Preliminary and Incomplete. Do not cite without permission of the authors.

Interim Report on: Estimating Consumer Surplus in e-Bay Computer Monitor Auctions.*

Tugba Giray
Economics Department
Bilkent University
Ankara, Turkey

Kevin Hasker
Economics Department
Bilkent University
Ankara, Turkey

Robin Sickles
Department of Economics
Rice University
Houston, Texas

October 15, 2005

Abstract

Using data from Computer Monitor Auctions on e-Bay collected in 2000, we estimate bidding functions by maximum likelihood using five different assumptions about the underlying distribution of independent private values. We assume these values come from the log-normal, the gamma, the Weibull, the Logistic or the Pareto distribution. We then estimate the consumer surplus in these auctions using two different methodologies. First we construct the expected consumer surplus and then we construct a lower bound for consumer surplus using a "rational reassignment" methodology. Using various distributions for the independent private values will allow us to see how much consumer surplus estimates depend on this assumption, and we will also use various statistical tests to find the best fit to the data among the various distributions considered.

1 Introduction

It is well established that eBay is a significant economic marketplace. Economists have long hailed the price discovery power of auctions, but unfortunately

*Paper prepared for the *FTC Roundtable: Economics of Internet Auctions*, October 27th, 2005, Washington, DC.

the cost of getting bidders together prevented their widespread usage. eBay overcame this problem by allowing people to auction items over the Internet. Because of this eBay has become a significant marketplace, and due to the economies of the marketplace they are likely to remain one for the significant future. However we are still unsure how much eBay benefits the economy. One measure of this benefit is the Consumer Surplus that eBay generates. This paper measures this important attribute in the market for Computer Monitors.

We estimate bidders' values and an exogenous entry process using maximum likelihood. Since we can not be certain of the true underlying distribution of bidders values we estimate our function using multiple distributional assumptions. This allows us to estimate Consumer Surplus under different distributional assumptions and test for sensitivity to distribution. It also allows us to test which distribution best fits the data and test how well each estimate performs against a non-specified non-parametric distribution. Our data set for this analysis is around 3000 PC color computer monitors with a screen size of between 14 and 21 inches which were auctioned between February 23, 2000 and June 11, 2000.

It is surprising how few attempts have been made to estimate Consumer Surplus. Song [10] estimates a semi-parametric model using both the second and third highest bids in university yearbook auctions. She constructs an innovative semi-parametric methodology using the second and third highest bids and estimates the median Consumer Surplus in university yearbook auctions at $25.54. In comparison our methodology is to search over parametric models using maximum likelihood. Not only does this dispense with the need to use the third highest bid (which is of questionable trustworthiness) but it allows us to suggest a best parametric model which might be applicable in other research. Bapna, Jank, and Shmueli [5] use a revolutionary new data collection technique that allows them to directly observe a bidder's stated value. While this is a brilliant technique unfortunately they have a very heterogenous data set and do not estimate a structural bidding function, however their median Consumer Surplus (in all categories) is \$3.53. Several other articles have touched on this subject: Bapna, Paulo and Gupta [2] , [3] and Bapna, Goes, Gupta, and Jin [4] but these papers are focused on mechanism design and look at first price or multi unit auctions and thus are outside of the scope of this paper.

eBay has two different auction formats. The common format is an English auction with a hard stop time. This is the type of auction used in 87 percent of our data set and the type of auctions on which we focus. When our data was collected bidding goes on from three to ten days and stops at a preset time.

Our estimation techniques are based on methods developed by Donald and Paarsch [7]. Unlike that paper we do not have to estimate the minimum or maximum value a bid can take since in our auctions the natural lower boundary is zero and there is no reasonable binding upper boundary—we assume it is infinity with extremely low probability. We also are able to estimate a full likelihood function since our data set includes all auctions where no one decided to bid. There are several other methodologies currently available in the literature. First of all is the semi-parametric technique found by Song [10]. This requires the use of some of the data from the third highest bid. While for clear theoretic reasons one can always assume that the second highest bid is a bidders' value these techniques do not provide the same guarantee for the third highest bid. Instead you have to rely on the bidders planning not to update their bid—which they frequently do. This could potentially bias the results. As well there are the usual problems with slow convergence of non-parametric techniques. Furthermore while our models are more restrictive finding the best fit is more informative than with non-parametric techniques. With non-parametric techniques comparing distributions for different goods is difficult, with parametric maximum likelihood we can easily compare our results across different product categories and observe if there is some fundamental underlying distribution of values. Another interesting technique is a Bayesian methodology developed in Bajari and Hortaçsu [1]. However these techniques require that the bidding functions are linearly scalable, a restriction unnecessary with our approach and violated by our structural form. A final technique is a Non-linear Simulated Least Squares methodology developed by Laffont, Ossard, and Vuong [9]. This approach overcomes the complexity of calculating the likelihood function by simulating the auctions, and is a flexible methodology that can be used for many models where bidders have private values. We have used this technique in previous papers (Gonzalez, Hasker, and Sickles [8]) but in this paper since maximum likelihood is feasible we prefer the more standard approach.

