Term Papers by Students of the Class

BMGT 808X "Applied Regression Analysis"

Prof. Wolfgang Jank

Robert H Smith School of Business
University of Maryland

Spring 2006

Authors:

- Mingfeng Lin
- Jie Mein Goh
- Eylem Koca
- Alicia Dixon
- Stacy Cuffe
- Natalia Lorinkova

Bidder Migration and Its Price Effects

Mingfeng Lin

Mingfeng@rhsmith.umd.edu
Department of Decision and Information Technologies
RH Smith School of Business
University of Maryland, College Park

Last updated: 05/14/2006

Abstract

This paper examines one type of bidder behavior on eBay auctions: bidder migration across auctions. Bidders are considered to have migrated when they lose in an earlier auction and then engage in a later auction for the same item. Several possible justifications for bidder migration Ire proposed, in particular, it is discovered that many migrating bidders are resellers – those who sell and buy the same product – and that some could be potentially shill bidders. I found that bidder migration occurs with considerable frequency in the sample I examined, and that a greater number of migrating bidders is associated with a higher level of auction price. However, interestingly, the existence of one type of migrating bidders – resellers – is found to be associated with lower auction prices.

Keywords: bidder migration, sequential auctions, resale, reseller, eBay, price

Introduction

Unlike traditional auctions, eBay enables bidders to engage in more than one auction at different points of time; unless they win, there is no cost associated with bidding on multiple auctions. For many products offered on eBay, consumers are able to attempt more than once to obtain the product he or she is trying to purchase. This in turn could lead to strategic behavior by bidders (such as trying out different auctions for the best price), and could also lead to other interesting phenomenon such as shill bidders – where bids are placed with the only purpose of pushing up the price – and resellers – who can potentially purchase from occasional low price auctions and re-sell the items on his or her own auctions.

I am interested in knowing more about what "bidder IDs" do after they lose in an auction. Note that I consider bidder IDs since it is always possible that multiple IDs correspond to one individual. More specifically, the purpose of this paper is to study the occurrence and effects of *migrating bidders*. In an auction, I consider a bidder to be "*migrating*" when he or she has lost previously on the same item (in earlier auctions). I will also call them "migrating bidders", or "migrabidders".

In sum, research questions I am attempting to address are:

- Do we observe bidder migrations on eBay?
- What are the possible explanations for bidder migrations?
- How does bidder migration affect the price outcome of auctions?

Literature

Economic theory on auctions has usually studied one-shot auctions (for a review, see Klemperer 1999). Multi-unit auctions have traditionally been limited to B2B settings, such as procurement auctions or telecommunication license auctions. In recent years there have also been some notable studies on sequential auctions (second-price, with restrictive market conditions), which appear to be quite close to migrating bidders that this paper is seeking to examine (e.g. Katzman 1999, Milgrom and Webber 2000).

Generally, studies on sequential auctions show that the price *decreases* as the auction progresses (e.g. Gail and Hausch 1994).

However, one critical difference between sequential auctions and bidder migration is that sequential auctions are intentionally spaced auctions hosted by the same seller, whereas in our study, migrating bidders are in no way restricted to bid again on a same seller – unless the bidder ID was specifically intended to do so (shill). Moreover, the above mentioned papers assume that bidders either have unit demand (Milgrom and Webber 2000) or unlimited demands (Katzman 1999). As an empirical study, we are not restricted by these assumptions.

Elmaghraby (2003) showed that different ordering of products can and does affect the outcome of auctions, but its focus is from the sellers' perspective; that is, how they should order their auctions to maximize profits or minimize costs. Our study is different since different auctions are hosted by different sellers; hence the order of auctions is not within the control of individual sellers. We will only be able to partially study the effect of auction ordering through the number of migrating bidders.

Jeitschko (1998) studied learning in sequential auctions, which suggests that bidders are in fact interacting with each other, and through results of earlier auctions, bidders will be able to infer other bidders' types (private values) and this will have an influence on their subsequent bids. It is shown that bidders unaware of informational effects place higher bids and thus have lower payoffs. In our study, bidder pools are dynamically changing over time, hence it is difficult to infer the learning process (since opponents are different). At the same time, if there is more than one migrating bidder in an auction, there could be potential mutual learning effects – if they all participated in one or more same auctions previously. We thus hypothesize that the higher number of migrating bidders, the lower the price. On the other hand, a greater number of migrating bidders also implies a higher level of competition among bidders. If this hypothesis is rejected by our data, it will indicate that the learning effect in our eBay setting is minimal, or is largely dominated by competition effects.

H1: The higher the number of migrating bidders in an auction, the lower the price (learning effect).

H1a: The higher the number of migrating bidders in an auction, the higher the price (competition effect).

Resale has also been studied in the economics literature. Generally speaking, it has been suggested that when bidders engage in an auction with an intention to resell the product, his or her willingness to pay in the current auction depends on the expectation of the secondary market, hence is posited to be increasing with the level of competition expected among resale buyers (Haile 2001). In our context, therefore, the highest bid of resellers should be positively related with the expected competition. Interestingly, this would appear to suggest that the larger number of bidders in the current auction (that the reseller is engaging in), the higher the bid he is willing to place. This would be especially true when there is more than 1 reseller, since they will be competing for the resale opportunity. This in turn would suggest that the presence of reseller would increase the price outcome of the focal auction. Unlike H1, this proposition seems to be consistent with a competition effect: the higher number of bidders (including resellers), the higher the price (H2).

H2: The higher the number of resellers in an auction, the higher the price.

Variables used in our study are mostly those suggested or confirmed by earlier studies, including starting bid, seller experience, number of bidders in an auctions, number of bids received, and so on. Since these are not the focal variables of our study, specific hypotheses are not listed for each of them.

Data

The data for this study is the bidding history of auctions on eBay, between August 10th and September 12th, 2001. Due to the huge amount of data in this period and the nature of our study, we focus on one particular product: Oakley sunglasses. As mentioned previously, in this study I define "*migrating bidders*" as those IDs who had lost in a previous auction (also for Oakley sunglasses) and subsequently placed bids in another auction for the same product. Table 1 gives summary statistics for the data set used.

Table 1: Summary Statistics

Number of auctions	884
Auctions without migrating bidders	486
Auctions with migrating bidders	398
Number of unique bidder IDs	3677
Number of auctions involving resellers	80
Maximum number of earlier losses (for a bidder)	38

Model and Variable Measurement

Table 1 gives the dependent and independent variables used in the current model.

Table 2: Model

Unit of Analysis	*Auction*	
Dependent Variable	*Log(price)*	
Independent Variables	*Seller Experience*	*Starting Bid*
	Duration of Auction	*Number of Bids*
	Bids per person	*Number of Migrating bidders*
	Number of Resellers	*Maximum number of previous losses*

Since we are focusing on one product only, we consider an ID to be a reseller if, during the time frame mentioned above, it appeared both as a seller and as a buyer (on different auctions). For each auction, we looked at the bidders who are involved, and the number of resellers is used as one of the independent variables.

Migrating bidders are not as straightforward to measure as it appeared to be. The following are the steps taken to decide if a bidder ID has been migrating.

First, for a given auction, all bids of a bidder ID are dropped except the one with his or her highest bid in the current auction.

Second, for a given auction, a bidder's maximum bid is compared with the highest bid of that auction. A dummy variable (*loss*) is created: 0 if he or she wins that auction; 1 otherwise.

Third, the dataset is sorted first by bidder ID and then by auctions' start date and time. For each bidder in each auction, a new variable is created to denote the running sum of the dummy variable *loss*.

Fourth, a new variable "number of previous losses" is created by subtracting the current value of *loss* from the running sum of *loss*. If this variable is equal to 0, then this bidder is *not* a migrating bidder for that auction. Non-zero values indicate migrating bidders.

Finally, the dataset is sorted by auction ID, and a new variable is generated to indicate the number of migrating bidder IDs in each auction. This concludes the data preparation.

Model Selection, Results, and Discussions

Based on review of prior research and motivations of this study, as well as availability of data, the following are the list of potential predictors in our model. Price is the dependent variable. Adding other variables, such as maximum bidder experience in an auction, did not significantly change the results.

- Seller Experience
- Starting Bid
- Duration
- Maximum # of Prev. Losses
- Number of resellers
- Number of Migrating Bidders
- Number of Bids
- Bids Per Person

Transformation of data

Many variables – in fact almost all those mentioned above – are highly skewed or not close enough to normal distribution. Therefore I attempted several methods to reduce skewness.

For price, several power transformations (cubic, square, log, etc) were attempted on price. However, none of them were able to retain the hypothesis of normality. I finally used

Box-Cox transformation on price to reduce skewness (L=.1539212). This further reduces the peak of distribution than logarithm (see the graph below for a side-by-side comparison between two transformations using kernel density distribution)

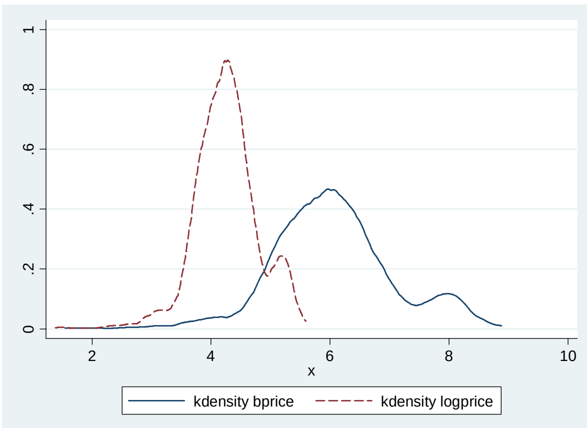


Figure 1: Comparison of Transformations

For number of bids, square root is used. Since no proper transformation could be found for starting bid and bids per person, these two variables enter the regression directly. Maximum number of losses was also transformed using Box-Cox transformation. Our focal variable, number of migrating bidders, were also transformed using Box-Cox method (after adding 1). Post-estimation analysis of residual plots also indicate that these transformations help reduce heteroskedasticity.

Resellers are either 0 or 1 in auctions so it is not transformed. Duration of auction also enters the regression without further transformation.

Table 3: OLS Regression Results

DEPENDENT VARIABLE:	Price (Box-Cox transformed)
INDEPENDENT VARIABLES	COEFFICIENTS
Seller Experience	-.00002 (.00002)
Starting Bid	.0223*** (.0009)
Duration	-.0137 (.0126)
Maximum # of Prev. Losses (B.C)	-.1029** (.0439)
Number of resellers	-.1217

	(.088)
Number of Migrating Bidders (Square root)	.2098*** (.0701)
Number of Bids (square root)	.6283*** (.0323)
Bids Per Person	-.0281* (.0161)
Adjusted R-square	46.69%
Number of Observations	884

***: p<0.01; **: p<0.05; *: p<0.1. Values in parentheses are standard errors.

Removing the three variables with non-significant coefficients does not significantly improve R-square; therefore we keep them in the regression. This model has an R-square of over 46% per cent with an observation number of 884. Starting bid, number of migration bidders (square root), number of bids (square root) and bids per person are all significant on various levels. Pair-wise Scatterplot of all variables in the regression is shown below.

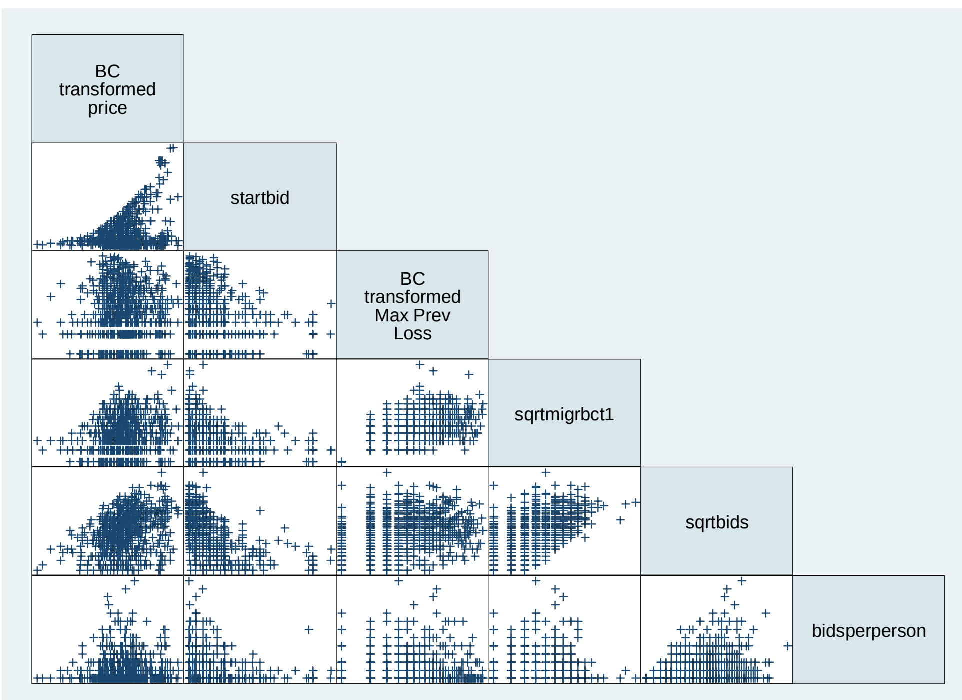


Figure 2: Scatter Plots

Post estimation diagnosis

1. Multicollinearity

VIF is used to detect multicollinearity and the result is show below. It indicates that multicollinearity is not a significant concern in our model.

Table 4: Analysis of Multicollinearity

Variable	VIF	1/VIF
sqrtmigrbct1	2.60	0.384479
sqrtbids	1.96	0.508977
bmaxpr1	1.94	0.514243
startbid	1.47	0.679264
bidsperper~n	1.09	0.920533
sellerexp	1.08	0.927392
duration	1.06	0.944286
resellerco~t	1.02	0.985109
Mean VIF	1.53	

2. Heteroskedasticity

I used residue-versus-fitted value plot to detect multicollinearity. From the graph below, it appears that there is still certain amount of heteroskedasticity; but this is much better than what it was prior to the above transformations. The residuals appear to be reasonably symmetric above and below the zero line, though dispersion seems much smaller at higher values of fitted price.

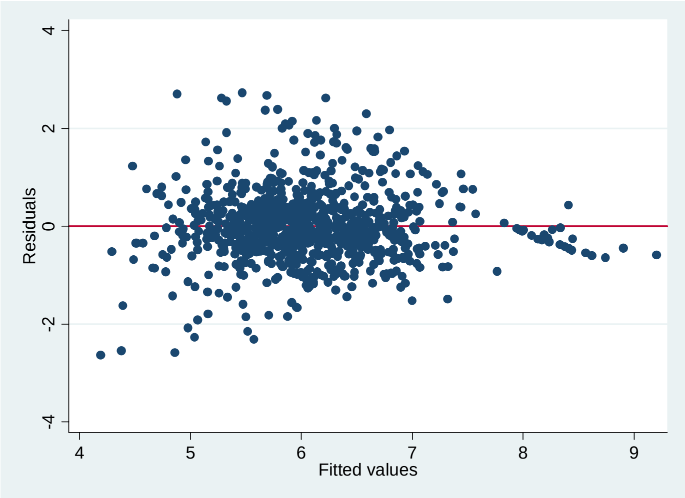


Figure 3: Residual vs. Fitted Value

3. Missing variable test

A missing variable (omitted variable) test shows that we do have the problem of missing value, which is consistent with the R-square of below 50%. However, I was unable to identify more predictors from the dataset that can help improve on this.

4. Normality

Predicted values of residual were used to test for normality. Although the histogram of r appears to be reasonably close to normality, further diagnostic plots seem to suggest that the residuals are not normality distributed. Q-norm plot indicates that the distribution diverges from normality especially at the tails (with especially high or low prices). Numerical normality test (e.g. Swilk or skewness-kurtosis tests) also rejects null hypotheses of normality.

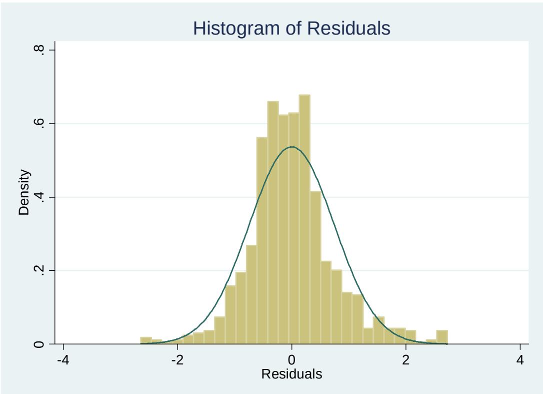


Figure 4

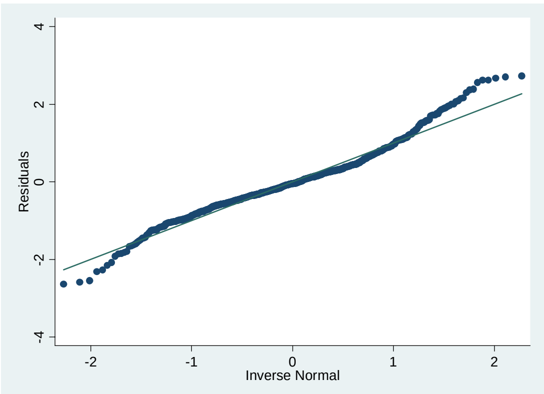
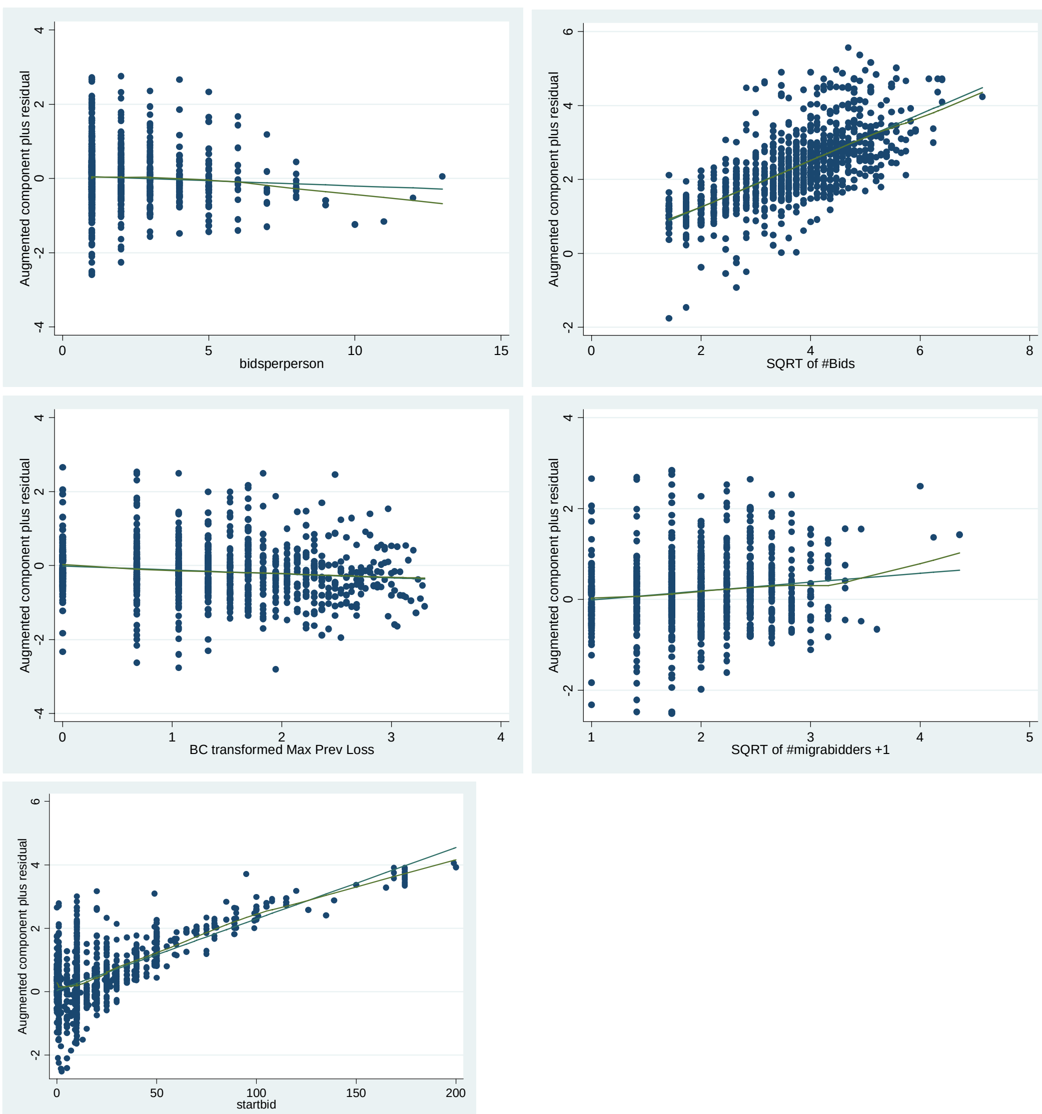


Figure 5

5. Linearity

Augmented component-plus-residual plot of all independent variables of the model shows that most of them, especially those statistically significant, do not deviate from linearity too much, except at extreme values of price. This is due to the fact that they have already been transformed prior to the regression. It also indicates that the transformation were valid. The following are the acprplots for the five significant regressors.



