An Automated and Data-Driven Bidding Strategy for Online Auctions

Wolfgang Jank* & Shu Zhang†‡

Abstract

The flexibility of time and geographical location as well as the availability of an abundance of both old and new products makes online auctions an important part of people's daily shopping experience. While many bidders rely on variants of the well-documented early or last-minutes bidding strategies, neither strategy takes into account the aspect of auction competition: at any point in time, there are hundreds, even thousands, of same or similar items up for sale, competing for the same bidder. In this paper, we propose a novel automated and data-driven bidding strategy. Our strategy consists of two main components. First, we develop a dynamic, forward-looking model for price in competing auctions. By incorporating dynamic features of the auction process and its competitive environment, our model is capable of accurately predicting an auction's price, taking into account information from simultaneous auctions. Then, using the idea of maximizing consumer surplus, we build a bidding framework around this model that determines a triumvirate of decision points: the best auction to bid on, the best bid-time and the best bid-amount. In simulations, we compare our automated strategy to early and last-minute bidding and find that our approach extracts considerably higher expected surplus. We also argue that our approach devotes significantly less effort to the process of bidding.

Key words: functional data, dynamics, online auction, bidding, forecasting, competition, eBay, electronic commerce, consumer surplus.

*Robert H. Smith School of Business, University of Maryland

†Applied Mathematics and Scientific Computation Program, Department of Mathematics, University of Maryland

‡We would like to thank Itir Karesmen, Louiqa Rashid, and Michael Fu for their input. This paper has benefited from feedback received at a research seminar at the RH Smith School of Business, University of Maryland.

1 Introduction

The flexibility of time and location as well as the abundance of both old and new products make online auctions an important part of people's daily shopping experience. On eBay, the largest C2C online auction site, millions of products are on sale every day. What complicates bidding decisions is that many of these auctions are not independent of one another: bidders who bid in one auction can also bid in another auction, and thus affect the price in two auctions simultaneously; products of same (or similar) type compete for the same bidder; and sellers compete with one another for the highest price. As a result, auctions form a rather complicated set of inter-linked information.

In this paper, we focus on this inter-linked information from the bidders' point of view. Different from the fixed-price environment, buyers in online auctions face many different decisions. They have to decide whether to bid early or late, whether to place a single bid or multiple updates, whether to bid high or low. Bidding is further complicated by the existence of many auctions that offer the same, or similar item simultaneously. In that case, one's bidding strategy has to be expanded to include decisions on which auction to bid on, when to bid on that auction, and how much.

There exist two very well documented (and researched) bidding strategies for online auctions, *early bidding* and *last-minute bidding*. By signaling their commitment early in the auction process, early bidders (Bapna et al., 2004) often discourage potential competitors from entering the same auction. In contrast, last-minute bidders (Roth and Ockenfels, 2002; Shmueli et al., 2007) wait until the very last moment since the chances of being out-bid decrease with the time left in the auction. However, there are two serious drawbacks to early and last-minute bidding. First, neither strategy takes into account the effect of competition. In other words, neither strategy considers the information from simultaneous auctions offering similar products. Moreover, by restricting the strategy to a rather limited time window (early *vs.* late), both strategies are inflexible since they require bidders to monitor an auction and wait for the right moment to come.

In this paper, we propose a novel automated and data-driven bidding strategy. Our strategy consists of two main components. First, we develop a dynamic, forward-looking forecasting model for price in competing auctions. Then, using the idea of maximizing consumer surplus, we build a bidding framework around this model that determines a triumvirate of decision points: the best auction to bid on, the best bid-time and the best bid-amount. We expand upon each of these components below.

The first component of our automated bidding strategy is a dynamic forecasting model for the price in competing auctions. There has been considerable amount of work on predicting an auction's closing price using *static* (or pre-auction) information. For instance, a seller's reputation (Bajari and Hortacsu, 2004), an auction's duration (Haruvy et al., 2007a), or an item's shipping costs (Haruvy et al., 2007b), all have been shown to affect the final price. Similarly, Lucking-Reiley et al. (2007) find that auction length, sellers' ratings, opening and reserve prices all are significant predictors of price. (See also Simonsohn and Ariely (2007).) One drawback of these approaches is that they only consider static information available *before* the start of the auction. In contrast, our approach captures the dynamic nature of the auction process.

Dynamics have only recently been found to exist and to affect the outcome of an online auction (e.g. Bapna et al., 2008b). Jank and Shmueli (2008) find that auctions selling identical products fall into one of three different segments of price dynamics. Wang et al. (2008b) show that an auction's price dynamics can be characterized well using a single class of functional differential equation models (see also Jank et al., 2008). Moreover, Wang et al. (2008a) show that the inclusion of price dynamics into forecasting models significantly improves the predictive capability of the outcome of an auction. In this paper, we study dynamics in the context of competing auctions. That is, we study the effect of dynamics generated in simultaneous auctions, selling the same or similar product as in the auction of interest.

We incorporate the dynamic nature of the auction process by employing a modern statistical approach called *functional data analysis* (FDA). FDA (see Ramsay and Silverman, 2005) allows us to incorporate real-time information of the ongoing auction rather than only static pre-auction information. Real-time information includes the current price level, or the current number of bidders. It also allows us to incorporate auction dynamics. By auction dynamics we mean, e.g., the velocity of price and its acceleration. We find that the predictive accuracy of our model is greatly improved by incorporating dynamic auction components.

While the information from within an auction matters, what happens in other, simultaneous auctions also has an effect on the outcome. Ariely and Simonson (2003) point out that concurrent auctions with similar items affect bidders' assessment of an item's value. Sun (2005) shows large price dispersion in sequential eBay auctions for identical items. In concurrent eBay auctions, Anwar et al. (2006) find that significant cross-bidding occurs and that bidders often switch to lower-priced auctions. More evidence of auction competition can be found in Haruvy et al. (2007a). Chan et al. (2007) consider two-dimensional market competition using market-breadth and -depth measures to characterize competition in online auctions. Jank and Shmueli (2007) propose a novel spatio-temporal model for auction competition. Their model considers temporal competition using an innovative measure to overcome the unevenly spaced time-series of closing auctions; it considers competition with similar (or substitute) products using a spatial component that measures the distance of substitute products in the associated feature space. In this paper, we propose several innovative measures for auction competition using the concept of functional data analysis. Our measures can be grouped into three conceptually different types: measures that capture competition due to static (or pre-auction) information; measures that capture competition due to evolving information; and measures that capture competition in the form of dynamics.

Our proposed forecasting model includes measures of dynamics and competition as predictors. We perform model selection to reduce the available set of candidate predictors to a parsimonious few and we compare its predictive capability to several alternate approaches such as General Additive Models (GAM) and Classification and Regression Trees (CART). Both GAM and CART are popular tools from the data mining literature since they automate the model selection process without assuming strict functional relationships between response and predictors. We find that our model can predict an auction's future price with high accuracy and outperforms alternate approaches.

In the second component, we build a comprehensive bidding strategy around our forecasting model. The idea is based on maximizing consumer surplus, which refers to the difference between the bidders' *willingness to pay* and the price *actually paid*. We formulate an automated algorithm for selecting the best auction to bid on, and for determining the best bid-time and -amount. The best auction provides bidders with the highest surplus, and the best bid-amount equals the predicted closing price. This strategy automates the entire decision-making process, and, in contrast to early or late-minute bidding strategies, it frees the bidder of time-constraints since bidding can occur immediately.

We conduct a simulation study where we compare our automated, data-driven bidding strategy with early bidding and last-minute bidding. In similar fashion to Bapna et al. (2004), we compare all bidding strategies on two different dimensions: the probability of winning an auction, and the surplus extracted. We find that although snipers have the highest probability of winning, our strategy results in a much higher surplus and considerably less amount of time devoted to the process of bidding. Moreover, early bidders have the lowest average winning probability and surplus. We also investigate the impact of the *prediction window* on the resulting surplus. The prediction window is equivalent to the given time frame within which a bidders wants to win an auction. Shorter time frames correspond to bidders that want to win more quickly; longer time frames correspond to bidders that allow more time for search and selection. We find that, as the width of the prediction window is increasing, surplus goes up while the probability of winning goes down.

The paper unfolds as follows: In Section 2, we introduce the data used in our study, and motivate the problem of auction competition. In Section 3, we propose our forecasting model. We discuss the creation of predictors related to dynamics and competition, the selection of important predictors, and the updating of our model to produce forecasts of future prices. Results of model fitting are discussed in Section 4 and we also compare the predictive capability of our model to alternate approaches. In Section 5, we discuss the framework for our automated bidding strategy, and the results of our simulation study. We conclude in Section 6 with further research directions.

2 Data Description

2.1 Palm M515 eBay Auctions

The data used in this study are the complete bidding records for a specific product (new Palm M515 handheld devices), auctioned between March 14, 2003 and May 25, 2003. The market price at the time of data collecting was $220 (based on `Amazon.com`). Each bidding record includes the auction number, starting- and closing-time and -prices, bids with associated time stamps, and other information, such as auction duration, shipping fee, seller's feedback score, whether the seller is a power seller, whether the product is from an eBay store, and whether auction's description includes pictures. A summary of this information can be found in Table 1.