2 The Data Set and Our Collection Techniques.

eBay saves all information about closed auctions on their website for a month after the auction closes. This allows people who participated in the auction to verify the outcome, and provides the source for our data set. Our data was collected using a "spider" program which periodically searches eBay for recently closed computer monitor auctions and downloads the pages giving the item description and the bid history. Software development was done in Python—a multi-platform, multi-OS, object-oriented programming language. It is divided into three parts. It first goes to eBay's site and collects the item description page and the bidding history page. It next parses the web pages, and makes a database entry for each closed auction. The final part iterates through the database entries stored, and creates a tab-delimited ASCII file. This method has allowed us to collect information on approximately 9000 English auctions of PC computer monitors.

The original data processing program did not process all of the data. It provided us with the core of the data which was augmented with further processing of the raw html files. Using string searches we have managed to collect extensive descriptive information for the entire data set. With further data processing we have managed to collect all of the bidding histories. This process provided us with information on the 6543 auctions that are used in the estimates.

Our data set consists of PC color computer monitors with a size between 14 and 21 inches which were auctioned between February 23, 2000 and June 11, 2000. All monitors are in working order, and we ignored touch screen monitors, LCD monitors, Apple monitors, and other types of monitors that are bought for different purposes than the monitors in our sample. Also, if there were any bid retractions or cancellations (this happened in 7.4 percent of the auctions) we dropped the observation because the retractions might indicate collusion.

Descriptive variables except for monitor size were constructed using string searches. In Gonzalez, Hasker, and Sickles [8] the strings that were used for each variable are detailed. This allowed us to collect data on whether there was a secret reservation price, whether it was met, monitor resolutions, dot pitch, whether a warranty was offered, several different brand names, whether the monitor was new, Like-New, or refurbished, and whether it was a flat screened monitor. "Brand name" is used for monitors that are from one of the ten largest firms represented in our data set. These firms are Sony, Compaq, NEC, IBM, Hewlett Packard, Dell, Gateway, Viewsonic, Sun, and Hitachi in order of size. Sony has around a 10 percent market share, the smallest are all around 3 percent, in total these 10 firms represent 57 percent of the market. Dot pitch and resolution are not reported in all of the auctions. Dot Pitch is reported in 35 percent of the auctions, resolution in 58 percent. In the appendix in Section 7 the descriptive statistics of variables of interest are presented.

3 Model and the Maximum Likelihood Functions.

We will use maximum likelihood to estimate bidders' values and an exogenous entry process in eBay auctions.

We assume that bidders' values are log-linear in a set of auction specific characteristics x_n and their private value ρ_i. The formula thus is:

$$\ln b_n^w = \max\left\{\ln r_n, x_n'\beta + \ln \rho_n^{(2:I)}\right\} \tag{1}$$

where $\rho_n^{(2:I)}$ is the value of the second highest bidder in auction n. We will allow for various models of the distribution of b_n^w.

Let $F_n(\beta)$ be the cumulative distribution function (CDF) of the bidders' values and $f_n(\beta)$ be the probability density function (PDF)—where β may include some distribution specific coefficients. Denote D_0 as the dummy which equals one if there are no bidders, and D_1 as the dummy which equals one if there is one bidder. Then if I is the number of bidders the likelihood of auction n given I is:

$$\begin{aligned} l_n(\beta|I) &= \left(F_n(\beta)^I\right)^{D_0} * \\ &\quad \left(I\left(1-F_n(\beta)\right)F_n(\beta)^{I-1}\right)^{D_1} * \\ &\quad \left(I(I-1)\left(1-F_n(\beta)\right)F_n(\beta)^{I-2} f_n(\beta)\right)^{1-D_0-D_1} . \end{aligned}$$

I will be a stochastic variable that can range from $\underline{I}_n$—the number of bidders who bid in this auction—to $\overline{I}$—an arbitrary upper bound on the number of bidders in any auction. Notice that just because we observe only $\underline{I}_n$ bidders in an auction does not mean there might not have been more. More bidders might have come to the auction but realized they did not want to bid.

The number of bidders in an auction will be determined by a Poisson entry process. The parameter of the entry process, λ_n, will be log-linear in a set of auction specific characteristics z_n—where $x_n \subseteq z_n$. Some auction characteristics might affect entry but not values, but we assume that if the auction characteristic affects values it must affect entry. The estimated functional form for entry is thus:

$$\ln \lambda_n = z_n' \gamma + \ln \nu_n$$

Letting T_n be the length of the auction ($T_n \in \{3, 5, 7, 10\}$) and D_{sr} be a dummy which is one if there is a secret reservation price. Then our total likelihood for auction n is:

$$l_n(\beta, \gamma) = \frac{\Sigma_{i=\underline{I}_n}^{\overline{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n} l_n(\beta | I)}{\Sigma_{i=D_{sr}}^{\overline{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n}} .$$

$\underline{I}_n$ is increased by one if there is a secret reservation price, thus we are following Bajari and Hortaçsu [1] in treating the auctioneer as another bidder if there is a secret reservation price.

Notice that we can use full maximum likelihood since our data collection technique captures all auctions that do not result in sales. In general data only includes auctions that result in a sale, making ours a rare example of full maximum likelihood estimation in auctions.

The choice of $\bar{I}$ is obviously arbitrary, to derive the estimates we choose $\bar{I} = 30$, and then tested the results when $\bar{I} = 50$. This change did not change the coefficients, thus it appears our choice of $\bar{I} = 30$ is sufficient.

