Exploratory Statistical Study of E-Bay Textbook Auctions

by

Wen MIN

A Major Paper
Submitted to the Faculty of Graduate Studies and Research
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master Of Science
at the University of Windsor

Windsor, Ontario, Canada
2005

Exploratory Statistical Study of E-Bay Textbook Auctions

by Wen MIN

M.Sc. Major Paper

APPROVED BY

Dr. M. Hlynka, Supervisor

Department of Mathematics & Statistics

University of Windsor

Dr. Karen Fung, Departmental Reader

Department of Mathematics & Statistics

University of Windsor

Date

Acknowledgements

I would like to thank my supervisor Dr. Myron Hlynka for his guidance, support and encouragement through my graduate study. I would like to thank my departmental reader Dr. Karen Fung for her valuable course instruction and her time in examining this major paper. I would like to thank my family, especially my daughter who had to learn to do many things by herself. I would like to thank all the faculty and staff in the department of Mathematics and Statistics at the University of Windsor for their friendship, care and help to me.

Abstract

This paper reports an exploratory analysis and modeling of online auction data for text books at the eBay website, and also provides a number of descriptive statistics on patterns with some facts and figures relating to eBay auctions. We examined 1485 different textbook auctions, each with at least 5 bids. We divided our auctions into categories based on the number of bids. Category 1 consisted of auctions with 5, 6, 7 bids, Category 2 consisted of auctions with 8, 9, 10 bids, … Category 8 consisted of auctions with 26, 27, 28 bids. Our data from eBay show that the category number follows a truncated geometric distribution.

Regression models were used in this work to explore, summarize and test hypotheses about relations between some of the variables. We examined relations between variables such as the average bid and the average high bid (each person has a highest bid), the winning bid, and the interaction of the number of bidders with the average bid.

Some variables were discovered to be approximately normally distributed. An example is the ratio of the winning bid to the product of the number of bids and the average high bid.

Table of Contents

Exploratory Statistical Study for E-Bay Auctions of Textbooks

by Wen MIN

Section 1. Introduction

eBay is a successful online auction website, where people from all over the world buy and sell thousands of items every day. Statisticians and scientists of human behavior have shown their interest in auctions in general, and in on-line auctions in particular, so that the literature on auctions and eBay is fairly extensive.

For instance, Baron (2002) gives a general discussion of eBay auctions, including economics, trust, reputation, feedback, accounting, punishment, privacy, and other issues. Lucking-Reiley et al. (2000) gave graphs and histograms on prices, volumes, seller reputations and other features from 20,000 auctions of (somewhat) rare United States pennies. Ockenfels and Roth (2002) gave explanations of bidder behaviour and also studied theory and evidence concerning different rules for ending an auction (Ockenfels and Roth (2003)). In Roth and Ockenfels (2002), there is a discussion of last minute bidding, which we later show makes up a large proportion of the total number of bids. Their paper also includes the timing of bids in auction, which is useful for market design. Shmueli and Jank (2004) modeled the dynamics of an online auction, applying a statistical approach. They used functional data analysis, cluster analysis and regression-type models to explore and summarize the data.

Most research is in multiple categories, in a macro sense. However, auctions for different categories have different patterns, and there are many variables to consider, such as the average bidding price, the average number of people being attracted by the auctions and the number of auctions going on everyday. For example, conditioned on the number of bids being at least five, the average number of bids for textbooks is close to 9 but the average number of bids for antiques is much larger. Moreover, bids on eBay antique auctions were even more concentrated near the end of the bidding period than those of eBay textbooks auctions, according to Lucking-Reley (2000). Considering those differences, this paper our aims to study the special features and structure of the restricted category of online textbook auctions.

The eBay auction as a kind of economic transaction was born in 1995 and is getting better for both sellers and buyers. Consumers are getting more experienced, and eBay policies have changed to attract more people. For example, since 2003, eBay began to allow sellers to specify a "buy-it-now" value. This allows buyers to buy at a fixed price, and if the seller has multiple copies of the same item, this gives the items more chance to be sold. Such changes will affect the future of eBay, and research in this field at a later time may show more improvements in eBay. This paper describes eBay bidding patterns for textbooks, attempts to deduce a reasonable explanation for the patterns, and presents some simple graphs that may aid in evaluating bidding performances. We also present some strategies that can help both sellers and buyers.

As a first step, a major effort was made to collect data on textbook auctions during part of 2004. We examined and assembled eBay records in order to collect a large text data set that could be used for analysis. Our sample can be considered to be a cluster sample of two months taken from the population of textbook auction over a large time period.

We approached the data analysis by simple exploration of the data before moving to more formal confirmatory analysis. We used graphical displays, summarization, and data reduction techniques to understand the data structure, features, and complexity. In order to describe existing bidding patterns, Excel, SPSS and SAS were used.

Some parametric and nonparametric models were used for this project. In this paper, we answer questions of the following type. Is there a relationship between the variables of interests? Do the data for eBay textbook auctions follow any particular distribution? It was observed that some linear relationships exist between some variables related with book auctions. Also we show that some variables approximately follow a normal distribution or a lognormal distribution.

Section 2. Details of an eBay Auction

Variable names that are used in this paper are given in table 1 as follows. The terms apply to a single auction. The bids are numbered 1,2,…,N over time. The bidders are numbered 1,2,…,M. The Winning Bid refers to the amount paid by the highest bidder, even though the actual bid may have been higher.

Table 1. Variables definition

Definition	Computation	Symbol
Amount of bid n		Bn
Highest bid for person m		Bm
Number of bids		N
Number of bidders		M
The average bid	$\Sigma(Bn)/N$	AverageBn
The average high bid	$\Sigma(Bm)/M$	AverageBm
Winning bid		WinningB
Second highest bid		Second-highB

Figure 1 displays an example of a bidding page for an ongoing eBay auction, which includes all the information for the buyers, the seller's information, the payment and the shipping.

Figure 1. eBay web page for an auction

In an actual eBay screen, if you click on "history" in Figure 1, Figure 2 will show up.

Figure 2. An actual Bid history for an on going auction

User ID	Bid Amount	Date of bid
dustin2158(0)	---	Aug-23-04 19:44:55 PDT
brenderc4(12)	---	Aug-23-04 18:46:34 PDT
dustin2158(0)	---	Aug-23-04 19:44:44 PDT
dustin2158(0)	---	Aug-23-04 19:44:27 PDT
dustin2158(0)	---	Aug-23-04 18:35:42 PDT
brenderc4(12)	---	Aug-23-04 16:09:54 PDT
dustin2158(0)	---	Aug-23-04 18:35:21 PDT
dustin2158(0)	---	Aug-23-04 15:16:13 PDT

Figure 2 displays bidding history such as the start price, the current price, the time remaining, the current number of bids n, the current number of bidders m, and the minimum allowable bid at the current time. Hence, Figure 1 and Figure 2 include all information that each bidder knows during the process of bidding. From Figure 2, we can see that a bid history is a time series. Figure 2 is no longer typical of what is displayed during an auction. Now the current bids of all but the highest are revealed as the auction continues.

The seller sets the "starting bid". If not, it is set to zero. An "increment" is also set for the bidding. The increment is set automatically by eBay as follows based on the current price.

Table 2: Increment Levels

Current Price	Bid Increment
$ 0.01 - $ 0.99	$ 0.05
$ 1.00 - $ 4.99	$ 0.25
$ 5.00 - $ 24.99	$ 0.50
$ 25.00 - $ 99.99	$ 1.00
$ 100.00 - $ 249.99	$ 2.50
$ 250.00 - $ 499.99	$ 5.00
$ 500.00 - $ 999.99	$ 10.00
$ 1000.00 - $ 2499.99	$ 25.00
$ 2500.00 - $ 4999.99	$ 50.00
$ 5000.00 and up	$ 100.00

The first actual bid must be at least as large as the starting bid. The "current price/bid" is set to the starting bid as soon as the first bid is made. After that, there is a "minimum allowable bid" for further bidding. How is the minimum allowable bid determined? If there has been only one bidder thus far, then the minimum allowable bid is set to the starting bid plus the increment. If there has been more than one bid, then the minimum allowable bid is set to the "current price" plus the increment. If the most recent bid is less than the highest bid by more than the increment, then eBay automatically increases the amount to be taken from the high bidder (if successful) and current price is the second highest bid plus the increment. If the most recent bid is less than the highest bid, but the difference between the highest bid and the most recent bid is less than increment, then the current price is the highest bid. If the most recent bid is the highest bid, then the current price is also the most recent bid. This can be expressed as the minimum of the second highest bid plus the increment and the highest bid. At the end of the auction, the bidder who submitted the highest allowable bid wins the auction and pays the "paying bid." The paying bid is not necessarily the actual bid of the highest bidder. It may be less. The paying bid is always the current price after the last bidder. Postage and other costs are extra.

eBay's Policy

Bidders are always informed about the current price as the auction progresses. The magnitude of the highest submitted bid is, however, never revealed to bidders, who only see the final sale price, or paying bid. The following Figure 3 shows the information available after an auction ends.

