existing auction theory.

Key words and phrases: functional data analysis, smoothing, online auctions, bid sniping, price velocity, price acceleration, autoregressive models, exponential smoothing.

*PhD student, STAT Program, Department of Mathematics. email: shanshan@math.umd.edu

†Corresponding Author, Assistant Professor, Decision and Information Technologies Department, Robert H. Smith School of Business. email: wjank@rhsmith.umd.edu

‡Assistant Professor, Decision and Information Technologies Department, Robert H. Smith School of Business. email: gshmueli@rhsmith.umd.edu

§This research was partially funded by the NSF grant DMI-0205489. The authors gratefully acknowledge support by the Center of Electronic Markets and Enterprises, U of Maryland.

1 Introduction

Electronic commerce, and in particular online auctions, have created a lot of public interest in recent years. One of the main drivers of this interest is eBay. On any given day, there are several million items across thousands of categories for sale on eBay. Ebay's popularity is evidently quantified in the following numbers: in 2003, $24 billion was reported in gross merchandise sales, up from $15 billion in 2002; the cumulative confirmed registered users at the end of 2003 totaled a record 94.9 million, which was a 54% increase over the 61.7 million users reported at the end of 2002; and eBay hosts approximately 154,000 stores worldwide (see *http://investor.ebay.com/news/Q403/EBAY012104-712351.pdf/*).

In this paper we develop forecasting models to predict the final price of online auctions. Predicting the final price is interesting, especially to the participants of an auction who compete against each other for the lowest price. Forecasting online auctions is also important for buyer agent companies who pick the most lucrative auctions for their clients (see e.g. *http://staging.argosydev.com/agentproxy/*) or sniping agents that provide last-moment bidding services (e.g. *http://cniper.com/*). Predicting an auction's outcome is also important for the seller who may have the option to sell the item before the auction is over.

While online auctions have created a lot of interest among the public, their popularity is also one of the reasons for an increasing amount of scholarly research. Another reason is the availability of a vast amount of bidding data which has stirred a great number of empirical studies in economics and information systems. For instance, Bapna et al. (2004) find that significant heterogeneity exists in the users of electronic markets like eBay and develop a stable taxonomy of bidding behavior in online auctions. The determinants of bidder and seller behavior are also explored by Bajari and Hortacsu (2003). The winner's curse is a common phenomenon in online markets. With the existence of a common value, the winner's curse occurs when bidders are not aware that they will only win the auction when they have the highest evaluation of the product and as a consequence, inexperienced bidders frequently overpay. Bajari and Hortacsu (2004) take this as manifestation of informational asymmetry in electronic markets. Bajari and Hortacsu (2003) use a structural econometric model of bidding on eBay to measure the extent of the winner's curse. Feedback mechanism is a popular feature of online auctions and can decrease the informational asymmetries between buyers and sellers. Ba and Pavlou (2002) demonstrate that proper feedback mechanisms can induce trust and trust can reduce information asymmetry by reducing transaction-specific risks. A detailed discussion of important differences between internet-based feedback mechanisms and traditional "work-of-mouth" networks can be found in Dellarocas (2003). The author also surveys important issues related to design, evaluation and use of online feedback mechanisms. Lucking-Reiley et al. (2000), investigating the determining factors of price, find that a seller's feedback rating has a measurable effect on auction prices, with a few negative ratings having a much greater impact than many positive ratings. They also find that the magnitude of the opening bid and the use of secret reserve prices tend to have a positive effect on the final auction price. Other empirical work observes the prevalence of "bid sniping" in eBay's auctions (see Roth and Ockenfels (2002), Bajari and Hortacsu (2004), Lucking-Reiley et al. (2000)). Bidders hold back their bids as long as possible, resulting in a huge amount of bids placed in the last moments of the auction. Last-moment bidding may be a response by rational bidders against naive bidders or a form of "tacit collusion" by the bidders against the seller. Generally, bidders feel that they increase their chances of winning by revealing their valuation as late as possible during the auction. Despite the prevalence of bid sniping, "early bidding" also exists. People may bid early to establish their time priority on multiunit auction sites like Ubid.com or perhaps to assess their competition (Bapna et al. (2004)). In either case, both "late bidding" and "early bidding" indicate that the dynamics change tremendously over the course of an online auction.

While online auctions experience an increasing amount of interest in the economics and information systems literature, relative little work, with a few recent exceptions, has been done from a statistical point of view. To deal with the overwhelming amount of data found on auction sites like eBay.com, Shmueli and Jank (2004b) introduce graphical methods such as profile plots and statistical-zooming to visualize online auctions in an informative way. Their visualizations allow for a straightforward inspection of bidding heterogeneity, manifested in "early bidding" and "sniping". Modelling bid arrivals during an auction, Shmueli et al. (2004) introduce a class of 3-stage non-homogenous Poisson processes to describe the heterogeneous stages of bid arrivals within a finite time period. Furthermore, recent research of Jank and Shmueli (2003) proposes the use of modern statistical methods, in particular functional data analysis, to investigate the dynamics of the bidding process rather than just looking at the auction statically. The authors utilize functional cluster analysis to find that the price-dynamics, like the price-velocity and price-acceleration, can be quite different for different auctions. In an extension of that work, Shmueli and Jank (2004a) employ functional regression to investigate the effect of covariates like the opening bid on the dynamics of the auction. Interestingly, they find that during the beginning of the auction, high opening bids are associated with faster acceleration in the bidding process while towards the auction end, high opening price are associated with a slow-down of the price-dynamics.

One fundamentally intriguing topic that has been absent from the literature to date is the forecasting of the outcome of an online auction. Predicting an auction's final price is important to the bidder, the seller or to service companies associated with bidding and selling. Predicting online auctions though is not an easy task. Online auction data have particular features that make traditional forecasting methods hard or even impossible to apply. Online auction data typically arrive as a sequence of bids placed over time. However, since the bid arrival is not evenly spaced and since the auction stops at a fixed ending time, traditional time series methods do not yield good predictive models. Moreover, Jank and Shmueli (2003) observe that the dynamics of the price change sharply throughout the auction, especially towards the end of the auction. Traditional forecasting methods, though, do not take these changing dynamics into account.

To overcome the difficulties induced by the features of online auction data we propose the use of functional data analysis methods. Functional data analysis (FDA) has at its heart the analysis of curves rather than points or vectors and its origin is often attributed to the work of Ramsay and Silverman (1997). Statisticians have found a number of ways to generalize statistical methodology to the functional data context. James et al. (2000), for instance, present a technique for principal component analysis of sparse functional data. While Faraway (1997) performs regression analysis for a functional response, Escabias et al. (2004) discuss logistic regression when the response variable is binary, and James (2002) considers generalized linear models for functional data. While there has been some work on forecasting via FDA (see Valderrama et al. (2002)), there has been no attempt to incorporate the dynamics. In this work we propose a new and dynamic approach to auction forecasting using an auction's velocity and acceleration, and we show its superiority over traditional methods.

The article is organized as follows. In Section 2 we briefly introduce the auction mechanism on eBay, data availability and the data that we use in this work. In Section 3 we derive our l-step forecasting model based on an FDA approach. Results of applying the method to our data are given in Section 4. Section 5 is devoted to conclusions and future extensions.

2 eBay's Auction Mechanism, Data Availability and Data Used

2.1 eBay's auction mechanism

eBay is the largest consumer-oriented auction site over the internet world. How eBay works is a prototype of internet auctions. All eBay auctions use an ascending-bid format and there is a fixed ending-time and -date set by the seller. This is in contrast to other auctions, like Amazon, where the auction continues until 10 minutes after the last bid arrival. eBay uses a so-called "proxy bidding" system, which updates the bids on the buyer's behalf. In this system, bidders only specify the maximum amount they are willing to pay for an item. The system keeps this amount secret, and increases the bid just one increment over the next highest bid. Thus, the person with the highest proxy bid is assured to be in the lead of the auction, and the winner's price is determined by the magnitude of the second-highest bidder. One side effect of this procedure is that it puts some incentive for bidders to submit bids early since the earlier bid wins in the case of a tie.

2.2 Data Availability on eBay

A great deal of information on eBay auctions is publicly available. eBay posts on its web site the complete bid histories of closed auctions for at least one month after the closing date. A typical bid history includes information about the magnitude of each bid and when this bid was placed (see Figure 1). Other information is also available. Besides a user's ID, eBay posts her feedback rating which reveals the difference between the number positively and negatively rated transactions. eBay also discloses information about the seller, the auction format and about the item sold.

Large amounts of data are available on eBay. While eBay data could, in theory, be collected "by hand" simply by browsing through individual web pages, in practice this can be very time consuming. Practically, large amounts of eBay data can be collected, in short time, using web agents or web crawlers. Web crawlers are pieces of softwares, typically produced in software like Java or Pearl, that visit a number of pages automatically and extract the required information. That way, high quality information on a large number of auctions can be gathered in a halter of only moments.