Table 1: Descriptive statistics of bidding records. The top panel reports statistics for continuous variables; the two bottom panels report statistics for discrete variables.

Variable	Mean	Stdev	Median	Min	Max
opening price	76.67	92.45	9.99	0.01	265.00
shipping fee	15.44	5.51	15.00	0.00	50.00
seller's feedback	545.73	1787.47	44.00	0.00	27652.00
closing price	229.45	22.00	232.50	172.50	290.00
number of bids	17.45	11.23	17.50	1.00	54.00

auction length	3 days	5 days	7 days	10 days
number of auctions (pct)	95(25.00%)	59(15.53%)	219(57.63%)	7(1.84%)

Variable	Yes	No
power seller	121(31.84%)	231(60.79%)
eBay store	117(30.79%)	235(61.84%)
picture	332(87.37%)	20(5.26%)

The top panel in Table 1 shows that the opening price varies significantly, between $265, which is higher than the market value, and $0.01, which is the default on eBay. Half of the sellers set the opening prices lower than $10 since lower opening bids often attract more bidders. The shipping fee includes shipping and handling that the seller charges, and it is often perceived as a hidden cost on eBay. We can see that shipping fees are as high as $50, which adds considerably to the total transaction costs. The sellers feedback rating is the difference between the number of positive and negative ratings. The higher the feedback, the more trustworthy a seller is often perceived, but there are also problems with feedback as trust signal (Dellarocas, 2003). In our data, seller feedback is extremely right-skewed with a mean and median of 545 and 44, respectively. The closing price is the price at which an auction transacts. It varies between $172 and $290 despite the fact that the market value is around \$220. The high variance in closing prices shows that there are significant opportunities for an informed bidder to make a bargain. The number of bids measures the amount of competition in an auction. Typically, auctions with a larger number of bids see higher closing prices. In our data, the average number of bids is 17 while some auctions receive as many as 54 bids.

From the bottom two panels in Table 1 we can see that most auctions are 7-days long, the most popular auction duration on eBay. We can also see that over 87% of all auctions feature a picture. Pictures carry visual information about products, thus enhance bidders' confidence in the quality of the item. Power sellers are sellers with consistently high volume of monthly sales, over 98% positive ratings, and PayPal accounts in good financial standing. We can see that 30% of sellers are power sellers in our data. And lastly, sellers with feedback scores of 20 or higher, verified ID and PayPal accounts in good financial standing are permitted to open stores on eBay, which provides easy management of accounts and brand boosting when the sellers have multiple items listed. Approximately 30% of all auctions are associated with an eBay store.

2.2 Competition between Auctions

We now provide a first, exploratory glance into the competitive nature between auctions. At any point in time, there are millions of auctions transacting simultaneously on eBay. Many of these auctions sell the same or similar items. This information overload causes a problem for bidders who have to decide which auction to bid on. Their bidding decisions, in turn, cause dependencies across auctions, since prices will stay low in auctions where bidders decide not to bid, and will be high in others. This simultaneousness, or *concurrency*, thus results in competition between auctions and their outcomes (Jank and Shmueli, 2007; Haruvy et al., 2008).

Consider Figure 1 (left panel) which shows a calendar plot (Jank and Shmueli, 2005) of closing prices vs calendar time for the data from Section 2.1. Each line in the graph represents an auction, where the length of the line corresponds to the length of the auction. The y-axis denotes the closing price, and the x-axis denotes calendar time. We can see that, at any given time point, there exist many auctions that offer the identical item; some auctions have only begun, while others are

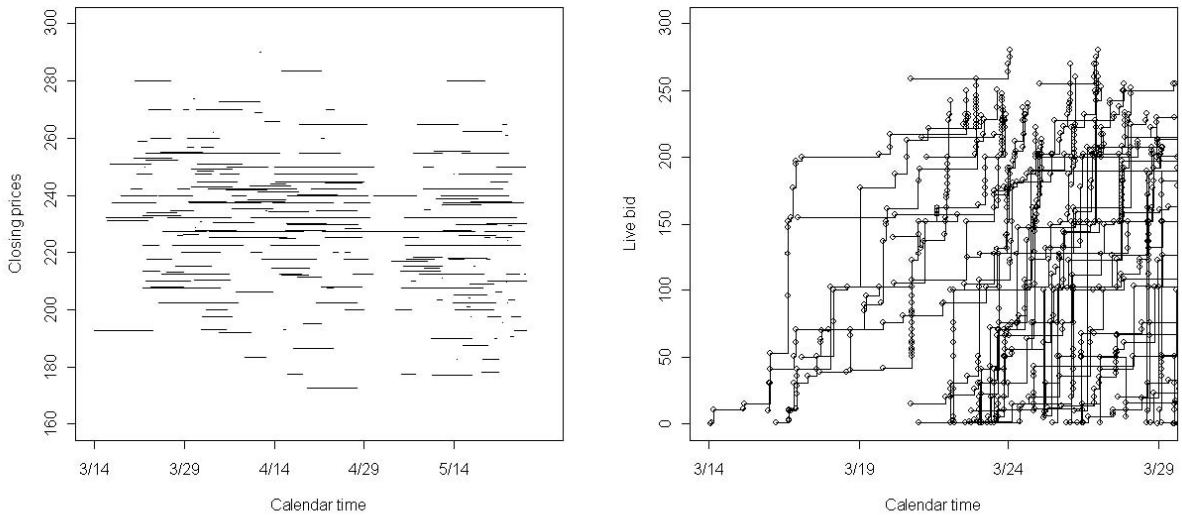

Figure 1: Calendar plot of Palm M515 eBay auctions (left panel); a snapshot of the associate live price curves (right panel).

about to close. Moveover, the closing prices vary greatly, so choosing the *right* auction is extremely important for the price-conscious bidder.

Another illustration of the information-overload that bidders face is shown in the right panel of Figure 1. It shows the price processes of ongoing auctions. In particular, it shows, for each individual auction, the *live price curve*, that is, the price that bidders see at any given time during the ongoing auction. We can see that the information can be quite overwhelming: the amount of concurrent auctions, the variation in prices and the fact that some auctions are only in the early stages, while others are about to end, all cause challenges for properly processing the given information. Moreover, we see that prices increase unevenly throughout most auctions. They increase fast in some auctions, but much slower in others. We refer to this as *price dynamics*, which will be an important factor in our modeling approach.

In the following, we propose a dynamic model for forecasting price in competing auctions. We then use this model to build a data-driven and automated bidding strategy around it. We start by describing the forecasting model.

3 A Model for Forecasting Competing Auctions

Our forecasting model has several features: We model the real-time price process of ongoing auctions using functional data analysis (FDA), which allows us to incorporate information about the dynamics of price. We also propose several innovative ways of incorporating competition across concurrent auctions, and then we suggest an innovative way to perform model selection and model updating. We describe these features in detail next.

3.1 Functional Data Analysis and Price-Dynamics

The price-process of online auctions is characterized by an extremely dynamic environment. One aspect of this environment is the changing bid density, where the number of bids per time unit changes constantly. The resulting unequally-spaced time-series of bids deem traditional models (which assume evenly spaced measurements) inadequate. Furthermore, the changing bidding patterns also result in varying price-dynamics. By price-dynamics we mean the change in price and the rate at which this change occurs. Traditional forecasting models, which do not account for such instantaneous change, fail to accurately predict auction prices (Wang et al., 2008a). To incorporate this dynamic environment, we take a functional data modeling approach.

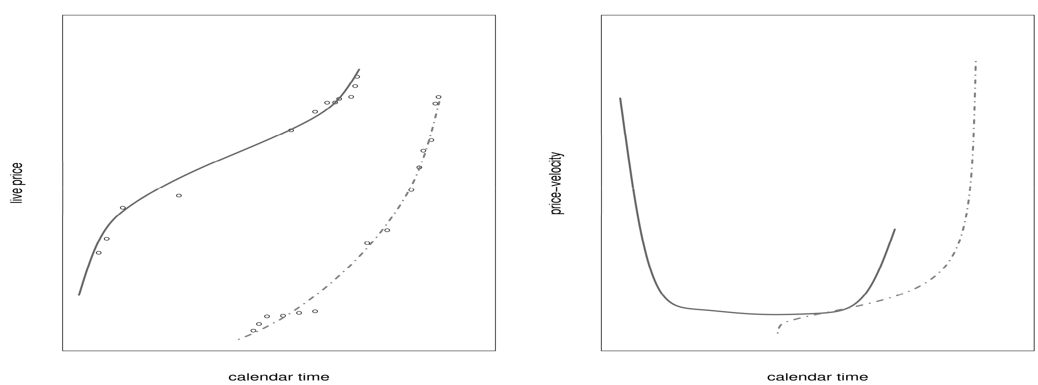

Figure 2: Smooth price curves (left panel) and corresponding velocities (right panel) for two sample auctions. The dots in the left panel denote the observed bids.