Figure 3. An example of a bidding history page for a finished eBay auction

User ID	Bid Amount	Date of bid
dustin2158(0)	US $60.00	Aug-23-04 20:09:50 PDT
brenderc4(12)	US $59.00	Aug-23-04 19:46:09 PDT
dustin2158(0)	US $59.00	Aug-23-04 20:09:41 PDT
dustin2158(0)	US $56.00	Aug-23-04 19:44:55 PDT
brenderc4(12)	US $55.00	Aug-23-04 18:46:34 PDT
dustin2158(0)	US $55.00	Aug-23-04 19:44:44 PDT
dustin2158(0)	US $52.50	Aug-23-04 19:44:27 PDT
dustin2158(0)	US $50.50	Aug-23-04 18:35:42 PDT
brenderc4(12)	US $50.00	Aug-23-04 16:09:54 PDT
dustin2158(0)	US $48.50	Aug-23-04 18:35:21 PDT
dustin2158(0)	US $46.00	Aug-23-04 15:16:13 PDT
brenderc4(12)	US $45.00	Aug-20-04 20:46:34 PDT

Other bidding restrictions also apply. Bids always had to meet or exceed the current minimum acceptable bid, which is one increment over the 'current price', if there is at least one acceptable bid. The increment size can be \$0.25, \$0.50, \$1.00 or something else. Since there may be several people bidding near the end of an auction, eBay may not be able to update itself fast enough to exclude some bids, which would otherwise not have been allowed. If more than one bidder submitted the highest bid, the bidder who submitted the high bid first becomes the high bidder, and wins the auction. If identical bids were submitted simultaneously, one bidder is randomly chosen to be the high bidder.

As already stated, the bid history such as that of Figure 3 will be shown after the auction ends. (A partial history is shown during the auction, see figure 2). Figure 3 shows the time and date of each bid, the bid amount and bidders' ID. At the end of each bid, the high bidder and current price are displayed to all. A great deal of information on eBay auction is publicly available. Anyone may view the listings of past auctions of items for sale on eBay's site up to a half month ago. Until the year 2002, eBay kept the past history information for one month ,according to Roth and Ockenfels (2002).

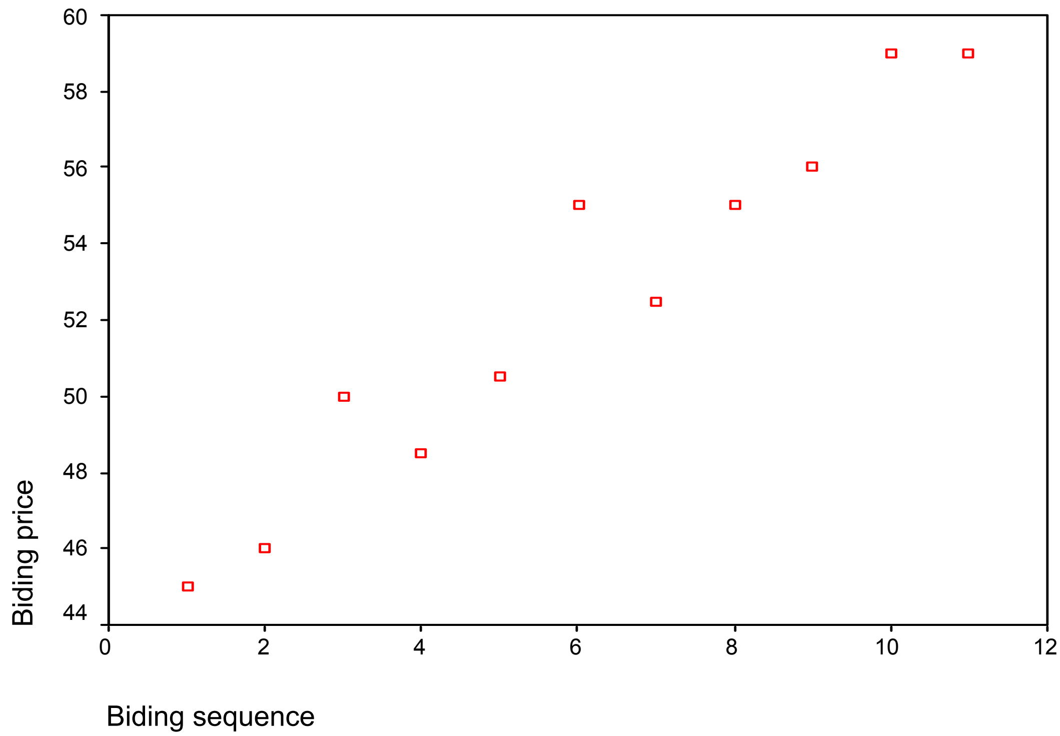
From Figure 3 we can see that only two people, brenderc4(12) and dustin2158(0) took part in this auction. brenderc4(12) bid 4 times and dustin2158(0) bid 8 times, with dustin2158(0) finally winning. The number (12) after brenderc4 indicates that brenderc4 has been the winner of 12 other auctions. Brenderc4 therefore has a history of payments to sellers, whose rating of brendercr4's reliability is available for all to see. On the other hand, dustin2158 has the number (0), which means that there is no successful buying history. Some sellers place restrictions on the bidders to exclude buyers with an unproven history.

The starting bid was probably set at 45. The first bid, by brenderc4, is 45. Then the current price is also 45. Since the increment is 1, the minimum allowable bid is 45+1=46. In fact, the second bid is 46, and it is acceptable. The current price becomes 46 and the minimum allowable bid is 46+1=47. The third bid (in time) exceeds the minimum allowable bid and the third bid is 50. At this point, the current bid is 47 and the minimum allowable bid is 48. The amount 50 is not known by anyone other than the person who bid that amount. The fourth bid is 48.50, which is above the minimum allowable bid. However, it is more than an increment below the (unobserved) maximum bid so the current price becomes the second highest bid plus the increment, namely 49.50. The minimum allowable bid is 1 increment above the current price, namely 49.50+1=50.50. The fifth bid is exactly 50.50, putting dustin2128 in the lead. The current bid is set to 50.50, and the minimum allowable bid is 50.50+1=51.50. The sixth bid is 55, made by brenderc4. The current price thus becomes 51.50 and the minimum allowable bid is 52.50. The seventh bid is 52.50, which fails and the current price is set to 53.50, while the minimum allowable bid is 54.50. The eighth bid exceeds the minimum allowable amount and has value 55. This bid ties the maximum bid but would lose if the

auction were to close at this point because the other 55 bid came earlier. The current bid is set to 55 and the minimum allowable bid is 56. dustin2128 then bids 56, followed by brenderc4 with 59, followed by dustin2128 with 59 (but in second place because of time), followed by dustin2128 at 60. It could be that brenderc4 gave up bidding at this point because the price was too high, or the auction closed before brenderc4 could submit a higher bid.

Online auction data usually arrive in the form of a set of bids recorded over the duration of an auction. Let Bn be (the amount of) bid n, for n=1, … , 12. Then Bn occurs before bid B(n+1) for n=1, … , 11. The integer n of Bn is the number of the bid. In Figure 4, we plot the bid versus the number of the bid.

Figure 4. Bid price with the bid sequence



Sellers' Policy

When a seller lists his/her goods or services for auction at eBay, she provides a short title and a long description of the item (she may place a photograph). The seller chooses a number of parameters to specify how the auction will run, such as the opening bid amount, reserve price, the length of the auction (one, two, three, five, seven, or ten days), and the time when the auction starts.

Sometimes sellers set a reserve price. This is defined to be the minimal amount that has to be bid in order that the seller concedes his/her property rights for the object. If the highest bid fails to reach at least the reserve price, the seller keeps the object (abstains from sale). Although a reserve price reduces the probability of a sale, it can improve the seller's expected return because it forces bidders with higher valuations of the item to bid more than they might bid otherwise. Sellers have the opportunity to modify the reserve price during the auction under certain circumstances, but such changes are rare.