6. Influential and observations and outliers

I used leverage-versus-squared-residuals plot to detect potentially influential observations. One auction turns out to be very different from the rest of the dataset:

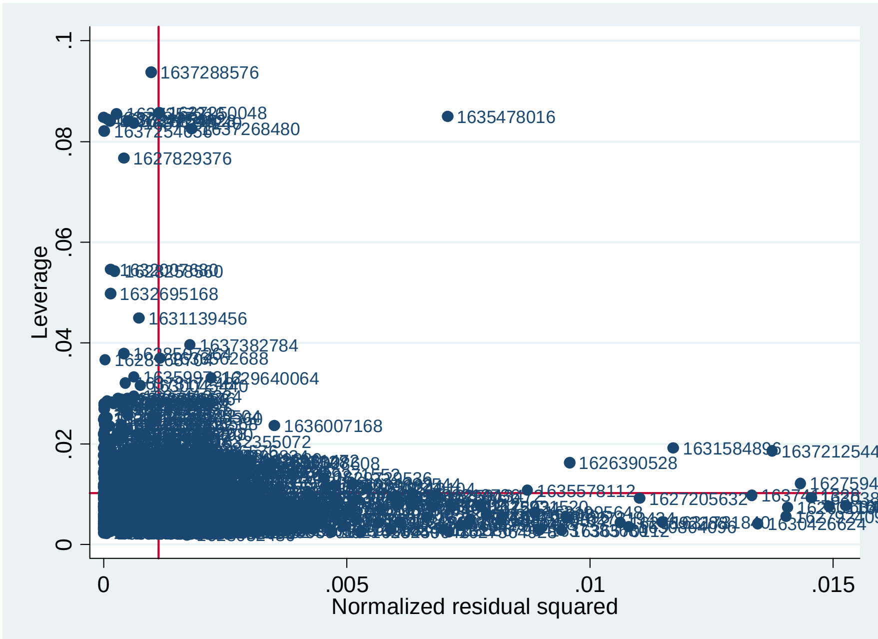


Figure 6: LVR2 plot

Since there appears to be only one auction (auction id = 1635478016), I further deleted it from the regression and redid the regression. It turned out that due to the relatively large sample size, regression results were not significantly different. Improvement of R-square is at about 0.2%, and for all variables, the coefficients' signs as well as their significance level were the same as before. Hence we will continue our discussion below using the results from table 3.

Possible Complications

One possible concern regarding the variables in this regression is the truncated nature of data. One solution is to use auctions that occur later than, say 7 days after the start date of this dataset. This reduces our data size, but it does not significantly alter our results. The revised results are indicated below.

Table 5: Revised Regression Results

DEPENDENT VARIABLE:	Price (Box-Cox transformed)
INDEPENDENT VARIABLES	COEFFICIENTS
Seller Experience	-.00002 (.00002)

Starting Bid	.0223*** (.0009)
Duration	-.0045 (.0154)
Maximum # of Prev. Losses (B.C)	-.1154** (.0493)
Number of resellers	-.1895 * (.0979)
Number of Migrating Bidders (Square root)	.2653*** (.0805)
Number of Bids (square root)	.6177*** (.0371)
Bids Per Person	-.0271 (.018)
Adjusted R-square	47.11%
Number of Observations	658

This model slightly increases R-square. Our focal variable, number of migrating bidders, are not affected. Number of resellers becomes significant, while the number of bids becomes insignificant.

Discussions

Migrating bidders

Migrating bidders has not yet attracted much attention in the academia, although this seem to be a common place practice among bidders. Using the "conservative" sample, that is, leaving out the first 7 days of our dataset, there are 48% of auctions involving migrating bidders. As mentioned earlier, migrating bidders are theoretically different from those bidders in a sequential auction setting. Hence it is important and worthwhile to further investigate their impact on the auction outcomes.

The above analyses show that as we expected, number of migrating bidders indeed have a statistically significant impact on the price outcome of an auction: the larger the number of migrating bidders, the higher the price of an auctions, ceteris paribus. However, due to the transformation used, it is difficult to numerically express this relationship. It appears that the presence of migrating bidders increases competition for an item, and leads to a higher price outcome. This is consistent with our hypothesis H1a, which indicates that the competition effect seems to dominate.

What are the possible motivations for bidders to migrate from an auction to another? Here I identify some of the possible reasons; due to limitation of paper length, further investigation is left for future research.

1. Unfulfilled demand. Migrating bidders could be those who lost due to technical reasons – e.g. they lost because a competing bidder used "sniping tools" to put in a higher bid at the last minute.
2. Migrating bidders could be resellers. These are the people who arbitrage among auctions, buying at lower price and sell later. Further discussion of this possibility follows.
3. Shill bidders. Although shill bidding is strictly forbidden according to eBay policy, there is still strong incentive for sellers to do so. Although there is still no universally accepted criteria about what constitutes shill bidding, some suspicious activities in our sample include:
    a. Some bidder IDs lose as much as 38 times without winning.
    b. Many migrating bidders have very low experience value, even zero.

Resellers

Reseller is an interesting issue in our dataset as well as general eBay research. Anecdotally these are the people who make a living on eBay. Our literature review indicates that there have been quite some studies on the potential effects of resale motive on auction outcomes. H2 proposes that larger number of resellers would lead to higher prices. However, no auction is observed that has more than 1 reseller. This might indicate that resellers are avoiding each other in auctions, but we leave this for future studies.

"Number of resellers" in our earlier regression only appears to be marginally significant, with a negative coefficient. To investigate further, and since we only have either zero or 1 reseller in auctions, I conducted simple hypotheses testing on sub-samples of zero resellers and 1 reseller (the maximum number of resellers in our dataset is 1). It is shown that:

1. Difference of mean and standard deviation of these two groups are both statistically different.
2. Presence of resellers is associated with lower standard deviation of price levels; and
3. Presence of resellers is associated with lower prices, on the average.

Therefore we reject H2. Note that we are not inferring causality: unless there is collusion among sellers (which certainly merits further investigation), resellers themselves will not be able to affect the price levels since they are competing with other bidders, who apparently do not have a resale motive.

It is possible that, resellers are indeed sophisticated enough to identify opportunities where they can secure this item at a lower price. A scatter plot between seller experience and the number of resellers indicate that, resellers appear to be more willing to participate in novice-sellers auctions – by taking greater risk, they are able to find a lower price.

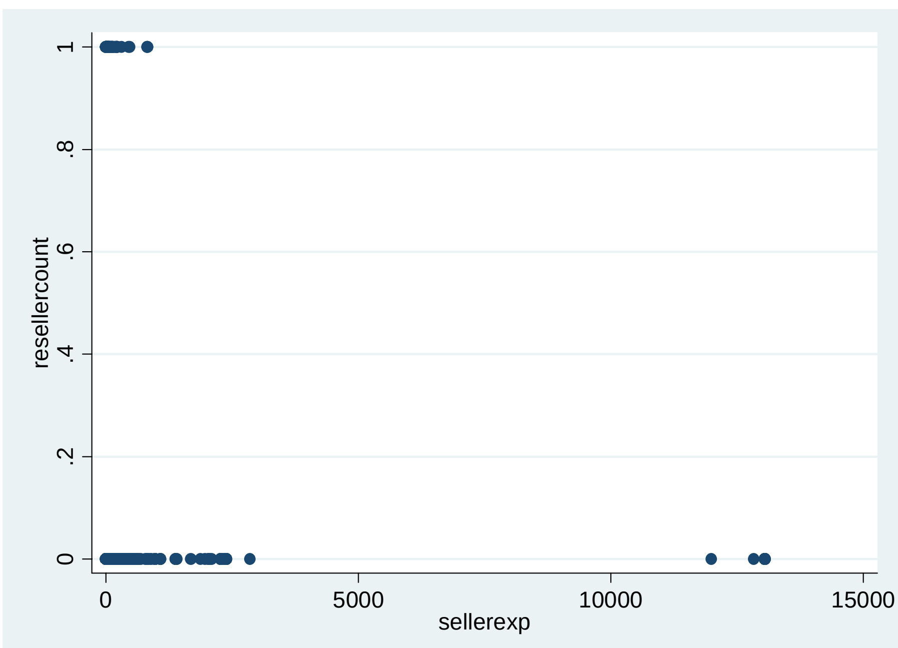


Figure 7: Number of resellers vs. Seller Experience

Potentially and hopefully, they will be able to sell these items later at a higher price so that their risk-taking can be awarded. However, since we do not have observations on shipping and handling costs (both when they are buying and when they are selling), it is unidentifiable whether these resellers were able to sell these items at a higher price later, but this would be an interesting topic to further explore.

Limitations, conclusions, and future research

There are two major limitations to our study. First, in order to make the data more tractable our study only used one product in the analysis. It would be of much interest to generalize the above results to other products.

Second, our dataset did not indicate the shipping and handling costs of auctions, we are not able to fully capture the difference across auctions. Given a more complete dataset, we would be able to further investigate this issue and validate our results in this study.

As indicated above, the resale motive would have interesting effects on the outcomes of auctions. Also due to the limitation of our dataset, this has not yet been investigated sufficiently but it would be an interesting topic to examine.

We are able to answer several of the questions stated at the beginning of this paper. We found that bidders do migrate across auctions, and that this increases the price outcome of later auctions. We find empirical support for H1a (competition effect) over H1 (learning effect). We also identified that resellers tend to participate in auctions hosted by novice sellers, and that these auctions tend to have lower price levels (in support of H2). This indicates that they seem willing to take risks to and were able to identify lower-cost opportunities. It would be a natural extension of the current study to identify whether these resellers were able to find higher prices in later auctions, which depends on data availability, especially shipping and handling costs.

Reference

Elmaghraby, Wedad 2003, The Importance of Ordering in Sequential Auctions, Management Science, Vol 49, No. 5, pp 673-682

Gail, I.L. and Hausch, D.B. 1994, Bottom-fishing and declining prices in sequential auctions, Games and Economic Behavior, Vol 7, pp 318-331

Haile,Philip A. 2001, Auctions with Resale Markets: An application to US forest service timber sales, American Economic Review, Vol 91, No. 3 (June 200), pp. 399 - 427

Jeitschko, Thomas D. 1998, Learning in Sequential Auctions, Southern Economic Journal, Vol 65, pp

Katzman, B., 1999, A two stage sequential auction with multi-unit demands, Journal of Economic Theory, Vol 86, no.1, pp 77-99

Klemperer, P. 1999, Auction Theory: A guide to the literature. Journal of Economic Surveys, Vol 13 No 3, pp 227-286

Milgrom P, R. Webber 2000, A theory of auctions and competitive bidding II. P. Klemperer, ed. The Economic Theory of Auctions, Edward Elgar Publishing, Inc, Cheltenham, UK

BMGT808X Research Paper

Professor Wolfgang Jank

Exploring Spacing as a Seller Strategy on E-Bay

Jie Mein Goh

jgoh@rhsmith.umd.edu

Exploring Spacing as a Seller Strategy on E-Bay

Introduction

With the advent of online auction like e-Bay, anyone can become a seller or buyer on the Internet. The growing number of transactions carried out on eBay has grown tremendously, propelling the net sales to reach $1.2 billion in the year 2003. The large number of transactions and revenues derived from eBay has also lured firms to sell their products on eBay. With increasing competition among the sellers, one question that arises is what strategy can they adopt to maximize their profit. Knowing which strategy affects the final bid value can help sellers on eBay to maximize their profit. Thus, this paper attempts to fill this gap by examining one form of seller strategy: spacing across similar auctions.

Research has explored bidders' characteristics and their effects on the final bid value. Other studies have examined the impact of auction design on revenues (Riley & Samuelson, 1981; Peters and Severinov, 2002). Studies that examine buyers' characteristics explored bidder reputation on auction outcomes (Wooders 2000). The paper by Anderson (2004) analyzes the seller characteristics (frequency of selling, reputation, qualities of the product) and options (length of auction, information provided about the product, stating the price, and whether the "buy it now" option is used). Resnick (2003) found seller reputation to have a positive impact on auction outcome. Additionally, it is valuable for sellers on online auctions to know whether strategies can be employed to maximize their profit. Although there are many sellers on eBay that are repeat sellers, their paper did not specifically address issues related to repeat sellers. To the best of our knowledge, previous literature focusing from the perspective of the repeat sellers and their strategies is lacking. One possible strategy for a repeat seller is to employ spacing across similar auctions. For instance, a repeat seller can space out his auctions such that similar items on sale do not overlap.

Previous literature has not specifically examined spacing as a seller strategy and this paper attempts to examine the feasibility of this strategy. For repeat sellers, a very important question has not been answered by previous literature: Does spacing have any impact on the final bid value? Therefore, this paper attempts to fill this lacuna by exploring the impact of spacing as a strategy by repeat sellers on e-Bay. Specifically, we would like to answer the following research questions:

1. Should repeat sellers employ spacing between two auctions?
2. What other factors affect the final bid value?

Theoretical & Hypotheses Development

Linking spacing and profits

Auctions selling the same items during the same period is likely to corrode the sellers' profits. This is because, buyers of an item has far more auctions to attempt to get the item that they want. This will result in a lower final bid value as buyers may perceive the flood of items as large supply thereby suppressing their desire to increase their bid prices. By spacing out their similar auctions, buyers perceive the supply of that item to be "limited" and thus will attempt to bid higher to obtain their desired item. Thus, I posit that,

> H1: For a repeat seller, spacing similar auctions across different days increase the final bid value.

Data Description

Bidding data (Bidder1.xls and Bidder2.xls) from eBay consisting of auctions of various categories was used in this study. From these data, I extracted the relevant subset containing only repeat sellers. A program was written for this purpose. This resulted in a total of 6164 data points. There are 10079 unique auctions and this indicates that that the 61% of auctions in the data set are repeat auctions, that is, auctions selling the same item by the same seller. Of these repeat auctions, 33% of these auctions had spacing between itself and other similar auctions. The total number of repeat sellers in the data set is 5100 and the number of repeat sellers who adopt spacing as a strategy is 1244. These figures indicate that only 24% of the total sellers in the data set are repeat sellers. 905 (72.7%) repeat sellers in this data set employed spacing as a strategy.

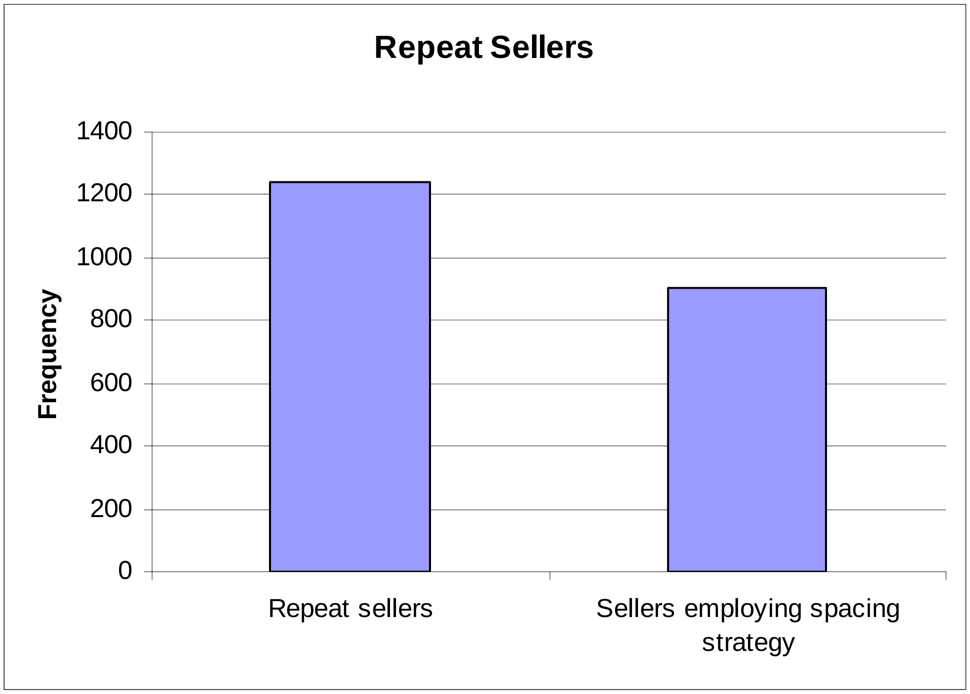


Figure 1: Sellers employing spacing as strategy

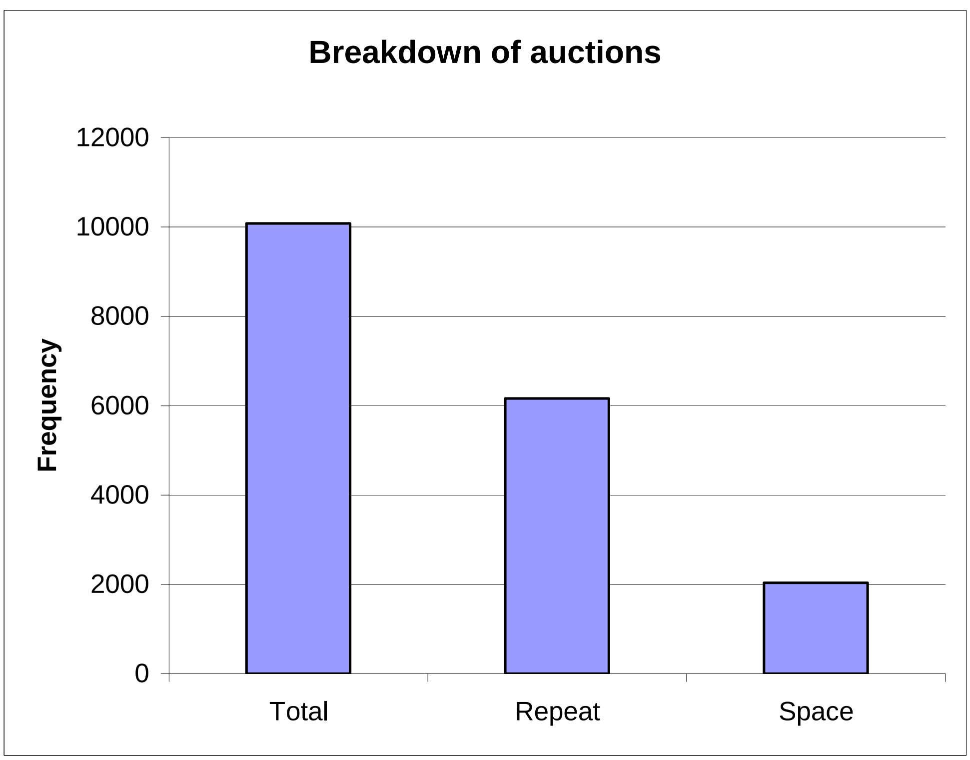


Figure 2: Breakdown of Auction Types

Data Manipulation

The variables used in our model are described in turn:

Dependent variable

Final bid value (FINALBID) refers to the resulting bid value of the auction item. This is obtained by getting the second highest bid of an auction.