Functional data analysis (Ramsay and Silverman, 2005) uses smoothing methods to recover (or estimate) the underlying price curve from observed bidding data. From the price curve, we then obtain estimates of the price dynamics via its first and second derivatives. Figure 2 illustrates the process of generating smooth price curves from observed data (left panel) and estimating the corresponding price-velocities (right panel). We see that the smooth curves capture the trend of the price increase due to the discrete bids; the velocity captures the instant change of price increase. For more details on the smoothing process, see Appendix A.

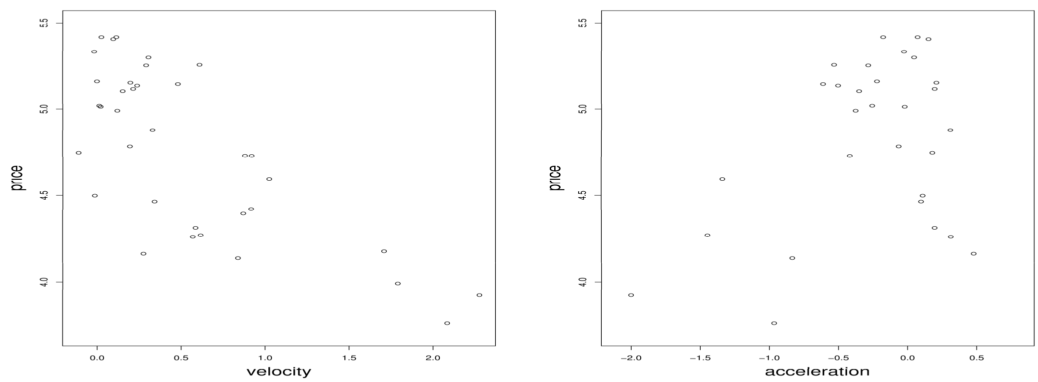

Figure 3: Relationship between dynamics and future prices: the x-axis denotes dynamics at time T; the y-axis denotes the auction price at time $T+1$.

The impact of price dynamics on the price of an auction is illustrated in Figure 3. In that Figure, we plot an auction's dynamics at time T (velocity, left panel; acceleration, right panel) on the x-axis, and we plot the price at time $T+1$ on the y-axis (on the log-scale). We can see a strong relationship between dynamics and future prices, for both velocity and acceleration. This gives an early indication that price dynamics will play an important role in our forecasting model.

3.2 Capturing Competition

One major component of our model is competition. That is, we want to capture the effect of what happens in other, simultaneous auctions. To that end, we must first define meaningful measures for competition. There are many different ways of defining competition measures and we explore several alternatives below. All measures are driven by the same general principle which is illustrated in Figure 4. We define a focal auction (indicated by the solid line in Figure 4) as the auction for

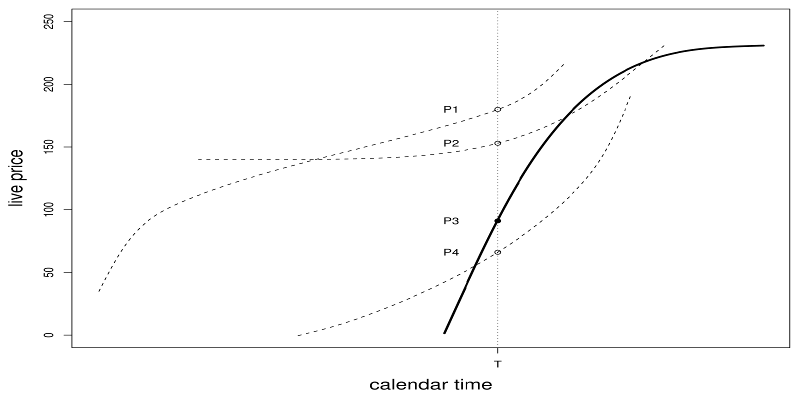

Figure 4: Illustrating competition: The sold black line denotes the focal auction; the dashed lines denote competing auctions; T denotes the time of decision-making.

which a bidder wants to decide whether or not to bid on. At time T of decision-making, there are several other auctions that take place simultaneous (indicated by the dotted lines). One meaningful measure of competition is the level of price in other auctions. In our example, there are four different prices levels at time T, varying from high (p1) to low (p4). The price level in the focal auction at that time is p3. Thus, a possible measure for the price competition is given by the *average price* in concurrent auctions (which we denote by *c.avg.price*), that is, by the average of p1, p2 and p4. In similar fashion, the *average price-velocity* (*c.avg.vel*) in concurrent auctions would be defined as the average of the corresponding price-velocities, and so on.

In this paper, we investigate several different competition features and their impact on the price of the focal auction. Table 2 categorizes these features by the information that they carry: *Static competition features* are known at the outset of the auction and do not change during the auction process; examples include the opening price or the shipping fees of concurrent auctions. *Evolving competition features* change during the auction process, such as the current price or the current number of bidders of concurrent auctions. And *price-dynamic competition features* capture the effect of changing dynamics in competing auctions.

In Figure 5, we explore the relationship between the competition features from Table 2 and the future price in the focal auction. We can see that some features (e.g. the average price-velocity

Table 2: Candidate Competition Features

Name	Description
	Static Features
c.openbid.avg	Average opening price of concurrent auctions
c.dura.avg	Average duration of concurrent auctions
c.ship.avg	Average shipping fee of concurrent auctions
c.feedback.avg	Average sellers' feedback of concurrent auctions
c.power.avg	Average number of power seller in concurrent auctions
c.store.avg	Average number of eBay stores in concurrent auctions
c.pic.avg	Average number of pictures in concurrent auctions
	Evolving Features
c.price.avg	Current average price in concurrent auctions
c.price.vol	Current price volatility (stdev) in concurrent auctions
c.price.disc	Price discount (difference) between focal auction and highest concurrent price
c.t.left.avg	Average time left in concurrent auctions
c.t.left.vol	Volatility (stdev) of time left in concurrent auctions
c.nbids.avg	Average number of bids in concurrent auctions
c.nbids.vol	Volatility (stdev) of number of bids in concurrent auctions
c.nbidders.avg	Average number of bidders common to focal and concurrent auctions
c.nbidders.vol	Volatility (stdev) of number of bidders common to focal and concurrent auctions
	Price-dynamic Features
c.vel.avg	Average price-velocity in concurrent auctions
c.vel.vol	Volatility (stdev) of price-velocity in concurrent auctions
c.acc.avg	Average price-acceleration in concurrent auctions
c.acc.vol	Volatility (stdev) of price-acceleration in concurrent auctions

and -acceleration of competing auctions) have a strong relationship with price, while others (e.g. the average number of bidders in simultaneous auctions) have a rather weak relationship. Pairwise correlation analysis (not reproduced here) also shows that, unsurprisingly, many of the features in Table 2 are multicollinear. Thus, a good modeling strategy will start with a suitable variable selection procedure. We will use the initial observations from Figure 5 for guidance when selecting the most relevant set of competition features in the next section.

3.3 Variable Selection

Many different pieces of information can affect price in online auctions. We differentiate between two main components, i.e. information from *within* the focal auction vs. information from other, *competing* auctions that take place simultaneously. Within each component, information can be further segmented into static, evolving and price-dynamic information, similar to Table 2. Table 3