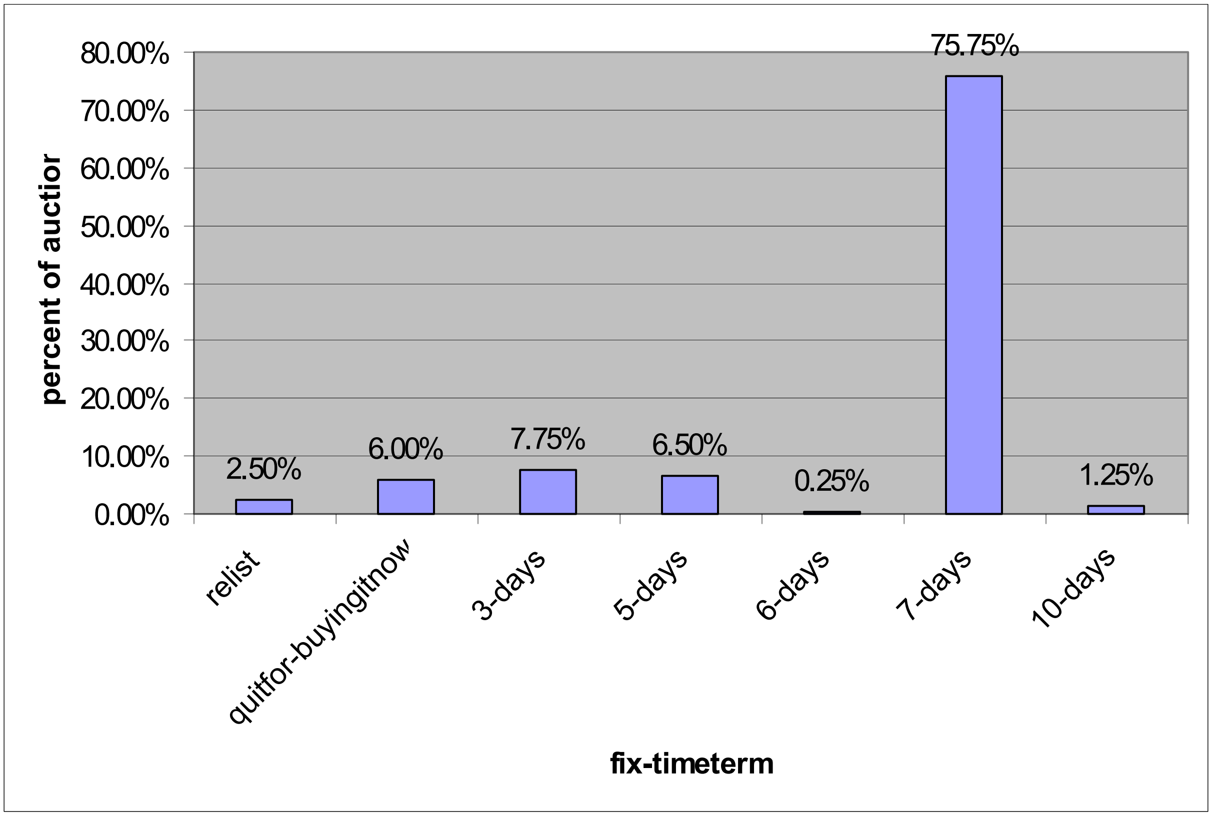
Sellers can also select the "Buy it now" option, together with the reserve price auction, which means that the item may be bought at "buy it now" price, or at the lower price of the winning bid.

Section 3. Description of the Data

The data on which this study is based are the bid history tabulations in the category of books for auctions that ended during June, July, August and September 2004. We started by plotting parts of the raw data accompanied by the summaries and characteristics.

We know that sellers decide the start time and the fixed end time, (a "hard close"), that is, eBay auctions end at a scheduled time. An individual auction on eBay lasts between one to ten days as selected by the seller. Most sellers choose seven days. Figure 5 displays a volume distribution of auctions by auction length. Most auctions, over 70%, were set seven days in length, while only about 7% each were set as three days or five days in length. The six days in length case rarely happened. Relist=2.5% means that the auction ended early before the fixed time, because the seller cancelled the auction and reedited his description and re-listed on the eBay later. This happened sometimes. "Quitfor-buyingitnow" means the auction ended before the fixed time because the item was bought by "buy it now"-buyers. 6% of auctions ended this way.

Figure 5. Histogram of auctions' percent by auction length

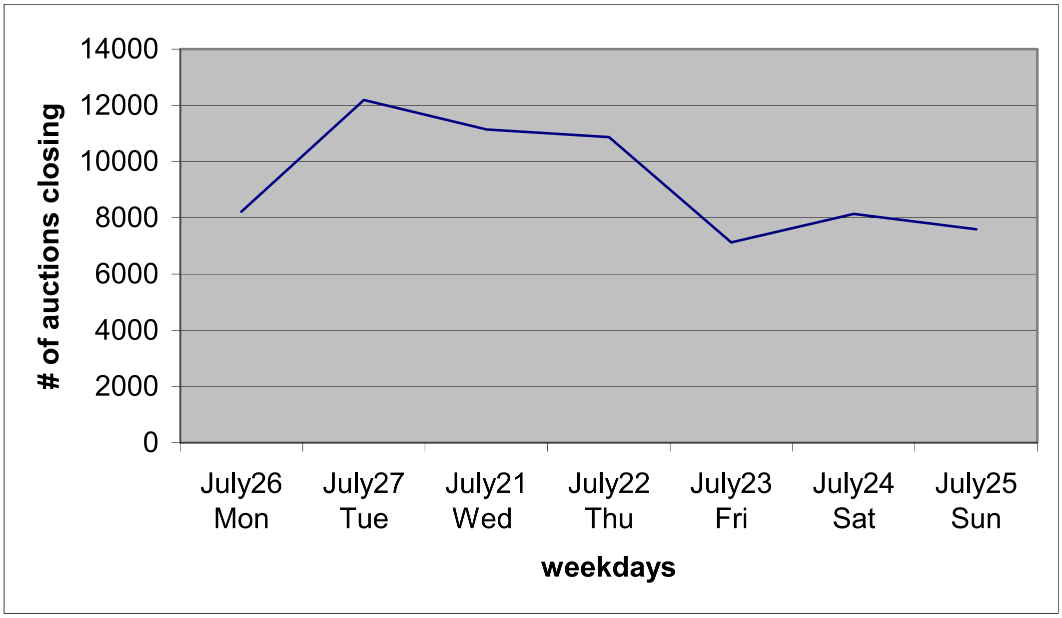


Lucking-Reley, Bryan, Prasad and Reeves (2000) observed that a higher proportion of auctions closed on a weekend day compared to a weekday for antique auctions. However in our research, there was no evidence that the sellers had a preference to close the auction on weekend days. One reason may be that eBay is getting global and Monday morning in Tokyo, Japan is Sunday night in New York, i.e. different time zones dim the weekend effect. Another reason for the difference between the two results is that our research used data collected during the summer, which can be a vacation season for many people. So the weekend effect of earlier researchers did not show up in our distribution of the auction closures. We can see detailed information on auction closures in Table 3. We examined 65273 textbook auctions that closed in the week from July 21 to July 27, 2004. The average number of textbook auctions closing each day is nearly 10000 in this week. Figure 6 shows the number of auctions closing by day-of-the-week. For these data, it is not true that the volume is heaviest on weekends.

Table 3. The number of auctions closing in one week

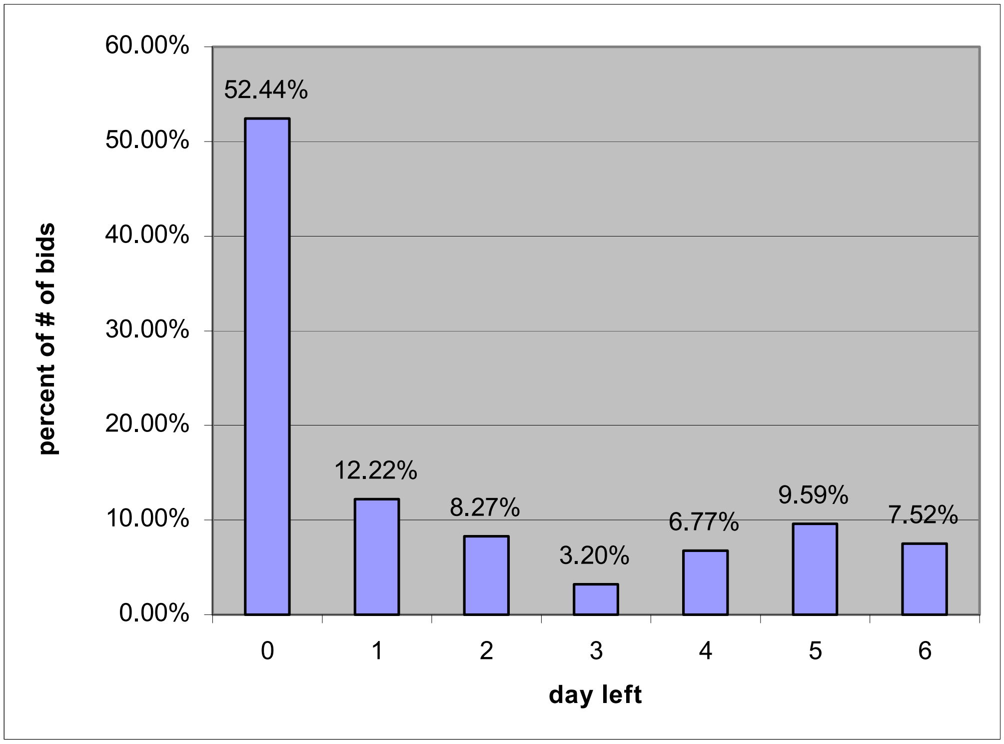
July26 Mon	8223	12.60%
July27 Tue	12187	18.67%
July21 Wed	11142	17.07%
July22 Thu	10866	16.65%
July23 Fri	7126	10.92%
July24 Sat	8137	12.47%
July25 Sun	7592	11.63%