Independent Variables

Space (SPACE) is a binary variable (0 or 1) that indicates whether a repeat seller employ spacing across days across all his auctions selling similar item. The notion of space refers to the existence of non-overlapping regions of auctions selling similar items by the same seller on the sae date. To illustrate the concept of space, see Figure 3. For the example shown, auction 1 will be coded as "1", that is have space, because there is a non-overlapping region. Similarly, auction 2 and 4 will be coded as "1". Auction 3 does not have any non-overlapping regions, thus, it is coded as "0".

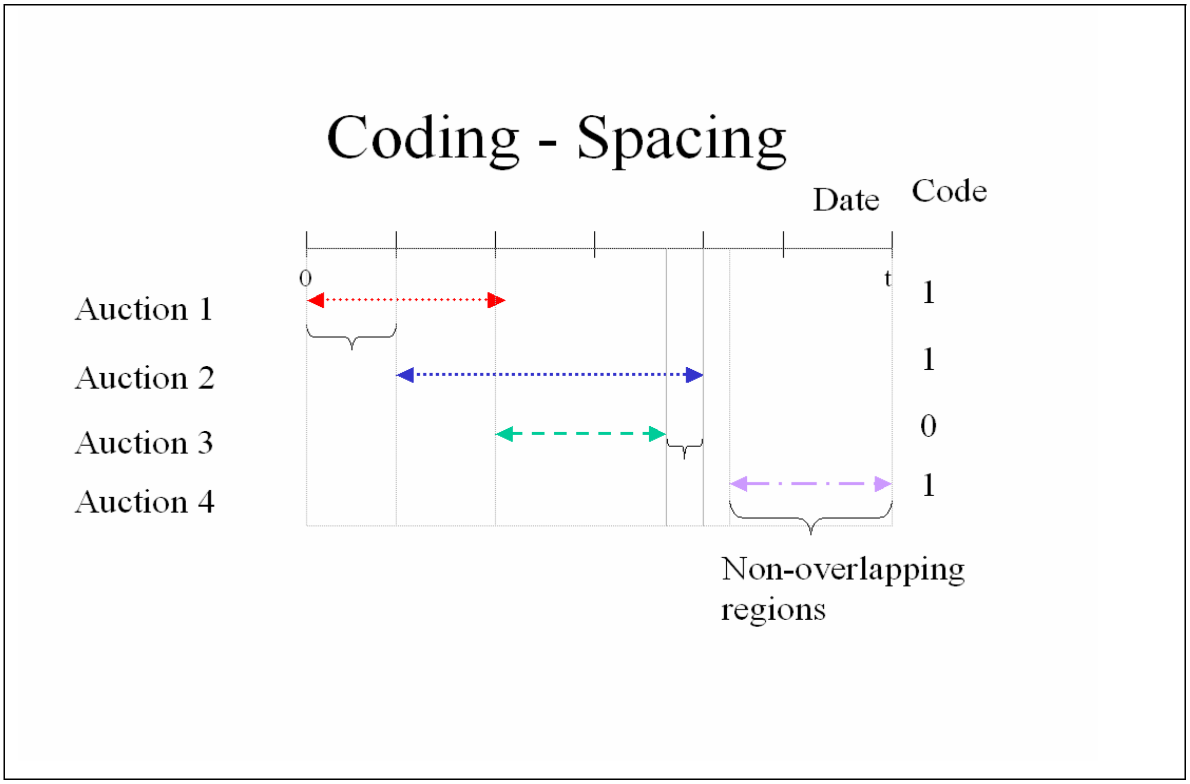


Figure 3: Illustration of Spacing

Seller experience (SELLEREXP) refers to the feedback received by other buyers. This was transformed using a log function.

Total number of bids for that auction (BIDS) was tracked as another independent variable. This was transformed using a log function.

Duration of auction (DURATION) refers to the number of days the auction is held.

Starting bid (STARTBID) refers to the starting bid for the auction.

Pairwise correlation is ran and it appeared that log(bids) and log(startbid) is highly correlated with the dependent variable.

	space	log(sellerexp+3)	log(bids)	duration	log(startbid)	log(finalbid)
space	1					
log(sellerexp+3)	-0.0109	1				
log(bids)	-0.0107	-0.0743	1			
duration	0.16307	0.09636	0.0668	1		
log(startbid)	0.07572	-0.0322	-0.2891	-0.0668	1	
log(finalbid)	0.08421	-0.0802	**0.47191**	0.08757	**0.27258483**	1

Table 1: Pairwise correlation table.

Methodology

The following linear regression models were tested:

Model 1

$$\text{Log(FINALBID)} = \alpha + \beta_1\text{SPACE} + \beta_2\text{Log(BIDS)} + \beta_3\text{DURATION} + \beta_4\text{Log(STARTBID)}$$

Model 2

$$\text{Log(FINALBID)} = \alpha + \beta_1\text{SPACE} + \beta_2\text{Log(SELLEREXP+3)} + \beta_3\text{Log(BIDS)} + \beta_4\text{DURATION} + \beta_5\text{Log(STARTBID)}$$

Model 3

$$\text{Log(FINALBID)} = \alpha + \beta_1\text{SPACE} + \beta_2\text{Log(SELLEREXP+3)} + \beta_3\text{Log(BIDS)} + \beta_4\text{DURATION} + \beta_5\text{Log(STARTBID)} + \beta_6\text{SPACE} * \text{Log(SELLEREXP+3)}$$

Analysis and Discussion

Based on the three models described above, linear regression models were tested. Table 2 shows the comparison of key metrics for model fit across these models. Based on the adjusted R-squared values, models 2 and 3 are a better fit compared to model 1. Model 3 is slightly better in terms of GCV. However, since model 3 has a larger number of terms, Model 2 is chosen over model 3 because of parsimony.

Model	Adjusted R-Sq	GCV	p-value
1	0.408	1.083	<2.2e-16
2	0.414	1.073	<2.2e-16
3	0.414	1.072	<2.2e-16

Table 2: Model fit

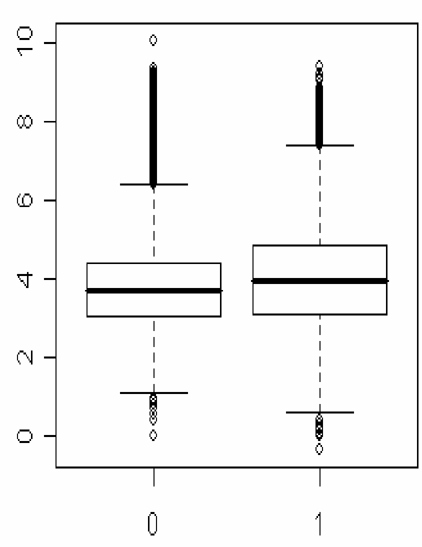


Figure 4: Boxplot showing auctions with space and those without spacing

From our results (Table 3), it is clear that the hypothesis is supported ($p<0.005$). The coefficient for space suggests that there is a positive relationship between spacing and final bid price and that using spacing will increase the log of final bid value by 0.128.

This implies that spacing across dates is a strategic move for repeat sellers and will help repeat sellers generate a higher final bid price. Therefore the results may also imply that

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.808795	0.071772	11.269	< 2e-16	***
space	0.12833	0.028543	4.496	7.05E-06	***
log(sellerexp + 3)	-0.02043	0.007108	-2.874	0.00407	**
log(bids)	1.120367	0.019325	57.976	< 2e-16	***
duration	0.047689	0.006517	7.318	2.84E-13	***
log(startbid)	0.304359	0.007013	43.399	< 2e-16	***

Table 3: Results

it is beneficial for a seller to limit the number of similar items that goes on sale together. Auctions that occur together have a reduced profit margin for each auction.

Other independent variables explored such as bids, duration and start bid all had a significant and positive relationship with the final bid value. The number of bids indicate the demand for the auction item thus it is quite intuitive that as the number of bids get higher, the bidding become more intensive and thus the final bid value will be higher. The effect for this variable is highest compared to all the other variables. For duration, the results seem to suggest that the longer the auction, the higher the final bid value. Start bid has a strong correlation with the final bid since the higher the start bid thus it is obvious that it has a significant and positive relationship with the final bid value.

However, the result for seller experience was quite unexpected. I was expecting that greater seller experience result in higher bid value because sellers who are more experienced will try their best to maximize profits and thus increase the final bid value for their auctions. However, the results suggest that seller experience is negatively related with the final bid value. This effect is less than the other variables that were studied. One possible reason could be that the transformed data was still skewed, however, upon close examination of the data, seller experience showed a normal distribution after transformation (Figure 5). Thus, this reason was ruled out. Although the results were surprising, there is evidence from prior literature that showed similar results. Previous work noted that eBay sellers are composed of a portion of people who want to get rid of their unwanted items. Therefore, this could lead to very low prices since sellers are willing to sell “junk” items at very low prices.

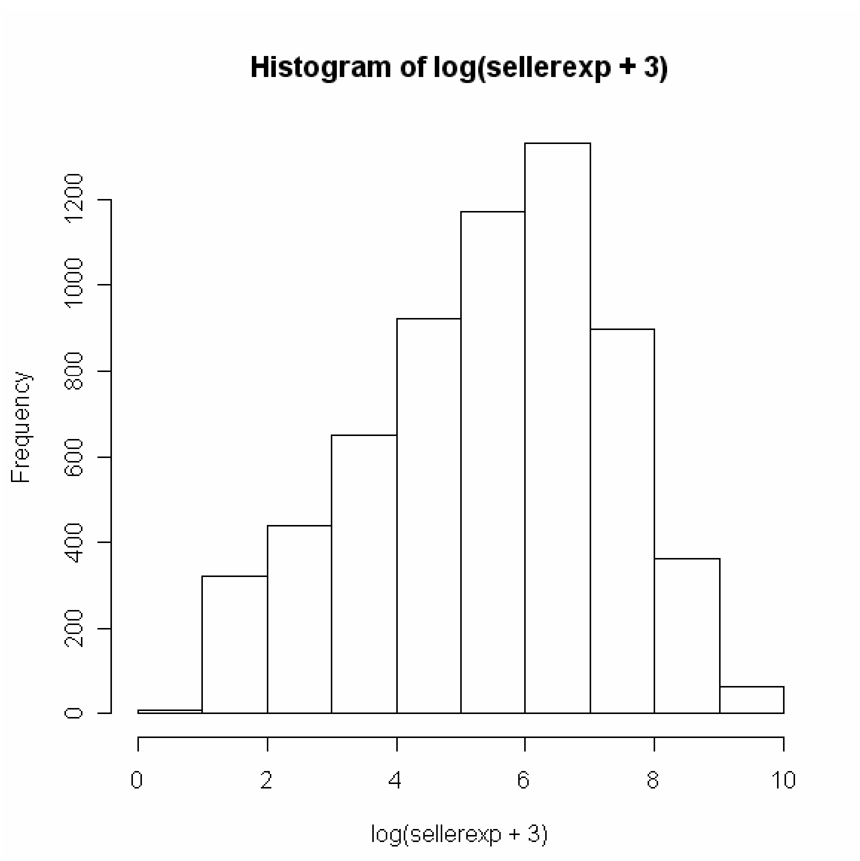


Figure 5: Distribution of log(sellerexp)

Future Work

The problem of spacing as a seller strategy can be viewed from many angles. Figure 6 shows the different ways of analyzing the strategy. For a seller, he can create an auction by considering similar products (e.g., category) that he is selling (ii), other sellers who sell similar items (iii), or sellers selling similar products (iv).

	Product	Category
Individual Seller	(i)	(ii)
Across sellers	(iii)	(iv)

Figure 6: Future work

Another important question that needs to be further investigated is the spacing between the auctions. To put it simply, how should a seller space out his auctions? The findings from such a study can help sellers understand how best to maximize their profit when they employ spacing as their strategy. Lastly, since eBay is made up of many different types of sellers which range from individuals to big firms, it is interesting to find out whether this strategy is valid for all the different types of sellers. Our current dataset does not permit such an analysis.

Conclusion

This paper addresses the question that has not been discussed in previous studies: Is spacing a strategic move for a repeat seller? Three models, which link the impact of spacing on final bid price, were tested. Findings from the analysis supported the hypothesis. The results from this study have a practical implication for repeat sellers on eBay, that is, spacing is a feasible strategy to maximize profits on online auctions. This paper also provides starting ground for further work for sellers who want to know how best to implement the spacing strategy.

References

Anderson, Steven, Friedman, Daniel, Milam, Garrett, and Singh Nirvikar, (2004), “Seller Strategies on eBay,” UCSC, Working paper.

Peters, Michael and Sergei Severinov, (2002), "Internet Auctions with Many Traders" University of Toronto Working Paper, http://www.economics.utoronto.ca/peters/papers/reserve_prices_4.pdf

Houser, Daniel, and John Wooders, (2000), “Reputation in Auctions: Theory, and Evidence from eBay,” http://bpa.arizona.edu/~jwooders/ebay.pdf

Riley, John and William F Samuelson, (1981), "Optimal Auctions," American Economic Review, 71 381-392.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood, (2003), “The Value of Reputation on eBay: A Controlled Experiment,” http://www.si.umich.edu/~presnick/papers/postcards/.

An Investigation of Regression Tools for Dynamic Price Predictions in Online Auctions

EYLEM KOCA

BMGT808X Project

May 15, 2006

Abstract: Accurately predicting the final price for an ongoing online auction is of high interest from both the sellers' and bidders' perspective. In this study, we are examining the potential of regression tools to predict the final prices during eBay auctions. Specifically, we are defining new variables as a proxy for price speed and acceleration and examining the significance of these price dynamics parameters in final price prediction using parametric, semi-parametric and non-parametric regression models. In our analyses, we find that incorporation of price dynamics information does not yield significant marginal improvement to the prediction performance of our models.

Keywords: *Online auctions, final price prediction, price dynamics, non-parametric regression, semi-parametric regression*

Table of Contents

List of Figures

List of Tables

1. Introduction

In the recent years, online auctions have received increasing amount of attention from researchers, inline with the fact that the public interest in electronic commerce and online auctions has been growing rapidly. Most of the research however has been done in the economics, marketing or information systems literature [1, 5, 6, 7], and little work has been done that investigates the potential of statistical tools in the field. Moreover, surprisingly enough, only a few studies look at the prediction of final prices for an online auction, which can present a highly beneficial tool for all parties in the market [3, 8]. Such a tool can be used to sort the auctions in which the same item is sold, so that the bidders can choose the minimum expected price auction. Additional services that answer “what if I bid this amount now?” may be offered to bidders. Another profitable service for an auction house could be including an online “insure-it-now” option that guarantees a minimum final price to the seller [8].

Ghani and Simmons (2004) use data-mining methods and predict end-of-auction prices statically, i.e., at the start of an auction. Their approach cannot incorporate online information and thus cannot yield the aforementioned benefits. Wang et al. (2006) use functional models that allow the incorporation of price dynamics to predict final prices dynamically. In this study, we investigate the potential of parametric, semi-parametric and non-parametric regression models to accurately predict the final prices during an online auction. We also question whether price dynamics parameters such as price speed and acceleration are significant determinants in such prediction tools.

The rest of the paper is organized as follows. In Section 2, we describe the available data and the steps taken to preprocess it for our analyses. Section 3 presents the regression models developed in this study. Section 4 describes the final price prediction studies and Section 5 makes concluding remarks.

2. Data Preprocessing

We use the eBay online auction data collected between July 29, 2001 and February 13, 2002 for a variety of products. This data includes information at the bid level (bid amount, bid date, bidder experience) as well as auction level information (product id, starting bid, winning bid, duration, seller experience). We divide this dataset in two: we use the "training data" to estimate our model parameters and test the prediction performance of our models on the "test data". The eBay auctions are generally second-price auctions, that is, the owner of the highest bid wins but pays the second highest bid plus a minimum bid increment. However, since the data did not include the bid increment information we used the current highest bid as a proxy for the current price. As a result, we omit the bids that do not increase the current highest price. This should not affect the results of our analyses because we do so for both training and test datasets.

Our objective is to develop regression tools that can use the live (bid level) information to predict the end-of-auction price. The question is: given the aforementioned data current as of prediction epoch, how much will the price increase over the current price? However, by a quick inspection of the data we observe that prices across auctions are highly variable: final prices range from \$0.75 to \$24,500. Therefore, we change the above question to address the *percent increase in the price over the current price* (RI) and define our dependent variable accordingly.

Note that we treat each bid point during an auction as a "prediction epoch". This implicitly assumes that the given the current information, bid point are independent from each other. Such an assumption can be justified despite the fact an auction will be represented by multiple observations: given the current state, another prediction is done at each bid point by looking only forward (to the end-of-auction), without using the information for subsequent bid points.

As a proxy for price dynamics parameters we define the following predictor variables. Average relative speed of price (AS) is the percentage increase of price since the start-of-auction until the prediction point divided by the time since the start-of-auction. Similarly, the current relative speed of price (CS) is the percentage increase in price *since the previous bid*. In addition to these, we include a lagged version of CS, CSlag. Note that the difference between CSlag and CS is a proxy for the price acceleration. Note further that, these price dynamics variables are not meaningful for the first and last bids during an auction, and that the second bids have to be deleted due to the incorporation of CSlag. Therefore, using the notation in Figure 1, we define:

$$AS=\frac{B-B_0}{B_0}\frac{1}{Now}$$

$$CS=\frac{B-B_2}{B_2}\frac{1}{Now-Prev_2}$$

$$CSlag=\frac{B_2-B_1}{B_1}\frac{1}{Prev_2-Prev_1}$$

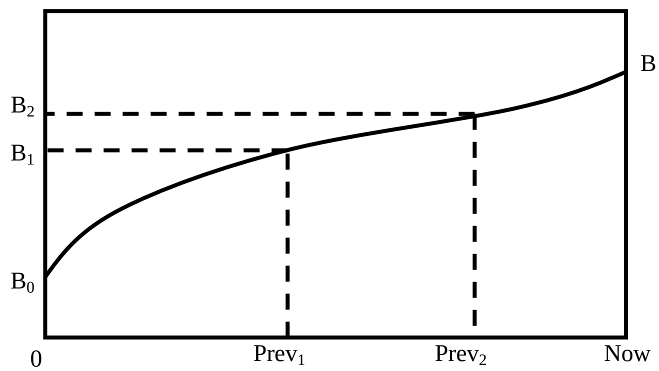


Figure 1 A depiction of price dynamics during an auction.

Typically duration of online auctions in eBay is 3, 5, 7, or 10 days. Besides their length, price patterns may display different behavior in these categories. Thus, dummy variables are devised: three dummies for 5, 7, and 10 days (D_i). The same argument is valid for different product categories: there are 15 product ID's and 14 dummies (P_j) are used to represent these. Other predictor variables that we employ in our analysis include starting bid (SB), seller experience (SE), average bidder experience until prediction point (ABE), and time left until the end-of-auction (TL).