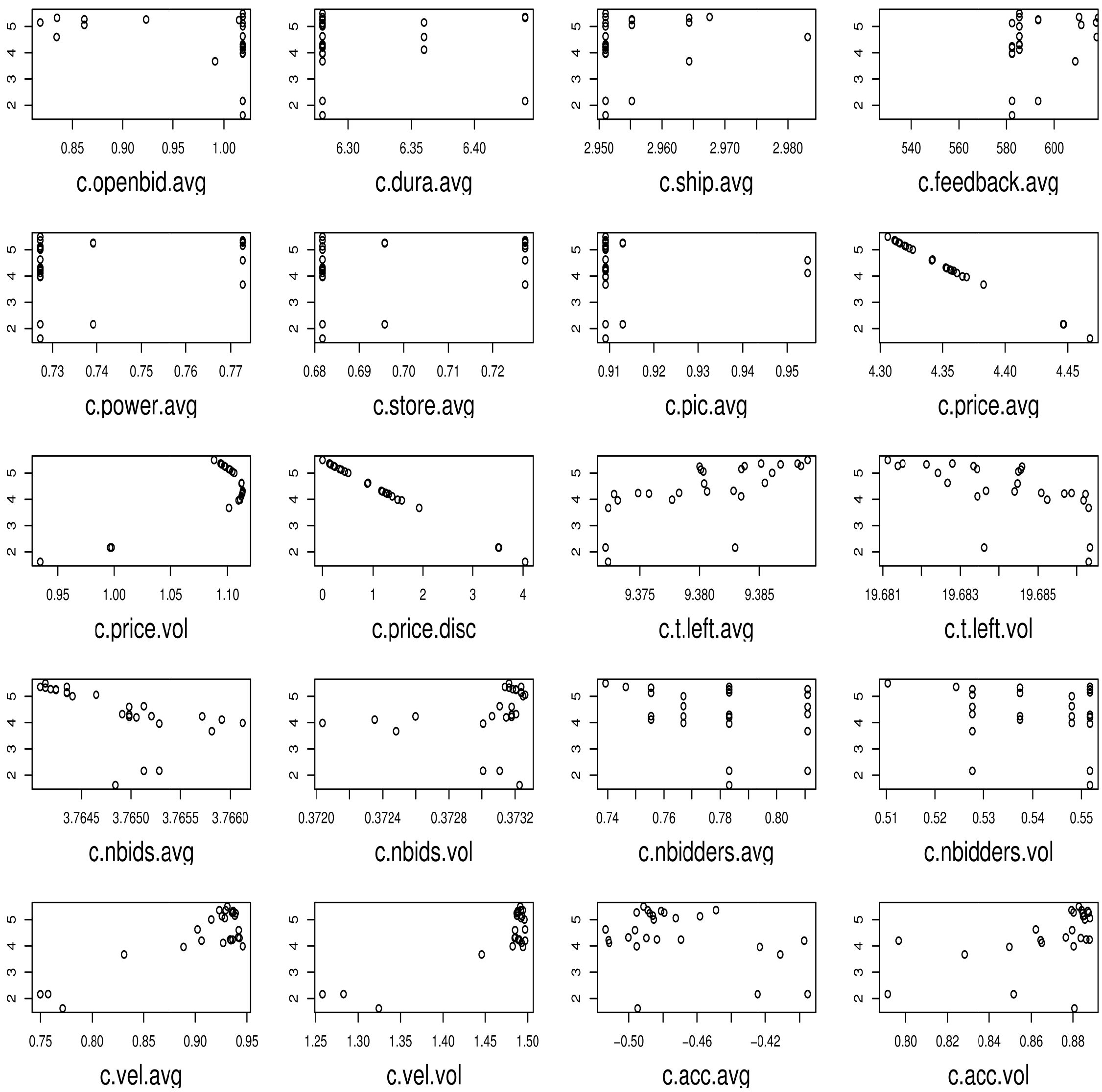

Figure 5: Pairwise relationships between all competition features from Table 2 (measured at time T) and the price (measured at $T+1$) in the focal auction.

Table 3: Candidate information for our forecasting model

	Information from within the focal auction
Static information	opening bid, auction duration, shipping fee, seller's feedback, power seller, eBay store, picture
Evolving information	current price, time left, current number of bids, current number of bidders
Price-dynamic information	price velocity, price acceleration
	Information from competing auctions
Static information	c.openbid.avg, c.dura.avg, c.ship.avg, c.feedback.avg, c.power.avg, c.store.avg, c.pic.avg
Evolving information	c.price.avg, c.price.vol, c.price.disc, c.t.left.avg, c.t.left.vol, c.nbids.avg, c.nbids.vol, c.nbidders.avg, c.nbidders.vol
Price-dynamic information	c.vel.avg, c.vel.vol, c.acc.avg, c.acc.vol

list all the different pieces of information that are candidates for our forecasting model.

Table 3 shows that there are over 30 different variables that are candidates for our forecasting model. Thus, an important first step in our modeling efforts is the selection of a parsimonious subset of relevant predictors. Variable selection has been researched in the statistics literature for a while (Berk, 1978) and it is receiving increasing attention today with the availability of more and more data sets featuring larger and larger number of variables (George, 2000). A complicating factor in our situation is the time-varying nature of our model. Our goal is to find a model that predicts well at time $T+1$, *universally* across all time periods $T = 1, 2, 3, \ldots, N_T$. Classical variable selection procedures focus on cross-sectional data, that is, on data corresponding to a *single* time period only. Since our data varies over time, it is quite plausible that there exists *one* model that best predicts at time $T + 1$, while *another* (different) model best predicts at a different time $T' + 1$. Our goal is to find a model that is not geared to a single time period only, but applies rather universally to the eBay market over a longer time window. To that end, we choose a model which has good *average* performance, averaged over all time periods T of interest. We describe this approach next.

Our model has the general form

$$\mathbf{y}_{T+1} = \boldsymbol{\beta}'_T \mathbf{x}_T \tag{1}$$

Table 4: Percentage of significant time points. The two leftmost columns refer to predictors from within the focal auction; the two rightmost columns refer to predictors from competing auctions.

Focal auction	*p.sig*	Competing auctions	*p.sig*
openbid	.199	c.openbid.avg	.193
duration	.032	c.dura.avg	.032
shipping	.039	c.ship.avg	.046
sellerfeed	.055	c.feedback.avg	.044
powerseller	.061	c.power.avg	.076
store	.092	c.store.avg	.104
picture	.028	c.pic.avg	.028
currenprice	1.00	c.price.avg	1.00
		c.price.vol	.886
		c.price.disc	1.00
timeleft	.775	c.t.left.avg	.771
		c.t.left.vol	.758
numbids	.780	c.nbids.avg	.777
		c.nbids.vol	.509
numbidders	.197	c.nbidders.avg	.188
		c.nbidders.vol	.086
price-velocity	.762	c.vel.avg	.762
		c.vel.vol	.624
price-acceleration	.308	c.acc.avg	.309
		c.acc.vol	.306

where $\mathbf{y}_{T+1}$ denotes the auction price at $T+1$, $\mathbf{x}_T = (x_{T1}, \ldots, x_{Tp})'$ is a vector of predictors, and $\boldsymbol{\beta}_T = (\beta_{T1}, \ldots, \beta_{Tp})'$ is a vector of coefficients to be estimated from the data. The goal is to select only those predictors that are important for predicting the price $\mathbf{y}_{T+1}$, across all time periods $T = 1, 2, 3, \ldots, N_T$.

We accomplish this in several steps. In the first step, we run simple regressions (i.e. $p = 1$) between each individual predictor from Table 3 and the response $\mathbf{y}_{T+1}$ at *each* time period $T, T = 1, 2, 3, \ldots, N_T$. We then calculate the percentage of time points a predictor is significant (at the 5% significance level). That is, for each predictor $x_k = (x_{1k}, \ldots, x_{N_T k})'$, $k = 1, \ldots, p$, we compute

$$p.sig_k := \frac{1}{N_T} \sum_T \mathbf{1}_{\{x_{Tk} \text{ significant at 5\% level}\}}. \tag{2}$$

Table 4 shows the results for a fine grid of hourly forecasts (i.e. $(T+1) - T = 1$ hour) which results in $N_T = 1,754$ different time periods. We can see that the predictors that *individually* have a strong effect on $\mathbf{y}_{T+1}$ (consistently across all time periods T) are the current price, price-velocity and -acceleration, time left and the number of bids (from within the focal auctions) and c.price.avg, c.price.vol, c.price.disc, c.t.left.avg, c.t.left.vol, c.nbids.avg, c.nbids.vol, c.vel.avg, c.vel.vol, c.acc.avg and c.acc.vol (from competing auctions). It is interesting that most of these variables relate to price (or price-movement) from the focal auction relative to competing auctions. This suggests that information about price and its dynamics effectively captures much of the relevant auction information such as information about the product, the auction format, the seller and competition between bidders. However, also note that the results so far are based only on simple regressions ($p = 1$) and thus may not fully reflect the *joint* effect of a predictor in the presence of other predictor variables. To that end, we investigate pairwise correlations (again, averaged across all time periods, $T = 1, \ldots, N_T$; correlation-table not reported here) and find high collinearity between ten pairs: the current price and c.price.avg, the current price and c.price.vol, the current price and c.price.disc, the current price and time left, the current price and c.t.left.avg, the current price and number of bids, the current price and c.nbids.avg, price-velocity and c.vel.avg, price-velocity and c.vel.vol, and price-acceleration and c.acc.avg. This high collinearity is not surprising since many of these predictors carry similar information, only coded in a slightly different way. We eliminate all highly collinear predictors; next we derive our final model using the Bayesian Information Criterion (BIC).

In similar fashion to (2), one can compute the *average* BIC across all time periods. That is, let avg.BIC $:= 1/N_T \sum_T \text{BIC}(T)$, where $\text{BIC}(T), T = 1, 2, 3, \ldots, N_T$, denotes the Bayesian Information Criterion (e.g. George, 2000) of a model computed at time period T. By comparing all possible subsets of non-collinear predictors, we arrive at our final forecasting model as

$$\mathbf{y}_{T+1} = \alpha_T + \beta_{1T}\text{current price}_T + \beta_{2T}\text{velocity}_T + \beta_{3T}\text{acceleration}_T + \beta_{4T}\text{c.acc.vol}_T. \qquad (3)$$

Table 5 shows the avg.BIC of our final forecasting model (3) compared to several competitor models. We can see that our model results in the lowest avg.BIC. It is also interesting to see that models with *only* information from competing auctions perform almost as well as models with the corresponding information only from within the focal auction. This is yet another piece of evidence

Table 5: Average BIC computed across all time periods T. The first row shows the value of $avg.BIC$ for our model in (3); the remaining rows show the corresponding values of several competing models.