Figure 6. Volume auctions closed by day of the week



We next consider the distribution of the time of bids for seven day auctions, measured from the start of Day 1. Figures 7, 8 and 9 give information about the distribution of the time of bids. eBay auctions begin to be more active a day before the scheduled end time, and half of the bids come on the final day. More than 25% of all bids in an eBay auction occur in the last hour. Figure 7 shows the bidding distribution by the day left. We see that over half of the bids were received on the last day.

Figure 7. Bidding distribution by the day left



Our sample to study the bidding times consisted of 589 bids. Figure 8 shows the conditional distribution for the last hour of the last day, among the bids on the last day. We observed that almost half of the last day bidders bid in the last hour. Figure 9 presents the conditional bidding distribution in the last hour by minute. 85% of last hour bidders bid at the last 30 minutes.

Figure 8. Bidding distribution by the hour left in the last day

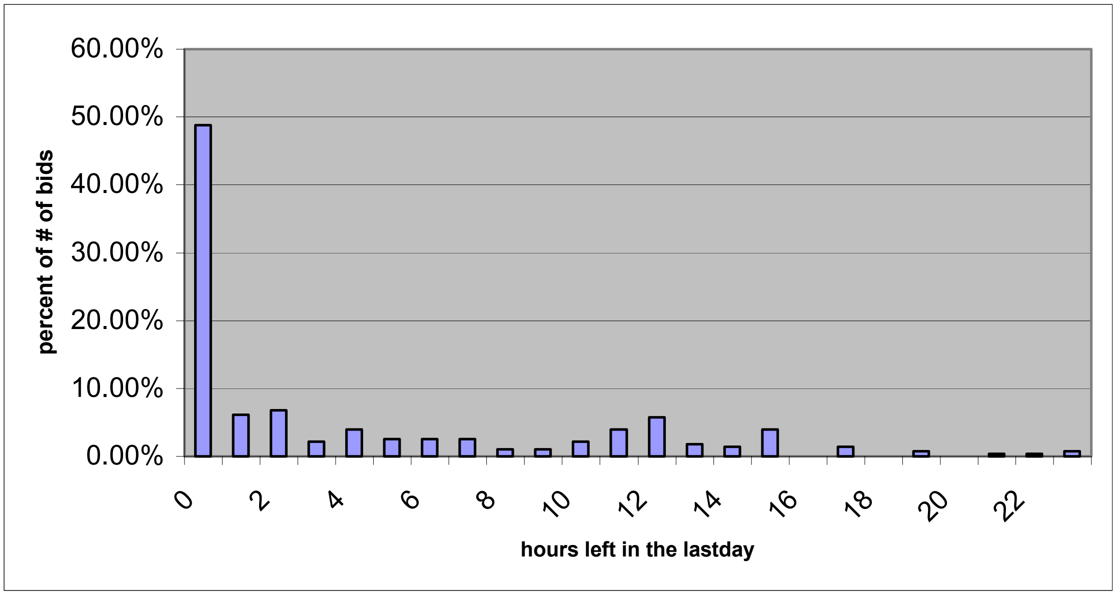
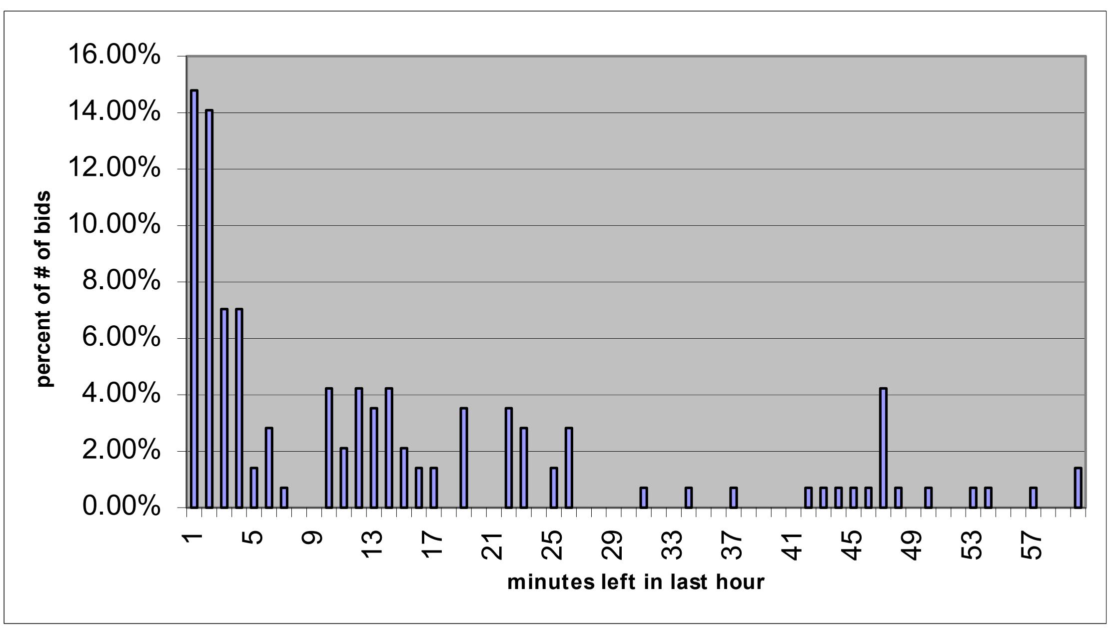


Figure 9. Bidding distribution in the last hour by minute



Why does this happen on eBay? Ockenfels and Roth (2002) give a behavioral explanation. One reason is a strategic response to incremental bidding. For a second price auction conducted over time, early bids give other bidders time to respond, but can be submitted with certainty, while very late bids do not give other bidders time to respond. However, late bids have a danger that they will not be successfully transmitted.

A summary of the distribution of bids over time in our sample (Education & textbooks) auctions is presented in Table 4.

Table 4. Frequencies of late bidding in eBay books' Auction

	Share of all bidders' last bids
In last day	52.4%
In last hour	25.6%
In Last 30 minutes	21.8%
In Last 20 minutes	19.1%
In Last 10 minutes	13.3%

Section 4. Bidding Patterns in eBay Auction

4.1 Geometric distribution of number of auctions grouped by number of bids

Different auctions might attract different numbers of bidders. While some auctions had no bid, we only considered auctions with at least 5 bids. A lower number of bids was not considered competitive enough to allow for a study of auction competition. Of the auctions we studied, the most frequent number of bids was five. In a random sample of 1485 textbook auctions with at least 5 bids, very few auctions attracted more than 34 bidders. Table 5 shows the original data and gives the number of auctions vs. number of bidders. Figure 9 shows this information graphically. Visual examination of these data (shaded part) gives the immediate impression that the number of bids decreases with number of auctions (having that number of bids).