3.1 The Distributions:

Since we can not be certain *a-priori* what the true distribution of bidders values is we test several different distributions: the Log-Normal, Weibull, Gamma, Logistic, and Pareto.

The PDF of the Log-Normal is:

$$f_n(\beta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\ln b_n^w - x_n' \beta\right)^2 / 2\sigma^2}$$

we write $\beta = \{\beta, \sigma\}$ for simplicity. The PDF of the Weibull is:

$$f_n(\beta) = \alpha \frac{(b_n^w)^{\alpha-1}}{\left(e^{x_n'\beta}\right)^{\alpha}} e^{-\left(\frac{b_n^w}{e^{x_n'\beta}}\right)^{\alpha}}$$

where $\beta = \{\beta, \alpha\}$. The parameter α is the shape parameter of the Weibull distribution, and note that we estimate $\ln \alpha$ to avoid this parameter being negative. The PDF of the Gamma is:

$$f_n(\beta) = \frac{1}{\Gamma(\alpha)} \frac{(b_n^w)^{\alpha-1}}{\left(e^{x_n'\beta}\right)^{\alpha}} e^{-\frac{b_n^w}{e^{x_n'\beta}}}$$

and again $\beta = \{\beta, \alpha\}$. Again the parameter α is the shape parameter of this distribution and we estimate $\ln \alpha$. The PDF of the Logistic is:

$$f_n(\beta) = \left(1 + e^{-\frac{b_n^w}{e^{x_n'\beta}}}\right)^{-1} \left(1 - \left(1 + e^{-\frac{b_n^w}{e^{x_n'\beta}}}\right)^{-1}\right) \frac{1}{e^{x_n'\beta}}$$

notice that this distribution does not have another parameter. The PDF of the Pareto is:

$$f_n(\beta) = \frac{\alpha}{e^{x_n'\beta}} \left(1 + \frac{b_n^w}{e^{x_n'\beta}}\right)^{-(\alpha+1)}$$

again $\beta = \{\beta, \alpha\}$ and α is the shape parameter of this distribution and we estimate $\ln \alpha$.

4 The Estimates

Estimates are based on a subset of the complete data set in order to allow for out of sample specification tests. These have not yet been carried out. We also will consider an alternative models of entry. These results can only be considered preliminary. Suggestions on alternative specifications would be greatly appreciated.

We first present the estimates of the exogenous values ($e^{x_n'\beta}$) and then we present our estimates of the parameter of the entry process (Lambda). The right hand side variables in our models are the size of the monitor, diagonal screen size, the dot pitch (the distance between dots on the screen), resolution, and size of picture seen on the monitor. We also have a series of dummies indicating whether or not the monitor is New, Like New, or Refurbished (the omitted category is Used) and whether or not the monitor has a Warranty, is a Brand Name, or is Flat panel. The final variable is the Seller's Feedback. This increases by one with every sale that results in a pleased customer, so is both an indicator of the Seller's experience and reputation.

<insert Table 1 about here>

Results are in Table 1 in the Appendix. We first note the general stability of coefficients across estimates. The only coefficients which seem to meaningfully vary are two dummies, the "Like New" dummy and the Warranty dummy. Both of these coefficients are significantly smaller with Logistic private values. In general it seems Refurbished monitors are no better than Used monitors. These are monitors that have been "rebuilt" by the auctioneer and apparently bidders do not trust that the auctioneer has done a good job. Also, Brand Name monitors appear to have little relative value, possibly because Brand Name really conveys to the bidder that the monitor is a common brand. The only coefficient that has a surprising sign is the coefficient on the log of seller's feedback. However, while the point estimate is negative in each regression the effect is neither statistically nor economically significant. In keeping with Song [10] we find that the total seller's feedback does not affect the value of the good. In contrast Bajari and Hortaçsu [1] find that it has a significant positive coefficient—however coins are a very different class of goods and thus this might explain the difference our results.

While the differences in coefficients are generally small the exogenous value of a computer monitor can be very different for a given auction using the different techniques and the summary statistics of these values differ significantly as indicated in Table 2.

Table 2-Predicted Exogenous Values

.	Log-Normal	Weibull	Gamma	Logistic	Pareto
Average	13.99	15.24	129.46	48.66	48.01
Median	10.17	11.25	93.1	37.19	36.97
Standard Deviation	10.55	11.7	103	35.76	35.05
Minimum	3.23	3.47	28.41	12.64	12.01
Maximum	69.33	74.25	672.59	202.9	229.07

The exponential is a function with a significant right skewness, and thus a few large observations will artificially inflate the average value. Therefore we prefer to look at the medians. The Log-Normal and Weibull distributions tend to produce very low estimates of exogenous value (around \$10-\$11), the Pareto and the Logistic produce a higher value (around \$36-\$37) and the Gamma produces a very high median value (\$93).

<Table 3 about here>

The coefficients of the entry process presented in Table 3 are much less stable. The new right hand side variables in this regression are a square term for Seller's Feedback—allowing for a decreasing marginal benefit of experience. We also have a series of category dummies—the default is the "general" classification but a seller is allowed to put the monitor into the $\leq 17''$ screen, $\geq 19''$ screen or the Monotonic sub-categories if they wish. Notice that all monitors that are put in the Monotonic sub-category are misplaced—all monitors in our data set are color monitors. The two final dummies are one if the auctioneer put a Secret Reserve on the item (an unobserved reservation price) and if this Secret Reserve was not met—or in our analysis the auctioneer "sold" the item to himself.