Further investigation of the data yields the fact that all of our continuous variables (dependent and predictor) are highly skewed and an appropriate log transformation is applied to all. Figure 2 shows the histograms for the dependent variable RI and TL. It is seen that a log transformation on RI solves the skewness issue, which was the case for all other variables. However, distribution of TL shows a different pattern and the log transformation does not yield a desired distribution. Still, it is preferred to the original variable.

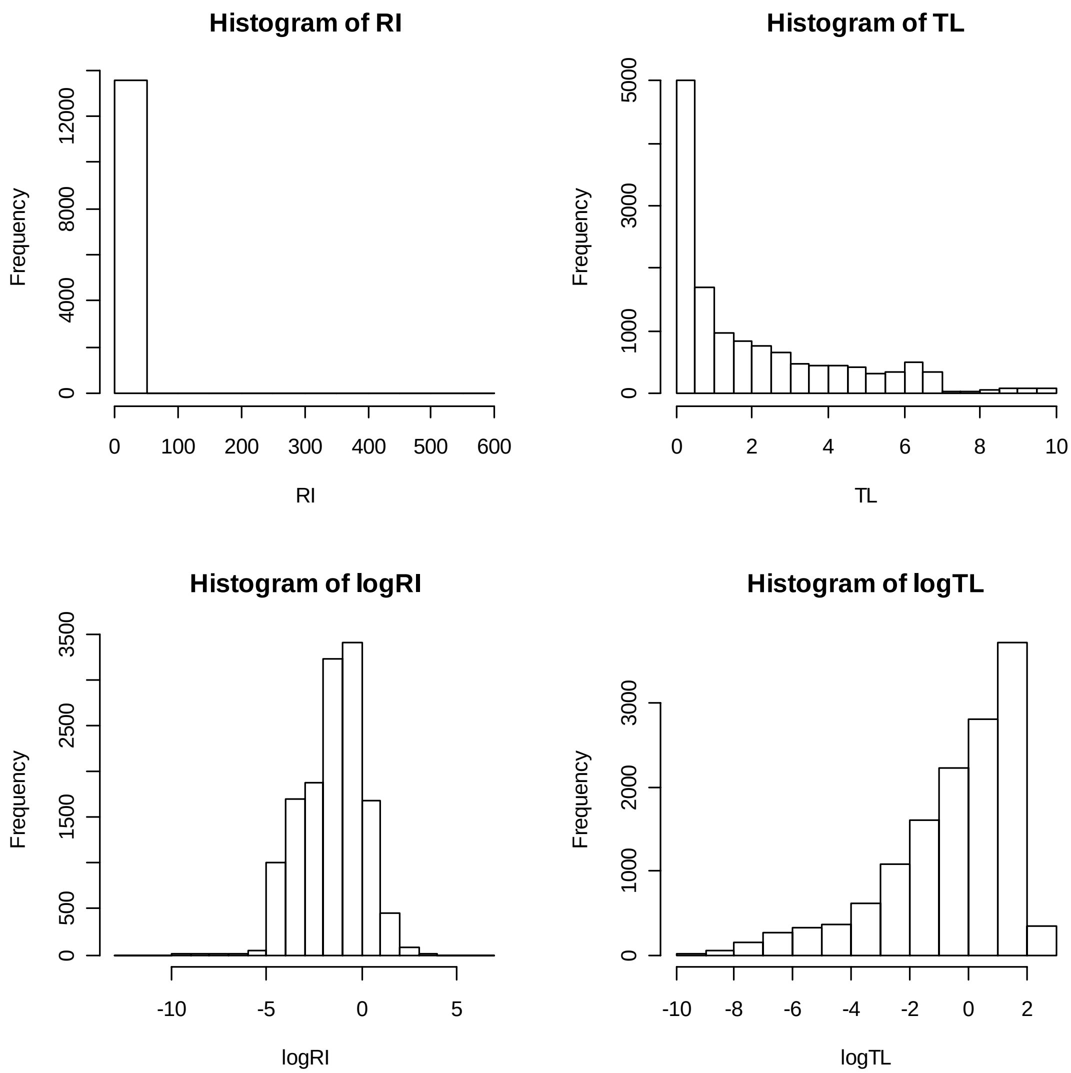


Figure 2 Histograms for RI, TL, and their log transforms.

The initial cumulative dataset included 10078 auctions and 99898 bid level observations. However, during the data preprocessing phase almost three fourths of this data has been deleted: the training dataset includes 3244 auctions and 13539 bid level observations, while the test dataset includes 3179 auctions and 13532 bid level observations. On the other hand, we analyze another price prediction scheme that predicts prices only on the final day during the auction. This analysis will provide us with more insights about the prediction performance of our models. For this last-day prediction scheme, the test dataset includes 2399 auctions and 6367 bid level observations. In the following, we present our regression models.

3. Regression Models

Before describing our regression models we give

Figure 3 to help visualize the data and the underlying relations. Note that logX denotes a log transformed X variable. It is seen from these scatterplots that logTL is the only variable that has an apparent correlation with the dependent variable (logRI). Table 1 confirms our observations and also indicates that logSB is negatively correlated with logAS which is positively correlated with logTL.

Table 1 Correlation coefficients between the variables of interest.

	logRI	logTL	logAS	logCS	logCSlag	logSB	logSE
logTL	0.51						
logAS	0.19	0.30					
logCS	0.04	-0.12	0.10				
logCSlag	0.15	0.07	0.23	0.14			
logSB	-0.15	-0.11	-0.74	-0.04	-0.12		
logSE	-0.02	-0.06	0.03	0.00	0.00	-0.09	
logABE	-0.06	-0.07	-0.03	-0.05	-0.05	0.02	0.21

3.1. The Parametric Model

The parametric model (Model 1) is constructed simply by adding all main effects, dummy variables and all possible paired interaction terms between the continuous variables. Although, our aim is to devise a prediction tool and interpretability of the model is not a primary goal, we drop the insignificant interaction terms from the model. The regression output for the resulting model is given in Table 2. We see that only a few product ID dummies are insignificant. The adjusted R-squared value (0.3521) and the F-statistic for the model (189.6) are satisfactory.

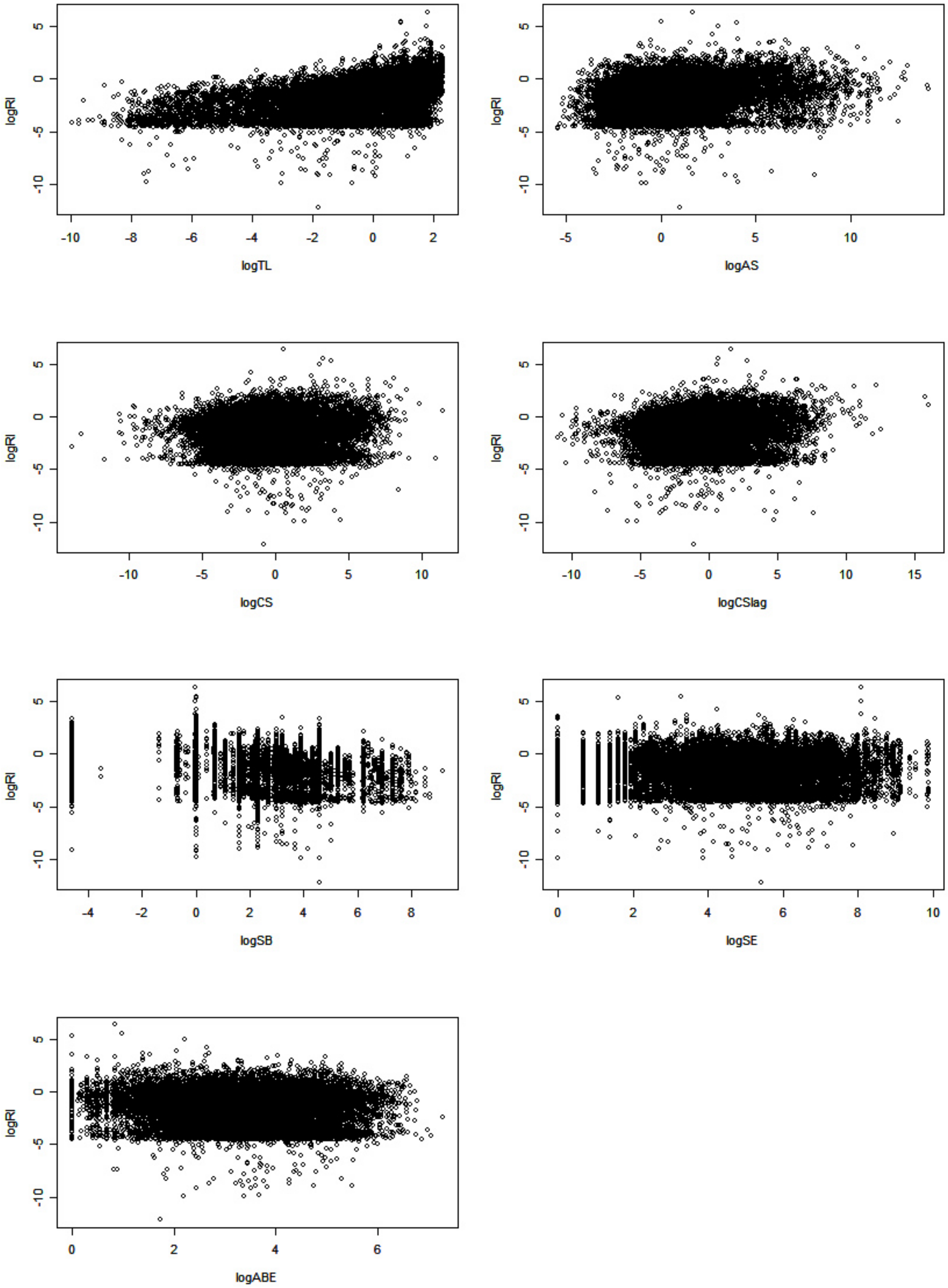


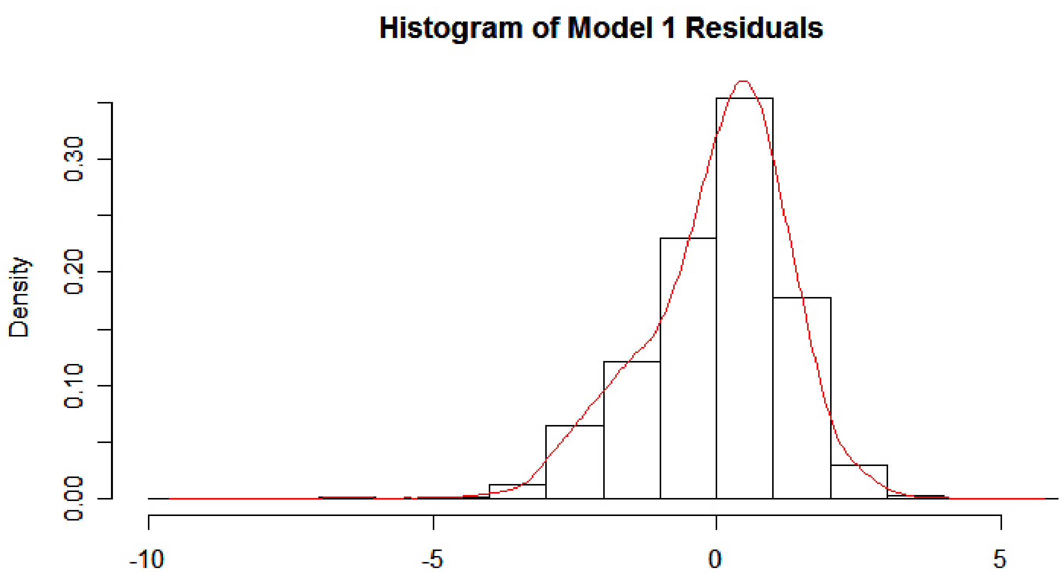
Figure 3 Scatterplot of predictor variables versus the dependent variable.

Table 2 Regression output for the parametric model (Model 1).

	Estimate	Std. Error	P-value
(Intercept)	0.5131	0.1357	0.0002
logTL	0.4675	0.0217	0.0000
logAS	-0.3275	0.0212	0.0000
logCS	0.0497	0.0120	0.0000
logCSlag	0.1012	0.0115	0.0000
logSB	-0.4405	0.0269	0.0000
logSE	-0.0459	0.0172	0.0076
logABE	-0.1651	0.0296	0.0000
logTL:logAS	0.0358	0.0032	0.0000
logTL:logCS	-0.0092	0.0019	0.0000
logTL:logCSlag	0.0062	0.0019	0.0009
logTL:logSB	0.0147	0.0039	0.0001
logTL:logSE	-0.0091	0.0026	0.0004
logTL:logABE	-0.0144	0.0045	0.0014
logAS:logCS	0.0137	0.0016	0.0000
logAS:logSB	0.0244	0.0015	0.0000
logAS:logABE	0.0162	0.0053	0.0024
logCS:logCSlag	0.0075	0.0014	0.0000
logCS:logABE	-0.0082	0.0033	0.0115
logCSlag:logABE	-0.0134	0.0032	0.0000
logSB:logSE	-0.0069	0.0028	0.0151
logSB:logABE	0.0216	0.0063	0.0006
logSE:logABE	0.0139	0.0045	0.0022
P2	0.5457	0.0896	0.0000
P3	-0.3552	0.1077	0.0010
P4	-0.1778	0.0849	0.0363
P5	-0.4702	0.0803	0.0000
P6	0.9244	0.0866	0.0000
P7	-0.1608	0.1072	0.1339
P8	-0.3357	0.0917	0.0003
P9	0.0313	0.0889	0.7250
P10	0.2548	0.0849	0.0027
P11	-0.0702	0.0775	0.3651
P12	-1.1074	0.1624	0.0000
P13	-0.3346	0.0916	0.0003
P14	0.3685	0.0795	0.0000
P15	-0.3501	0.1088	0.0013
D1	-0.2521	0.0403	0.0000
D2	-0.4310	0.0347	0.0000
D3	-0.6235	0.0492	0.0000

```
Residual standard error: 1.305 on 13499 DF
Multiple R-Squared: 0.354
Adjusted R-squared: 0.3521
F-statistic: 189.6 on 39 and 13499 DF
p-value: < 2.2e-16
```

The histogram of residuals and the QQ plot given in Figure 4 suggest that the model tends to overpredict more than it underpredicts the dependent variable. Combining this with Figure 5, we conclude that the model tends to underpredict when there is a long time to end-of-auction, while it overpredicts when there is relatively a short time until end-of-auction.



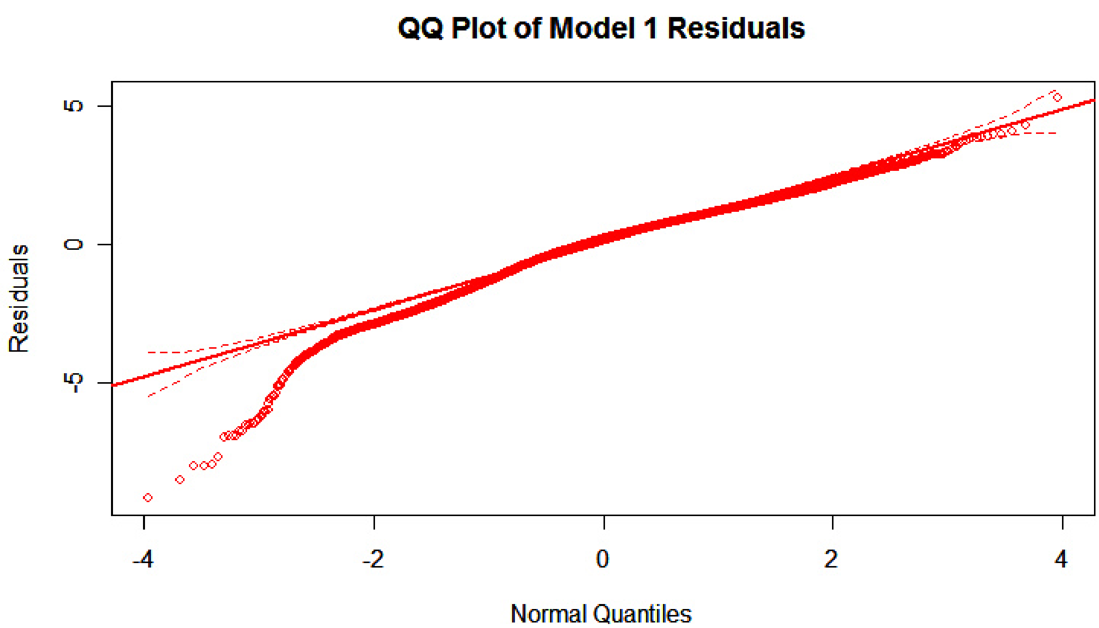


Figure 4 Histogram and QQ plot of Model 1 residuals.

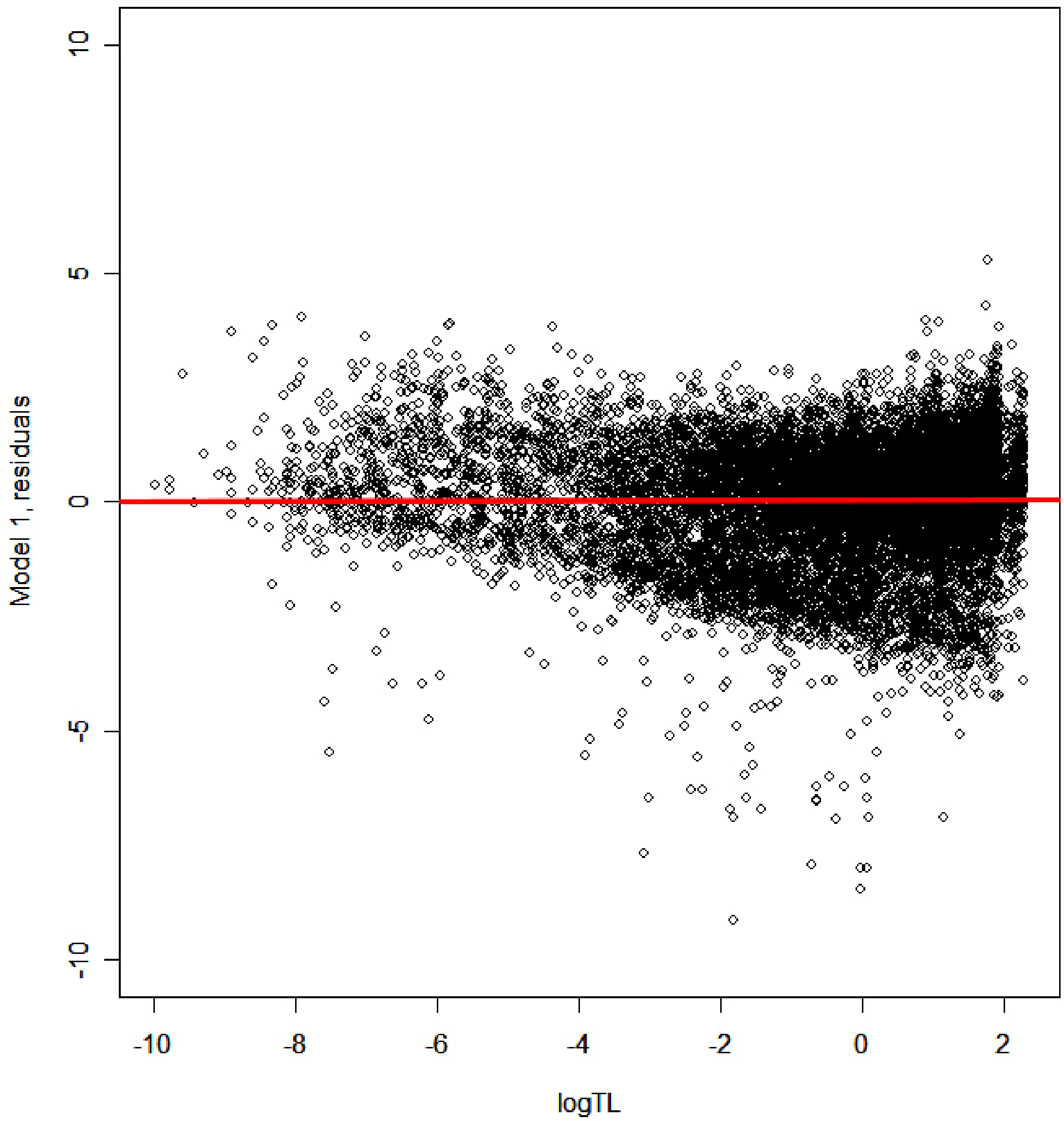


Figure 5 Plot of Model 1 residuals versus logTL.