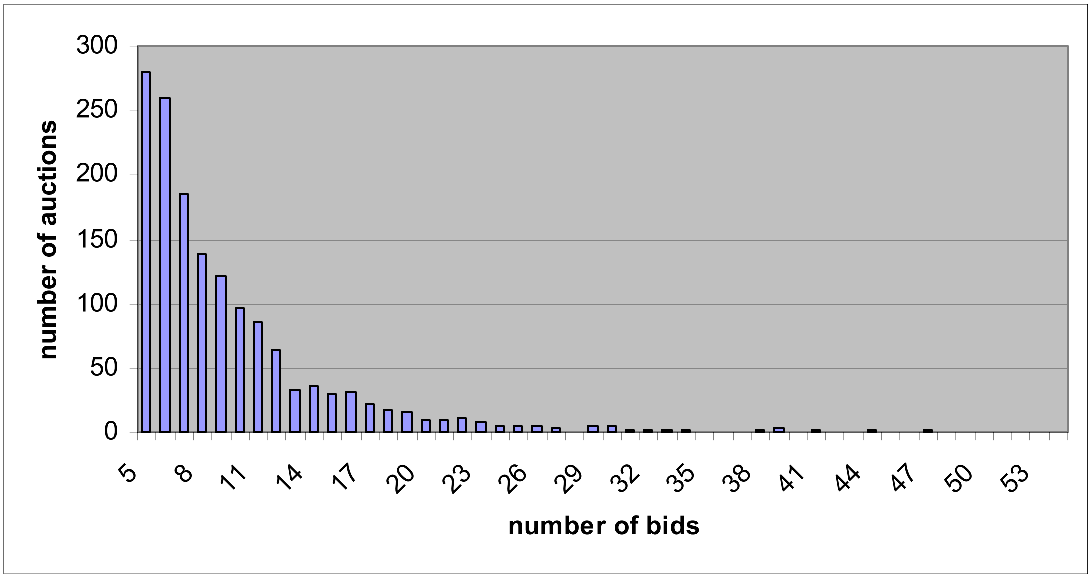
Table 5. The number of auctions grouped by the number of bids

i bids	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
# of auctions with i bids	**280**	**259**	**185**	**138**	**121**	**96**	**85**	**64**	**32**	**36**	**29**	**31**	**21**	**17**	**16**	**10**	**10**

i bids	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**
# of auctions with i bids	**11**	**8**	**4**	**4**	**4**	**3**	**0**	**5**	**4**	**2**	**1**	**1**	**1**	**0**	**0**	**0**	**1**

i bids	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**
# of auctions with i bids	**3**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Figure 9. The distribution of auctions by number of bidders



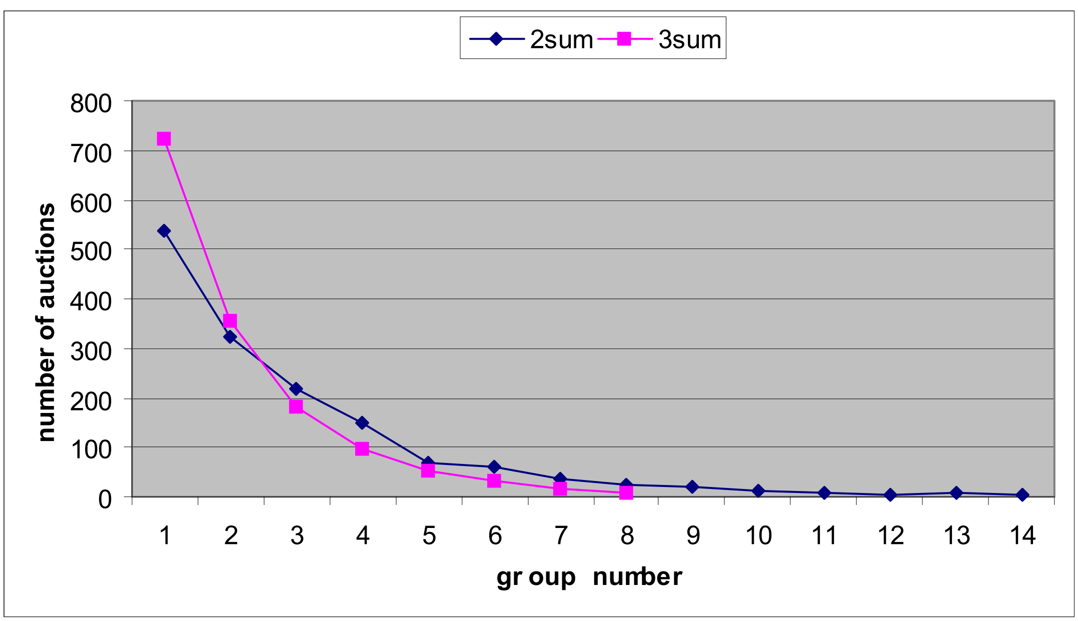
In order to estimate unknown parameters of the population, the sample data had to be processed first. It is very obvious that there are few auctions when the number of bids is greater than 34, and the data appear to behave very well if the number of bids is less than 29, so we truncate the data and considered the data with the number of bids less than 29.

First, the data were smoothed. Table 6 shows the smoothed data and indicates the method applied. The methods included a "two step sum" and a "three step sum". The "two step sum" summed counts of the number of auctions with the lowest two number of bids, 5 and 6, then summed the number of auctions with the next two lowest numbers of bids, 7 and 8, and so on up to 30. This gave 13 groups. The "three step sum" summed the number of auctions with the three lowest numbers of bids, 5 and 6 and 7, then summed the number of auctions with the next three numbers of bids, 8 and 9 and 10, and so on up to 28. This gave us 8 new groups. From Figure 10 we found the three step sum is smoother than the two step sum and we chose to work with the three step grouping in what follows.

Table 6. Data smoothing data with "3-step sum" and "2-step sum"

Category X	# of bids	# of auctions	# of bids	#of auctions
1	5,6,7	724	5,6	539
2	8,9,10	355	7,8	323
3	11,12,13	181	9,10	217
4	14,15,16	96	11,12	149
5	17,18,19	54	13,14	68
6	20,21,22	31	15,16	60
7	23,24,25	16	17,18	38
8	26,27,28	7	19,20	26
			21,22	21
			23,24	12
			25,26	8
			27,28	3

Figure 10. Comparing methods of data smoothing, 3 step sum and 2 step sum



So far, we divided our auctions into categories based on the number of bids.

Category 1 consisted of auctions with 5, 6, 7 bids, Category 2 consisted of auctions with 8, 9, 10 bids, etc. Let X be the category number. Almost 1/2 of auctions happened in the X=1 category.

We next try to find the distribution of X where X is the category number X=1, 2, 3, 4, 5, 6, 7, 8. The decreasing form of Figure 10 suggested that we try to fit a truncated geometric distribution. We choose a truncated geometric distribution rather than a geometric distribution because examination of the data suggests that we get an unusually high number of bids more frequently than would be predicted under a standard geometric distribution.

Suppose X has a truncated Geometric distribution with parameter p. Then $f(x)=p(1-p)^{(x-1)}/\sum_{i=1}^{8}[(1-p)^{(i-1)}\ p]$, for x=1,…,8. We give several methods to estimate p.

Method 1. Simple Method

$f(1)=(1-p)^{0}p/\sum_{i=1}^{8}[(1-p)^{(i-1)}\ p]=1/\sum_{i=1}^{8}[(1-p)^{(i-1)}\ ]$

$f(2)=(1-p)\ f(1)$

$f(3)=(1-p)^{2}\ f(1)$

…

$f(8)=(1-p)^{7}\ f(1)$

If we had a geometric (rather than a truncated geometric) then

$f(1)=p$

$f(2)=(1-p)\ f(1)$

$f(3)=(1-p)^{2}\ f(1)$

…

$f(8)=(1-p)^7 f(1)$

So one estimate of p would be the relative frequency of category 1 among the original 1485 data points.

Thus $\hat{p}$ =724/1485=0.487542. We use a goodness of fit test to see how closely the data follow a truncated geometric distribution. See table 7.

H0: X follows a truncated geometric distribution

H1: X does not follow a truncated geometric distribution.

Table 7. Estimate p Using simple method

X	Frequency	p(1-p)^(x-1)	f(x)	Expect frequency	Chi-sq
1	724	**0.487542088**	0.489872047	717.172676820	0.064995
2	355	0.249844800	0.251038806	367.520812506	0.426563
3	181	0.128034945	0.128646823	188.338948194	0.285975
4	96	0.065612520	0.065926082	96.515784140	0.002756
5	54	0.033623655	0.033784342	49.460277215	0.416679
6	31	0.017230708	0.017313054	25.346310389	1.261099
7	16	0.008830013	0.008872211	12.988917299	0.698027
8	7	0.004525010	0.004546635	6.656273438	0.017750
	1464	**0.995243740**	**1.000000000**	**1464.000000000**	**3.173844**

When alpha=0.05, the Chi-sq test value 3.173844< $\chi_{.05,6df}^{2} = 12.59$.

So our goodness–of-fit test from this method shows that H0 cannot be rejected.