Model	*avg.BIC*
Our model from eq. (3)	**-381.59**
Full model (all 33 predictors from Table 3)	-147.08
All 13 predictors from the focal auction (Table 3)	-319.82
Only 2 focal auction dynamics	37.77
Only 4 focal auction evolving predictors	-83.87
All 20 predictors from competing auctions (Table 3)	-313.52
Only 4 competing auction dynamics	37.58
Only 9 competing auction evolving predictors	-84.29

for the tight connectivity of the auction marketplace.

We also compare the performance of our final (parametric) model (3) to flexible alternate approaches. Alternatives to parametric models which have received attention in the data mining literature are the General Additive Model (GAM) (Hasti and Tibshirani, 1990) and Classification and Regression Trees (CART) (Breiman et al., 1984). Instead of assuming a sometimes overly restrictive linear relationship between the response and predictors, GAM allows for flexible non-parametric relationships. On the other hand, CART provides a data-driven and tree-structured way to partition the variable-space and is thus often viewed as an alternate way to formal variable selection. We contrast the results of GAM and CART to our model in Section 4.

3.4 Model Updating

The goal of our model is to predict price at a future time $T+1$ using only information from the present (i.e. time T) and the past ($T-1$, $T-2$, etc.). We accomplish this by estimating the functional relationship between $T-1$ and T and then applying this relationship to predict $T+1$ from T. Figure 6 illustrates this updating scheme.

At time T (present), we wish to make a prediction about the future price at time $T+1$. Per our model, $\mathbf{y}_{T+1}$ is given by $\boldsymbol{\beta}_T'\mathbf{x}_T$, where $\mathbf{x}_T$ contains information observed in the present (or past). Note that we cannot estimate $\boldsymbol{\beta}_T$ directly since the response ($\mathbf{y}_{T+1}$) is yet unobserved. We

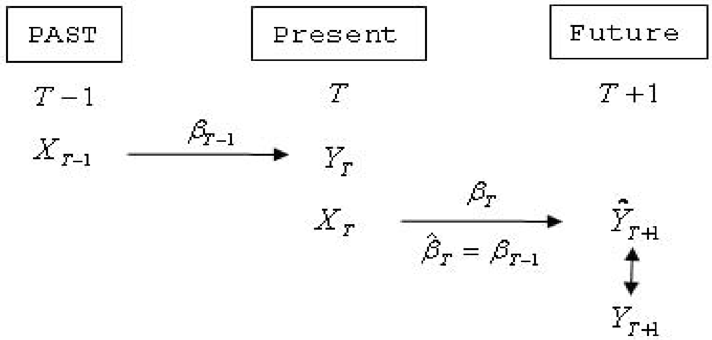

Figure 6: The illustration of the update scheme in the forecasting model

therefore estimate the relationship from the past: We estimate $\boldsymbol{\beta}_{T-1}$ for the price at T ($\mathbf{y}_T$) and then estimate $\boldsymbol{\beta}_T$ via $\hat{\boldsymbol{\beta}}_T := \boldsymbol{\beta}_{T-1}$. In that sense, we "roll" the relationship from the past one time period forward. We also investigated alternate updating approaches (such as estimating $\boldsymbol{\beta}_T$ via a moving average (MA) of prior relationships, $\hat{\boldsymbol{\beta}}_T := \text{MA}\{\boldsymbol{\beta}_{T-1}, \boldsymbol{\beta}_{T-2}, \boldsymbol{\beta}_{T-3}, \ldots\}$) but did not find significant improvements in model performance.

4 Estimation and Prediction Results

In this section we discuss estimation and prediction of our forecasting model in (3). We also compare its predictive capabilities to alternate forecasting approaches.

To that end, we divide our data set into a training set (80% of the data), and a validation set (remaining 20% of the data). Since our data varies over time (and since we are primarily interested in making accurate predictions of the future), our training set consists of all auctions that complete during the first 80% of our data's time span (i.e. between March 14 and May 10); the validation set contains all remaining auctions (i.e. between May 11 and May 25). In that sense, we first estimate our model on the training set; results of model-estimation and -fit are discussed below. We then apply the estimated model to the validation set to gauge its predictive capabilities; this is discussed in the second half of this section.

4.1 Model Estimation

Figure 7 shows the estimated coefficients for the parameters of our forecasting model (3). Recall that we estimate the model at every time point $T, T = 1, 2, 3, \ldots, N_T$ in the training set. In our application, we consider time intervals of one hour over the time period between March 14 and May 10, hence the coefficients also vary over that time period. Figure 7 shows the resulting trend of the coefficients together with 95% confidence bounds.

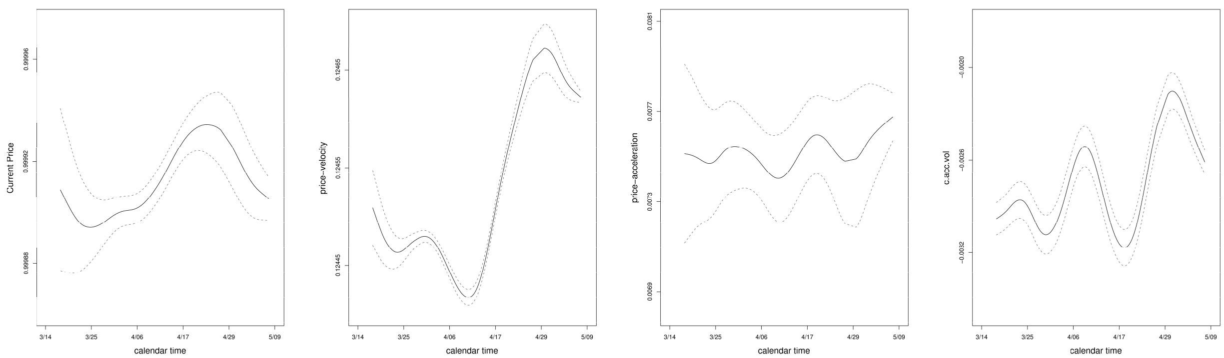

Figure 7: Estimated coefficients of model parameters, together with 95% confidence bounds. The x-axis denotes calendar time; the y-axis denotes the magnitude of the coefficient. The panels show (from left to right) current price, price-velocity, price-acceleration and the acceleration volatility of competing auctions (c.acc.vol).

We can see that information from within the focal auction (current price, price-velocity and -acceleration) has a positive relationship with the future price $\mathbf{y}_{T+1}$; in contrast, information from competing auctions (c.acc.vol) has a negative relationship. In other words, both the current level of price and its dynamics are positive indicators of future price. On the other hand, the volatility of price-acceleration in competing auctions is a negative indicator. Price-acceleration in competing auctions will be high if many bidders bid in auctions *different* from the focal auction. A high volatility in price-acceleration may suggest high uncertainty in the marketplace, with some auctions experiencing large price jumps and others experiencing no price-movements at all. This high uncertainty results in depressed prices of the focal auction.

4.2 Model Fit and Varying Time Intervals

Figure 7 shows the estimated model coefficients for one hour time intervals; that is, for $(T+1)-T=$ 1 hour. Alternatively, one could also consider models with a larger time intervals; that is, models that forecast further into the future. Intuitively, since forecasting further into the future is harder, such models should not perform as well. Figure 8 (left panel) shows the model-fit for the time intervals $(T+1)-T=1,2,3,\ldots,14$ hours. We measure model-fit by the average R^2 value, avg.$R^2 := 1/N_T \sum_T R^2(T)$, where $R^2(T), T=1,2,3,\ldots,N_T$, denotes the R^2 of a model computed at time period T. We can see that, as expected, the model-fit decreases as the time intervals get larger. Notice though that even for the largest time interval (14 hours), the value of avg.R^2 is still larger than 99%.

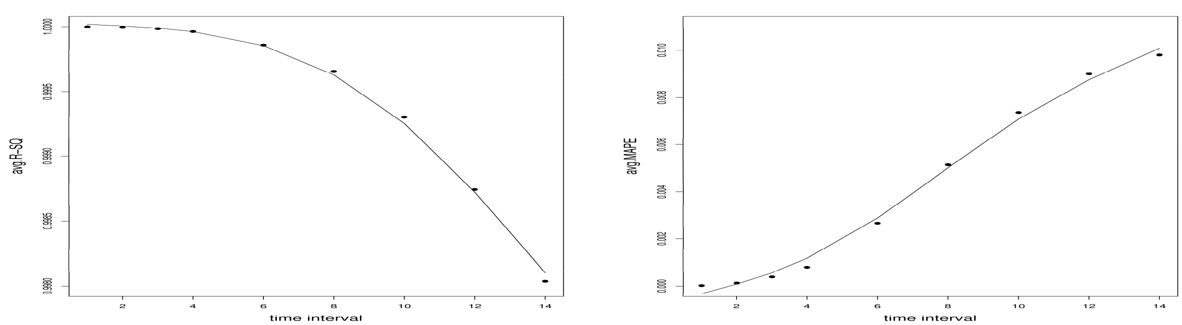

Figure 8: Model-fit and prediction accuracy for different time intervals. The x axis represents the time interval (in hours; ranging from 1 hr to 14 hrs); the y axis represents the value of avg.R^2 (left panel) and the value of avg.MAPE (right panel).