2.3 Data used in this Study

The data in this study are the bid histories of 185 closed 7-day auctions for the Palm M515 Personal Digital Assistant (PDA). Figure 2 shows a scatter plot of the bids in one of these auctions. We can see that bids arrive at unevenly spaced time intervals. While the number of incoming bids is sparse during some periods of the 7-day auction, it can be very dense other times such as at the very beginning of the auction and during the ending period.

Figure 3 provides the aggregated data for all N=185 Palm auctions. Notice that most of the bids arrive in the last minutes of the auction, which, as we have pointed out earlier, is a typical feature of eBay's auctions.

In this work we focus on developing forecasting methods for auctions of a specific product, like the Palm M515 PDA. Having the capacity to forecast an auction for a specific product allows, e.g. the bidder, to focus her bidding-efforts on a select sub-sample from a potentially large population of auctions for the identical good. Forecasting auctions for a specific product also allows for an exact measure of the forecasting-accuracy since the product value is relatively well-known. And finally, the lessons learned from this task can be used to derive more complex forecasting models for auctions of diverse product types.

3 Auction Forecasting via Functional Data Analysis

The focus and goal of this paper is to put online auction data into a context suitable for statistical analysis and forecasting. Auction data typically arrive as a sequence of bids over a period of time. Taking a functional data analysis approach, we treat the bids from a single auction as realizations from a continuous price curve. Our population of interest then consists of a large and flexible class of curves that can capture the heterogeneity across auctions. Our first step is then to estimate, or, recover the underlying curve for each auction by means of smoothing. One of the advantages of functional data analysis is that it can take into account the price dynamics of the auction. That is, we model not only the price position, but also the price velocity and the price acceleration. As our example will show, using the price dynamics significantly improves upon the predictability of an auction price.

While functional data analysis allows for an estimation of the price dynamics, it has additional advantages over traditional time series methods. Traditional methods (such as autoregressive models, or exponential smoothing) are not directly applicable to the auction context. One reason for this is the unevenly spaced arrival of bids. Another reason is that traditional methods do not take into account the finite time horizon of the auction. Traditional time series models typically apply to data like the everyday stock price of Microsoft, the weekly sales revenue at Macy's, the monthly temperature of a climato-observation on Iceland, all of which are evenly spaced time series whose ending points are unknown. As our study will show, having a finite time horizon results in surprising new statistical insight into forecasting methodology.

Our approach consists of three parts: First, we smooth the data to recover the underlying price curve. Second, we use this curve to model and predict the price dynamics. Finally, using the predicted dynamics, together with other relevant covariates, we estimate an econometric model and use it for prediction.

3.1 Smoothing and Recovery of the Price Curve

Our analysis begins, as it is often the case in functional data analysis (e.g. Ramsay (2000)), with some data pre-processing steps. We denote the time that the ith bid was placed, $i = 1, \ldots, n_j$, in auction j ($j = 1, \ldots, N$) by t_{ij}. Note that due to the irregular spacing of the bids, the t_{ij} vary for each auction. In our data, $N = 185$ and $0 < t_{ij} < 7$. Let $\hat{y}_i{}^{(j)}$ denote the bid amount placed at time t_{ij}. We treat these values as noisy realizations of a continuous price function and our goal is to reconstruct this function using appropriate smoothing techniques. To better capture the bidding activity especially at the end, we transform the bids into log-scores. In order to account for the irregular spacing, we linearly interpolate the raw data and sample it at a common set of time points t_i, $0 \leq t_i \leq 7$, $i = 1, \ldots, n$. Then we can represent each auction by a vector of equal length

$$y^{(j)} = (y_1^{(j)}, \ldots, y_n^{(j)}), \tag{1}$$

where $y_i^{(j)} = y^{(j)}(t_i)$ denote the value of the interpolated bid sampled at time t_i.

We use smoothing splines (see Ramsay and Silverman (1997) or Simonoff (1996)) to recover the underlying price curve. Smoothing splines are often recognized to be a good choice that achieve both small local variability and overall smoothness. Smoothing splines readily yield several derivatives of the targeted curve.

Consider a polynomial spline of degree p

$$f(t) = \beta_0 + \beta_1 \times t + \beta_2 \times t^2 + \cdots + \sum_{l=1}^{L} \beta_{pl}(t-\tau)_+^p, \tag{2}$$

where $\tau_1, \ldots, \tau_L$ is a set of knots and $u_+ = u_{[u \geq 0]}$ (Ruppert et al., 2003). The choices of L and p strongly influence the departure of the f from a straight line. The degree of departure can be measured by a roughness penalty

$$PEN_m = \int D^m f(s)^2 ds. \tag{3}$$

Larger values of PEN_m indicate a locally more varying curve.

The penalized smoothing spline for auction j, $f^{(j)}$, minimizes the penalized residual sum of squares

$$PENSS_{\lambda,m}^{(j)} = \sum_{i=1}^{n} (y_i^{(j)} - f^{(j)}(t_i))^2 + \lambda \times PEN_m^{(j)}, \tag{4}$$

where λ, the smoothing parameter, controls the trade-off between data fit and the variability of the function f.

In our application, we use the knots $0, 1, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 6.8125, 6.875, 6.9375, 7$. This choice reflects the distribution of bid arrivals over the 7-day auction. Indeed, as we have seen in Figure 3, about 50% of the bids arrive within the last 6 hours of the auction. Choosing a larger proportion of knots at the end of the auction allows the spline more flexibility to capture the intensive bidding activity at the end. (See Shmueli et al. (2004) for a related result on the distribution of bid arrivals.) We use smoothing splines of order $m = 5$ since this choice allows for a smooth estimation of the first three derivatives of the smoothing spline. To balance the tradeoff between data-fit and smoothness, we choose a smoothing parameter $\lambda = 20$ using MSE optimization (see Appendix for details on the choice of λ).

Figure 4 shows $f(t)$ for auction number 19. The plot on the upper left shows the curve pertaining to the log-price and the rest of the three plots display the corresponding first three derivatives of $f(t)$, respectively. Simply looking at the price position may lead us to consider this to be an ordinary linear regression problem since the bids “look” like a rather monotonically increasing straight line.

However, the dynamics show how much more variability there really is. It can be noted that the price dynamics are quite different across the entire course of the auction. For instance, while the price velocity (first derivative) increase at the beginning of the auction, it drops down in the third day and keeps declining till the end of the fifth day, then it rises again and increases towards the end of the auction. Specifically, a huge amount of bidding activity at the closing of the auction is demonstrated by the steady increase of the bid velocity during the last two days[1]. Meanwhile, the drastic changes in dynamics across the entire course of auction can be also seen in the curves of price acceleration (second derivative) and jerk (third derivative). Similar changes in auction dynamics have also been noted by Jank and Shmueli (2003).

3.2 Forecasting Auction Dynamics

We pointed out in the previous section that one of the main characteristics of online auctions is their rapid change in dynamics. Since changes in derivatives precede changes in the price curve (e.g., acceleration precedes increase in speed), we make use of derivative information for forecasting. In the following, we develop a model to estimate and forecast the auction dynamics.

Let $D^{(m)}y_t$ denote the m^{th} derivative of the price-position y_t at time t. Notice that in our notation, $y_t = f(t)$. We model the derivative curve as a polynomial in time with AR residuals,

$$D^{(m)}y_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k + u_t, t = 1, \ldots, T, \tag{5}$$

where u_t follows an autoregressive model of order p:

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \varepsilon_t, \varepsilon_t \sim i.i.d.N(0, \sigma^2). \tag{6}$$

First, we estimate the parameters $a_1, \ldots, a_k$. Then, using the residuals $\hat{u}_t$, we estimate the values of $\phi_1, \ldots, \phi_p$. The forecasting is also a 2-step procedure: we forecast the next residual via

$$\tilde{u}_{T+1|T} = \tilde{\phi}_1 u_T + \tilde{\phi}_2 u_{T-1} + \cdots + \tilde{\phi}_p u_{T-p+1}.$$

Using this forecast, we can predict the derivative at the next time point as

$$D^{(m)}\tilde{y}_{T+1|T} = \hat{a}_0 + \hat{a}_1 (T+1) + \cdots + \hat{a}_k (T+1)^k + \tilde{u}_{T+1|T}. \tag{7}$$

[1]Since bids increase by a single increment with every incoming high enough bid, a big jump in price velocity reflects heavy bidding activity

In a similar fashion, we can predict the residual l stages into the future:

$$D^{(m)}\tilde{y}_{T+l|T} = \hat{a}_0 + \hat{a}_1(T+l) + \cdots + \hat{a}_k(T+l)^k + \tilde{u}_{T+l|T}, l = 1, 2, \ldots. \quad (8)$$

3.3 Forecasting Model

After forecasting the price dynamics, we use these forecasts to predict the actual final auction price. There are several factors that can affect the final price of an auction (see e.g. Lucking-Reiley et al. (2000)). Among these factors are a bidder's feedback rating, a seller's feedback rating, the opening bid and the current number of bids and bidders in the auction. A bidder's rating can have notable impact on the auction price. Bidder feedback ratings proxy for experience and bidders with higher experience may extract a better price. On the other hand, seller ratings can also affect the auction outcome. Experienced sellers make better auction design choices and a higher seller rating results in more trust. The number of bidders also influences price since a larger amount of participators implies more competition. Shmueli and Jank (2004a) find that the effect of the opening price on the auction dynamics decreases over time, which would suggests an interaction term between opening price and time. Another important forecasting element is the highest price amount prior to time t. This results in a dynamic forecasting model as follows:

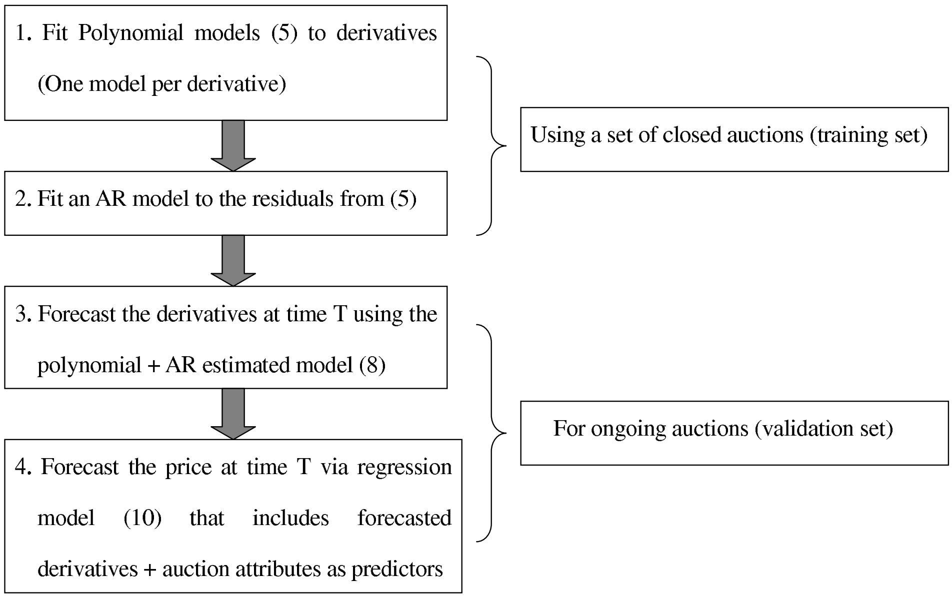
$$y_t = \alpha + \beta_1 X_1 + \cdots + \beta_q X_q + \gamma_1 D^{(1)} y_t + \gamma_2 D^{(2)} y_t + \gamma_{(3)} D^{(3)} y_t + \gamma_4 y_{t-1} \quad (9)$$

where $\{X_1, \ldots, X_k\}$ is the set of explanatory variables discussed above. We will select the best subset of explanatory variables via stepwise model selection.

The resulting l-step ahead prediction of the final price based on the bid history observed up to time T is given as:

$$\tilde{y}_{T+l|T} = \hat{\alpha} + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_q X_q + \hat{\gamma}_1 \tilde{D}^{(1)} y_{T+l|T} + \hat{\gamma}_1 \tilde{D}^{(2)} y_{T+l|T} + \hat{\gamma}_3 \tilde{D}^{(3)} y_{T+l|T} + \hat{\gamma}_4 y_{T+l-1|T}. \quad (10)$$

3.4 Overview of the Forecasting Algorithm



4 Results

4.1 Model Estimation

We randomly split the data into two sets, a training set (60%) and a validation set (40%). The first set includes data from 111 auctions while the latter contains data from 74 auctions. We fit the model to the training set and use the validation set to measure forecasting accuracy.

4.1.1 Estimating the Auction Dynamics

The model fitting and prediction are implemented using modules in the R software package. Figure 5 gives the graphs of the auto correlation function (ACF) as well as the partial auto correlation function (PACF) of the time series model residuals in Equation (6). Both graphs suggest an AR(1) model for the residuals, and using several estimation methods yields an estimated coefficient of 0.85 with standard error of less than 0.2 (see Table 1).

We can apply the forecasting model (8) in two different ways. If information about the average dynamics is desired, then we fit the model to the average of the derivative curves in the training set. On the other hand, we can also forecast individual auction dynamics by applying (8) to the derivative curves of individual auctions. The first operation allows us to forecast the overall performance of the dynamics in the validation set and bootstrapped confidence intervals can provide a measure of the forecasting uncertainty. In the latter approach we work on individual velocity curve such that the information of dynamic variability carried by each single auction is not lost.

We use model (8) to forecast the dynamics over the last day of the auction. First, we estimate the parameters based on the data in the training set. Then, assuming that the last day is not yet observed, we apply this model to the data in the validation set. We can thus evaluate forecasting accuracy by comparing the true dynamics with the forecasted dynamics on the last day.

Figure 6 shows the results of forecasting average and individual velocities. 95% bootstrap confidence intervals are given in the former case. The forecast of the average velocity catches the upward trend of the velocity towards the end of the auction and the 95% prediction interval is strictly positive, showing that bid-acceleration occurs in the last moment, which is compatible with previous observations. In the lower panels of Figure 6, forecasts of three individual auctions are given. We note the larger discrepancy between the individual forecasts and their true velocities compared with the forecast of the average velocity. These differences reflect the huge amount of variability occurring during the closing period of the auction. In the following, we use the estimated mean velocities since it concentrates on forecasting the overall performance of the dynamics.

4.1.2 Estimating the Price

We use stepwise model selection to find a good set of predictor variables for the model (9). Potential factors considered in the full model are current price, opening price, interaction between the opening price and time, price-velocity, price-acceleration and price-jerk , current average bidder rating, and current total number of bids. Table 2 provides the results from the stepwise procedure. The procedure starts with the full model (Model I), and proceeds by successively dropping the variable with the largest p-value. The final model (Model VII) contains current price, interaction between opening price and time, and price-velocity. Thus our model for the average auction curve can be expressed in the following recursive manner

$$y_t = \gamma_1 D^{(1)} y_t + \gamma_2 y_{t-1} + \beta X_{Openbid} \times t \tag{11}$$

where $X_{Openbid}$ is the average opening price of the auctions. This model tells us that, what we need to know in order to forecast the price at time t, is the starting price of the auction, the most recent price, the time when you want to forecast, and the price velocity. Notice that no external information such as the seller rating is included in this model.

4.2 Forecasting the Average Closing Price of n Auctions

The goal is to forecast the price over the last day of the auction. We first estimate models (7) and (9) on the training data. Then we apply the models to the validation data, assuming that the last day of the auction is unobserved. Thus we can measure prediction accuracy by comparing the actual price over the last day with its forecast. We split the last day into 4 equally spaced time intervals, [6.00, 6.25], [6.25, 6.50], [6.50, 6.75] and [6.75, 7.00]. For instance, the interval [6.00, 6.25] denotes the first 6 hours of the last day. Then, assuming that the last observation occured at the end of the day 6, we forecast day 6.25. Using this forecast, we predict day 6.50 and so on. The final forecast, day 7.00, predicts the closing price. Notice that the dependent variable in our model is the smoothed auction curve, that is, $y_t = f(t)$. We will refer to this curve as the *observed* or *true* curve, which is in contrast to the *predicted* curve based on our forecast. Recall that the true curve is an approximation to the actual bids. So we compare our forecast with the true curve as well as with the actual bids.

Figure 7 provides comparisons among the predicted average price curve, the average true price curve and the average actual bids. The graph on the left panel of Figure 7 gives us a complete scenario of the entire duration of the auction, and the one on the left just focuses on the last day to give a better view of details. The green crosses represent the average of the actual bids for the 74 auctions in the validation set, the solid dark line is the average of the true price curves in the validation set, and the red dotted line stands for the forecast of average price curve for day 7. From the graph, we can see that the predicted curve is very close to the true curve as well as to the actual bids. The forecasted value of the final price is 224.37 dollars, while the mean of the actual final price is 228.34 dollars. These accurate results are very encouraging. In the following we measure prediction uncertainty via confidence intervals.

A good forecast generally aims at less forecasting uncertainty or higher possibility of shooting straight to the true value. One way of measuring uncertainty is via prediction intervals. While analytical prediction intervals are cumbersome to derive due to our 2-stage modelling approach of the derivatives followed by price, we can obtain empirical prediction intervals using the bootstrap. The bootstrap is a re-sampling method which computes valid statistical measures like confidence intervals by means of computer intensive simulations using only the observed data. A thorough introduction to the bootstrap methodology can be found in Efron and Tibshirani (1993). To extract the prediction intervals for our forecast, we randomly select 74 auctions from the validation set with replacement. Taking these 74 bootstrapped auctions as the new validation set, we predict the price on the last day. Repeating this procedure 1000 times, we obtain 1000 predicted curves as shown in the left panel of Figure 8. We use these curves to obtain 95% prediction intervals (see the right panel of Figure 8). It should be noted that there exist two sources of error, one is from the model estimation and another one is from the smoothing. The bootstrap takes both of these two sources of error into account. From the graphs, we note that both the average true price curve and the actual bids fall inside the 95% prediction intervals and they are both fairly close to the the predicted values.