Method 2. Maximum Likelihood Estimation (MLE)

$f(x)=p(1-p)^{(x-1)}/\sum_{i=1}^{8}[(1-p)^{(i-1)} p]= p(1-p)^{(\Sigma x-1)}/[1-(1-p)^8]$, for x=1,…,8

The likelihood function is

$L(p)= \prod_{i=1}^{n} f(X_i; p)= \prod_{i=1}^{n} p(1-p)^{(X_i-1)}/[1-(1-p)^8]=p^n(1-p)^{[\Sigma Xi-n]}/[1-(1-p)^8]^n$

The log-likelihood is

$$\ln L(p)=n\ln p+(\sum_{i=1}^{n} X_i-n)\ln(1-p)-n\ln[1-(1-p)^8]$$

The maximum likelihood equation is

$$\frac{d}{dp}\ln L(p)=\frac{n}{p}-\frac{\sum_{i=1}^{n}(Xi-1)}{1-p}-8n(1-p)^7/[1-(1-p)^8]=set=0.$$

Then $\overline{X}=1/\hat{p}-8(1-\hat{p})^8/[1-(1-\hat{p})^8]$

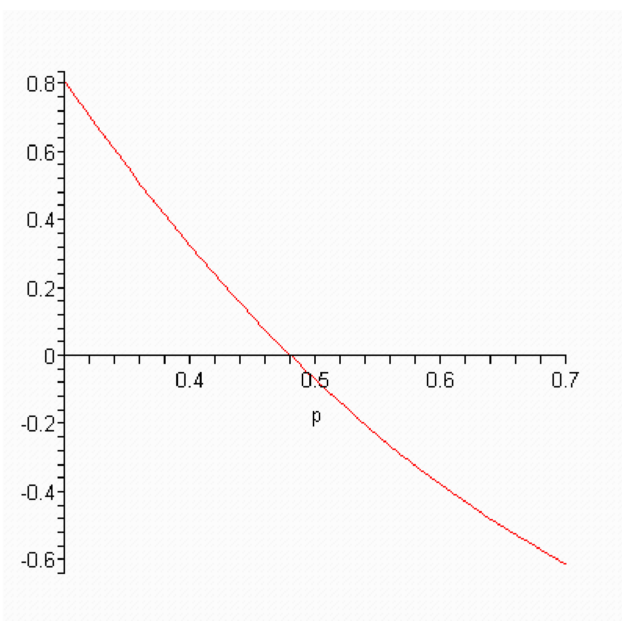
where $\overline{X}$=(1*724+2*355+3*181+4*96+5*54+6*31+7*16+8*7)÷(724+355+181+96+54+31+16+7)=2985/1464.

We solve this graphically using MAPLE with the command

➢ **plot(1/p-8*(1-p)^8/(1-(1-p)^8)-2985/1464,p=.3..0.7);**

From Figure 11, we obtain the solution $\hat{p}$=0.48 .

Figure 11: MLE Estimation of p



We perform a goodness of fit test in Table 8.

H0: X follows a truncated geometric distribution

H1: X does not follow a truncated geometric distribution.

Table 8. Estimate p Using MLE & MME Methods

X	Frequency	p(1-p)^(x-1)	f(x)	Expect frequency	Chi-sq
1	724	**0.48000000**	0.48257986	706.49691332	0.43362970
2	355	0.24960000	0.25094153	367.37839493	0.41707586
3	181	0.12979200	0.13048959	191.03676536	0.52731556
4	96	0.06749184	0.06785459	99.33911799	0.11223886
5	54	0.03509576	0.03528439	51.65634135	0.10633227
6	31	0.01824979	0.01834788	26.86129750	0.63767800
7	16	0.00948989	0.00954090	13.96787470	0.29564507
8	7	0.00493474	0.00496127	7.26329485	0.00954445
	1464	**0.99465403**	**1.00000000**	**1464.00000000**	**2.53945976**

When alpha=0.05, the Chi-sq value 2.53945976< $\chi_{.05,6df}^{2} = 12.59$.

So our goodness–of-fit test from this method shows that H0 cannot be rejected.

Method 3. Method of Moments Estimate (MME)

$f(x)=p(1-p)^{(x-1)}/\sum_{i=1}^{8}[(1-p)^{(i-1)}p]= p(1-p)^{(x-1)}/\{\sum_{i=1}^{8}[1-(1-p)^8]\}$, for x=1,…,8

$E(X)=\sum_{x=1}^{8}[xf(x)]=\sum_{x=1}^{8}\{x\,p(1-p)^{(x-1)}\}/\{\sum_{i=1}^{8}[1-(1-p)^8]\}=p\sum_{x=1}^{8}[x(1-p)^{(x-1)}]/[1-(1-p)^8]$

Let $S=\sum_{x=1}^{8}[x(1-p)^{(x-1)}]=1+2(1-p)+3(1-p)^2+\ldots+8(1-p)^7$

Then $(1-p)S=\sum_{x=1}^{8}[x(1-p)^x]=(1-p)+2(1-p)^2+\ldots+8(1-p)^8$

So $S-(1-p)S=pS=1+(1-p)+(1-p)^2+\ldots+(1-p)^7-8(1-p)^8$

$=[1-(1-p)^8]/p-8(1-p)^8$

Then

$$E(X) = pS/[1-(1-p)^8] = [1-(1-p)^8 - 8p(1-p)^8]/\{[1-(1-p)^8]p\}$$
$$= 1/p - 8(1-p)^8/[1-(1-p)^8]$$

We compute $\bar{x} = \Sigma x f(x)$ where f(x) is the relative frequency. We find $\bar{x} = 2985/1464$.

Set E(X)= $\bar{x}$ and solve for $\hat{p}$.

This gives the same equation as obtained with the MLE estimate so the solution must also be the same, i.e. $\hat{p} = .48$.

Method 4. Minimum Chi-square Estimate (MCE)

Chi-Sq value= Σ ((frequency of category i)-1464P_i)^2/(1464P_i)

We systematically searched over 16 different estimates of p, each time checking their Chi-square values. In this way, we found the value of p with the minimum Chi-square value. Then the value of p which gave the minimum Chi-square value is used to estimate p. Here taking $\hat{p}$ =0.48 gave the minimum Chi-square value 2.5395. Some results from our search are shown in Table 9.

Table 9. Estimate p using Minimum Chi-square Method (MCE)

P	0.4700	0.4750	0.4780	0.4785	0.4790	0.4795	**0.4800**	0.4805
Chi-sq	3.5933	2.8030	2.5807	2.5623	2.5493	2.5471	**2.5395**	2.5427
P	0.4810	0.4830	0.4860	0.4880	0.4890	0.4895	0.4900	0.4905
Chi-sq	2.5513	2.6407	2.9409	3.2534	3.4439	3.5477	3.9500	3.7729

Since we obtain the same value as obtained for the MLE of p, it is not necessary to repeat the goodness of fit test.

Our same test value from MLE, MME and MCE means that H0: truncated geometric distribution fits well, cannot be rejected.

In fact, the test statistic is considerably below the mean of a chi-squared random variable with 6 degrees of freedom.

Simulation

A long term goal in the study of eBay auctions is to understand them thoroughly and to be able to study different bidding strategies.

If we could accurately simulate the entire eBay auction system, then we could use the simulation to study different bidding strategies. Since the entire system simulation involves many distributions, we simulate only a small segment of the system to illustrate the type of methodology that would be required.

To simulate this x distribution, x=1, 2, 3, 4, 5, 6, 7, 8 we used the inverse cumulative sum method as follows:

Step1. Generate u from uniform(0,1);

Step2. If $F(i-1)<u<F(i)$, set X=i,

Where F(i) is the truncated geometric distribution accumulated value for I=1,2,3,4,5,6,8.

1464 random numbers were generated, and the simulation result is shown as follows, where p was estimated by the MLE/MME/min Chi-Square Method.

X	1	2	3	4	5	6	7	8
Frequency of X (simulate)	740	359	172	92	46	32	16	7
Frequency of X (sample)	724	355	181	96	54	31	16	7

4.2 Variables are Normal or Lognormal Distributed

The eBay textbook dataset auction over a 7 day period during July of 2004 contains 376 observations, which means the data were collected from 376 auctions. Table 10 displays a subset of these observations. The data have been sorted by averageBn. Each row record on this table represents a different auction.

Table 10. Original data from eBay textbook auction

averageBn	averageBm	n	m	WinningB	Second-highB
0.77	1.17	16	5	2.80	2.55
1.20	2.64	9	2	2.76	2.51
1.75	2.08	8	5	2.80	2.55
2.17	2.88	6	2	3.00	2.75
2.20	2.63	5	2	2.75	2.50
2.49	3.18	6	2	3.30	3.05
2.50	2.75	6	3	5.00	4.00
2.54	4.01	7	2	4.01	4.00

During an auction, a bidder may bid one time or several times. The highest bid of a bidder would be close to that bidder's estimate of the value of the object. We define averageBm to be the sum of the highest bid for each bidder divided by the number of bidders.