4.3 Prediction Performance

As pointed out above, we estimate the model on the training set; then we gauge its predictive performance on the validation set. We measure predictive performance of a model in terms of its *mean absolute percentage error* (MAPE). For each time period $T, T=1,2,3,\ldots,N_T$, in the validation set, we compute

$$\text{MAPE}(T) = \frac{1}{m_{T+1}} \sum_{i=1} \frac{|y_{T+1,i} - \hat{y}_{T+1,i}|}{|y_{T+1,i}|} \tag{4}$$

where $y_{T+1,i}$ and $\hat{y}_{T+1,i}$ denote the true and predicted values of auction #i at time $T+1$, respectively, and m_{T+1} denotes the number of auctions available at time $T+1$. We compute the *average MAPE* across all time periods as avg.MAPE $:= 1/N_T \sum_T$ MAPE(T). In similar fashion to Section 4.2, we investigate avg.MAPE for different time intervals, $(T+1) - T = 1, 2, 3, \ldots, 14$ hours. The right panel in Figure 8 shows the results.

Unsurprisingly, we see that as we predict further into the future (i.e. as time interval gets larger), the predictive performance decreases (i.e. avg.MAPE increases). It is interesting to see that for predictions up to 4 hours into the future, the prediction error is less than 0.1%. For time intervals larger than 4 hours, the prediction error increases at a faster rate. However, even for predictions as far as 14 hours into the future, the error is still less than 1%. This predictive accuracy is quite remarkable as we will see in the next subsection where we benchmark our approach against several competitor approaches.

4.4 Comparison with Alternate Models

We benchmark our model against four alternate models, the generalized additive model (GAM), classification and regression trees (CART), and two simper linear models: a purely static and an evolving linear model.

GAMs relax the restrictive linear model assumption between the response and predictors by a more flexible nonparametric form (Hasti and Tibshirani, 1990). CARTs (Breiman et al., 1984) provide a data-driven way to partition the variable-space and are thus often viewed as alternatives to formal variable selection. In addition, we consider two linear models that use a subset of the variables from Table 3: one model that uses only static information from the focal auction and another one that uses static and evolving information from the focal auction; we refer to these two models as “STATIC” and “EVOLVING,” respectively. The static model corresponds to the information of many prior eBay studies (e.g. Lucking-Reiley et al., 2007) in that it only considers pre-auction information. The evolving model accounts for changes due to the process of bidding, but it does not account for price dynamics or competition.

Figure 9 shows avg.MAPE (similar to Figure 8) for time intervals $(T+1) - T = 1, 2, 3, \ldots, 14$

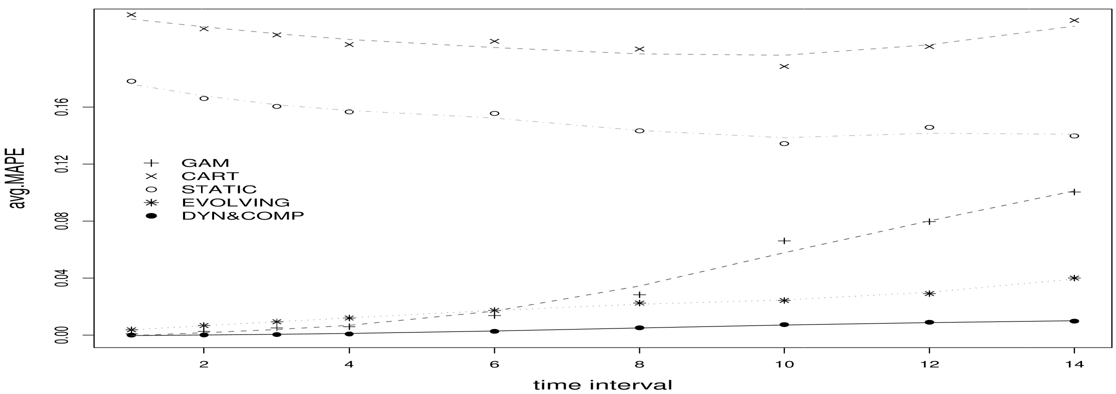

Figure 9: Prediction accuracy for competing models. The x axis represents the time interval (in hours); the y axis represents the value of avg.MAPE.

hours, for all 5 different models. We refer to our model (3) as “DYN&COMP,” since it incudes dynamics and competition features. We can see that STATIC and CART have the worst prediction performance, with an error universally larger than 10%. While our model performs the best, GAM and EVOLVING are competitive, at least for smaller time intervals. In other words, for predicting less than 4 hours into the future, both GAM and EVOLVING pose alternatives with prediction errors not too much larger compared to DYN&COMP. However, their predictive performance breaks down for larger time intervals. In fact, the error of GAM is as large as 10% for predicting 14 hours into the future, which is 10-times larger than the corresponding prediction error of DYN&COMP. While the performance of EVOLVING is somewhat better, its prediction error is 4-time as large as DYN&COMP for 14-hours time intervals. In the next section, we use the excellent forecasting performance of our model and build an automated bidding strategy around it.

5 An Automated Data-Driven Bidding Decision Rule

We now discuss the second component of our bidding strategy, building an automated and data-driven decision rule around our forecasting model. The decision rule provides answers to three basic bidding questions: which auction to bid on, when to bid on it and how much.

5.1 Decision Framework

Our decision framework is built upon the principles of maximizing consumer surplus (e.g. Bapna et al., 2008a). Consumer surplus is the difference between the actual price paid and the consumer's willingness to pay for an item, CS = WTP - Price, where CS denotes consumer surplus, and WTP denotes willingness to pay. Therefore, the lower the price, the higher is a bidder's surplus.

For each individual auction, our forecasting model (3) provides bidders with that auction's estimated future price; combining this with a bidder's WTP leads to an auction's estimated surplus. For a set of competing auctions, a plausible decision rule is to bid on that auction with the highest estimated surplus. Moreover, in order to avoid a negative surplus, a bidder should only bid on an auction if the predicted price is lower than their WTP.

Note that our forecasting model depends on the length of the time interval $(T+1)-T$ (which we also refer to as the *prediction window*). Our model can only predict the final price of an auction that ends at or before time $(T+1)$. Therefore, longer prediction windows will result in a larger number of candidate auctions, that is, in a larger supply of *potential* auctions to bid on. On the other hand, we have also seen in Section 4 that a larger time interval leads to an increased prediction error. Therefore, our decision rule faces a trade-off between supply of candidate auctions and prediction accuracy for each individual auction. We will investigate this trade-off in detail below.

Our decision rule picks that auction with the highest estimated surplus, as long as the surplus is positive. After picking an auction, the next two questions are with respect to the *time* and *amount* of the bid. Since our forecasting model is based on a fixed time interval, nothing is gained by waiting. So we suggest placing the bid as soon as an auction is picked. This frees bidders of time-constraints usually experienced with early- or last-minute bidding. Moreover, since our model predicts an auction's closing price at $\hat{y}_{T+1}$, we would expect to lose for bids lower than $\hat{y}_{T+1}$. Similarly, bids higher than $\hat{y}_{T+1}$ are expected to overpay. Therefore, we suggest to bid *exactly* the expected (or predicted) closing price $\hat{y}_{T+1}$.

In summary, our decision rule picks the auction with the highest predicted surplus, it bids the predicted price, and it places the bid immediately. We note that since bidding occurs immediately (and automatically), bidders do not have to wait for a certain time of the auction, as is often the case with last-minute bidders (or snipers).

5.2 Experimental Set-Up

We conduct a simulation study to compare our automated bidding strategy to two alternate (and popular) bidding approaches: *early bidding* (Bapna et al., 2004) and *last-minute bidding* (Roth and Ockenfels, 2002). Early bidding is often used as a signal for a bidder's commitment and intends to deter other bidders from entering the auction. Last-minute bidding is a popular strategy because it does not allow enough time for other bidders to react. In our simulation, we assume that bidders' willingness-to-pay (WTP) are drawn from a uniform distribution (Adamowicz et al., 1993) that is symmetrically distributed around the market value. At the time of data-collection, the market value of the item was around \$230, so we assume a uniform WTP distribution, Uniform(\$220, \$240). In the following, we describe the set-up of each simulated bidding strategy.