We compare our results with the forecasts provided by standard double exponential smoothing. Double exponential smoothing is one of the popular short term forecasting methods for time series. It assigns exponentially decreasing weights as the observation get elder and takes into account a trend. We apply double exponential smoothing to the linearly interpolated bids. The forecasted values as well as 95% confidence intervals are given in Figure 9. Comparing the true value (solid line) with the forecasted value (dotted line), we see that forecast based on exponential smoothing is not satisfying at all. In fact, the true data fails to fall inside the 95% confidence intervals. The discrepancy between the true value and the forecast gets larger and larger as the auction reaches the end. While the forecast somewhat captures the upward trend of the bids, it fails to account for the dramatic acceleration pulse of the price increase towards the closing of the auction. In contrast, our dynamic forecasting model produces forecasts that capture not only the upward trend but also the huge acceleration spurt towards the end very well. The advantage of our model stems from its adaptivity to changes in the price dynamics, which can lead to sharp changes in the price curve that are hard to capture with methods such as exponential smoothing.

4.3 Insights from Auction Forecasting

Figure 8 provides some interesting statistical insight into auction forecasting. Recall that our goal is to predict the last day of the auction. That is, assuming that the last observed valued occurs at day 6, we forecast the price towards the auction end. Notice that the prediction intervals of the forecast in Figure 8 are wider on day 6 than on day 7. That is, the prediction intervals *decrease* at the auction end. In other words, as we predict further into the future, the prediction uncertainty decreases!

This result is interesting (and at first counter-intuitive) from a forecasting perspective. Prediction intervals typically get larger with increasing t. That is, as we attempt to predict further into the future, the uncertainty increases. This is the case in many applications of time series methods like the stock market or a company's sales.

However, when forecasting online auctions, the uncertainty decreases towards the auction end. One explanation for this rather surprising result can be found in Bajari and Hortacsu (2003). The authors find that a bidder's prior information about the item's value is typically quite diffuse. At the start of the auction, uncertainty exists about the valuation for the item. This uncertainty decreases as bidders spend time and effort to research the item, making their estimate considerably more precise. On the other hand, as the auction progresses, information becomes available due to the information acquisition cost argument in auction theory. For example, Shmueli and Jank (2004a) point out that the auction start does not contain much information on the value of the auctioned item. But as the auction approaches its end, more and more information is being revealed with the increasing number of incoming bidders. Thus, the amount of uncertainty about the outcome of the auction decreases towards the auction end.

5 Conclusion and Future Directions

In this work we derive a dynamic forecasting model for online auction. We use functional data analysis to represent and model the price dynamics of a set of closed auctions. We then incorporate these dynamics into a forecasting model for an ongoing auction. Our application shows that the model manages to predict the auction price well. This is an contrast to standard forecasting methods like double exponential smoothing which severely under-predict the price. This shows that online auction forecasting is not an easy task. While traditional methods are hard to apply, they are also inaccurate since they do not take into account the dramatic change in auction dynamics. Our model, on the other hand, achieves high forecasting accuracy and accommodates the changing price-dynamics.

This work can be extended in several ways. In this work, we focus on auctions for the same product. The lessons learned from this work can be used to extend the model to different product types. Similarly, while we consider 7-day auctions only, more thought is necessary to combine auctions of different length. Combining auctions of different lengths is challenging since it involves registration of misaligned curves (see e.g. Ramsay and Silverman (1997) or James (2004)). However, in the auction context the misaligned curves are of different length which poses additional difficulties.

APPENDIX

A Choice of the smoothing parameter λ

Accuracy of predictions is important to the forecasting model. When a functional data analysis approach is used, the accuracy of predictions depends on the underlying smoothing spline. A smoothing spline is controlled by the smoothing parameter λ, the knots, the number of basis functions and the derivative that is penalized. A good choice of λ balances data-fit and smoothness. We can measure data-fit via the MSE between the smoothing spline and the data. On the other hand, the MSE between the spline and a straight regression line through the data measures the degree of smoothness. Figure 10 shows the MSEs for data-fit and smoothness for different values of λ. We can see that, not surprisingly, data-fit increases with λ while smoothness decreases with λ. On the upper left panel of the graph, data-fit grows very fast when λ is small, but it slows down around a value betweeen 10 and 200, and then it levels off as λ gets even bigger. It's a similar case with smoothness except that smoothness runs on the opposite direction. To find this exact turning point, we graph the first derivatives of both measurements. (Lower panel of Figure 10.) We can see that change rate of accuracy remains big when $\lambda > 20$ (decreasing though), drops below 0 when $\lambda = 20$ and goes back above 0 but keeps to be small after that. The same thing happens to the velocity of smoothness on the other direction. Thus, $\lambda = 20$ proves to be a good choice of our smoothing parameter that controls the trade-off between data fit and the variability of the smoothing spline.

References

Ba, S. and Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26:269–289.

Bajari, P. and Hortacsu, A. (2003). The winner's curse, reserve prices and endogenous entry: Empirical insights from ebay auctions. *Rand Journal of Economics*, 3:2:329–355.

Bajari, P. and Hortacsu, A. (2004). Economic insights from internet auctions. Nber working paper, no:w10076.

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

Dellarocas, C. (2003). The digitization of word-of-mouth: Promise and challenges of online reputation mechanisms. *Management Science*, October Issue.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman&Hall.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004). Modelling climatological data by functional logistic regression. Technical report, "http://isi-eh.usc.es/resumenes/186_103_abstract.pdf".

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, 39:254–261.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society, Series B*, 64:411–432.

James, G. M. (2004). Curve alignment by moments. Under review, "http://www-rcf.usc.edu/~gareth".

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

Jank, W. and Shmueli, G. (2003). Dynamic profiling of online auctions using curve clustering. Working paper, "http://www.rhsmith.umd.edu/dit/wjank/AuctionProfiling.pdf"/.

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. (2000). Pennies from ebay: The determinants of price in online auctions. Technical report, University of Arizona, "http://www.vanderbilt.edu/econ/reiley/papers/PenniesFromEBay.pdf".

Ramsay, J. O. (2000). Function components of variation in handwriting. *Journal of the American Statistical Association*, 95:9–15.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional data analysis*. Springer-Verlag New York, Inc., 1st edition.

Roth, A. E. and Ockenfels, A. (2002). Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon on the internet. *American Economic Review*, 92:1093–1103.

Shmueli, G. and Jank, W. (2004a). Modelling the dynamics of online auctions: A modern statistical approach. *New Empirical Methods in eCommerce (forthcoming), Editor: Kauffman.*

Shmueli, G. and Jank, W. (2004b). Visualizing online auctions. *Journal of Computational and Graphical Statistics(forthcoming).*

Shmueli, G., Russo, R. P., and Jank, W. (2004). Modelling bid arrivals in online auctions. Working paper, "http://www.smith.umd.edu/faculty/wjank/BidArrivals.pdf".

Simonoff, J. S. (1996). *Smoothing methods in statistics*. Springer-Verlag New York, Inc., 1st edition.

Valderrama, M. J., Ocana, F. A., and Aguilera, A. M. (2002). Forecasting pc-arima models for functional data. Technical report, "http://www.quantlet.de/scripts/compstat2002_wh/paper/invited/F_valderrama.pdf".

Table 1: Output for the AR model in Section 4.1.1 using different method.

Method	Yule-walker	Burg	OLS	MLE
Estimate of Coefficient	0.7511	0.8922	0.8012	0.9630
Standard Error	0.2396	0.1581	0.1257	0.1512

Table 2: The driving factors of price in eBay Palm M-515 auctions

Dependent variable: ln(Price)

Model	I	II	III	IV	V	VI	VII
Intercept	-0.065	-0.041	-0.0185	–	–	–	–
	(0.0039)	(0.0071)	(0.0026)	–	–	–	–
Current	1.017*	1.01*	1.004*	-0.999*	0.999*	0.996*	0.997*
ln(Price)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)
$ln(Openbid)$	-0.004	-0.0008*	-0.0008*	-0.0001	–	0.00085*	0.00082*
$\times Time$	(0.002)	(0.0004)	(0.0006)	(0.0117)	–	(2.28e-12)	(6.54e-13)
Velocity	0.262*	0.261*	0.2598*	0.259*	0.259*	0.261*	0.261*
of ln(Price)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)	(<2e-16)
Acceleration	0.004	–	–	–	–	–	–
of ln(Price)	(0.124)	–	–	–	–	–	–
Current Ave.	-0.00006*	-0.00007*	-0.00008	-0.00006*	-0.00005*	0.000025	–
Bidders' Rating	(1.15e-05)	(1.21e-07)	(1.85e-11)	(1.67e-10)	(1.42e-11)	(0.182)	–
Current Total	-0.0008	-0.0005	–	–	–	–	–
NumBids	(0.026)	(0.09212)	–	–	–	–	–

eBay.com Item Bid History

home | pay | register | sign out | services | site map

Start new search | Search

Advanced Search

Buy | Sell | My eBay | Community | Help

Powered By IBM

Back to item description

Bid History

Item number: 57226520

Hello springnass! (Not you?)