3.2. The Semi-Parametric Model

In the semi-parametric model (Model 2), nonparametric functional terms for logTL, logAS, logCS and logCSlag are included along with main effects for logSB, logSE, logABE, and product ID and duration dummies. This way we aim to allow the model to freely depend on the live information: time left and price dynamics. We allow 7th degree splines to enable maximum freedom on the form of these non-parametric terms. The results of this model plotted in Figure 6 indicate that the underlying dependency is indeed not linear. From these plots we observe that the confidence bound for logTL becomes tighter for large values of logTL: the model can fit a functional form with almost certainty. The confidence bound for logAS is also relatively tight; however, this is not the case for logCS and logCSlag. We also give a summary of output for linear terms in Table 3. A quick comparison of this table to Table 2 yields that signs of all coefficients are identical and their magnitude is similar; however we observe that in Model 2, logSE is not significant.

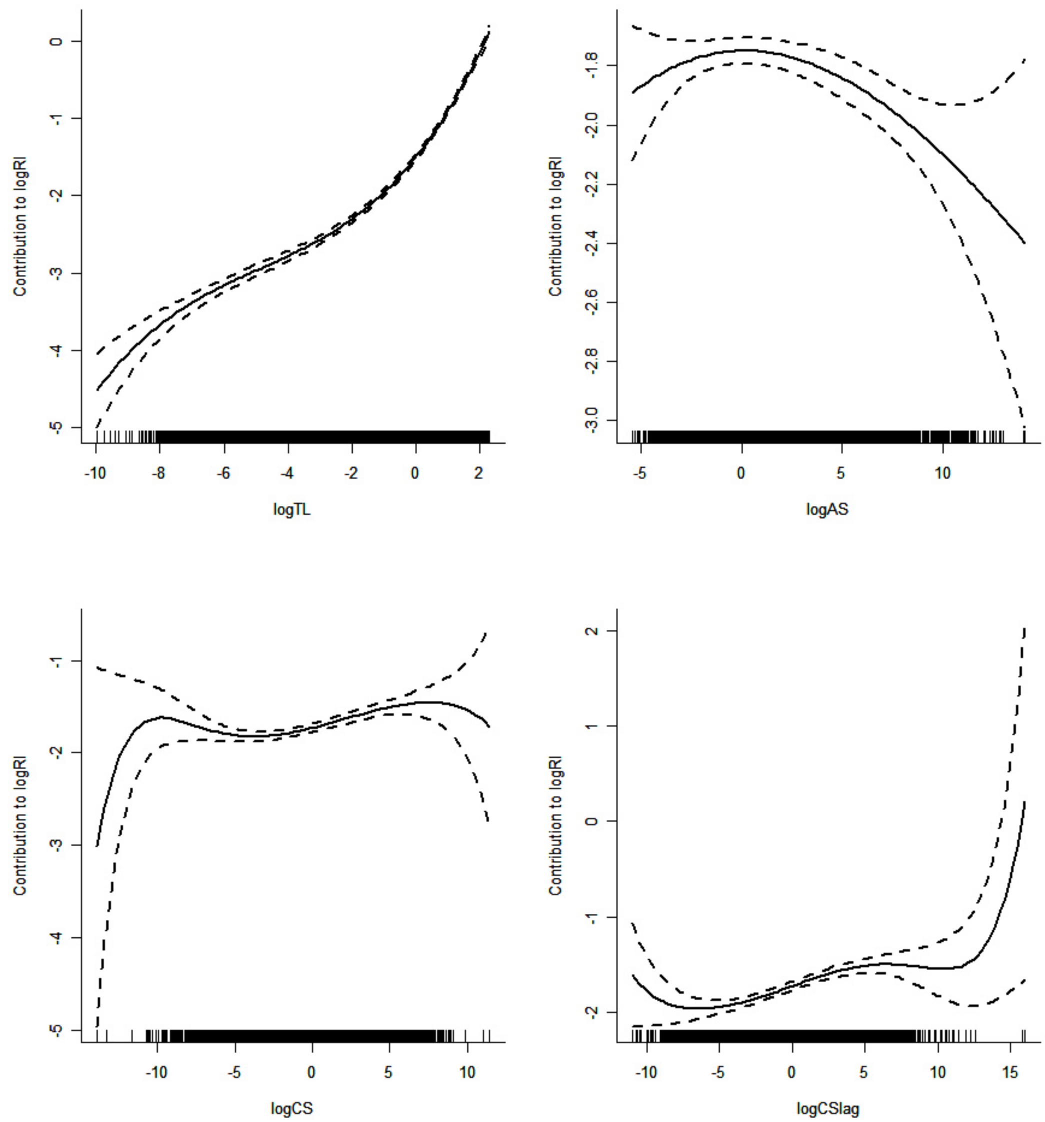


Figure 6 Plot of non-linear components in Model 2.

Table 3 Summary of output for linear components in Model 2.

	Estimate	Std. Error	P-value
(intercept)	0.2036	0.0922	0.0272
logSB	-0.4707	0.0124	0.0000
logSE	-0.0081	0.0059	0.1740
logABE	-0.0255	0.0098	0.0092
P2	0.6259	0.0874	0.0000
P3	-0.4312	0.1045	0.0000
P4	-0.2328	0.0826	0.0048
P5	-0.4682	0.0779	0.0000
P6	1.1980	0.0853	0.0000
P7	-0.1999	0.1043	0.0554
P8	-0.3946	0.0892	0.0000
P9	0.0382	0.0865	0.6588
P10	0.3227	0.0825	0.0001
P11	-0.1056	0.0752	0.1600
P12	-1.1000	0.1580	0.0000
P13	-0.3604	0.0891	0.0001
P14	0.5018	0.0772	0.0000
P15	-0.4172	0.1059	0.0001
D1	-0.4297	0.0395	0.0000
D2	-0.7517	0.0354	0.0000
D3	-1.1630	0.0512	0.0000

3.3. Non-Parametric Models

3.3.1. Model 3

In this model we omit the parametric terms in Model 2 and write it as an additive model with non-linear components for logTL, logAS, logCS and logCSlag. Comparison of Models 2 and 3 should give us an idea about the value of including static auction information in the model. Again, we allow for splines of 7th degree. Model 3 can be written as follows:

$$\mathrm{logRI} = f_1(\mathrm{logTL}) + f_2(\mathrm{logAS}) + f_3(\mathrm{logCS}) + f_4(\mathrm{logCSlag})$$

Surprisingly enough, the resulting non-linear components were (seemingly) identical to those in Model 2. We plot them in Figure 7.

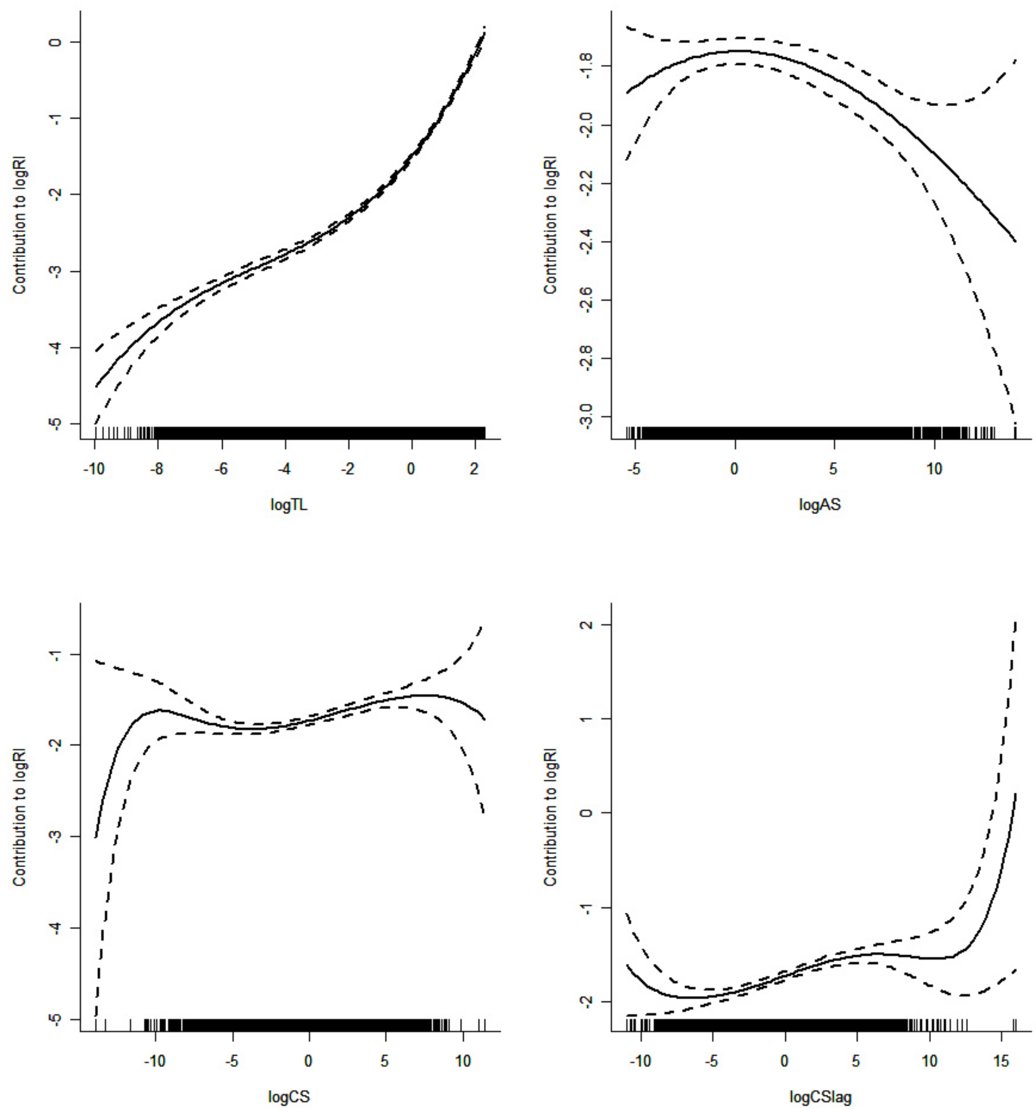


Figure 7 Plot of non-linear components in Model 3.

3.3.2. Model 4

This is the most parsimonious model in that the only predictor is logTL. Comparison of Model 4 to other models should let us evaluate the incorporation of additional predictors and in particular price dynamics parameters. Model 4 can be written as:

$$logRI = f(logTL)$$

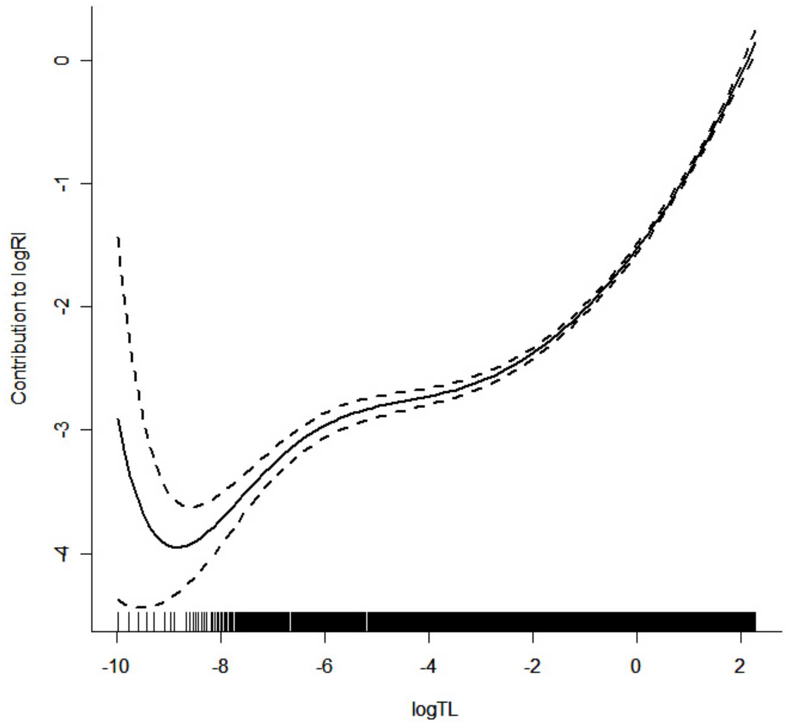


Figure 8 Plot of the non-linear term in Model 4.

Comparison of Figure 8 to Figure 6 and Figure 7 shows that the non-linear component in logTL is different in Model 4 from the previous models.

4. Final Price Prediction

Note that all model parameters are estimated using only the training dataset. In this section, we apply our models to the test dataset and obtain the fitted values. Since the dependent variable is a log transformed percentage increase over the current highest bid, we do an additional calculation to obtain the final price prediction as follows, where fitRI denotes the fitted values from the models, Bid is the current highest bid, and preP is the predicted final price:

$$preP = Bid\ (1+\exp(fitRI))$$

Then, we calculate absolute relative differences (ARD) between the observed final prices (actP) and predicted, and use this to evaluate and compare the prediction performance of the models:

$$\text{ARD} = \frac{\text{abs(actP-preP)}}{\text{actP}}$$

As previously mentioned we do two types of prediction analysis. In the first, we predict final prices for the whole duration of auctions. In the second, we focus at only the last day of auctions.

4.1. Whole-Auction Prediction

Implementing the above described procedure we obtain the ARD vector for each model and summarize the results in Table 4. The results are very interesting as they lead to the following conclusions:

1. The median error for all models in predicting the final prices is less than 15%. We consider this a quite accurate prediction performance.

2. Although Model 2 performs the best in terms of median score, it can fail by more than 200%.

3. Although Model 4 performs the worst in terms of median score, it fails only by 100% at most. Both non-parametric models are much safer in terms of worst case performance.

4. The worst case performance of the parametric model (Model 1) is the worst among all.

Table 4 Summary of whole-auction absolute relative differences for each model.

	Min.	1st Q.	Median	Mean	3rd Q.	Max.
Model 1	0.00%	5.75%	14.06%	19.79%	27.81%	278.70%
Model 2	0.00%	5.12%	12.72%	19.02%	27.40%	209.50%
Model 3	0.00%	5.89%	14.01%	20.40%	29.69%	129.30%
Model 4	0.00%	5.77%	14.09%	20.37%	29.92%	100.60%

4.2. Last-Day-of-Auction Prediction

In this case, we obtain fitted values only for the last day of each auction in the test dataset and compute ARD vectors accordingly. The results are summarized in Table 5, from which we make the following observations:

1. The median error of all our models is less than 8% and the worst case error is less than 80%.

2. The semi-parametric model (Model 2) has the overall best performance.

3. However, the performance of non-parametric models and is not substantially worse. Specifically, the simplest model (Model 4) performs very well.

Table 5 Summary of last-day-of-auction absolute relative differences for each model.

	Min.	1st Q.	Median	Mean	3rd Q.	Max.
Model 1	0.00%	3.50%	7.94%	11.74%	16.02%	75.12%
Model 2	0.00%	3.40%	7.61%	11.64%	16.00%	74.05%
Model 3	0.00%	3.82%	7.62%	12.08%	16.43%	79.82%
Model 4	0.00%	3.92%	7.65%	12.09%	16.33%	79.93%

5. Conclusions

In this study, we developed statistical tools to predict the final prices from an ongoing online auction. This problem is inherently interesting and yet challenging in that the live information has to be taken into account. While constructing our models, we devised novel parameters to account for the price dynamics (price speed and acceleration), and investigated the value of incorporating such parameters in a statistical prediction tool. We looked at the prediction performance of our models in two cases: (1) prediction for the length of an auction, and (2) prediction only at the last day of an auction.

Our analyses indicated that statistical tools can be developed with quite accurate final price prediction performance. Furthermore, we find that even a simple model (Model 4) can achieve such results. This leads us to conclude, contrary to the findings and general belief in the literature [2, 4, 8], that incorporation of price dynamics parameters does not bring substantial benefits in terms of prediction performance.

References

[1] Bajari, P. and A. Hortacsu (2003). “The Winner’s Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions,” *Rand Journal of Economics*, **3**(2):329–355.

[2] Bapna, R., W. Jank and G. Shmueli (2004). “Price Formation and its Dynamics in Online Auctions,” Technical Report, D&IT Department, University of Maryland, CP, “http://www.smith.umd.edu/faculty/wjank/auctionDynamics.pdf”.

[3] Ghani, R. and H. Simmons (2004). “Predicting the End-price of Online Auctions,” International Workshop on Data Mining and Adaptive Modelling Methods for Economics and Management held in conjunction with the 15th European Conference on Machine Learning (ECML/PKDDD 2004), Pisa, Italy, “http://www.accenture.com/NR/rdonlyres/78F2E7B5-FA33-48F1-857E-53CF9EB91D45/0/priceprediction.pdf”.

[4] Jank, W. and G. Shmueli (2005). “Studying Price-Dynamics and Auction-Energy in Functional Models,” Technical Report, D&IT Department, University of Maryland, CP, “http://www.smith.umd.edu/faculty/wjank/WISE-Abstract.pdf”.

[5] Lucking-Reiley, D., D. Bryan, N. Prasad and D. Reeves (2000). “Pennies from eBay: The Determinants of Price in Online Auctions,” Techical Report, University of Arizona, “http://www.vanderbilt.edu/econ/reiley/papers/PenniesFromEBay.pdf”.

[6] Ow, T. T. and C. A. Wood (2004). “Trust Versus the Winner’s Curse: The Effects of Bidder Experience on Prices in Online Auctions,” Working Paper, Mendoza College of Business, University of Notre Dame, “http://www.nd.edu/~cwood1/research/BidderX.pdf”.

[7] Roth, A. E. and A. Ockenfels (2002). “Last-Minute Bidding and the Rules for Ending Second Price Auctions: Evidence from eBay and Amazon on the Internet,” *American Economic Review*, **92**(4):1093–1103.

[8] Wang, S., W. Jank and G. Shmueli (2006). “Dynamic Forecasting of Online Auction Prices using Functional Data Analysis,” Under review, “http://www.smith.umd.edu/faculty/wjank/Predicting%20Online%20Auction.pdf”.

The Impact of the Use of a Time-Based Bid Strategies on the Odds of Winning for eBay Auctions

Alicia F. Dixon
May 2006
BMGT808X
Prof. Wolfgang Jank

Introduction to the topic

Electronic commerce has grown tremendously with the adoption of the Internet by consumer. One form of electronic commerce in particular, online auctions, has become established as a shopping channel alternative to traditional brick and mortar stores and catalog shopping (Lucking-Reiley 1999; Reindorp et al.). Online auctions have become popular because of they provide access to items not readily available through other shopping channels and because users often enjoy the experience of auction participation (Bajari and Hortacsu 2004). Arguably the most well-known online auction site is www.ebay.com (Bajari and Hortacsu 2003). Found in September of 1995, eBay has grown to have a sizeable install base, with over 181 million registered users as of December 2005 (eBay Inc. 2005).