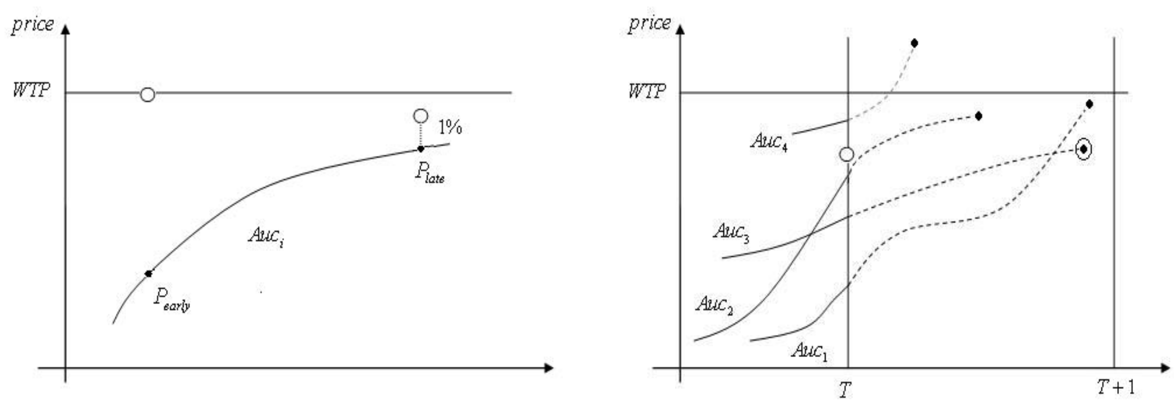

Figure 10: Illustration of bidding strategies. The left panel illustrates early and last-minute bidding; the right panel illustrates our automated bidding strategy.

5.2.1 Early-Bidding Strategy

We assume that early bidders bid at the end of the first auction day (Bapna et al., 2004). In fact, we find that slightly earlier or later bid times barely affect the outcome of the experiment. The process of early bidding is illustrated in the left panel of Figure 10. A bidder compares his WTP with the auction's current price at the end of the first day (p_{early}); if his WTP is higher, then he places a bid; otherwise, he does not place a bid and moves on to another auction. If he does place a bid, then the bid-amount equals his WTP since p_{early} carries little to no information about the auction's final price. In that sense, early bidders (if they are successful) will always incur a zero-surplus.

5.2.2 Last-Minute-Bidding Strategy

We assume that last-minute bidders place their bid one minute before the auction closes (Roth and Ockenfels, 2002). Last-minute bidding carries the danger that the bid does not go through due to network congestion, but we will not explicitly consider this disadvantage in our simulations. The process of last-minute bidding is again illustrated in the left panel of Figure 10. A bidder compares his WTP with the auction's current price one minute before closing (p_{late}). Similar to early bidding, if his WTP is higher, then he places a bid; otherwise, he does not place a bid and moves on to another auction. If he does place a bid, then the bid-amount is only incrementally higher than the current price, since the chances of being outbid within the last 60 seconds are small. In our simulations, we bid an increment of 1% over the current price p_{late} [1]. Note that last-minute bidders typically incur a positive surplus.

5.2.3 Our Automated Data-Driven Bidding Strategy

Our automated bidding strategy is conceptually different from early and last-minute bidding. Instead of making a bidding decision for each auction individually, our strategy requires a bidding decision for each time interval. Consider the right panel of Figure 10. At time T of decision making, there are four competing auctions, denoted by Auc_1-Auc_4, which all close before time $T+1$. The solid lines correspond to the observed part of the auction history; the dotted lines denote the future (and yet unobserved) price path. Since all auctions close before $T+1$, our model yields predictions of their final prices (denoted by the solid black circles). Note that Auc_4 has the highest predicted price; moreover, its predicted price is higher than the bidder's WTP; hence the bidder will never consider this auction. Auc_3 has the smallest predicted price; since the predicted price is also smaller

[1] We also study the robustness to different increments in Appendix B.

than the bidder's WTP, he places a bid on this auction. He bids the predicted price and he bids immediately, i.e. at time T. If he wins, then the bidder's surplus will be the difference between his WTP and the *actual* closing price.

5.3 Simulation Results

Similar to Bapna et al. (2004), we compare all bidding strategies on two different dimensions: the probability of winning the auction and the average surplus accrued. We compute the probability of winning (p.win) as the number of auctions won divided by the total number of auctions that the bidder placed a bid on. We compute the average surplus (avg.sur) as the corresponding average difference between WTP and actual price for all auctions won. We repeat the experiment for all auctions in the validation set for 20 different random draws from a bidder's WTP distribution. The results are shown in Table 6.

Table 6: Comparison of different bidding strategies. The first row corresponds to last-minute bidding; the second row corresponds to early bidding; and the last row corresponds to out automated bidding strategy. We report the mean estimates (as well standard errors in parentheses).

	p.win	avg.sur
Last-moment bidding	92% (.5%)	\$18.47 (\$0.35)
Early bidding	62% (1%)	\$0.00 (\$0.00)
Automated bidding	61% (1%)	\$32.33 (\$1.95)

We see that last-minute bidders have the highest probability of winning (92%, compared to 61% for our automated bidding strategy). This is not surprising, since last-minute bidding is geared at out-witting the competition in the last moment. However, we also see that last-minute bidding accrue a significantly lower surplus compared to our automated bidding strategy (\$18 vs. \$32). Another way of comparing the two bidding strategies is via their *expected surplus*, i.e. the product ($p.win \times avg.sur$). We find that last-minute bidding yields an expected surplus of \$16.99 while that of our automated bidding strategy is higher: \$19.72. Moreover while early bidders have a probability of 62% of winning the auction, their expected surplus equals zero since bidders always bid their WTP.

We close the comparison of bidding strategies on a practical note. In contrast to early bidding or last-minute bidding, the bidding decision in our strategy is free of time, i.e. the bidder bids *immediately* without having to wait for the end of the first auction-day or for the last minute of the auction. While this advantage may not mean much to bidders who use automated bidding agents such as `www.cniper.com`, not all eBay bidders are aware of this technology. Moreover, sniping technology is only available for the eBay market and not for smaller, more specialized auction markets (such as those for Indian art (e.g. Dass et al., 2007)). Therefore, while our automated bidding strategy yields real monetary benefits in terms of a higher expected surplus, it also yields less tangible benefits such as more convenience in terms of a truly automated bidding process.

5.3.1 Effect of the Prediction Window

We have pointed out earlier that the length of the prediction window (i.e. the length of the time interval $(T+1)-T$) has an effect on the outcome of our automated bidding strategy in that longer windows result in a larger supply of candidate auctions, but at the same time reduce the prediction accuracy of each individual auction. The results from the previous section (Table 6) are based on a prediction window of 12 hours and we have seen that it yields an expected surplus of \$19.72 for our automated bidding strategy. Longer prediction windows yield a larger number of candidate auctions and as such a larger probability of including an auction with a lower price (and hence a higher surplus). On the other hand, longer prediction windows also lead to less accurate predictions. Less accurate predictions can either lead to overpayment (if the predicted price, and hence our bid, are higher than the actual price); overpayment leads to a lower surplus. Less accurate predictions can also lead to a reduced probability of winning the auction (if the predicted price, and hence our bid, are lower than the actual price). Thus, a change in the prediction windows affects both the probability of winning as well as the average accrued surplus and it is not quite clear how it affects the overall *expected surplus.* To that end, we repeat the simulation study from Table 6 for prediction windows of different lengths. Table 7 shows the results.

We can see that larger prediction windows result in a larger average surplus which suggests that the effect of having a larger pool of candidate auctions outweighs the effect of overpayment. But we

Table 7: Tradeoff between the width of the prediction window and expected surplus. The first column denotes the width of the window; the second column denotes the probability of winning; the third column denotes the average accrued surplus; and the last column denotes the expected surplus, i.e. exp.sur = (p.win × avg.sur).

prediction window	p.win	avg.sur	exp.sur
14hrs	59.01%	$35.29	$20.82
12hrs	61.44%	$32.33	$19.80
9hrs	67.82%	$30.99	$21.02
6hrs	69.43%	$29.60	$20.55
3hrs	75.77%	$27.21	$20.62

also see that larger windows result in a smaller probability of winning since the less accurate predictions more frequently yield bids below the auction's actual closing price and hence an unsuccessful auction. Interestingly, the *expected surplus* is maximized for a prediction window of 9 hours. While our results do not prove optimality, they suggest a very interesting global optimization problem for future research.

6 Conclusions

The increasing popularity of online auctions puts more and more pressure on bidders of making informed bidding decisions in the face of competition. While classic bidding strategies such as early bidding or last-minute bidding are well-understood in the academic literature, they do not account for the component of competition originating from simultaneous auctions selling the same or similar item. Moreover, while it is unlikely that every bidder uses early or last-minute bidding in exactly the same way, to date they can only augment and adapt these strategies with gut-feeling or intuition stemming from past experience. We propose a novel automated and data-driven approach that provides bidders with valuable *objective* information about an auction's projected price in the face of competition.