Watch this item in My eBay

Item title: Palm m515 Handheld PDA (revised)

Time left: Auction has ended.

Only actual bids (not automatic bids generated up to a bidder's maximum) are shown. Automatic bids may be placed days or hours before a listing ends. Learn more about bidding.

User ID	Bid Amount	Date of bid
brcmhouse (73)	US $53.00	Sep-28-04 11:55:00 PDT
mulufuf (2)	US $52.00	Sep-28-04 11:54:55 PDT
rlkscott (13)	US $40.00	Sep-28-04 11:54:10 PDT
rlkscott (13)	US $38.00	Sep-28-04 11:53:18 PDT
mulufuf (2)	US $36.00	Sep-28-04 11:50:56 PDT
rlkscott (13)	US $35.00	Sep-28-04 11:48:52 PDT
mulufuf (2)	US $32.00	Sep-28-04 11:46:21 PDT
pksclucfp (11)	US $31.00	Sep-28-04 11:21:18 PDT
chinafishgene (1)	US $30.00	Sep-27-04 19:00:49 PDT
pksclucfp (11)	US $30.00	Sep-28-04 11:21:05 PDT
fredjd74 (72)	US $25.00	Sep-27-04 17:01:28 PDT
candyshop54 (1)	US $15.00	Sep-27-04 18:22:22 PDT
candyshop54 (1)	US $10.00	Sep-27-04 18:21:50 PDT

If you and another bidder placed the same bid amount, the earlier bid takes priority.
You can retract your bid under certain circumstances only.

About eBay | Announcements | Security Center | Policies | Site Map | Help

eBay official time

http://offer.ebay.com/ws3/eBayISAPI.dll?ViewBids&item=57226520119/28/2001 12:21:08 PM

Figure 1: *Partial bid-history for an eBay Palm-515 auction. On the left-most side of the table we can see a bidder's user ID, followed by the bidder's rating. The stars indicate that this eBay member has achieved 10 or more feedback points. The amount and time of the bids appear on the right.*

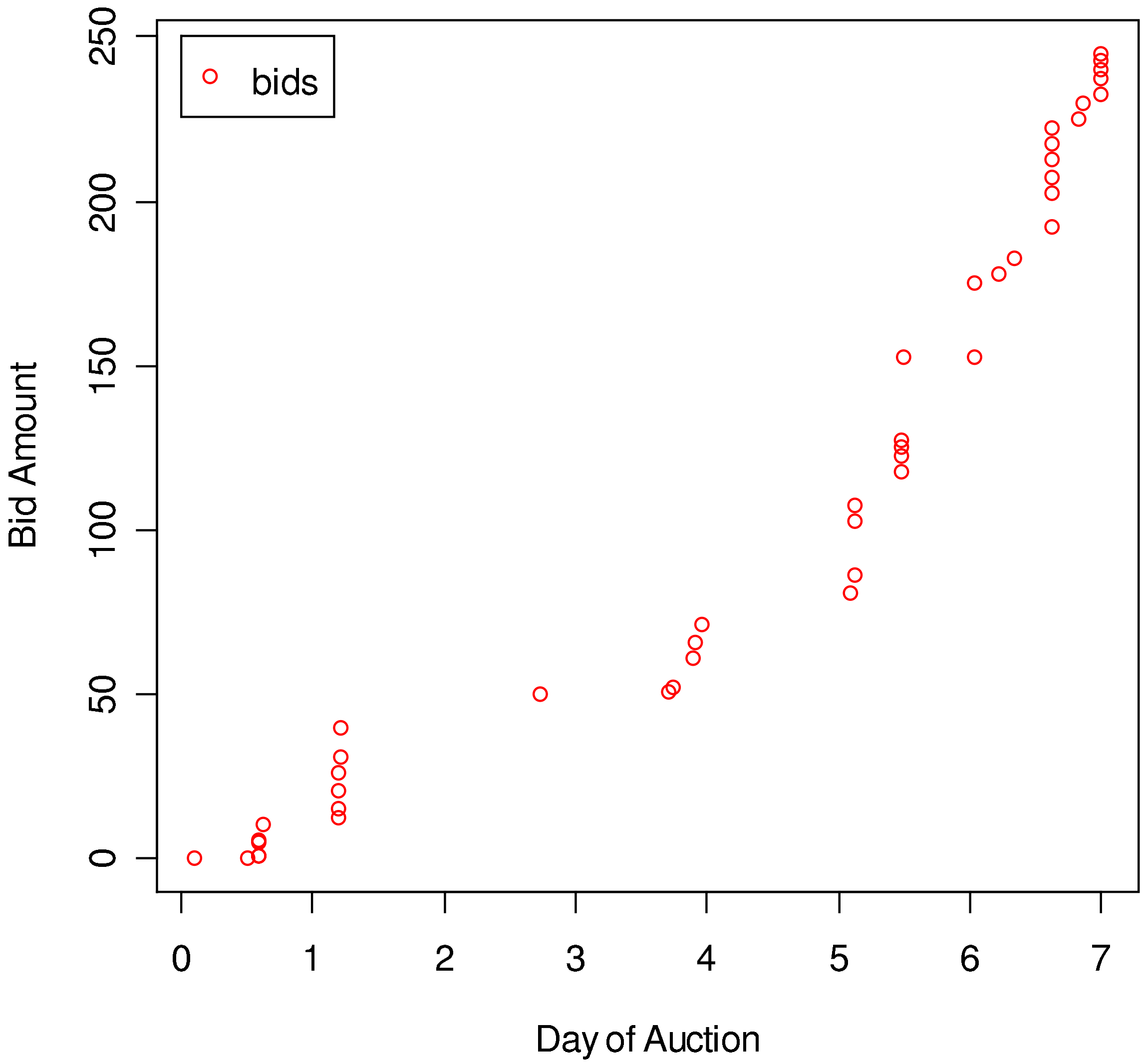


Figure 2: *The bids placed in auction number 63 of a Palm M515 auction*

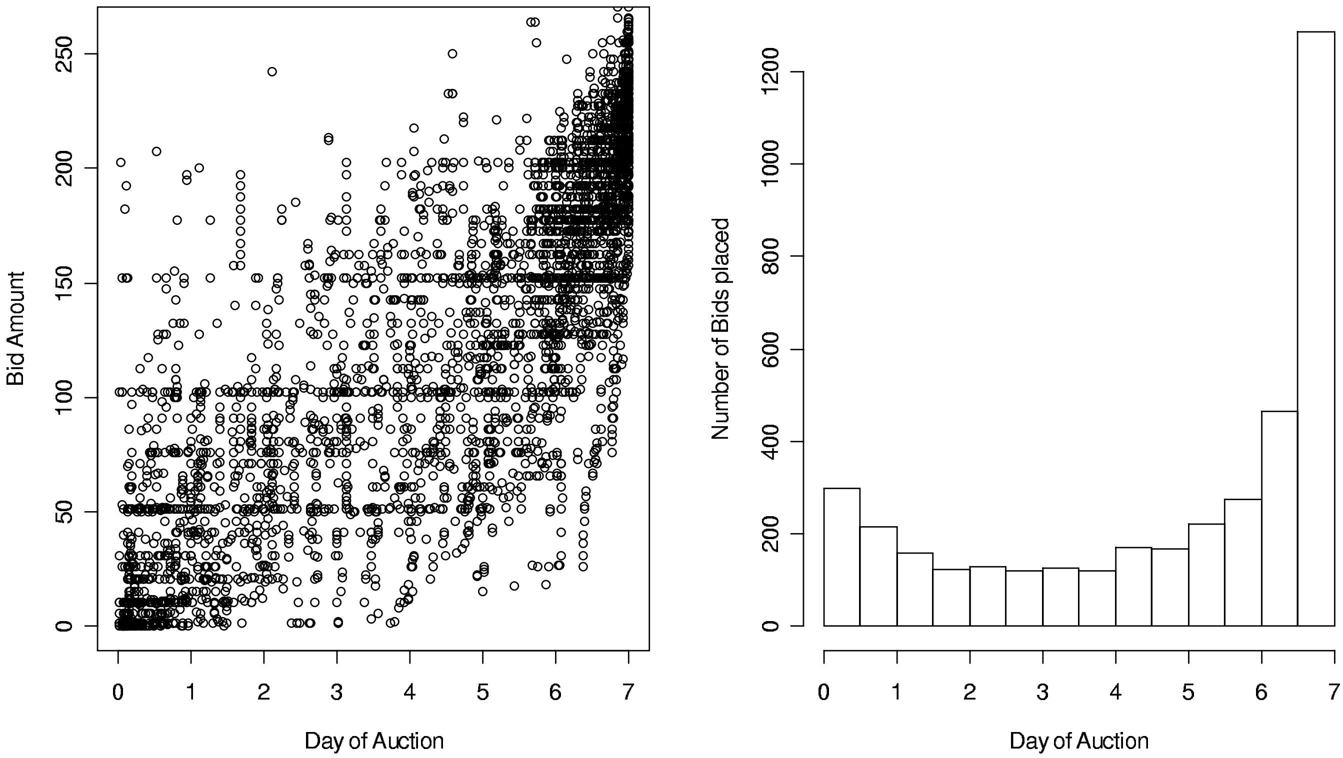


Figure 3: *Data for the 185 Palm M515 7-day auctions: The graph on the left shows the amount of the bid vs the time of the bid, aggregated across all auctions. The graph on the right shows a histogram of the distribution of the number of bids over 7 days. Each bin corresponds to a 12-hour time interval.*