This popularity has raised the question as to what makes a bid most likely to win. A common belief is that strategic bidding, particularly time-based bidding, can improve the chances of winning an auction. Intuitively this belief is plausible but empirical evidence can confirm or dispute the claim. In this paper actual data from eBay auctions is used to evaluate whether or not time-based bidding strategies, particularly bidding at the end of an auction, increase the chances that the bid will win the auction.

Positioning among related research

Online auctions have be the subject of numerous scholarly studies, including the determinants of price (Lucking-Reiley et al. 2000); final price predictions (Reindorp et al.); and the determinants of bidder and seller behavior (Bajari and Hortacsu 2003). Analyses of the timing of bids, last minute bidding in particular, have also been conducted. Last minute bidding is the phenomenon of placing a bid at a time that is approaching the close of an auction, for auctions with a defined ending time (Bajari and Hortacsu 2004). Several authors have provided explanations for why last-minute bidding occurs (Bajari and Hortacsu 2004; Bajari and Hortacsu 2003; Ockenfels and Roth 2003; Roth and Ockenfels 2002). These authors have found that the winning bids tend toward the end of auctions (Bajari and Hortacsu 2004; Ockenfels and Roth 2003); however, less than 20% of all bids are placed during the last few minutes of an auction. Though last minutes bids are in the minority, it appears that, when placed, last minute are more likely to win auctions (Ockenfels and Roth 2003). However, no assessments of the effect of the timing of a bid on the probability of winning have been presented in the extant literature. Thus, thee purpose of this paper is to evaluate whether or not the time of bid placement impacts the probability of winning an auction.

Data Description & Data Manipulations

Over 100,000 observations of bid history data were acquired from www.ebay.com using a web crawler. Data were collected between July 2001 and February 2002. Products sold in the data set included a variety of categories, such as sunglasses, golf clubs, wristwatches, luggage, calculators and vases. The use of a web crawler to form the data set allowed for the creation of a highly robust data set where no missing data was found. Auction lengths in the given data set ranged from 1 day to 10 days. Being that the standard auction lengths on eBay are 3, 5, 7, and 10 days, observations that represented any other auction length were removed from the data. Data was then sorted by auction ID and separated into four samples by auction length. Also, a sub-set of the original data, called a holdout sample, can be set aside during model fitting then later used to validate the fitted logistic regression models for accuracy (Hosmer and Lemeshow 1989). Over 40,000 observations from the original data were used to create a holdout sample, including all of the auction lengths, to for this purpose.

The variables are interest, as shown in Table 1, are bidder experience, bids per person, time left in minutes, and bid magnitude, as the independent variables and winner as the dependent variable. Two of these variables, bidder experience and bids per person, were given in the data set; the remaining three variables were created through formula manipulations of the given data.

Bidder Experience: Intuitively it seems that users who have participated in several auctions might have learned how to place more strategic bids, with higher probabilities of winning. To account for this possibility the variable bidder experience is included in mode. Bidder experience is the number of auctions in which the user has placed a bid.

Bids per Person: A use that places more than one bid in one auction should have a higher probability of winning the auction. Bids per person is the number of bids that a user placed in a specific auction observed in the data.

Time Left in Minutes: A time-based bid strategy is defined as purposefully selecting to place a bid at a certain point in the auction based on the time remaining in the auction, relative to the predetermined auction length, with the intent of improving the likelihood of winning. The variable time left in minutes is used to determine when in the auction a bid was placed. This variable was created by converting the given variable time left (tleft) from days to minutes by multiplying tleft*24 hours*60 minutes.

Bid Magnitude: A bid that is considerably larger than the previous bids would increase the total value of the auction. When bid values are significantly increased over the previous bid the auction value approaches the willingness to pay amount for average users, which could discourage other users from placing a competing bid. Bid magnitude is a change ratio formed as follows:

$$\frac{\text{(bid current observation - bid previous observation)}}{\text{bid previous observation}} \tag{1}$$

Thus, bid that significantly increase the magnitude of the bid have higher probabilities of winning the auction.

Bids per Person by Time Left in Minutes: To control for the impact of repeat bidding on the effect of the time-based bid strategy, the interaction term bids per person by time left in minutes is included in the model.

Winner: The response variable, winner, was created as a dummy variable for each observation by using two given variables, bid and high bid. When bid equaled high bid the observation was deemed the winning bid, represented by a 1. The bid did not win the auction was coded as 0, making the response variable a dichotomous variable. Therefore, logistic regression is used for the analysis, since the response variable is binary (Hosmer and Lemeshow 1989; Kleinbaum 1994).

Summary statistics for the variable of interest are given in Table 2, while histograms of these variables are illustrated in Figures 1 through 4. The mean values for all of the variables of interest are greater than the median values, indicating that the distributions are skewed. Also, all of the graphs show that these variables are right skewed; however, no variable transformations were performed since normality is not an assumption of logistic regression (Hair et al. 1998; Hosmer and Lemeshow 1989). Figure 3 shows the correlation matrices for the four samples. While there are significant relationships between many of the independent variables, the correlations are relatively low; thus, all four independent variables are included in the logistic regression model.

Methodology and Model Description

As mentioned previously, logistic regression is used to test the null hypothesis since the response variable is either 0 or 1. Thus, bidder experience, bids per person, time left in minutes, bid magnitude and the interaction between time left in minutes and bid magnitude are modeled by a assuming a generalized binomial logit model for a the probability of winning an auction. Let Y=1 or 0, denoting the probability of winning an auction, as given in the following expression:

$$P(Y=1)=\ln\left(\frac{P}{(1-P)}\right)= f(\alpha+\beta x)=\frac{e^{(\alpha+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\beta_4 X_4+\beta_5 X_3 * X_4)}}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\beta_4 X_4+\beta_5 X_3 * X_4)}} \quad (2)$$

where α is the intercept, β_is are the coefficients to be estimated, and X_is are the predictor variables. In this case, the estimated logit is given by equation 3:

$$\ln\left(\frac{P(Y=1)}{P(Y=0)}\right)=\alpha+\beta_1 bidderexperience+\beta_2\times bidsperperson+ \quad (3)$$
$$\beta_3\times tleftmns+\beta_4\times bidmagn+\beta_5\times tleftmns * bidmagn)$$

Using the software application SPSS version 14.0 to generate the analysis, the maximum likelihood coefficients, standard errors, and Wald statistic for the four samples result in the logit models expressed Table 4. Individual significance for the independent variables is judged based on the Wald statistic. With the exception of bidder experience, all of the Wald test statistic scores are significant on all of the four samples. Since the test statistic is insignificant bidder experience is removed from the model for the sake of parsimony. Revised logistic regression models are in Table 5 and confidence intervals for the point estimates of the coefficients are listed in Table 6.

Estimated experimental error (residuals) in the model is obtained by subtracting the actual observed responses values from the predicted (fitted) response values (Weiss 2005). An examination of the residuals is helpful to determine whether or not the chosen model is appropriate (Hair et al. 1998; Weiss 2005). Figure 6 shows the residuals versus fitted values for the 3, 5, 7 and 10 day auctions. The scatterplots for all four samples are of a similar nature, showing two funnel shapes that taper off on the right side of the graph. Such a pattern in the residual, where the residuals have decreasing scatter as the value of the response increases, indicates that the residuals are not normally distributed and that the error variance is not constant. In addition, the normal probability (QQ) plots for the four samples are illustrated in Figure 7. The slight "S" shaped curve on this graph and the break in the graph suggest that the residuals follow a bimodal distribution. Thus, Figure 7 also indicates that the variance is not constant and that the errors are not normally distributed. Since normally distributed residuals and constant variance, however, are not required assumptions of logistic regression, the model fit is deemed to be acceptable.

Residuals are also useful in determining if there are any outliers in the data. A few extreme points are shown in the scatterplots in Figure 6, but to determine whether or not these are leverage points, tests of influence are run based on the hat matrix. Influence plots, which are a bubble plot of studentized residuals by hat values, where the size of the bubbles represent the observations proportional to Cook's distances for the models, are shown in Figure 8. While each of these graphs illustrates that there are observations with hat values above the cut off ($2p/n$) (Hair et al. 1998), the Cook's distance values (Figure 9) for these observations are all below 1.0 (Hair et al. 1998) so the observations are not deemed to have leverage and are kept in the model. Being that no leverage points are identified and the model fit is appropriate, the significance of the model is now judged.

The overall significance of the model is determined using the likelihood ratio test statistic, also referred to as deviance D, which follows a chi-squared (χ^2) test (Hosmer and Lemeshow 1989). Deviance values, as given in Table 6, are 2053.31, 1838.29, 3921.11, and 1617.73 for Model 3, Model 5, Model 7 and Model 10 respectively, while the critical value for χ^2 ($p = 0.05$, $df = 4$) is 9.49, making all of the deviance values significant. Significant deviance indicates that the model adequately fits the data. Additional fit statistics, the Cox & Snell R^2, and Nagelkerke R^2, are presented in Table 7; with both indicating that less than approximately a third of the variance in the probability of winning is explained by the independent variables.

Results and Discussion

Once the fit of the model is found to be appropriate, then inferences can be made (Hosmer and Lemeshow 1989). Interpretation of the fitted models in multiple logistic regression focuses on the coefficient estimates for the model (Hair et al. 1998; Hosmer and Lemeshow

1989). Logistic coefficients calculate a comparison the probability of an event occurring with the probability of the event not occurring, which is called the odds ratio (Hair et al. 1998). In this case the hypothesis is tested using time left in minutes as the experience variable, adjusting for bids per person and bid magnitude.

Applying observational data to the model permits the computation of the odds ratios for a particular bid. Once the odds ratio for a particular bid is computed, a comparison of competing bids can be made in order to evaluate the effect of time left in minutes on the probability of winning. To illustrate such a comparison, two competing bids, Bid A and Bid B, are analyzed. For Bid A, a bidder places a single bid that raises the magnitude of the auction by 10%, at the first minute of the auction. Meanwhile, Bid B is also a single bid placed in the same auction that raised the value of the auction by 10%, but is placed at the last minute of the auction. Equations 4 and 5 calculate the odds ratios for Bid A and Bid B respectively.

$$e^{\alpha+\beta x} = 0.184 + -0.301(1) + 0.000(0.1) + -9.463(10079) + 0.001(10079 * 0.1) = 0.01385 \qquad (4)$$

$$e^{\alpha+\beta x} = 0.184 + -0.301(1) + 0.000(0.1) + -9.463(1) + 0.001(1 * 0.1) = 0.34542 \qquad (5)$$

The difference in the odds ratios for these two bids is 0.33158, suggesting that the odds ratio increases for bids placed closer to the end of the auction. Furthermore, the predicted probability of winning for these two bids can be compared, as computed in equation 6 for Bid A and in equation 7 for Bid B.

$$\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{e^{0.184+-0.301(10079)+0.000(0.1)+-9.463(1)+0.001(10079*0.1)}}{1+e^{0.184+-0.301(10079)+0.000(0.1)+-9.463(1)+0.001(10079*0.1)}} = 0.01366 \qquad (6)$$

$$\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{e^{0.184+-0.301(1)+0.000(0.1)+-9.463(1)+0.001(1*0.1)}}{1+e^{0.184+-0.301(1)+0.000(0.1)+-9.463(1)+0.001(1*0.1)}} = 0.25674 \qquad (7)$$

Bid B has a higher predicted probability of winning than Bid A which also suggests that later bids have greater chances of winning an auction. Additional bid comparison examples are shown in Table 8. Both the odds ratios and the predicted probabilities of winning increase for bids that are placed nearer to the close of the auction, which further suggests that last-minute bids have greater chances of winning an auction than bids placed early in the auction duration. Thus it can be inferred that using a last-minute bidding strategy has increase the probability of winning an auction.

Next, the holdout sample data is applied to the models to validate the model predictability. The results of the holdout sample validation are listed in Table 9. Using the variables of interest from the holdout sample to predict whether or not a bid won the auction, the models corrected predicted the response of winning or not winning at 13.19% for Model 3, 20.69% for Model 5, 14.95% for Model 7, and 13.04% for Model 10. Hence the overall predictive ability of the models is low. It is surmised that this lack of predictability is the result of the fact that many competing bids in an auction employ the same last-minute bidding strategy, negating its effect.

In summary, adjusting for bids per person and bid magnitude, there is a positive effect of time left in minutes on the chances of winning an auction. The results of fitted logistic regression models show that bids placed at times approaching the end of the auction have higher odds ratios than bids placed at the beginning of the auction duration. Higher odds ratios indicate increased chances of winning an auction and an increased probability of winning. Therefore it is concluded that bidding at the end of the auction is found to increase the probability of winning an auction, although if competing bids use the same strategy other variables could influence the auction outcome. Whenever possible a bidder should use a last-minute bidding strategy to improve the chances of winning the auction.

References

Bajari, Patrick and Ali Hortacsu (2004), "Economic Insights from Internet Auctions," Journal of Economic Literature, 42 (2), 457.

---- (2003), "The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions," RAND Journal of Economics, 34 (2), 329.

eBay Inc. (2005), "Annual Report 2005." San Diego, CA.

Hair, Jr., Joseph F., Rolph E. Anderson, Ronald L. Tatham, and William C. Black (1998), Multivariate Data Analysis (Fifth ed.). Upper Saddle River, New Jersey: Prentice-Hall.

Hosmer, David W. and Stanley Lemeshow (1989), Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Kleinbaum, David G. (1994), Logistic Regression: A Self-Learning Text. New York: Springer-Verlag.

Lucking-Reiley, David (1999), "Auctions on the Internet: What's Being Auctioned, and How?" Department of Economics, Vanderbilt University.

Lucking-Reiley, David, Doug Bryan, Naghi Prasad, and Daniel Reeves (2000), "Pennies from eBay: the Determinants of Price in Online Auctions."

Ockenfels, Axel and Alvin E. Roth (2003), "Late and Multiple Bidding in Second Price Auctions: Theory and Evidence Concerning Different Rules for Ending an Auction," Max Planck Institute for Research into Economic Systems.

Reindorp, Matthew, Wolfgang Jank, and Louiqa Rashid, "The Right Auction at the Right Price," Robert H Smith School of Business, University of Maryland.

Roth, Alvin E. and Axel Ockenfels (2002), "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet," American Economic Review, 92 (4), 1093.

Weiss, Neil A (2005), Introductory Statistics (Seventh ed.). Boston: Pearson Education.

Stacy Cuffe
5/5/06
BMGT 808X

A Time to Shill:
A look at bid shilling in the Ebay Market for Rolex wrist watches.

Introduction:

Millions of transactions take place electronically each day on eBay. A plethora of items, ranging in value from mere pennies to thousands of dollars, are being auctioned off daily in online markets. Extensive data has been gathered on these auctions, and a wealth of information about bidder and seller behavior can be gleaned from these auction histories. One phenomenon that manifests itself in eBay bidding data is the act of bid shilling. A shill bid is a bid placed by the auction's seller operating under another name, or by an agent of the auction's seller, in order to elevate the transaction price of his or her item. Since the seller's objective is to make as large a profit as possible, at times he or she may actually bid up the price of his or her own item, so that other bidders have to pay more in order to win the auction. This deceptive practice is illegal in the eBay market, yet practically impossible to regulate or identify with any degree of certainty. This particular paper focuses on the presence of bid shilling in the Rolex watch market and its possible effect on final selling prices.

Substantial research has been done on the practice of bid shilling. Sinha and Greenleaf (2000) discuss how bidder aggressiveness can actually motivate sellers to shill. Kaufman and Wood (2004) look at the effect of reserve price shilling, a type of shilling where the objective is not simply to raise the final transaction price, but rather to avoid having to pay auction house fees. In eBay the seller is typically charged a fee based on what he or she sets as the starting bid level. The seller can effectively save money by setting this level artificially low and then placing shill bids until the bid level is up to their true starting bid level. This paper concludes that reserve price shilling has a positive impact on final bid prices in the online rare coin market. Unlike Kaufman and Wood, we wish to examine the effect of traditional shilling on an auction's final transaction price.

Data Manipulation:

We begin our research with a rich database of both high and low value auctions in a wide variety of eBay categories including sunglasses, ties, golf balls, watches, vases, luggage, etc. collected over a period of time from 2001-2002. The database contains a bid history for each auction which displays the bidder rating, bid placed, time of bid, and remaining time in auction. For each auction, the database also features several unchanging parameters including auction id, product id, category id, winning bid, starting bid, auction duration, seller id and experience, number of bids placed, number of participating bidders, number of bids per participant, and the auction starting and ending dates.

Before we can attempt to identify shill bids in these auctions, we must first sort the bid histories into chronological order. For this study, we opt to focus on a single database file containing histories for several thousand different auctions. Once the bids are placed in sequential order, we keep track of the leading bid in each auction over time. It is important to remember that since the leading bid is not visible to auction participants, not all bids actually increase the final auction price. We then compute the percent increase over the previous leading bid for each bid that actually does change the current leading price. For now, we will ignore those bids that are lower than the present leading bid. The basic summary statistics for the average percent increase in bid is shown in Table 1 below.

Table 1:

Mean	Median	Std. Deviation	Min	Max
.8551	.0955	19.012	.000001	1499

For the sake of this paper, we choose to define a shill bid as a bid placed by someone who does not win the auction whose percentage increase is at least 1 standard deviation above the mean percent increase. We essentially flag all bids placed by participants who did not go on to win the auction that are over 1986% higher than the previous winning bid. The variance and standard deviation are extremely large due to some instances where the bid level jumped from one cent to several dollars. Although we initially consider deleting such instances from our study in order to obtain a more reasonable standard deviation, upon further reflection we decide that this type of increase is particularly representative of possible shiller behavior and we do not wish to remove any instances of potential shilling from our data set. We do not assert that all of these flagged bids are shill bids, nor can we state that any other bids are not. It is not possible for us to ever know for certain whether a bid is actually being placed by the seller or not. We are simply looking for unusually large increases in bid price by participants who ultimately drop out of the auction. The bid history shown below illustrates an example of an auction where a shill bid occurred.

Table 2:

Auction id	Winning bid	Bid	Leading bid	% increase	Shill?
1645271527	1600	26	26		
1645271527	1600	27	27	.038462	
1645271527	1600	46.5	46.5	.722222	
1645271527	1600	1000	1000	20.50538	YES
1645271527	1600	50.01	1000		
1645271527	1600	260	1000		
1645271527	1600	500	1000		
1645271527	1600	600	1000		
1645271527	1600	1000	1000		
1645271527	1600	1200	1200	.20	
1645271527	1600	1050	1200		
1645271527	1600	1250	1250	.041667	
1645271527	1600	1250	1250		
1645271527	1600	1300	1300	.04	
1645271527	1600	1501	1501	.154615	
1645271527	1600	1501	1501		
1645271527	1600	1551	1551	.033311	
1645271527	1600	1600	1600	.031593	

We find 53 auctions that contain potential shill bids. However, surprisingly, 33 of these auctions are in the Rolex wristwatch market. The 20 remaining potential shill bids are spread out thinly over several other markets. Since we traditionally need a sample size of at least 30 for the CLT to hold and for regression analysis to be meaningful, we decide to focus solely on the effects of shilling in the Rolex market from this point on. We now restrict our data set to the 734 auctions for Rolex watches.

Rather than have bid histories for each auction, we now wish to collapse each auction into a single row of data. We delete the bidding history for each auction and keep only those variables that we feel may have an impact on final price. For each auction, we opt to retain the following variables: auction id, winning bid, starting bid, auction duration, seller experience, number of bids, number of bidders, and a binary variable indicating whether or not a possible shill bid was detected. Now that the data is in a suitable format, we are ready to proceed with finding a model that attempts to explain the variation in final transaction price for Rolex watches.