Results for the entry process appear to be stable across distributions for the coefficients on Size and the dummies for Refurbished, Warranty, and Brand Name. Notice that while a high Resolution raises the item's value it seems to lower the expected number of bidders. This indicates some heterogeneity in our bidders. A high resolution means that for given screen size you can see a larger picture. This means that the details of the picture are smaller to the naked eye, and it is reasonable that some bidders do not want to pay more for such a monitor. Our results illustrate this, some bidders do not value Resolution and thus are not willing to bid on items with a high Resolution. The same tendency (though to a lesser degree) is found with flat screen monitors. While this is clearly a positive aspect not everyone will be willing to pay for it. In general this has a small negative effect on entry but a small positive effect on the monitor's value. Notice as well that in this regression both Dot Pitch and Seller's Feedback have a much more significant effect on entry than they do on the monitor's value. The coefficients on Seller's Feedback seems to suggest that "trust" is binary for our bidders. If a seller is not experienced then bidders might not want to buy his or her monitor, but if they decide to try and buy it they discount the seller's lack of experience.

The most troubling parameter in these regressions is the coefficient on the Secret Reservation Dummy. Bajari and Hortaçsu [1] found it had a negative coefficient and while it is theoretically possible that this could have no effect there is no theoretical explanation for it having a positive coefficient. A potential problem is that this variable is significantly correlated with the error term. Notice from the correlation table that there is a strong positive correlation between having a secret reservation price and both the sales price and number of bidders. The causation is probably that auctioneers who have a valuable monitor want to put a high reserve on it, but think this will drive bidders away. This leads them to use a secret reservation price. Unfortunately all of the variables that can not be captured in a regression—the brand name, the specific monitor model—will cause the auctioneer to use a Secret Reservation Price and cause bidders to be eager to buy that item. We will solve this problem in the future by instrumenting the Secret Reservation Price on the other right hand side variables and other variables that we have not used in our regressions. We also will give the same treatment to the Open Reservation price—the traditional reservation price—and include this variable in our regression.

There is significant heterogeneity in the expected number of bidders—though the variation is not large in absolute terms.

Table 4-Predicted Values for Lambda, the Entry Parameter

.	Log-Normal	Weibull	Gamma	Logistic	Pareto
Average	54.6	6358.57	7941.4	1263.9	3001.23
Median	5.63	7.83	9.89	3.67	5.23
Standard Deviation	203.79	58569.11	71354.81	9042.86	24508.77
Minimum	0.01	0.11	0.01	0.01	0.01
Maximum	4189.82	2452401.44	3017199.93	391642.72	1074452.44

These estimates seem to suggest that the median number of bidders is between 3.5 and 10 in Internet auctions with an average across models of around 5. This is not a large variation in absolute terms but significant in percentage terms. The averages vary more widely, but again this is an exponential function so it has a large right skewness.

5 Consumer Surplus

. While the *a-priori* Consumer Surplus is a function of I the *ex-post* Consumer Surplus is not and thus estimating *ex-post* Consumer Surplus is relatively straightforward exercise. This is in part because we do not have to calculate a summation over the possible values of I. Consumer surplus in auction n is:

$$E\left(v_n^{(1:I)}|v_n^{(2:I)} = \frac{b_n^w}{e^{x'_n\beta}}|I \geq 2\right) e^{x'_n\beta} - b_n^w$$

The expectation is (for $I \geq 2$):

$$\begin{aligned} E\left(v_n^{(1:I)}|v_n^{(2:I)} = \frac{b_n^w}{e^{x'_n\beta}}|I \geq 2\right) &= \frac{\int_{v_n^{(2:I)}}^{\infty} (I)(I-1)\, z f_n(z,\beta)\, f_n\left(v_n^{(2:I)},\beta\right)\left(F_n\left(v_n^{(2:I)},\beta\right)\right)^{I-2}}{(I)(I-1)\left(1-F_n\left(v_n^{(2:I)},\beta\right)\right) f_n\left(v_n^{(2:I)},\beta\right)\left(F_n\left(v_n^{(2:I)},\beta\right)\right)^{I-2}} \\ &= \frac{\int_{v_n^{(2:I)}}^{\infty} z f_n(z,\beta)}{1-F_n\left(v_n^{(2:I)},\beta\right)} \end{aligned}$$

when $I = 1$ this is:

$$E\left(v_n^{(1:I)} | v_n^{(2:I)} = \frac{r_n^w}{e^{x_n'\beta}} | I \geq 1\right) = \frac{\int_{v_n^{(2:I)}}^{\infty} (I)\, z f_n(z,\beta) \left(F_n\left(v_n^{(2:I)},\beta\right)\right)^{I-2}}{(I)\left(1 - F_n\left(v_n^{(2:I)},\beta\right)\right)\left(F_n\left(v_n^{(2:I)},\beta\right)\right)^{I-2}}$$

$$= \frac{\int_{v_n^{(2:I)}}^{\infty} z f_n(z,\beta)}{1 - F_n\left(v_n^{(2:I)},\beta\right)}$$

thus it is independent of $I \geq 1$.