Our approach consists of two main components. In the first component, we derive a novel dynamic forecasting model for price in competing auctions. We show that out model outperforms several competitor models. In the second component, we build a comprehensive bidding strategy around out forecasting model, using the idea of maximizing consumer surplus. We find that our strategy outperforms classical bidding strategies such as early bidding or last-minute bidding in terms of expected surplus accrued. We also argue that our approach simplifies the bidding process since bidding occurs immediately without the need to wait for a particular time in the auction process which adds convenience to the bidder.

There are several ways in which this research can be expanded. We have already pointed to the problem of selecting the optimal prediction window in Section 5.3.1. Another way to expand this research is via allowing for closing *and* continuing auctions. Recall that our current approach only consider auctions that close within the given prediction window. The reason is that our forecasting model is geared to the fixed time interval $(T+1)-T$ so we can only predict the final price of auctions that end within that interval. Of course, one can roll the model one additional time period forward to make predictions at $T+2$, based on the predicted values at $T+1$; however, predictions two time periods into the future (i.e. $T \rightarrow T+2$) are more uncertain than predictions only one step forward (i.e. $T \rightarrow T+1$). It is not quite clear how to discount the additional prediction uncertainty in our decision framework. Another way to expand this research is via allowing for variable and adaptive WTP distributions. In our simulations, we assume that both early and last-minute bidders have the same WTP distribution. It may be possible that bidders with different strategies also have different product valuations. Moreover, we assume that the WTP distribution remains constant over our prediction window. While this may be realistic for short windows over only several hours, a bidder that wants an item immediately may have a different valuation compared to a bidder that is willing to wait several weeks. All-in-all, there are many opportunities for future research and we hope to inspire some of it with this paper.

References

Adamowicz, W., Bhardwaj, V., and Macnab, B. (1993). Experiments on the difference between willingness to pay and willingness to accept. *Land Economics*, 69(4):416–427.

Anwar, S., McMillan, R., and Zheng, M. (2006). Bidding behavior at competing auctions: Evidence from ebay. *European Economic Review*, 50(2):307–322.

Ariely, D. and Simonson, I. (2003). Buying, bidding, playing, or competing? value accessment and decision dynamics in online auctions. *Journal of Consumer Psychology*, 13(1&2):113–123.

Bajari, P. and Hortacsu, A. (2004). Economic insights from internet auctions. *Journal of Economic Literature*, 42(2):457–486.

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):21–43.

Bapna, R., Jank, W., and Shmueli, G. (2008a). Consumer surplus in online auctions. *Information Systems Research*, 19(2):June Issue.

Bapna, R., Jank, W., and Shmueli, G. (2008b). Price formation and its dynamics in online auctions. *Decision Support Systems*, 44(3):641–656.

Berk, K. (1978). Comparing subset regression procedures. *Technometrics*, 20(1):1–6.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth.

Chan, T. Y., Kadiyali, V., and Park, Y.-H. (2007). Willingness to pay and competition in online auctions. *Journal of Marketing Research*, 44(2):324–333.

Dass, M., Jank, W., Reddy, S., Shmueli, G., and Wang, S. (2007). Dynamic price forecasts in online indian art auctions. *In the proceeding of the Third Symposium on Statistical Challenges in eCommerce Research*, May 19-20, 2007, University of Connecticut.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online reputation mechanisms. *Management Science*, 49(10):1407–1442.

George, E. (2000). The variable selection problem. *Journal of the American Statistical Association*, 95(452):1304–1308.

Haruvy, E., Katok, E., and Leszczyc, P. T. P. (2007a). Consumer optimization, switching and search in pairs of online auctions for identical items. Working Paper, University of Texas at Dallas.

Haruvy, E., Katok, E., and Leszczyc, P. T. P. (2007b). Individual choice among charity auctions. Working Paper, University of Texas at Dallas.

Haruvy, E., Popkowski Leszczyc, P., Carare, O., Cox, J., Greenleaf, E., Jap, S., Jank, W., Park, Y., and Rothkopf, M. (2008). Competition between auctions. *Marketing Letters*, Forthcoming.

Hasti, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Jank, W. and Shmueli, G. (2005). Visualizing online auctions. *Journal of Computational and Graphical Statistics*, 14(2):299–319.

Jank, W. and Shmueli, G. (2007). Modelling concurrency of events in on-line auctions via spatiotemporal semiparametric models. *Journal of the Royal Statistical Society: Series C*, 56(1):1–27.

Jank, W. and Shmueli, G. (2008). Studying heterogeneity of price evolution in ebay auctions via functional clustering. *In Adomavicius and Gupta (eds.): Handbook of Information Systems Series: Business Computing, Elsevier*, (Forthcoming).

Jank, W., Shmueli, G., and Wang, S. (2008). Modeling price dynamics in online auctions via regression trees. *In Jank and Shmueli (eds.): Statistical Methods in eCommerce Research, Wiley & Sons*, (Forthcoming).

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. (2007). Pennies from ebay: the determinants of price in online auctions. *Journal of Industrial Economics*, 55(2):223–233.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis, (Second edition)*. Springer-Verlag, New York.

Roth, A. and Ockenfels, A. (2002). Last-minutes 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., Russo, R., and Jank, W. (2007). The barista: A model for bid arrivals in online auctions. *The Annals of Applied Statistics*, 1(2):412–441.

Simonsohn, U. and Ariely, D. (2007). When rational sellers face non-rational buyers: Evidence from herding on ebay. Working Paper, Available at SSRN: http://ssrn.com/abstract=722484.

Sun, E. (2005). The effects of auctions parameters on price dispersion and bidder entry on ebay: A conditional logit analysis. Working Paper, Stanford University.

Wang, S., Jank, W., and Shmueli, G. (2008a). Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business and Economic Statistics*, 26(2):144–160.

Wang, S., Jank, W., Shmueli, G., and Smith, P. (2008b). Modeling price dynamics in ebay auctions using principal differential analysis. *Journal of American Statisticl Association*, Forthcoming.

Appendix A: Generating Smooth Functional Objects

There exist a variety of methods for recovering an underlying smooth functional object from the data. One particular method that provides good flexibility and computational efficiency is the polynomial smoothing spline.

Consider a polynomial spline of degree p:

$$f(t) = \tilde{\beta}_0 + \tilde{\beta}_1 t + \tilde{\beta}_2 t^2 + \cdots + \tilde{\beta}_p t^p + \sum_{l=1}^{L} \tilde{\beta}_{pl}[(t - \tau_l)_+]^p, \tag{5}$$

where the constants $\tau_1, \ldots, \tau_L$ are a set of L knots and $u_+ = uI_{[u \geq 0]}$ denotes the positive part of the function u. The choices of L and p strongly influence the local variability of the function f, with larger values resulting in a rougher f, exhibiting larger deviation from a straight line. While this may result in a very good data fit, a locally very variable function may not recover the underlying trend very well. One can measure the degree of departure from a straight line by defining a roughness penalty $\text{PEN}_m = \int \{D^m f(t)\}^2 dt$, where $D^m f$, $m = 1, 2, 3, \ldots$, denotes the mth derivative of the function f. For $m = 2$, for instance, PEN_2 yields the integrated squared second derivative of f which is sensitive to the curvature of the function f.

Fitting a polynomial smoothing spline to the observed data $\tilde{y}_1, \ldots, \tilde{y}_n$ involves finding the coefficients $(\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_p, \tilde{\beta}_{p1}, \ldots \tilde{\beta}_{pL})^T$ of (5) that minimize the penalized residual sum of squares

$$Q_{\lambda,m} = \lambda \times \text{PEN}_m + \sum_{t=1}^{n} \{\tilde{y}_i - f(t_i)\}^2, \tag{6}$$

where the smoothing parameter $\lambda \geq 0$ controls the trade-off between the data-fit, as measured by the summation on the right-hand side of (6), and the local variability of the function f, measured by the roughness penalty PEN_m. Minimization of the penalized residual sum of squares (6) is done in a way very similar to the minimization of the least squares operator in standard regression analysis.

Appendix B: Robustness of Last-Minute Bidding

Last-minute bidders place an incremental bid over the current high-bid and we assume in Section 5.2.2 that this increment equals 1%. In practice, this increment could be larger or smaller; it could also be that some last-minute bidders increment not by a percentage of the current price but rather by a fixed amount. Table 8 investigate that robustness of last-minute bidding to different increment strategies. We can see that the expected surplus is rather unaffected by the increment strategy. Moreover, regardless of the actual strategy chosen, the expected surplus is significantly lower than that of our automated bidding strategy in Table 7.

Table 8: Robustness of last-minute bidding to different increments.

Increment	p.win	avg.sur	exp.sur
$1	89.38%	$18.61	$16.63
$5	96.43%	$17.39	$16.77
1%	92.34%	$18.47	$17.05
5%	98.03%	$15.97	$15.65