Methodology & Model Description:

Before we attempt to determine the effect of a potential shill bid on the final auction transaction price, we must first take a look at the distributions of the individual variables that may be included in our model and investigate any possible relationships between these variables. Figures 1.1-1.6 show histograms for many of the variables in our transformed data set. Notice that most of these histograms indicate that their respective variables are heavily skewed. We compute a few basic summary statistics for these variables in Table 3 below.

Table 3

	Mean	Median	Min	Max
Winning Bid	2334	1550	52	24500
Starting Bid	627.10	100	.01	9500
Duration	6.512	7	0	10
Seller Exp.	439.3	101	-2	9055
Number bids	15.52	14	2	59
Num. bidders	7.792	7	2	21

Notice that the mean is substantially larger than the median for the variables winning bid, starting bid, and seller experience, and somewhat larger for number of bids and number of bidders. This provides further evidence of positive skewness and prompts us to take log transformations of these variables in order to reduce the skewness of the data. Figures 1.7-1.11 show histograms for these variables once log transformations are taken. Note that the variables are now much more normal-looking in appearance. We will consider the use of these transformed variables in our regression model.

We now shift our focus towards investigating which variables have an impact of winning bid. The results of a basic correlation analysis are shown in Table 4 below.

Table 4:

	Win bid	Start bid	Duration	Seller Exp	# bids	#bidders
Win bid	1	.4096	-.0451	-.0663	.0861	.0185
Start bid	.4096	1	-.1413	-.1093	-.3766	-.4286
Duration	-.0451	-.1413	1	.1883	.1518	.2287
Seller Exp	-.0663	-.1093	.1883	1	.1042	.1757
# bids	.0861	-.3766	.1518	.1042	1	.8298
#bidders	.0185	-.0285	.2287	.1757	.8298	1

Based on the correlations shown in the above table, it appears that Starting bid has a fairly strong positive relationship with winning bid, and that there is a somewhat weaker positive relationship between winning bid and number of bids. The scatterplots in Fig 2.1 and 2.2 further support the existence of such a relationship. The correlation coefficients between winning bid and the other variables are substantially lower, indicating that they may not be as useful in explaining variation in winning bid.

We obtain a preliminary model by performing basic stepwise regression techniques. Our first model, a fully parametric model henceforth referred to as Model A, explains log(Winning bid) as a function of starting bid, duration, log(number of bids), and the binary shill variable. However, the adjusted R-squared for this model was remarkably low (.295), and the residuals v. fits plot for this model, seen in figure 2.3 exhibits some abnormal tail behavior. Also, the q-q plot for Model A, seen in Fig 2.4, shows that the model does a very poor job of predicting winning bid values on the tails.

The regression output for this model is shown in table 5 in the results section. Notice, however, that the shilling variable is significant at the .01 level.

Unhappy with our results thus far and disparaged that some of the central assumptions of linear regression, namely the independence and constant variance of errors, do not appear to hold for our model, we look back at the scatterplots in figures 2.1 and 2.2 and notice that the relationships between winning bid and starting bid and between winning bid and number of bids do not exactly appear to be linear. We realize that our basic multiple regression model is likely missing a nonlinear term, and for this reason opt to experiment with nonparametric regression. Nonparametric regression makes minimal assumptions about the relationships between dependent and independent variables and about the distribution of error terms. Because of the lack of assumptions, nonparametric regression is often able to uncover structure in the data that linear regression misses.

In order to investigate this notion of a non-linear relationship between some of the variables and winning bid, we decide to employ semi-parametric regression techniques. Since the relationships between winning bid and starting bid, and between winning bid and number of bids appear to be non-linear, we estimate these functions nonparametrically using a cubic smoothing spline. However, we continue to estimate the binary shill variable parametrically. This semi-parametric model, henceforth referred to as Model B, does appreciably better than the strictly parametric model. It has an adjusted R-squared of .371. We look at several other semi-parametric models including some with nonparametrically estimated interaction terms, but none of these outperform the aforementioned Model B. It is important to remember that since we are now using a nonparametric model, there are no assumptions about the distribution, independence, and variance of error terms that must be satisfied in order to preserve the validity of the model. The fundamental difference between Model A and Model B is that Model A carries with it the assumption that winning bid is a linear function of starting bid, duration, number of bids, and shilling, while Model B does not assume any such linearity. The regression output for this semi-parametric model is shown in Table 6 in the results section.

Results:

Regression output for Model A.

Table 5:

MODEL A	Estimate	Std. error	p-value	Significant at .01
(Intercept)	6.6065	.1456	2e-16	Yes
Start bid	.00052	.00003	2e-16	Yes
Log(numbids)	.4279	.04355	2e-16	Yes
Duration	-.02276	.01354	.0933	
shill	.4012	.1451	.0059	Yes

Residual Std. Error	R-squared	Adj. R-squared
.7929	.2989	.295

Regression output for Model B.

Table 6:

Parametric:	Estimate	Std. Error	P-value	Significant at .01
(Intercept)	8.1196	.0929	<2e-16	Yes
Shill	.5211	.1419	.00026	Yes
Approximate significance of smooth terms:				
Nonparametric:	edf	F	P-value	Significant at .01
s(startbid)	8.738	47.40	<2e-16	Yes
S(numbids)	7.170	16.22	<2e-16	Yes

Adjusted R-squared	Deviance Explained
.371	38.6%

Conclusions:

Since the adjusted R-squared value is higher for the semi-parametric model than for the fully parametric model, we are able to conclude that the semi-parametric model does a somewhat better job of explaining the variation in winning bid amount for the eBay Rolex market. Notice that regardless of which model we use, we find that the binary shill variable is significant at the alpha=.01 level. This allows us to conclude that shill bids, or perhaps legitimate bids simply placed in the same manner as shill bids, do have a positive impact on final auction transaction price in the Rolex market. We expect that similar results would hold in the eBay market for most other products. This is particularly disturbing news for officials attempting to regulate internet fraud and for eBay itself. It appears that it is actually in a seller's best interest to place a shill bid in his or her own auction. Since it is practically impossible to positively identify a shiller, this practice of bidding up one's own item could potentially become a huge problem for online market industry and its participants in the near future.

References:

1. Kauffman, Robert J. and Wood, Charles A., "The Effects of Shilling on Final Bid Prices in Online Auctions," April 2004, 1-26.

2. Sinha, A. and Greenleaf, E., "The Impact of Discrete Bidding and Bidder Agressiveness on Seller's Strategies in Open English Auctions: Reserves and Covert Shilling," July 2000., 1-21.

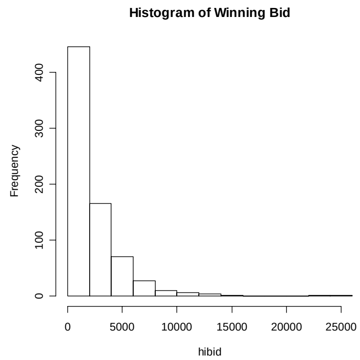


Fig .1.1

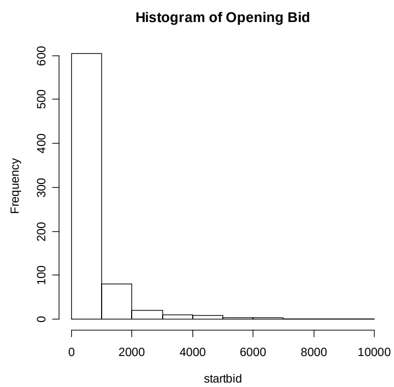


Fig 1.2

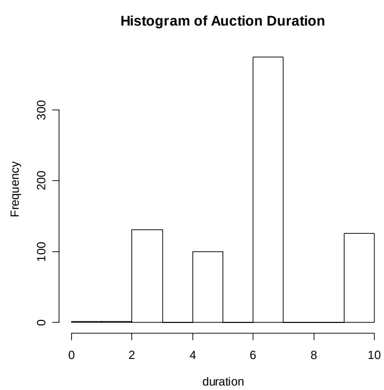


Fig 1.3

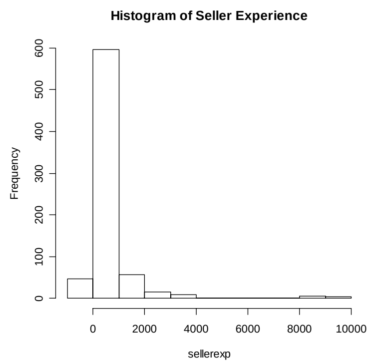


Fig 1.4

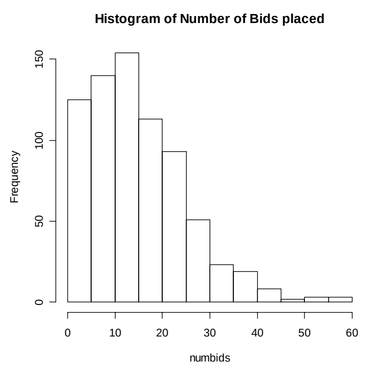


Fig 1.5

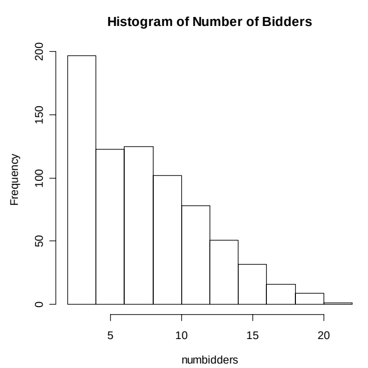


Fig 1.6

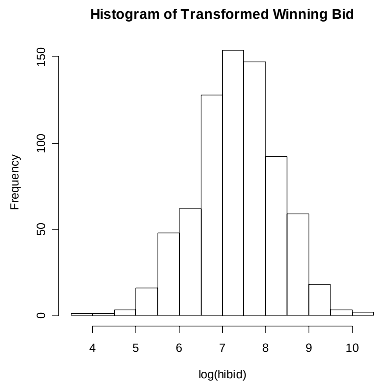


Fig 1.7

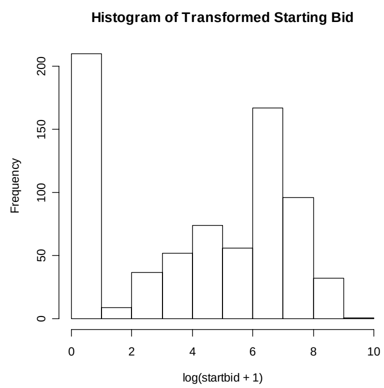


Fig 1.8

Fig 1.9

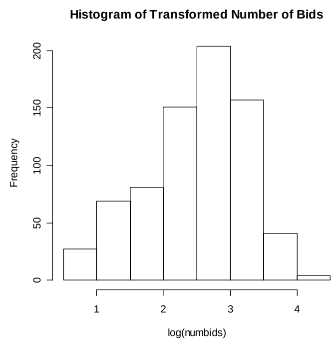


Fig 1.10

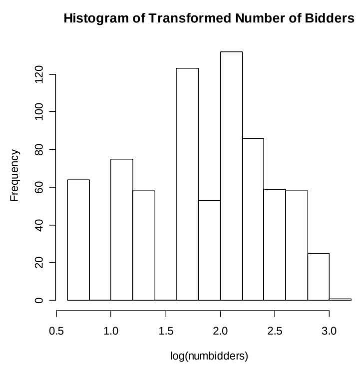


Fig 1.11

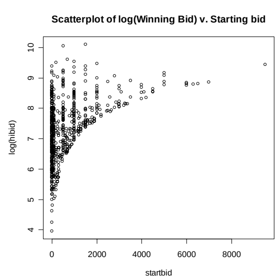


Fig 2.1

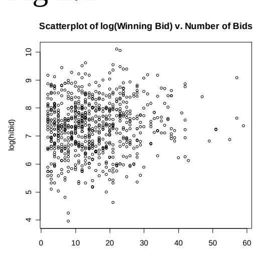


Fig. 2.2


Residuals v. Fits plot for Model A
g$res
g$fit
7
8
9
10
11
-3
-2
-1
0
1
2

Fig 2.3


QQ plot for Model A
Studentized Residuals(g)
t Quantiles
-3
-2
-1
0
1
2
3
-4

Fig. 2.4

CONCURRENCY OF LIVE AUCTION PRICES

Natalia Lorinkova

University of Maryland

Abstract

It has long been hypothesized that bidders do not only look to one auction for decision making about the magnitude of their bids. Rather, they also take into account prices from other, simultaneously occurring auctions.

Thus the research goal of this paper is to investigate how the price and bid-level in one auction is affected by the price and bid-level in other auctions - auctions that occur during simultaneously or that have closed just a short time ago. The paper finds out that the bid-level (or winning bid) of online auctions affect each other concurrently – that is the bid-level of the current auction is affected by that of a simultaneous or shortly ago closed auction.

I. Introduction

According to the Langenscheidt's New College Merriam-Webster English Dictionary "concurrence" means "the simultaneous occurrence of events or circumstances". Concurrency can occur on a temporal or spatial scale. A temporal concurrent model is one in which events from the past affect the outcome of present events. Examples of such concurrent events in our daily lives and in the business world are plentiful. Even in the legal world we can observe temporal concurrent events since most of the legal decisions and verdicts are based on precedents – that is the outcome of a case today depends on the outcome of a similar case in the past. Another marketing example, which is very close to the average consumer, is the market value of a real state. What is the value of the house a person owns? Well, if we use the subject term of our paper – it is concurrent on the market value of the neighborhood houses sold not long ago. The market value of a house though, is an example of not only temporal concurrency but also of spatial concurrency since the value of a property in question depends on the value of a property with similar features.

However, despite the omnipresence of concurrency in marketing and consumer events, the concurrency of the online markets has hardly been studied. In a 2005 article Jank and Shmueli (working paper) examine a concurrency model consisting of three components, basing on online eBay auctions. Following in this path, the current research seeks to re-examine the concurrency of live auctions and the main research question of this research is "Is there concurrency in online auctions?"

II. Ebay online book auctions and concurrency

2.1 Data

In order to answer the research question data from eBay book auctions is used. The whole data set contains data for a variety of book auction that took place online on Ebay between October 17th and October 26th 2005. Since books (including variety of subcategories such as non-fiction, textbooks, and children's books) are commonly traded online we deliberately focus on these data as a basis for analysis.

2.2 Data frame

Determining the frame of the data to be used in the analyses is one of the most important tasks of a researcher, since a properly defined data base is a solid foundation for valid research results. This paper is no exception. Time consuming and repetitive data modeling was done until we reached our final set of data, containing 1330 observations. As mentioned above, the initial data contained information about 10 days online book auctions. Keeping in mind our task – to support or reject the proposition that concurrency exist in online auction, we focus on auctions that finish on the same day – namely October 24th. Our rationale here is that books are inexpensive, commonly traded online items, and a consumer is highly unlikely to spend a week, or even a couple of days monitoring the "stock market price" of a less than 10 –dollar item. (Table 1 contains the descriptive statistics of the variables, including that of the wining bid and the mode for the data set is less than 10 USD. The reason that we use the mode and not the mean as a descriptive statistic is the presence of a couple of extreme outliers, which highly tilt the price). We chose the 24th since it contained the highest number of observations and the most diverse dataset (range of prices and winning bid varied most for 10/24th). We also deleted some observations due to missing data, and a number of observations were deleted for simplicity purposes, since they were absolutely identical with the observation before or after them and there was no "action" – for example wining bid $ 0.99, starting bid $0.99 and 1 bidder. Currency differences were also transformed to USD based on the 24th October exchange rate.

Thus our task is to propose a statistical model, which would capture a possible time series in the data set and we believe that auctions which finished a bit earlier (like a couple of minutes or hours) influence the price of current auctions. This line of reasoning is in unison with the data frame we use, since consumers may not monitor the online book market for a week, but this does not mean they are completely unaware of what is going on at the same time (or a little while ago), at the same market (online) with a similar item. We suspect that shortly before making a bid consumers monitor similar items (books) auctions and this influences the final bid they make, and consequently the winning bid for a book.

III. Data analyses

3.1. Determining significant variables

In order to answer our research question we need to construct a statistical model that explains the relationships between the variables. Our first task is to examine the data set and determine which the important variables are. For this, we use different tools – visualization, t-tests, and intuitions. Starting with intuition we determine what might be the variables that play a role in influencing the final price (winning bid) of an online-traded book. The first influential variable should be "Starting bid" since a common logic is the higher the starting bid the higher the winning bid. Total number of bidders (coded as No.of.bids) should also be a factor, since the more the bidders, the higher the competition for an item and the higher the price (at least according to the demand and supply model). For the initial analysis Length (of auction), Winner rating and Seller rating are also included as possibly influential variables. Table 1 contains the descriptive statistics of the variables and gives initial idea about which of them may need transformation due to skewness or kurtosis of the distribution. Attachment 1 contains the histograms and scatterplots of the variable distribution. One variable – positive feedback of seller – is excluded from the analysis, based on the descriptive statistics. The range of this variable is 100 – it ranges from 0 to 100, however its mean is 99.6 and its mode is 100, which means that most of the sellers get the highest possible feedback. Thus, this variable is not likely to add any explanatory value to our model and is dropped from the analyses. All other variables are transformed through log transformation (Attachment 2).

However, for two of the variables the transformation does not improve the distribution – for No.of.bids and Length. As a consequence, the base line statistical model has Winning bid (log-transformed) as the dependent variable and 5 independent variables: Starting bid, Number of bids, Length, Seller rating and Winner rating.

3.2. Initial analysis

The relation between Winning.bid and Starting.bid approximates linear relationship and can be explained by OLS, therefore we use linear regression analysis to prove/disprove the research hypothesis.

Equation 1 shows our baseline regression model, used to determine significant variables:

Winning bid = α +b1Starting.bid + b2Length + b3No.of.bids + b4Seller.rating + b5Winner.rating (1)

Table 1: Descriptive Statistics

Length		*No of bids*		*Winner rating*		*Seller rating*	
Mean	7.067568	Mean	3.936186	Mean	222.72073	Mean	1413.406
Standard Error	0.012744	Standard Error	0.131449	Standard Error	12.757583	Standard Error	73.09335
Median	7	Median	2	Median	75	Median	531
Mode	7	Mode	1	Mode	1	Mode	19
Standard Deviation	0.4651	Standard Deviation	4.797443	Standard Deviation	461.21795	Standard Deviation	2664.651
Sample Variance	0.216318	Sample Variance	23.01546	Sample Variance	212721.99	Sample Variance	7100365
Kurtosis	35.28367	Kurtosis	12.22315	Kurtosis	95.020585	Kurtosis	42.76787
Skewness	5.977321	Skewness	3.078269	Skewness	7.6464407	Skewness	5.424068
Range	4	Range	38	Range	7474	Range	34513
Minimum	6	Minimum	1	Minimum	0	Minimum	0
Maximum	10	Maximum	39	Maximum	7474	Maximum	34513
Sum	9414	Sum	5243	Sum	291096	Sum	1878416
Count	1332	Count	1332	Count	1307	Count	1329

Winning bid		*Starting bid*		*Positive feedback of seller*	
Mean	34.46497	Mean	10.25979	Mean	99.628937
Standard Error	5.113142	Standard Error	2.187432	Standard Error	0.0782278
Median	7	Median	4	Median	99.9
Mode	9.99	Mode	0.99	Mode	100
Standard Deviation	186.5421	Standard Deviation	79.80381	Standard Deviation	2.8496816
Sample Variance	34797.95	Sample Variance	6368.648	Sample Variance	8.1206854
Kurtosis	358.4529	Kurtosis	1126.523	Kurtosis	1129.6226
Skewness	16.64469	Skewness	32.47698	Skewness	-32.47083
Range	4749.99	Range	2800	Range	100
Minimum	0.01	Minimum	0	Minimum	0
Maximum	4750	Maximum	2800	Maximum	100
Sum	45872.88	Sum	13655.78	Sum	132207.6
Count	1331	Count	1331	Count	1327

Table 2 presents the results from the regression analysis. Since both Seller.rating and Winner.rating appear to be insignificant we drop them from further statistical models

Table 2: Regression output

	Coefficients:	Est. Std. Error	t-value	Pr(>|t|)
(Intercept)	-7.746e+02	6.721e+01	-11.525	<2e-16 ***
Starting.bid	8.710e-01	4.796e-02	18.159	<2e-16 ***
Length	1.063e+02	9.729e+00	10.924	<2e-16 ***
No.of.bids	1.195e+01	9.000e-01	13.276	<2e-16 ***
Seller.rating	-6.713e-04	1.390e-03	-0.483	0.629
Winner.rating	3.457e-03	8.044e-03	0.430	0.667

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 133.8 on 1301 degrees of freedom

Multiple R-Squared: 0.4288, Adjusted R-squared: 0.4266

F-statistic: 195.3 on 5 and 1301 DF, p-value: < 2.2e-16

And use the model suggested in equation 2 as our bas model:

Winning.bid= α+ b1 Starting.bid + b2Length +b3 No.of.bids+ἐ (2)

Attachment 3 contains the output summary of the base model, which proves superior to the previous model, which included seller.rating and winner.rating as dependent variables. To better account for the non-even distribution and for observations with value 0 the variable Startign.bid is transformed as (log(Starting.bid + 10) and this transformed variable is used for the regression analyses. (Please see Attachment 3 for the regression output).