Lemma 1 *If $I \geq 1$ then* ex-post *Consumer Surplus is independent of I, and thus independent of the entry process.*

Proof. See above. ■

Using this insight we can estimate ex-post Consumer Surplus. Summary statistics for estimates of Consumer Surplus for the various distributions are given in Table 5.[1]

Table 5-Expected Consumer Surplus

	Log-Normal	Weibull	Gamma	Logistic	Pareto
Average	\$131.36	\$718.41	\$5210.20	NA	\$1815.60
Median	\$97.21	\$560.88	\$3959.60	NA	\$1434.90
Standard Deviation	\$118.56	\$563.99	\$4146.20	NA	\$1320.90
Minimum	\$7.91	\$102.97	\$757.82	NA	\$433.16
Maximum	\$977.78	\$3572.30	\$27494.00	NA	\$8583.4

These preliminary estimates of Consumer Surplus vary widely and are unreasonably high. Given that our median computer monitor sold for \$100 even the most conservative estimate asserts that consumers are taking 49% of the available surplus. It is likely that these estimates are sensitive to the upper tail of the distribution. While we assume for estimation that a bidder can have an extremely private large value for the computer monitor there should be some reasonable upper bound. We are currently checking the sensitivity of these estimates to upper tail properties of the various distributions and to upper truncation of the distributions.

[1] The estimates for the Logistic regression are currently not available due to calculation difficulties.

Given the wide range and large size of these estimates this makes it of great interest to try to estimate Consumer Surplus using an alternative methodology. We can also construct a "lower bound" estimate for Consumer Surplus by assuming that $v_n^{(1:I)} = v_n^{(2:I)}$ in every auction. We can then reassign bidders to see how much someone with the value of $v_n^{(2:I)}$ could win in other auctions in our data set. If I was a constant in our regressions this would be a precise lower bound. As it is it provides an estimate for Consumer Surplus that is independent of the tails of our distributions.

Table 6-Lower Bound Estimates of Consumer Surplus

.	Log-Normal	Weibull	Gamma	Logistic	Pareto
Average	\$57.78	\$56.771	\$55.78	\$57.15	\$57.75
Median	\$36.49	\$36.229	\$35.85	\$36.48	\$36.39
Standard Deviation	\$73.97	\$72.798	\$71.12	\$73.8	\$75.05
Minimum	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
Maximum	\$1914.10	\$1919.10	\$1910.00	\$1917.80	\$1946.20

These lower bound estimates of Consumer Surplus are in Table 6. The estimates are quite comparable across different distributions and are much more stable than those from Table 5. These result point to a significant amount of consumer surplus captured in these auctions. Compared with the median sales price of \$100 the results indicate that consumers are capturing at least 26% of the total surplus which is quite significant considering the conservative assumption made to derive these last estimates.

We are also constructing estimates of *a-priori* consumer surplus but as of yet those results are not available.

6 Finding the best distribution.

We can use two different types of tests for the preferred distribution of private values.

6.1 Tests based on the Likelihood.

Since we have multiple distributional assumptions we will choose the distribution that minimizes one of three different information criteria, which are all very similar given our models. These criteria were intended to compare models which are very different in the number of parameters or observed variables. Since we hold these numbers nearly constant across our different models the test becomes essentially which distributional assumption yields the largest likelihood value, or alternatively, the smallest negative log-likelihood divided by the number of observations—the objective function we use in our estimation. The first criterion we use is the Akaike Information Criterion:

$$AIC = -\frac{1}{N}\log L\left(\hat{\beta},\hat{\gamma}\right) + \frac{1}{N}k$$

where N is the sample size, $L\left(\hat{\beta},\hat{\gamma}\right)$ is the likelihood, and k is the number of parameters. Another criteria that puts more of a penalty on complexity is the Bayesian Information Criterion or the Schwartz Criterion:

$$BIC = -\frac{1}{N}\log L\left(\hat{\beta},\hat{\gamma}\right) + \frac{\log(N)}{2N}k$$

Our final statistic will be the Browne–Cudeck Criterion [6] which is:

$$BCC = -\frac{1}{N}\log L\left(\hat{\beta},\hat{\gamma}\right) + \frac{1}{N-p-2}k$$

where p is the number of observed variables. Given that $N = 2934$, $p = 20$, and $k \in \{33, 34\}$, the difference in these measures is small.

Table 7

	Log-Normal	Weibull	Gamma	Logistic	Pareto
Objective Function	1.1039	3.7662	3.7245	3.8077	3.896
AIC	1.115 5	3.777 8	3.736 1	3.8190	3.907 6
BIC	1.150 2	3.812 5	3.770 8	3.852 6	3.942 3
BCC	1.115 6	3.777 9	3.736 2	3.819 1	3.907 7

The results from the different distributions for the AIC, the BIC, and the BCC criteria are in Table 7. The Log-Normal is clearly the best distribution by all measures while the second best is the Gamma. Notice that the order of all the criteria is the same as the order of the objective functions.[2]

6.2 Tests against the Non-Parametric distribution of Third Highest Values.

In future analysis we will compare the distribution of third highest values to the uniform distribution in order to test whether any of our parametric distributions are close to the non-parametric true distribution of bidders' values.

