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 25, 2005

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

1 Introduction

- It is well established that eBay is a significant economic marketplace.

- Economists have long hailed the price discovery power of auctions but the cost of getting bidders together prevented their widespread usage.

- eBay overcame this problem by allowing people to auction items over the Internet.

- We are still unsure how much eBay benefits the economy.

- One measure of this benefit is the Consumer Surplus that eBay generates.

- Our paper measures this important attribute in the market for Computer Monitors.

- We estimate bidders' values and an entry process using maximum likelihood.

- Instead of nonparametric approaches we estimate our function using multiple distributional assumptions. This allows us to:

  - estimate Consumer Surplus under different distributional assumptions

  - test for sensitivity to distribution.

  - test which distribution best fits the data

  - test how well each estimate performs against a non-specified non-parametric distribution.

- Our data set is almost 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.

- Few attempts have been made to estimate Consumer Surplus.
  - Song (2004) estimates a semi-parametric model using both the second and third highest bids in university yearbook auctions using innovative semi-parametric methods.
  - Bapna, Jank, and Shmueli (2005) use a revolutionary new data collection technique that allows them to directly observe a bidder's stated value.
    - They do not estimate a structural bidding function
  - Bapna, Paulo and Gupta ((2003a, b) and Bapna, Goes, Gupta, and Jin (2004) focused on mechanism design and look at first price or multi unit auctions.

- The common format for eBay 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 (1993).
  - We do not have to estimate the minimum or maximum value a bid can take
  - 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 can 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.
  - Semi-parametric techniques such as in Song (2004).
  - The Bayesian methodology developed in Bajari and Hortaçsu (2003)
  - Non-linear Simulated Least Squares methodology developed by Laffont, Ossard, and Vuong (2005).

2 The Data Set and Our Data Collection Protocols

- 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.
  - It then iterates through the database entries stored, and creates a tab-delimited ASCII file.
- We collected 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.

- 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.

- Gonzalez, Hasker, and Sickles (2005) detail the strings used for each variable

- We were able 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
- whether it was a flat screened monitor.

- Descriptive statistics are in Tables A-1 and A-2.

3 Model and the Maximum Likelihood Functions

- We use maximum likelihood to estimate bidders' values and begin with an exogenous entry process.

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

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

  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.

- $F_n(\beta)$ is the cumulative distribution function (CDF) of the bidders' values

- $f_n(\beta)$ is the probability density function (PDF)—where β may include some distribution specific coefficients.

- D_0 as the dummy which equals one if there are no bidders

- D_1 as the dummy which equals one if there is one bidder.

- I is the number of bidders

- Likelihood of auction n given I is:

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

- I is 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.

- Although we observe only $\underline{I}_n$ bidders in an auction there might 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, are log-linear in a set of auction specific characteristics z_n

- The estimated functional form for entry is:

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

- T_n is the length of the auction ($T_n \in \{3, 5, 7, 10\}$)

- D_{sr} is a dummy which is one if there is a secret reservation price.

- The likelihood for auction n is:

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

- $\underline{I}_n$ is increased by one if there is a secret reservation price following Bajari and Hortaçsu (2003) in treating the auctioneer as another bidder if there is a secret reservation price.

- 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.

- Since we can not be certain *a-priori* what the true distribution of bidders values is we test several different distributions:

  - Log-Normal
  - Weibull
  - Gamma
  - Logistic
  - Pareto

4 Preliminary Estimates

- The right hand side variables in our models are:
  - are the size of the monitor
  - diagonal screen size
  - dot pitch (the distance between dots on the screen)
  - resolution
  - size of picture seen on the monitor
  - dummies indicating whether or not the monitor is New, Like New, or Refurbished, whether or not the monitor has a Warranty, is a Brand Name, or is Flat panel.
  - Seller's Feedback-increases by one with every sale that results in a pleased customer-is both an indicator of the Seller's experience and reputation.

- Results are in Table 1
  - General stability of coefficients across estimates.
  - Significant variation in two dummies, the "Like New" dummy and the Warranty dummy.
  - Both of these coefficients are significantly smaller with Logistic private values.
  - Refurbished monitors appear to be no better than Used monitors.
  - 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 but it is insignicant which is consistent with findings in Song (2004).

- Total seller's feedback does not affect the value of the good. In contrast Bajari and Hortaçsu [**?**] 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.

- The coefficients of the entry process are presented in Table 3.

- 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.

- 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.

- While a high Resolution raises the item's value it seems to lower the expected number of bidders, indicating some heterogeneity in our bidders.

- Our results illustrate that 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.

- The coefficients on Seller's Feedback seem 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-if they decide to try and buy it they discount the seller's lack of experience.

- Estimates of the secret reservation dummy are problematic. Bajari and Hortaçsu [**?**] found it had a negative coefficient.

- There is no theoretical explanation for it having a positive coefficient.

- A potential problem is that this variable is endogenous. We are currently examining methods to deal with this.

- Estimates 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.

5 Consumer Surplus

- While the *a-priori* Consumer Surplus is a function of I the *ex-post* Consumer Surplus is not

- 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 $I \geq 2$):

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

when $I = 1$ this is:

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

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

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.*

- Using this insight we can estimate ex-post Consumer Surplus. We 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 which appear to casue problems in more direct estimates of Consumer Surplus..

- These lower bound estimates of Consumer Surplus are in Table 6.

- Estimates are quite comparable across different distributions and are quite stable.

- Results point to a significant amount of consumer surplus captured in these auctions.

- Compared to 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.

- We use the Akaike Information Criterion, Bayesian Information Criterion (Schwartz Criterion) and the Browne–Cudeck Criterion.

- Results from the different distributions for the AIC, the BIC, and the BCC criteria are in Table 7.

- Log-Normal is the best distribution by all measures while the the other distributions cluster together in their statistical validity.

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 $v^{3:I} = \frac{b_n^{3:I}}{e^{x_n'\beta}}$:

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

and:

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

- 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.