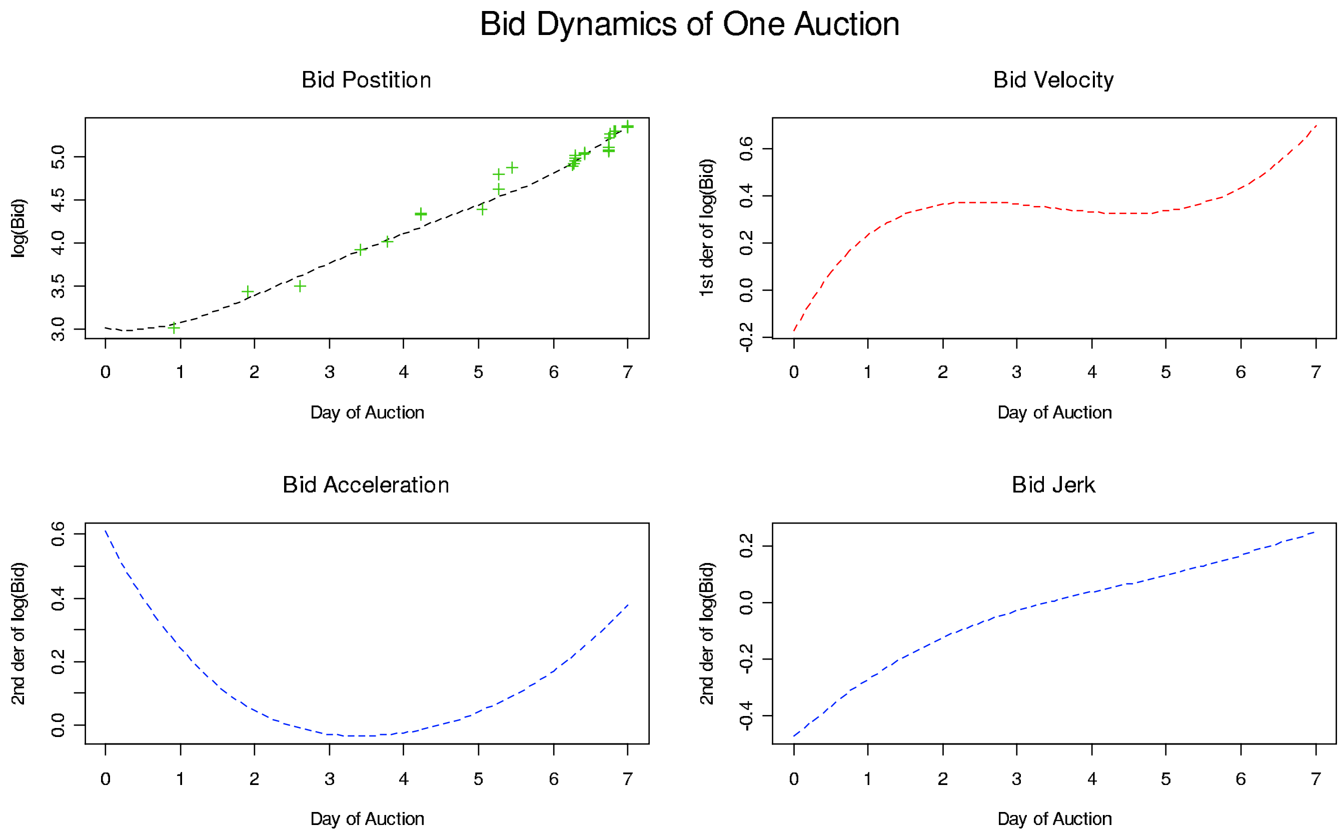


Figure 4: *Bidding dynamics for auction number 19.*

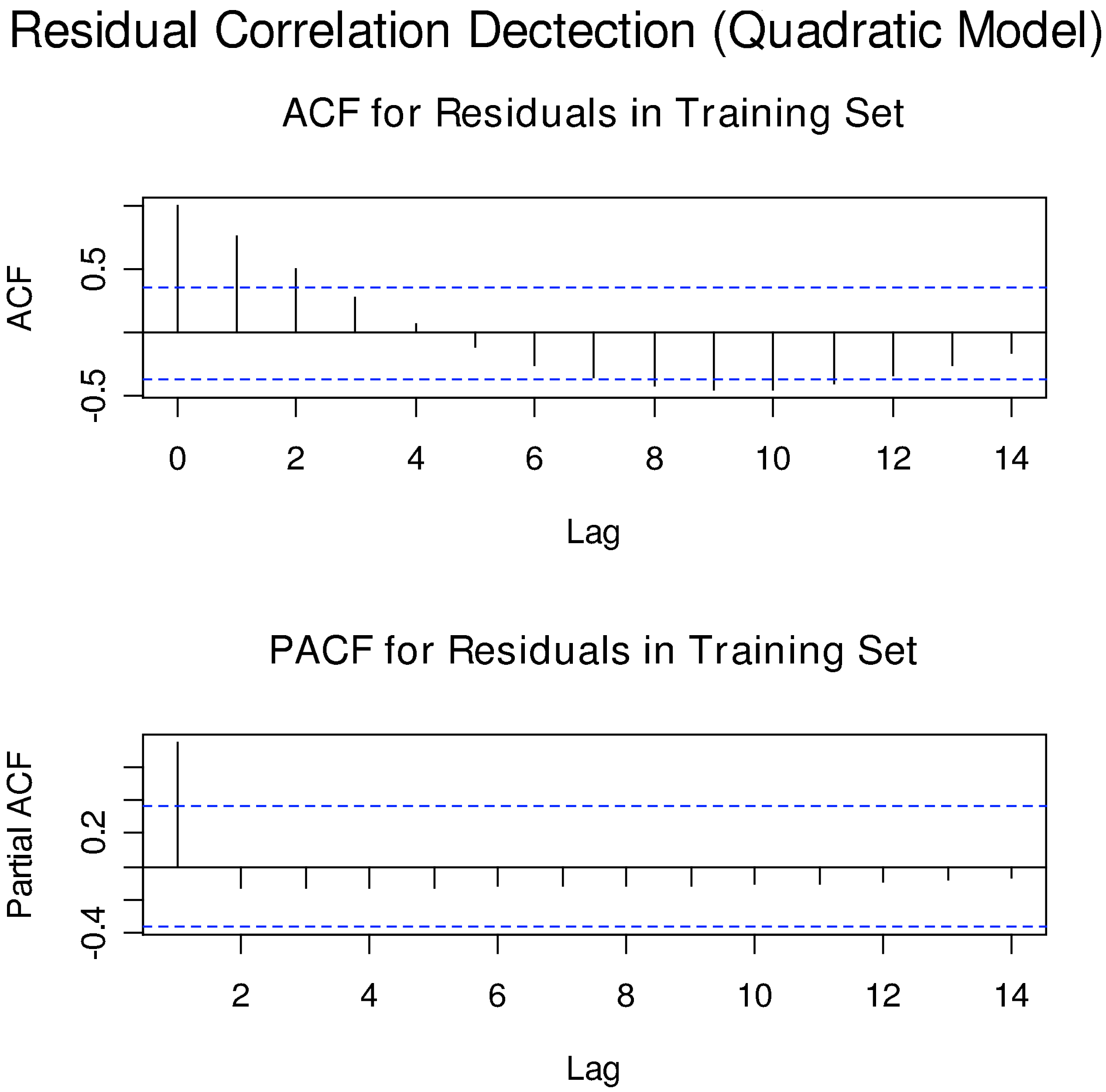


Figure 5: *ACF and PACF plots for the residuals in the training set.*

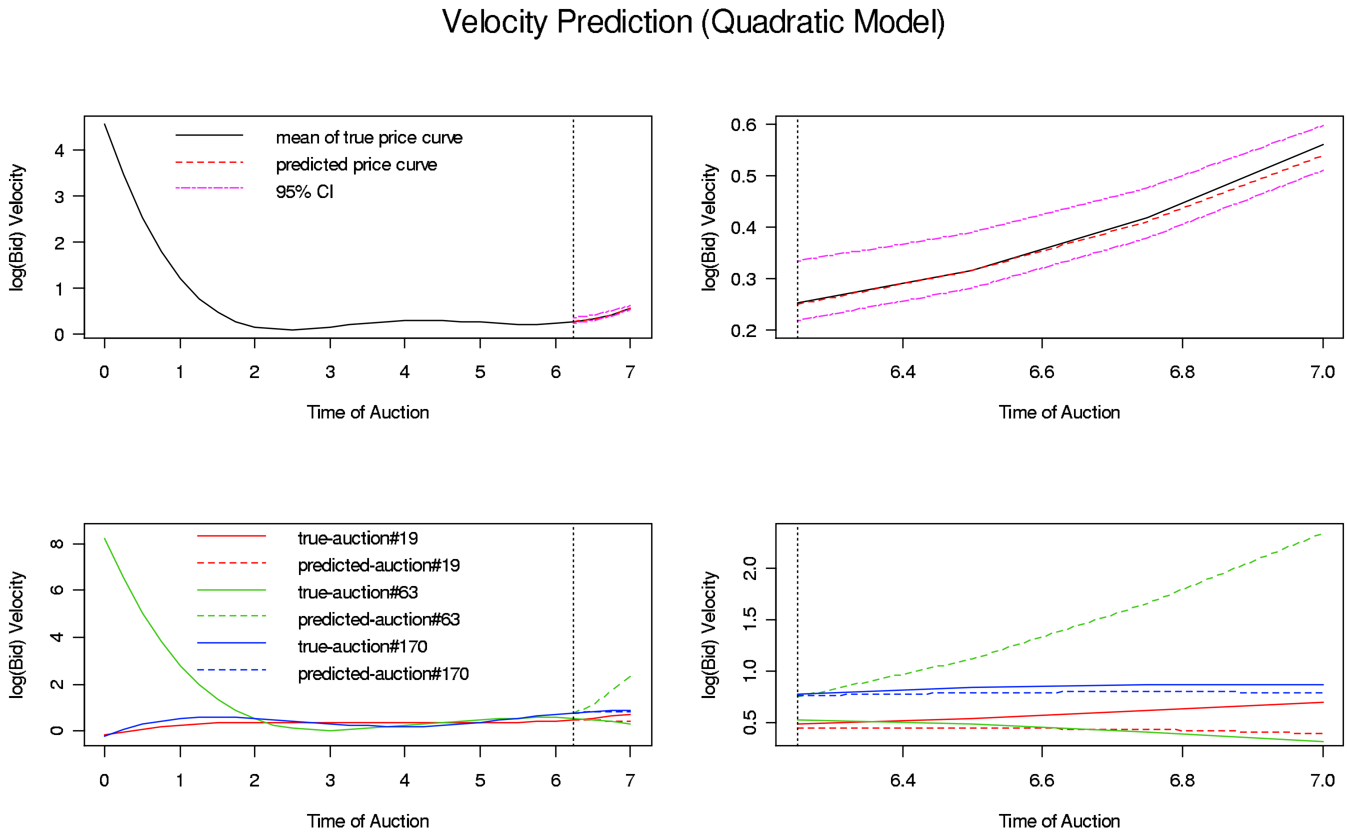


Figure 6: *Upper panel: Forecasts of the mean velocity of log(Bid) for day 7 using quadratic time series model with AR(1) error and 95% confidence intervals by bootstrapping. Lower panel: Individual forecasts of the velocities of auction #19, #63 and #170.*

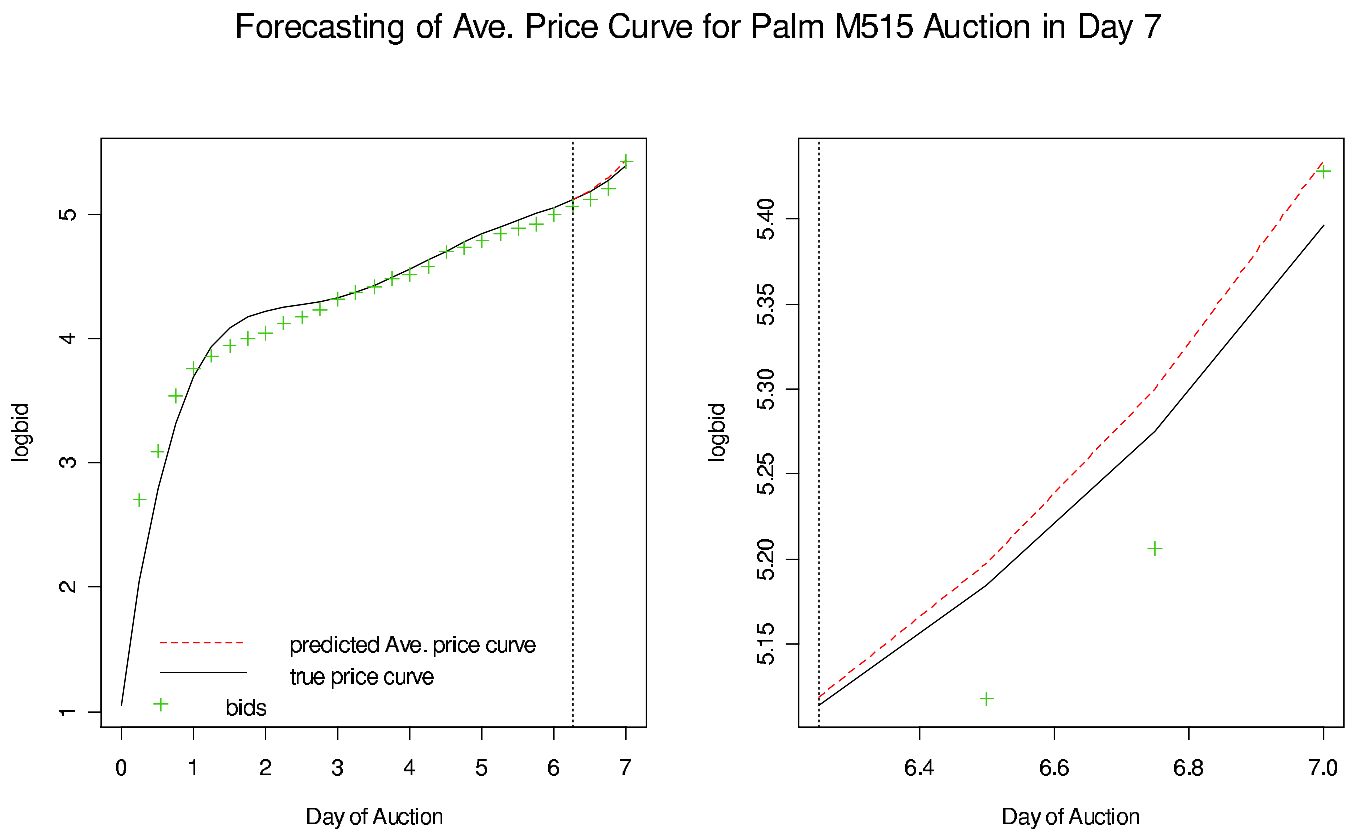


Figure 7: *Forecast of the average price curve in validation set. The graph on the left is drawn on the entire auction period, and the graph on the right is drawn on the last day.*

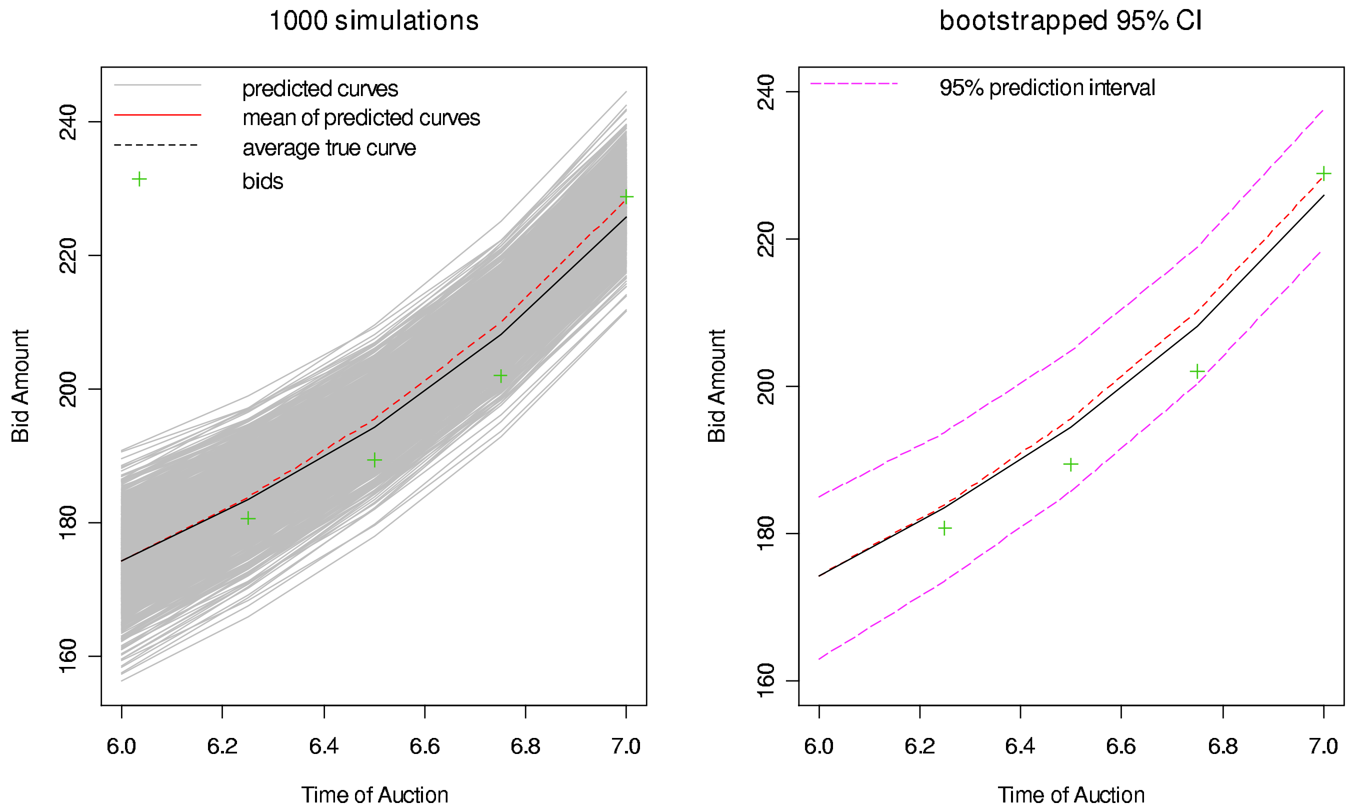


Figure 8: *bootstrapped forecasts and corresponding 95% prediction intervals based on un-logged scale.*