Using the distribution of the third order statistic we can map each observation to the uniform, using the following function where $v^{3:I} = \frac{b_n^{3:I}}{e^{x_n' \beta}}$:

$$G_n\left(v^{3:I}, \beta|I\right) = \int_0^{v^{3:I}} \frac{I!}{(I-3)!2!} F_n(z,\beta)^{I-3} \left(1 - F_n(z,\beta)\right)^2 f_n(z,\beta)\, dz$$

and:

$$G_n\left(v^{3:I}, \beta, \gamma\right) = \frac{\Sigma_{i=\underline{I}_n}^{\overline{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n} G_n\left(v^{3:I}, \beta|I\right)}{\Sigma_{i=3}^{\overline{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n}}\ .$$

We can then compare the resulting distribution to the uniform to see which is the best fit, and if we can accept the null that the distribution is the true underlying distribution of values.

[2]While these tests all resoundingly accept the Log-Normal we would like to caution this conclusion with some reduced form logic. The Log-Normal is the only distribution where we minimizes the log of the error terms. If our criteria was minimizing the sum of square errors clearly for reasonable β the sum of log errors would usually be smaller than the sum of absolute errors.

References

[1] Bajari, Patrick and Ali Hortaçsu . (2003). "Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions." *RAND Journal of Economics*, pp. 329-355.

[2] Bapna, Ravi; Paulo Goes, Alok Gupta (2003). "Analysis and Design of Business-to-Consumer Online Auctions" *Management Science*, 49:(1), 85-101.

[3] Bapna, Ravi; Goes, Paulo; Gupta, Alok (2003). "Replicating Online Yankee Auctions to Analyze Auctioneers' and Bidders' Strategies," *Information Systems Research*, 14:(3), 244-268.

[4] Bapna, Ravi; Paulo Goes, Alok Gupta, and Yiwei Jin (2004). "User Heterogeneity and its Impact on Electronic Auction Market Design: An Empirical Exploration," *MIS Quarterly*, 28:1, pp. 21-43.

[5] Bapna, Ravi; Wolfgang Jank, Galit Shmueli. (2005) "Consumer Surplus in Online Auctions." mimeo, available at: http://www.rhsmith.umd.edu/ceme/statistics/consumerSurplusOnlineAuctions%20March2005.pdf

[6] Browne, M. W., & Cudeck, R. (1992). "Alternative ways of assessing model fit." *Sociological Methods and Research.* 21(2), 230-258.

[7] Donald, Steven and Harry Paarsch. (1993). "Piecewise Maximum Likelihood Estimation in Empirical Models of Auctions." *International Economic Review*, 34:121-148.

[8] Gonzalez, Raul; Kevin Hasker, and Robin C. Sickles. (2005) "An Analysis of Strategic Behavior in eBay Auctions." mimeo, available at: http://www.bilkent.edu.tr/~hasker/Research/Gonzalez-Hasker-Sickles-eBay-04-05-21.pdf

[9] Laffont, Jean-Jacques; Herve Ossard and Quang Vuong. (1995). "Econometrics of First-Price Auctions." *Econometrica*, 63:953-980.

[10] Song, Unjy. (2004) "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders." mimeo, available at: http://faculty.arts.ubc.ca/usong/eBay.pdf

7 Appendix: Tables, Descriptive Statistics and the Correlation Matrix

Table 1

.	Estimates of the Exogenous Value.				
.	Log-Normal	Weibull	Gamma	Logistic	Pareto
Constant	-14.1874***	-14.2123	-11.9023***	-11.712	-12.4335***
.	(4.94)	(25.58)	(1.38)	(50.17)	(2.41)
Log, Size	4.581	4.6543	4.7877***	4.6516***	4.3902***
.	(7.28)	(90.46)	(0.70)	(0.45)	(0.42)
Log, Dot Pitch	-0.8868	-0.9692	-0.9479	-0.7436	-0.9826
.	(1.37)	(1.89)	(0.95)	(15.94)	(0.96)
Dummy, No Dot Pitch	1.1696	1.2171	1.1543	0.9294	1.228
.	(1.75)	(2.49)	(14.43)	(3.19)	(1.56)
Log, Resolution	0.3465	0.3203	0.2487	0.1834	0.3388*
.	(2.02)	(0.39)	(0.81)	(0.31)	(0.23)
Dummy, No Resolution	2.591	2.4002	1.8994	1.3759	2.5007***
.	(3.88)	(2.03)	(1.80)	(32.20)	(0.87)
Dummy, New	0.3671	0.3031	0.3279	0.238	0.2609
.	(58.22)	(6.48)	(7.35)	(7.02)	(1.11)
Dummy, Like-new	0.2354	0.2198	0.2432	0.0751	0.1985
.	(6.17)	(16.25)	(7.54)	(161.40)	(1.68)
Dummy, Refurbished	0.0127	0.0357	0.0512	0.0203	0.0044
.	(9.78)	(26.44)	(6.91)	(8.31)	(2.44)
Dummy, Warranty	0.1259	0.1238	0.1101	0.0567	0.1384
.	(7.96)	(6.28)	(3.18)	(0.74)	(0.39)
Dummy, Brand Name	-0.0055	0.0026	0.0043	0.0065	-0.0061
.	(1.49)	(3.30)	(0.64)	(1.84)	(4.10)
Dummy, Flat Screen	0.237	0.2298	0.2527	0.1812	0.2127
.	(23.60)	(5.45)	(1.21)	(6.87)	(35.54)
Log, Seller's Feedback +1	-0.0222	-0.0204	-0.0263	-0.0252	-0.0146
.	(1.51)	(0.42)	(1.43)	(0.56)	(0.19)
Distribution Variable+	1.3528	-0.6531	-1.7121	NA	0.8402
.	(4.91)	(1.00)	(2.60)	NA	(0.72)
Number of Auctions	2934	2934	2934	2934	2934
-Log Likelihood/Number of Auctions	1.1039	3.7662	3.7245	3.8077	3.896