Similarly, define averageBn to be the sum of the all bids divided by the number of bids. Usually, averageBm was greater than averageBn. Curiously, it is possible for averageBm to be greater than averageBn. When M=2, N=2 which means only two people took part in the bidding and they each bid once. Thus averageBm=averageBn in this case.

Table 11 below gives descriptive statistics for the above dataset.

Table 11. Summary statistics for eBay textbook auctions in the dataset

	Mean	Min	Max
averageBn	20.46	0.77	578.57
averageBm	23.05	1.17	672.5
n	9	5	35
m	4	2	11
WinningB	28.8	2.75	835
Second-highB	28.16	2.55	825

From table 11, it is known that for our sample of eBay textbook auctions, conditioned on the number of bids being greater than 4, the average number of bids was 9. The average number of bidders per auction was 4. In other words, for a single auction, there are averaged four bidders who made an average 9 bids. Cases with more than 11 bidders in one auction were rare. Similarly, auctions with more than 35 bids were also rare.

In examining auction data, it was apparent that the distribution of bids depended on n (the number of bids) and m (the number of bidders) and the perceived value of the item. To find patterns involving the winning bid, it was clearly necessary to cancel the effect of (the perceived value). Some examples of measures which cancel the effect of the perceived value are ratios WinningB/averageBn=y1, WinningB/averageBm=y11, Second-highB/averageBn=y2 and Second-highB/averageBm =y12.

The ratio makes the work simple because the units are no longer considered, and the ratio is unit free. Another reason that ratios are appropriate is that although bid values are expected to be different for different items, the value of the products that are auctioned are highly variable and range from only a few dollars to several hundreds of dollars (for example, a board book for children costs $2.75 and a set of textbooks for nurses costs $835), the ratio is almost consistent.

We define a series of ratio variables:

WinningB/(averageBm), Winning*B*/(averageBn), WinningB/(averageBm), WinningB/(averageBn), Second-highB/(averageBm), Second-highB/(averageBm), Second-highB/(averageBn), Second-highB/(averageBn).

These variables were studied, and further study indicated that multiplying the ratio by the factor1/m, 1/n, 1/sqrt(n) or 1/sqrt(m) would yield some random variables that are close to being lognormal or normal. About 40 variables were created and tested (shown in Table 12). Five of them followed a normal or lognormal distribution. Table 13 which follows is a table of those variables (listed using the variables' names in SPSS and SAS) with some summary statistics values.

Hence, y3 and y4 and are seen to be approximately normal while y5, y6, y8 and y20 are approximately lognormal.

Table 12. The variables composed

Variables Definition	Names In SAS	Variables definition	Names in SAS
WinningB/averageBn	y1	ln(y1)	lny1
WinningB/averageBn/n	y3	ln(y3)	lny3
WinningB/averageBn/m	y5	ln(y5)	lny5
WinningB/averageBn/SQRT(n)	y7	ln(y7)	lny7
WinningB/averageBn/SQRT(m)	y9	ln(y9)	lny9
WinningB/averageBm	y11	ln(y11)	lny11
WinningB/averageBm/m	y13	ln(y13)	lny13
WinningB/averageBm/SQRT(m)	y15	ln(y15)	lny15
WinningB/averageBm/n	y17	ln(y17)	lny17
WinningB/averageBm/SQRT(m)	Y19	ln(y19)	Lny19
Second-highB/averageBn;	y2	ln(y2)	lny2
Second-highB/averageBn/n;	Y4	ln(y4)	lny4
Second-highB/averageBn/m	y6	ln(y6)	lny6

Second-highB/averageBn/SQRT(n)	y8	ln(y8)	lny8
Second-highB/averageBn/SQRT(m)	y10	ln(y10)	lny10
Second-highB/averageBm	y12	ln(y12)	lny12
Second-highB/averageBm/m	y14	ln(y14)	lny14
Second-highB/averageBm/SQRT(m)	y16	ln(y16)	lny16
Second-highB/averageBm/n	y18	ln(y18)	lny18
Second-highB/averageBm/SQRT(n)	y20	ln(y20)	lny20

Table 13. The variables' statistics that follow normal/lognormal distributions

Variables	**Names in SAS**	$\overline{X}$	**S**	**Skewness /Kurtosis**	**Shapiro-Wilk /P value***
WinningB/(averageBn*n)	Y3	0.18	0.003	0.06/-0.35	0.994/0.155
Second-highB/(averageBn*n)	Y4	0.18	0.003	0.05/-0.41	0.994/0.167
Ln(WinningB/(averageBn*m)	Lny5	-0.95	0.192	-0.02/-0.47	0.994/0.109
Ln(Second-highB/(averageBn*m))	Lny6	-0.99	0.190	0.007/-0.5	0.992/0.052
ln(Second-highB/(averageBn*SQRT(n)))	lnY8	-0.75	0.009	-0.07/0.53	0.993/0.075
ln(Second-highB/(averageBm*SQRT(n)))	lnY20	-0.88	0.009	-0.03/0.03	0.998/0.921

The Shapiro-Wilk statistics, W, is the ratio of the test estimator of the variance (based on the square of a linear combination of the order statistics) to the usual corrected sum of squares estimator of the variance. W must be greater than zero and less than or equal to one, with small values of W leading to rejection of the null hypothesis.

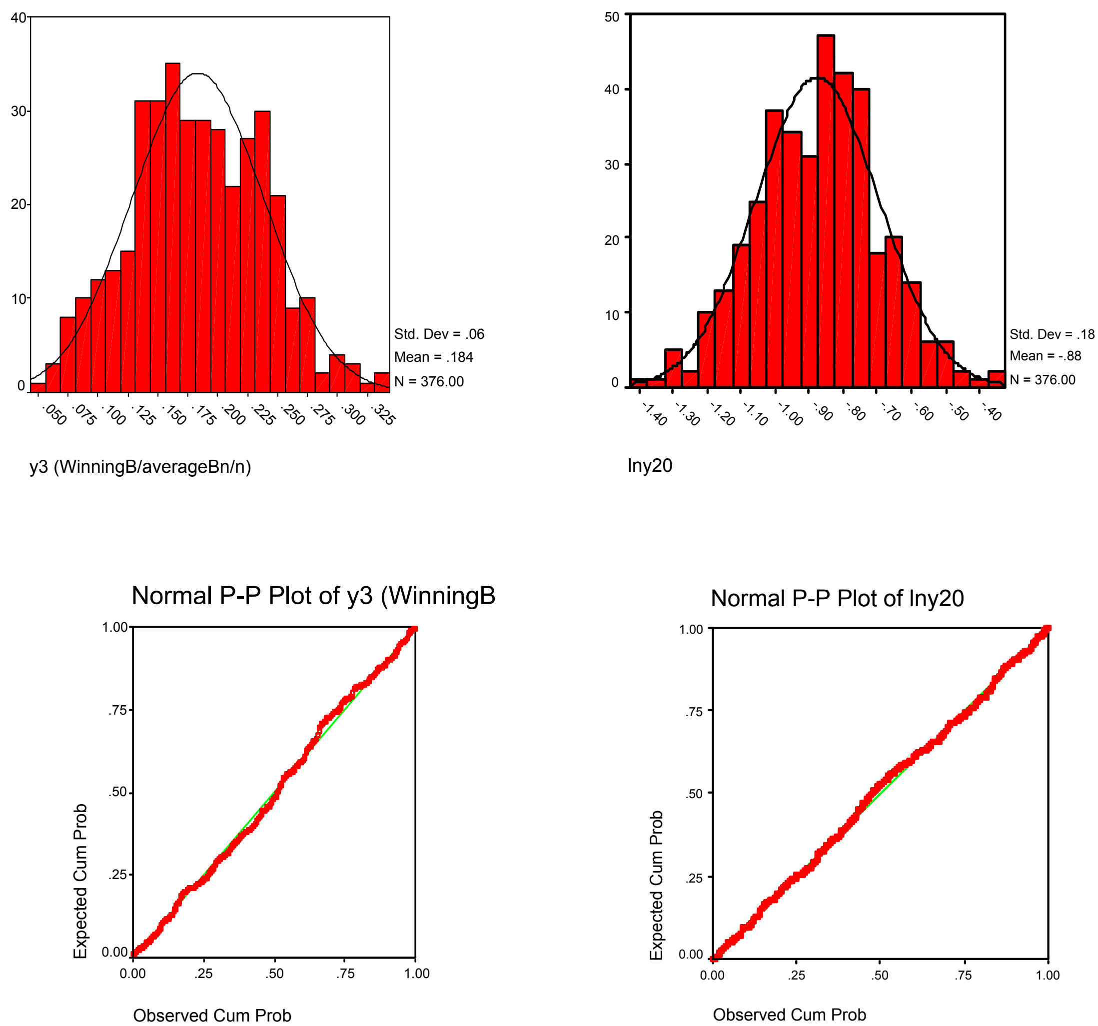
P_P Probability Plots: These plot variable's cumulative proportions against the expected cumulative proportions of any of a number of test distributions. Probability plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line. (from www.science.uwaterloo.ca/course-notes/biology/bio1361/lecture09.ppt)