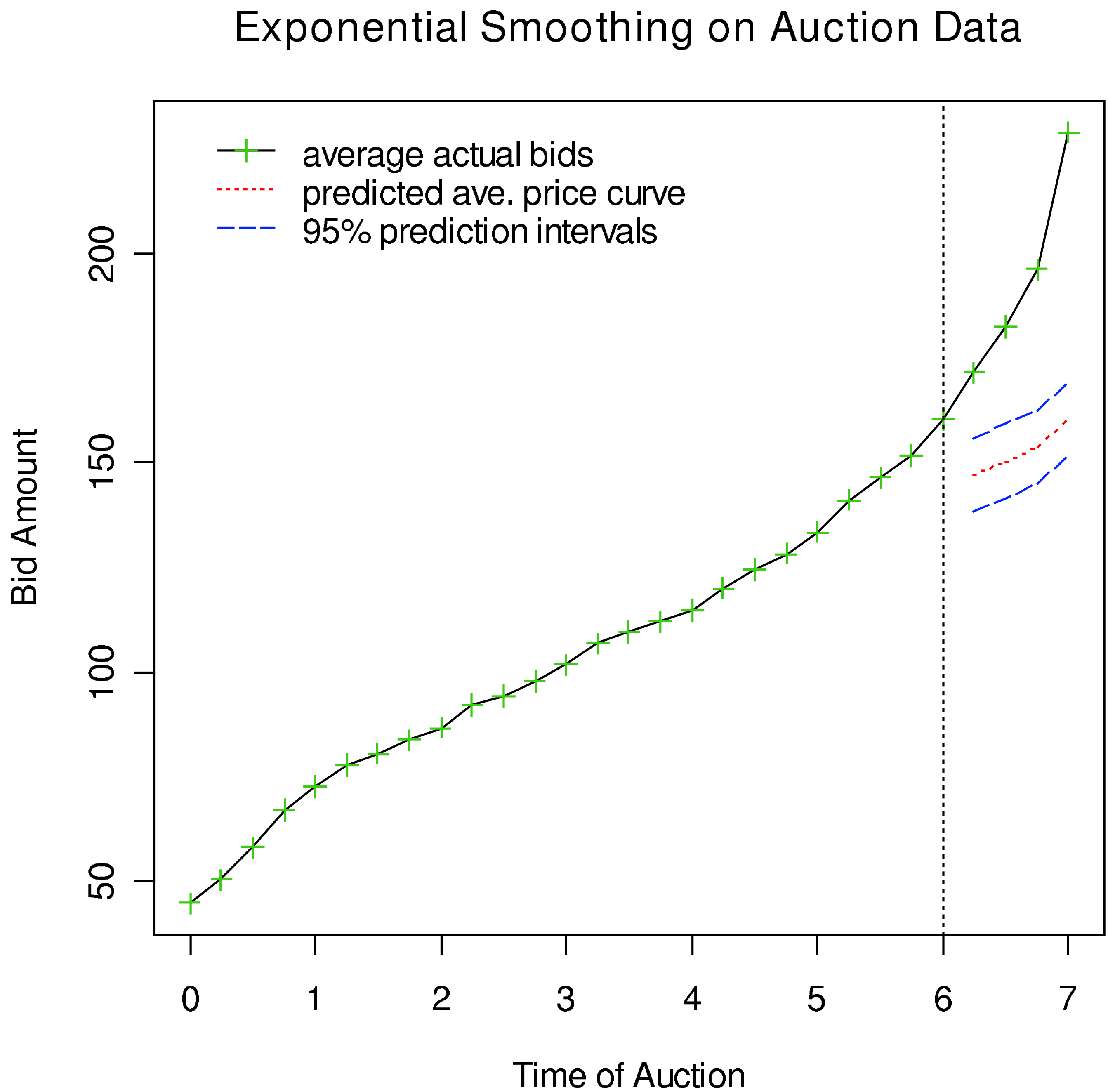


Figure 9: *predictions and 95% confidence intervals using double exponential smoothing.*

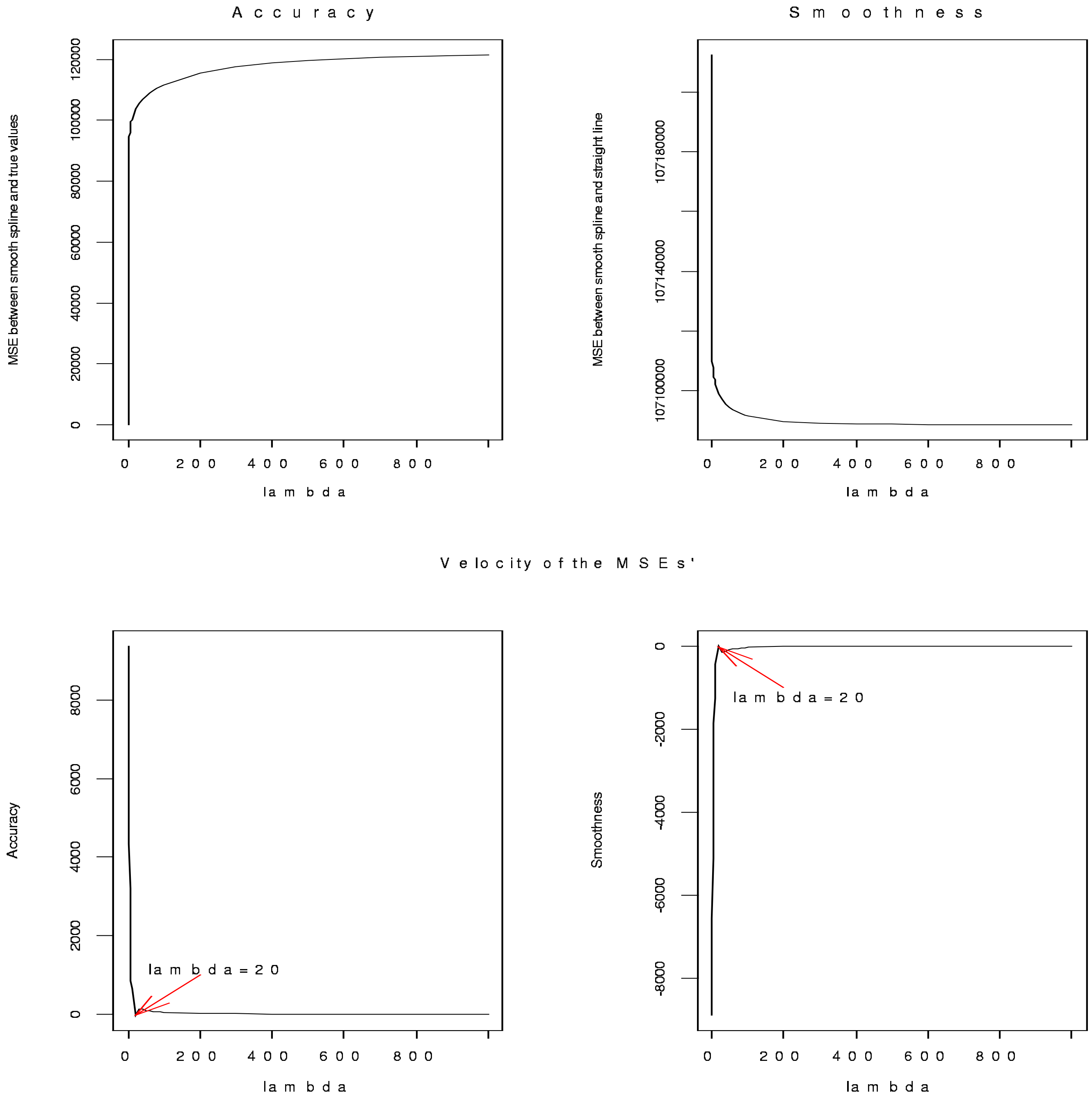


Figure 10: *Decision of the best smoothing parameter lambda.*