+ For the Log-Normal this Parameter is the Standard Deviation. For the Weibull, Gamma, and Pareto this is the log of the shape parameter. Standard deviations are reported in parentheses below the coefficients.

* Coefficient is significant at the 10% Level. ** Coefficient is significant at the 5% Level. *** Coefficient is significant at the 1% Level.

Table 3

.	Estimates of the Entry Parameter.				
.	Log-Normal	Weibull	Gamma	Logistic	Pareto
Constant	-18.7408***	-11.2964***	-8.9873***	-14.2597***	-13.8536***
.	(8.76)	(1.59)	(1.78)	(1.75)	(6.42)
Log, Size	4.751**	4.6517***	4.5342***	4.516***	4.6355**
.	(2.66)	(0.51)	(0.54)	(0.75)	(2.72)
Log, Dot Pitch	-13.3547	-10.4419	-10.6719***	-6.9808	-7.4747***
.	(29.33)	(233.34)	(4.06)	(42.23)	(0.71)
Dummy, No Dot Pitch	15.7051***	11.9219***	12.2254***	8.2244***	8.4918
.	(6.50)	(1.21)	(1.75)	(3.27)	(13.04)
Log, Resolution	-1.4549**	-1.9028	-2.1986	-0.9195*	-1.0805
.	(0.76)	(25.85)	(3.62)	(0.62)	(1.86)
Dummy, No Resolution	-10.6756	-13.8705***	-15.9685***	-6.6961***	-7.9835***
.	(31.17)	(1.48)	(1.59)	(1.15)	(1.11)
Dummy, New	0.6895	7.5107	7.4086	7.3883	7.5322
.	(4.86)	(1625.75)	(943.09)	(168.29)	(201.18)
Dummy, Like-new	1.4005	6.0785	6.5766	6.5504	6.6402
.	(7.22)	(278.19)	(1119.23)	(144.42)	(288.70)
Dummy, Refurbished	0.0667	0.0394	0.0053	0.0888	0.1226
.	(45.67)	(7.20)	(7.43)	(85.63)	(1.24)
Dummy, Warranty	0.7048	0.8485	0.7892	1.0237	1.0245
.	(21.09)	(95.26)	(75.16)	(3.99)	(1.24)
Dummy, Brand Name	0.0447	0.0317	0.0218	0.0107	0.0506
.	(3.31)	(1.62)	(1.46)	(4.34)	(1.62)
Dummy, Flat Screen	-0.3016	-0.2701	-0.3433	-0.0075	-0.1825
.	(51.59)	(17.55)	(7.11)	(2.81)	(1.20)
Log, Seller's Feedback +1	2.0296	1.9942	2.1961	0.9276	1.4721
.	(8.12)	(2.66)	(3.49)	(3.51)	(1.21)
Log, (Seller's Feedback $+1)^2$	-0.9316	-0.9262	-1.027*	-0.4004	-0.6663*
.	(1.66)	(2.66)	(0.72)	(0.34)	(0.42)
Category Dummy, $\leq 17''$ Screen	0.7227	0.7289	0.727	0.4473	0.7222
.	(10.60)	(4.92)	(2.01)	(1.54)	(9.50)
Category Dummy, $\geq 19''$ Screen	-0.5309	-0.4429	-0.4344	-0.6287	-0.4136
.	(18.00)	(2.50)	(1.00)	(1.81)	(2.35)
Category Dummy, Monotonic	-3.455	-1.2607	-3.772	-3.6809	-3.7338
.	(73.33)	(5.44)	(79.11)	(63.76)	(26.21)
Dummy, Secret Reserve	1.444	1.4702	1.7837	0.9709	1.3793
.	(8.53)	(9.76)	(8.72)	(1.94)	(56.80)
Dummy, Secret Reserve not met.	-0.705	-0.3936	-0.6446	-0.4306	-0.3846
.	(11.86)	(161.97)	(7.79)	(1.82)	(2.15)
Number of Auctions	2934	2934	2934	2934	2934
-Log Likelihood/Number of Auctions	1.1039	3.7662	3.7245	3.8077	3.896

Standard deviations are reported in parentheses below the coefficients.

* Coefficients are significant at the 10% Level. ** Coefficients are significant at the 5% Level. *** Coefficients are significant at the 1% Level.