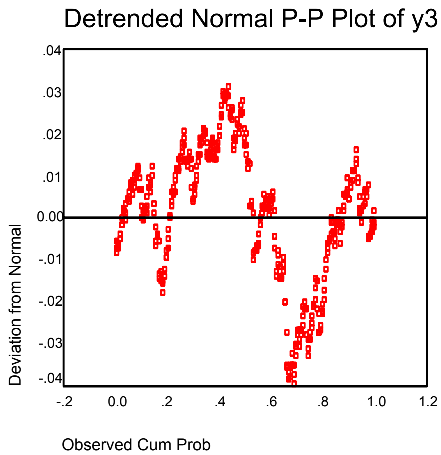
Detrended normal P-P plots depict the actual deviations of the data points from the straight horizontal line. No specific pattern in a detrended plot indicates normality of the variables.

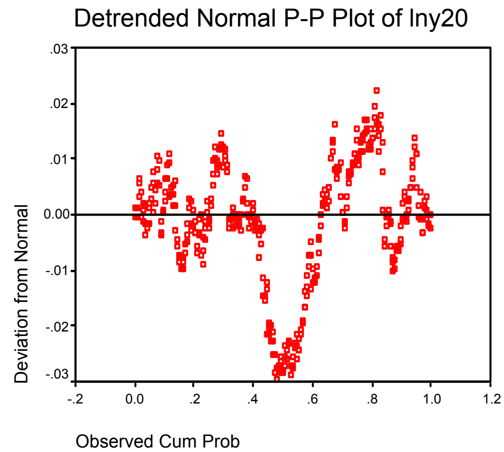
(www.indiana.edu/nstatmath/stat/all/normality/testing-normality.pdf)

Figure 12 shows normality of Y3, lnY20 in SPSS graphs

Figure 12. Graphs of Y3 and ln y20







4.3 Regression analysis

It is natural to use regression models to find and explain relationships between variables for online auction.

The averageBm is the price, which is most close the actual value of the item. It is difficult to calculate this value because we have to find out how many bidders and their highest bid amount. We find averageBn is easy to calculate, so we start by fitting a linear model that regresses averageBm on averageBn. The estimated coefficient for averageBn is 1.17(statistically significant, p-value<0.000). averageBm=-0.8+averageBn*1.17. Table 14 shows that the simple linear regression model is useful in explaining the variability of averageBm.

Table14. Linear Model of averageBm with averageBn

Dependent variable: averageBm

Independent Variables	**Coefficients Estimators**	**t-value**	**Signif Pvalue**	**R-Sq R-Sq***
averageBn	1.17	143.575	0.000	0.982
Intercept	–0.8	-2.203	0.03	

We also found that the WinningB has a strong linear relationship with the product of averageBm and m. The relationship between them is positive.

WinningB=6.74+0.22*averageBm*m.

The associated tests of table 15 show that this model helps explain the variability of the dependent variable.

Table 15. Linear Model of WinningB with averageBm*m

Dependent variable: WinningB

Independent Variables	**Coefficients Estimators**	**t-value**	**Signif Pvalue**	**R-Sq**
AverageBm*m	0.22	44.685	0.000	0.842
Intercept	6.74	5.128	0.000	

Section 5. Tips for Sellers and Buyers

Some tips for both sellers and buyer are summarized as follow.

For sellers:

- A high opening price will attract fewer bidders, and may lead to a lower final price. (Bajari & Hortacsu 2002, Roth & Ockenfels 2002, Lucking Reiley et al 2000.)
- A longer fixed auction time generally will attract more bidders.
- “buy it now” will increase the chance of selling an item.
- Do not end the auction between 12pm and 6am. According to the late bidding feature, this time will not encourage more bidders.
- If you want your items to be sold above a minimum level, use the price reserve feature.

For buyers:

- Some times if you do not want to risk paying much, you can try to bid small increments above the allowable bid.
- Bidding late is effective since most bidding occurs late.
- Buy the item from sellers with a good reputation.
- Use bids like $22.03 to avoid ties.

Section 6. Conclusions

The work in this paper is based on the eBay auction from May to September 2004. It is a sample representation of the all auctions on web auction. The results do suggest the following conclusions.

1. About 3/4 of book auctions had a one week length and there was no observed tendency that auctions closed on any special day.
2. Around 1/2 of bids happened on the last day for seven day auctions, and almost 1/2 of the last day bids happened at the last hour. 85% of last day bids happened at the last 30 minutes.
3. The category number of bids follow a truncated geometric distribution.
4. WinningB/(averageBn*n), WinningB/(averageBm*n),
   Second-highB/(averageBn*n) follow normal distributions, while
   Second-highB/(averageBn*SQRT(n)) and
   Second-highB/(averageBm*SQRT(n)) follow a lognormal distribution.
5. There are some linear relations between variables such as averageBm with averageBn, WinningB with averageBm*m, and n with m.

Section 7. Further Discussion

eBay is developing, customers are becoming more experienced, and eBay rules are changing to attract more people. The results obtained here are a tiny part of of eBay research. We not only found information about eBay textbook auctions, but also applied statistics in the analysis. After this project ended, eBay's policy changed to become more open to customers. Now during an auction, the bids are shown rather than hidden. This makes customers feel more knowledgeable since they know their opponents' bid. Under this new policy, some new phenomena may appear.

For eBay data, there might be other variables following normal, lognormal, or other kinds of specified patterns. Many other variables could be examined.

References

1. Dan Ariely, Axel Ockenfels and Alvin E. Roth, "An Experimental analysis of late-bidding in internet auctions", 31pp (2001).
2. Ariely, Dan, Axel Ockenfels and Alvin E. Roth. "An Experimental Analysis of Ending Rules in Internet Auctions." Working paper, Harvard University, 2002.
3. Bajari, Patrick and Ali Hortacsu, "The winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions." Rand Journal of Economics, 50pp, 2002.
4. David P. Baron, "Private Ordering on the Internet: The eBay Community of Traders," Business and Politics, Volume 4, 245-274, 32pp, 2002. http://www.bepress.com/bap/vol4/iss3/art1
5. W. J. Conover, "Practical nonparametric statistics" 1980.
6. Lucking-Reiley, David. "Auction on the Internet: What's Being Auctioned, and How?" Journal of Industrial Economics, 55pp, September 2000, 48(3).
7. David Lucking-Reiley, Doug Bryan, Naghi Prasad and Daniel Reeves, "Pennies from eBay: the Determinants of Price in Online Auctions", Technical Report. 24 pp, 2000. http://www.vanderbilt.edu/econ/reiley/papers/PenniesFromEBay.pdf
8. Raymond H. Myers, "Classical and modern Regression with Applications"
9. Axel Ockenfels and Alvin E. Roth. "The Timing of Bids in Internet Auctions: Market Design, Bidder Behavior, and Artificial Agents." Artificial Intelligence Magazine, Fall 2002, 79-88.
10. Axel Ockenfels, Alvin E. Roth, "Late and Multiple Bidding in Second Price Internet Auctions: Theory and Evidence Concerning Different Rules for Ending an Auction", CESifo Working Paper, No.992, 2003, 36 pp.

11. Roth, Alvin E. “The Economist as Engineer.” Fischer-Schultz Lecture, Working paper, 54 pp, 2001.
12. Roth, Alvin E., and Axel Ockenfels. “Last-minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet.”, American Economic Review, 2002, 92(4), 1093-1103.
13. Galit Shmueli and Wolfgang Jank, “Modeling the Dynamics of Online Auctions: A Modern Statistical Approach”, 34 pp, March 5,2004. http://www.rhsmith.umd.edu/ceme/statistics/FDA_eCOM.pdf
14. Wilcox, Ronald T. “Experts and Amateurs: The Role of Experience in Internet Auctions,” Marketing Letters, 11(4), 363-374, 2000. http://faculty.darden.virginia.edu/wilcoxr/pdf_docs/expertsamateurs.pdf