3.3 Exploratory analysis

Since we are concerned with concurrency in the data sample and concurrency indicates events that are dependent on each others outcome, it is highly likely that there is autocorrelation between variables in our data. The items that are being traded online in our data are similar (books) and are not affected by geographical features; hence we are more concerned with correlation in temporal context. To check for temporal correlation we use plots of autocorrelation function ACF and create time-series plot.

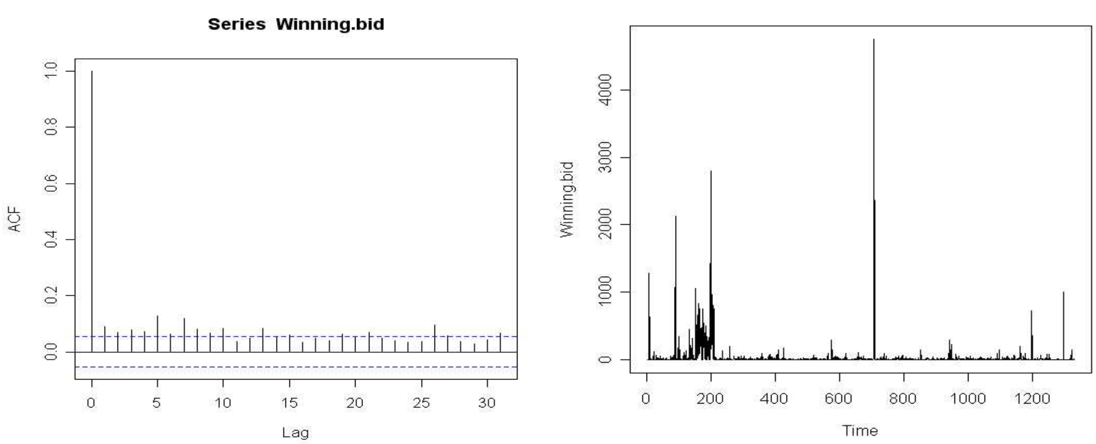


Figure 1: Autocorrelation plots

Fig. 1 plots do not show any clear trend in the time series correlation.
To investigate the matter further a lagged variables is introduced and the winning.bid observations are lagged one period. Thus the DV is Winning.bid and captures observations 1:1329, while the IV in the lagged regression equation is Win.Y.lag capturing observations [2:1330]. Table 3, Model 1 shows the results of the above explained regression.

Table 3: Summary of lagged models

	Model 1	Model 2	Model 3
Residual St. error	186.1 on 1326 d.f.	264.6 on 63 d.f.	158.4 on 937 d.f
Adjusted R-sq.	0.007151	-0.01578	-0.001067
Multiple R-sq.	0.007899	9.391e-05	4.998e-08

As shown in table 3, **Model 1 does not support the research proposition that online book auctions are influenced by concurrency** – both the R-sq. and the Adjusted R-sq. values are very low, which does not suggest a good fit of the model to the data. When the lagged variable is included in the linear regression model, together with the other significant variables (Starting.bid, Length and No.of.bids) the adjusted R-sq. is almost identical to that of the transformed model (see attachment 3) but the lagged variable appears to be non-significant (p-value = 0.825). Thus, the proposition that concurrency influences online book auctions remains unsupported.

Keeping in mind, that time-lag summarizes event distribution of events over equally spaced time periods, it is possible that the above proposed lagged model is wrong, since the events are not completely equally spaced. In order to create a lagged model with equally spaced events the concentration is on shorter time periods in which events occur at almost equal time-spaces. For example, from observation 28 until observation 93 online auctions end almost every minute. Thus, we create a new lagged variable with a time lag equal to 1-2 minutes and regress it on the original variable for the specified observation. The variables we use in this regression, which give us Model 2 (see table 3) are defined as follows:

```
Win.Y<-Winning.bid[28:92]
Win.Y.lag<-Winning.bid[29:93]
```

Model 2 appears to be a worse fit to the data than Model 1 and again fails to show support for the presence of concurrency in online book auctions.

In a final attempt to test the research proposition with lagged variable we focus on a different part of the data set, which contains auctions ending every 3-20 sec. Thus, the variables we use are defined as follows:

```
> Y<-Winning.bid[207:1145]
> Y.lag<-Winning.bid[208:1146]
```

Model 3 in Table 3 summarizes the results of these variables regressed and **again fails to suggest concurrency.**

IV. Additional Analyses

Although not supported by the above analyses, the question remains whether online auctions exhibit concurrency in their winning bids. Maybe the models suggested above failed to capture concurrency? Maybe additional factors play a role in determining winning bids and how they influence each other. One possible explanation is that although online book auctions trade similar items these items still exhibit spatial components. To test whether almost identical items auctions are concurrent on each other we split the data set and focus on items very close in features and therefore in prices. We focus on high-end items, with price 50 USD and higher, which are mostly antiquarian books and attract more bidders. The final data-set for this analysis contains 103 observations with 14.3 average numbers of bids. The average book price is $ 336.5.

Equation 3 presents the baseline regression equation:

$$\log(\text{Winning.bid}) = \alpha + b1\log(\text{Starting.bid}) + b2\text{Length} + b3\text{No.of.bids} + \acute{\varepsilon} \quad \textbf{(3)}$$

Model 1 in Table 4 summarizes the results of the regression.

Table 4: Additional Models summaries

	Model 1	Model 2	Model 3
Residual St. error	0.8769 on 99	599.6 on 100 d.f	0.9803 on 97
Adjusted R-sq.	0.2426	-0.009414	0.0626
Multiple R-sq.	0.2649	0.0005799	0.09972,

Model 2 summarizes the results for the temporal model, which regresses the Winning bid on its lagged values. The lag is specified as time period between 10-60 minutes. Model 2, although consisting of similar items, fails to show support for the presence of concurrency in online book auctions. Model 3 is our last step in addressing the research question, namely including the lagged variable in the base line regression equation (Equation 4):

$$\text{Log(Winning.bid)} = \alpha + \text{b1Winning.bid.lagged} + \text{b2log(Starting.bid)} + \text{b3Length} + \text{b4No.of.bids} + \acute{\varepsilon} \quad \textbf{(4)}$$

The inclusion of the lagged variable does not improve the model data fit over the baseline model, which again fails to show support that concurrency exist in online book auctions.

V. Conclusion

Although our results did not support the initial assumption that concurrency exists in online book auctions we do not believe that these research paper gives answers to all question about online auctions concurrency.

One possible explanation for the negative results is the nature of the data itself. Since, as already mentioned, books, are mostly inexpensive items their traders may not take the time to monitor the market. Why we failed to show support for concurrency in expensive antiquarian books is one of the enigmas, but our inclination is to explain the result with the inadequacy of the chosen data. It is possible that online concurrency for expensive book items exists in days, not minutes and hours lag. A suggestion for future research is to focus on a week, or even a month long observations of expensive book items and see how similar items' auctions influence each other.

Another possible explanation for the negative results is that this reflects the true picture of the online book auctions – maybe buyers and sellers are influenced by other items characteristics when determining their final bids, and not by how much somebody paid for a similar item.

In conclusion, we believe that, as suspected, the major reason for the non-concurrency of the online book auctions is (as a whole) the inexpensive price of the items traded and the influence of other factors (such as number of bidders and specific item characteristics on the final price).

References:

Draper, N.R., & Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). A Wiley-Interscience Publication

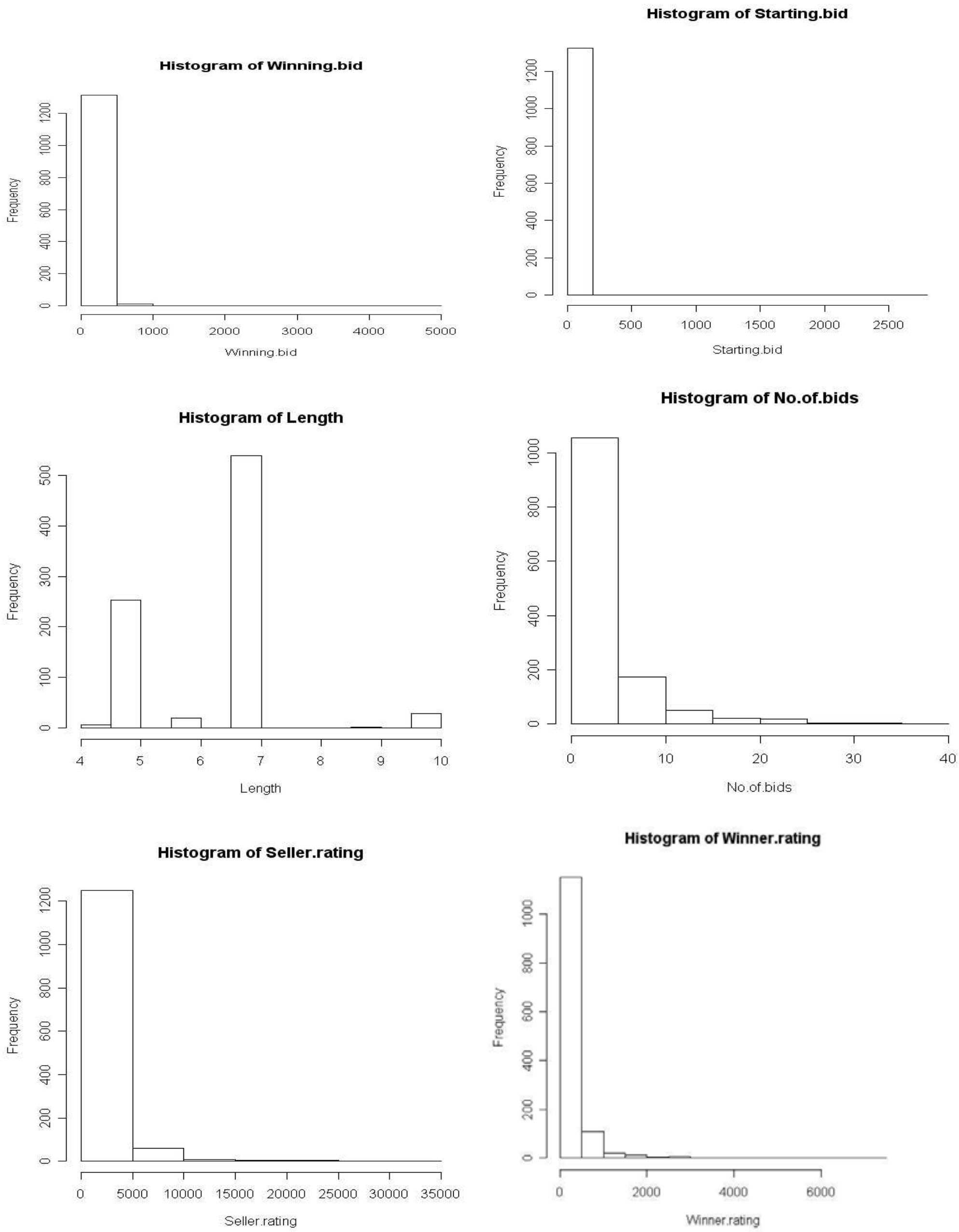
Jank, W., & Shmueli, G. (2005). Modeling Concurrency of Events in Online Auctions via Spatio-Temporal Semiparametric Models. *Working Paper*

Johnson, A.G. (1988) *Statistics.* Orlando, Fl: Harcourt Brace Jovanovich, Inc.

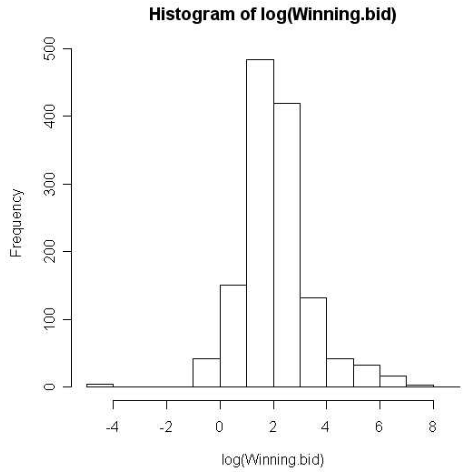
Neville, J., Simsek, O., & Jensen, D. Autocorrelation and Relational Learning: Challenges and Opportunities

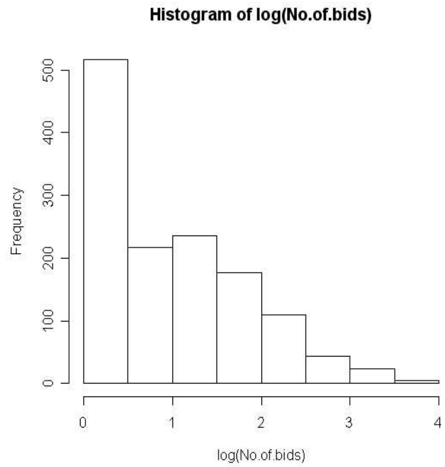
Reddy, S. K., & Dass, M. (2005). Modeling Online Art Auction Dynamics Using Functional Data Analysis – *submitted to the Statistical Science*

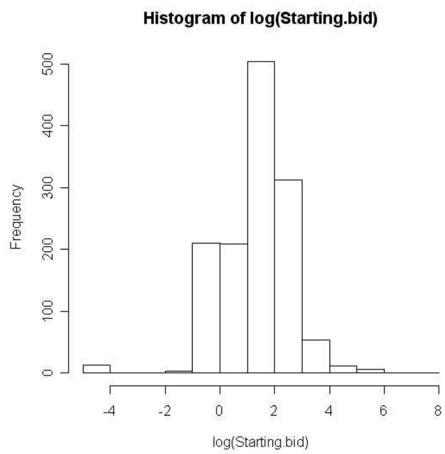
Attachment 1: Histograms of variables distribution

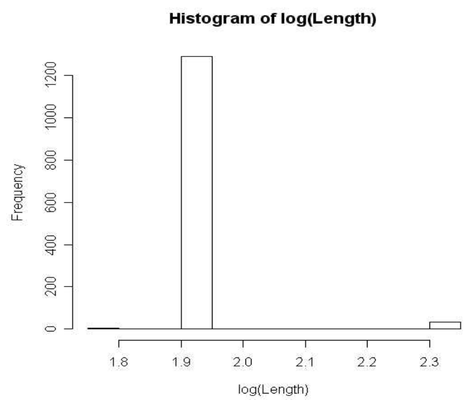


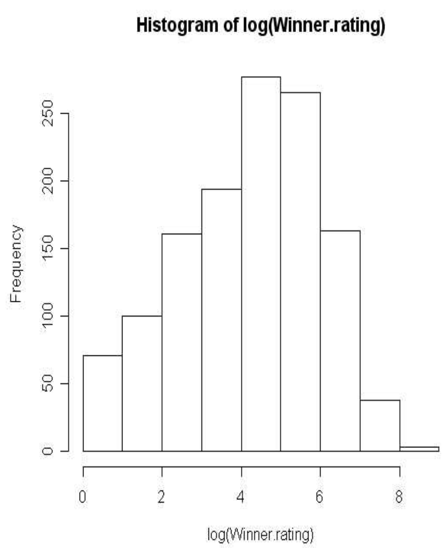
Attachment 2: Histograms of log-transformed variables

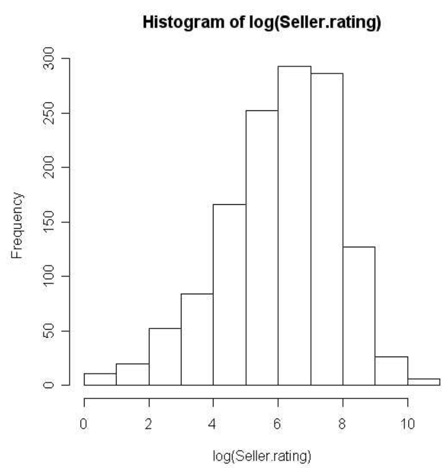












Attachment 3:

Base model output:

Winning.bid= α+ b1 Starting.bid + b2Length +b3 No.of.bids+έ

Residuals:

Min	1Q	Median	3Q	Max
-552.142	-16.277	7.927	20.417	3963.356

	Coefficients	Estimate Std. Error	t value	Pr(>|t|)
(Intercept)	-759.6694	65.9712	-11.52	<2e-16 ***
Starting.bid	0.8832	0.0498	17.73	<2e-16 ***
Length	103.8315	9.5605	10.86	<2e-16 ***

Transformed Model output:

log(Winning.bid) = a+b1log(Starting.bid + 10) +b2 Length + b3No.of.bids

	Coefficients:	Est. Std. Error	t-value	Pr(>|t|)
(Intercept)	**-4.776128**	**0.309762**	**-15.419**	**<2e-16 *** **
Log(Starting.bid +10)	**1.726583**	**0.042678**	**40.456**	**<2e-16 *** **
Length	**0.207132**	**0.044880**	**4.615**	**4.31e-06 *** **
No.of.bids	**0.162028**	**0.004283**	**37.828**	**<2e-16 *** **

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6672 on 1325 degrees of freedom

Multiple R-Squared: 0.7433, Adjusted R-squared: 0.7428

F-statistic: 1279 on 3 and 1325 DF, p-value: < 2.2e-16

Lagged model:

log(Y) = a+ b1Y.lag + b2log(Starting.bid + 10) + b3Length + b3 No.of.bids

Y= Winning.bid

Y.lag = Winning.bid.lagged

	Coefficients:	Estimate Std. Error	t value	Pr(>|t|)
(Intercept)	-4.762e+00	3.155e-01	-15.096	< 2e-16 ***
Y.lag	2.229e-05	1.009e-04	0.221	0.825
log(Starting.bid + 10)	1.725e+00	4.283e-02	40.286	< 2e-16 ***
Length	2.055e-01	4.553e-02	4.514	6.94e-06 ***
No.of.bids	1.620e-01	4.291e-03	37.761	< 2e-16 ***

Residual standard error: 0.6676 on 1323 degrees of freedom

Multiple R-Squared: 0.7434,   Adjusted R-squared: 0.7426

F-statistic: 958.1 on 4 and 1323 DF, p-value: < 2.2e-16