A.1-Descriptive Statistics

	Number	Mean	Std. Dev.	Skewness	Median	Maximum	Minimum
Sales Price	(1)	135.7	132.85	2.01	100	1430	0.01
Log, Size	(2)	2.82	0.14	0.28	2.83	3.04	2.64
Log, Dot Pitch	(3)	-.49	0.65	-0.58	0	0	-1.61
Dummy, No Dot Pitch	(4)	0.64	0.48	-0.56	1	1	0
Log, Resolution	(5)	4.32	3.4	-0.48	6.68	7.38	0
Dummy, No Resolution	(6)	0.38	0.49	0.49	0	1	0
Dummy, New	(7)	0.07	0.26	3.31	0	1	0
Dummy, Like-new	(8)	0.03	0.17	5.38	0	1	0
Dummy, Refurbished	(9)	0.13	0.33	2.23	0	1	0
Dummy, Warranty	(10)	0.03	0.17	5.51	0	1	0
Dummy, Brand Name	(11)	0.59	0.49	-0.36	1	1	0
Dummy, Flat Screen	(12)	0.18	0.38	1.71	0	1	0
Log, Seller's Feedback +1	(13)	3.9	1.95	-0.44	4.36	8.38	0
Log, (Seller's Feedback +1)^2	(14)	7.7	4	-0.44	8.7	16.75	0
Length of Auction	(15)	5.08	2.16	0.65	5	10	3
Category Dummy, <=17“ Screen	(16)	0.62	0.49	-0.5	1	1	0
Category Dummy, >=19“ Screen	(17)	0.27	0.45	1.02	0	1	0
Category Dummy, Monotonic	(18)	0	0.05	22.06	0	1	0
Dummy, Secret Reserve	(19)	0.18	0.38	1.68	0	1	0
Dummy, Secret Reserve not met.	(20)	0.1	0.3	2.69	0	1	0
Number of Bidders	(21)	3.92	4.06	1.06	3	22	0
Dummy, No Bidders	(22)	0.27	0.45	1.02	0	1	0
Dummy, One Bidder	(23)	0.13	0.34	2.2	0	1	0

A.2-Correlation Matrix of Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1)	1																
(2)	0.74	1															
(3)	-0.43	-0.31	1														
(4)	-0.4	-0.29	1	1													
(5)	0.16	0.13	-0.29	-0.29	1												
(6)	-0.14	-0.12	0.29	0.29	-1	1											
(7)	0.22	0.07	-0.09	-0.09	-0.01	0.01	1										
(8)	0.02	-0.01	0.01	0.01	-0.01	0	-0.05	1									
(9)	-0.02	-0.01	-0.02	-0.01	-0.04	0.04	-0.11	-0.06	1								
(10)	0.17	0.12	-0.13	-0.13	0.07	-0.06	0.13	-0.03	-0.01	1							
(11)	0.01	0.03	0.01	0.01	-0.01	0.01	0.01	-0.02	0	0.01	1						
(12)	0.28	0.2	-0.31	-0.31	0.21	-0.2	0.11	-0.03	-0.01	0.14	-0.01	1					
(13)	-0.08	-0.05	-0.06	-0.06	0.15	-0.15	-0.1	-0.01	-0.03	-0.01	-0.03	-0.01	1				
(14)	-0.08	-0.05	-0.06	-0.05	0.15	-0.15	-0.1	-0.01	-0.03	-0.01	-0.03	-0.01	1	1			
(15)	0.06	0.04	0.02	0.02	-0.19	0.18	0.12	0.04	0.05	-0.01	0	-0.01	-0.36	-0.36	1		
(16)	-0.53	-0.68	0.21	0.19	0	-0.01	-0.03	0	-0.01	-0.11	-0.01	-0.09	0.05	0.05	-0.07	1	
(17)	0.62	0.79	-0.19	-0.17	0.05	-0.04	0.07	0.02	0	0.13	0.04	0.11	-0.07	-0.07	0.05	-0.78	1
(18)	-0.02	0.04	0.03	0.03	-0.01	0.01	0.02	-0.01	0.01	-0.01	-0.01	0	-0.02	-0.02	0.05	-0.06	-0.03
(19)	0.27	0.23	-0.14	-0.14	0.05	-0.04	0.13	0.03	-0.01	0.05	0.01	0.26	-0.23	-0.23	0.05	-0.09	0.15
(20)	0.14	0.14	-0.08	-0.08	0.03	-0.03	0.11	0.01	-0.01	0.04	-0.02	0.21	-0.16	-0.16	0.02	-0.04	0.07
(21)	0.24	0.23	-0.16	-0.16	0.01	-0.01	0.08	0.11	-0.02	0.03	0	0.19	-0.05	-0.06	0.07	-0.09	0.18
(22)	-0.03	-0.04	0.06	0.07	0.08	-0.08	-0.03	-0.06	-0.01	-0.02	-0.01	-0.09	0.09	0.1	-0.14	0.01	-0.07
(23)	-0.08	-0.08	0.03	0.04	-0.02	0.02	-0.03	-0.05	0.02	0.03	0.03	-0.07	0.04	0.05	0.01	0.02	-0.05

	(18)	(19)	(20)	(21)	(22)	(23)
(18)	1					
(19)	-0.02	1				
(20)	-0.01	0.71	1			
(21)	-0.04	0.28	0.16	1		
(22)	-0.03	-0.23	-0.13	-0.59	1	
(23)	0.07	-0.13	-0.09	-0.28	-0.24	1