Estimating Consumer Surplus in eBay Computer Monitor Auctions.*

Tugba Giray
Economics Department
Bilkent University
Ankara, Turkey

Kevin Hasker
Economics Department
Bilkent University
Ankara, Turkey

Bibo Jiang
Economics Department
Rice University
Houston, Texas

Robin Sickles
Economics department
Rice University
Houston, Texas

July 28, 2009

Abstract

Our paper utilizes semi-nonparametric and nonparametric methods to directly estimate consumer surplus in eBay computer monitor sales. We also use nonparametric methods to select among alternative parametric models wherein the private values are assumed to have log-normal, gamma, Weibull, half-logistic, and Pareto distributions. We find substantial comparability between the estimated consumer surplus based on nonparametric methods and the parametric model selected using the nonparametric testing criteria. Although nonparametric estimators have their advantages in terms of robustness to distributional misspecification they require substantial sample sizes in order for standard asymptotic results to hold when observed covariate effects must be estimated. This suggests that at least for our eBay data, parametric models chosen in this way may offer advantages that can overcome their potential lack of distributional robustness.

JEl classification: C22, C51

Key words and phrases: eBay auction, semi-nonparametric estimation, nonparametric estimation, consumer surplus, consumer share of surplus

*We would like to thank Sandra Campo and participants at the Conference on Auctions and Games, Virginia Polytechnic Institute and State University, Blacksburg, Va., October 12- 14, 2007 as well as Ravi Bapna and participants of the *FTC Roundtable: Economics of Internet Auctions*, October 27th, 2005, Washington, DC. for insightful and needed criticism on earlier drafts of this paper. The usual caveat applies.

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 the cost of establishing a cohesive electronic market place 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 it is likely to remain one in the future. However 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 economic fundamental in the market for computer monitors. There are, however, important methodological issues that must be addressed in any empirical study of auctions. As has been shown for second price private-value auctions considered in this paper, the standard model is nonparametrically identified and nonparametric methods can be employed whose asymptotic properties of consistency and normality remain unchanged under alternative distributions of private values. However, when auction models such as these are faced with the necessity of controlling for observed heterogeneity with covariates, parametric models have an advantage. Although semi-nonparametric and nonparametric methods may accommodate regressors, the rates of convergence are slowed substantially by the curse of dimensionality problem inherent in applied nonparametric analyses. Our paper suggests an alternative middle ground that empirical researchers may find appealing and which ultimately may address both the lack of robustness of parametric methods and the curse of dimensionality problem. We propose a testing strategy in which nonparametric methods are used as model selection criteria for a finite set of parametric submodels. This has a link to the semiparametric efficiency bound literature which we do not formally address in this paper.[1]

In addition to nonparametric and semi-nonparametric methods to estimate consumer surplus, the nonparametric model selection criteria we use chooses among a finite set of alternative parametric submodels and the consumer surplus measures estimated therein. The distributions of private values that complete our family of parametric submodels, are the log-normal, gamma, Weibull,

[1] Models in which semiparametric efficiency bounds and adaptive estimators for order statistic models, such as those used in the auction literature to identify an equilibrium observation can found in,among others, Adams, Berger, and Sickles (1999), Adams and Sickles (2007), Park and Simar (1994), and Park, Sickles and Simar (1998, 2003, 2007). These papers consider such estimators for samples with repeated measures such as those used in the empirical auction literature.

half-logistic, and Pareto distributions–all with positive support. We find substantial comparability between the estimated consumer surplus based on semiparametric and nonparametric methods and the parametric submodel selected using the nonparametric testing criteria. Although nonparametric estimators have their advantages in terms of robustness to distributional misspecification they require substantial sample sizes in order for standard asymptotic results to hold when observed covariate effects must be estimated. This suggests, at least for our eBay data, parametric models selected using this model selection criteria may offer advantages that can overcome their potential lack of distributional robustness.

Our data set for this analysis contains 2934 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. Although recent methods for accessing data via "spider" programs have become commonplace[2] we believe ours was the first use of such a program to construct such large scale data sets (Gonzales, 2002, Gonzalez, et al. 2007, revised). We also discuss the data collection techniques that allowed us to construct our relatively large set of auction data.

Relatively few attempts have been made to estimate consumer surplus in auction models, although this is presumably one of the arguments in favor of such mechanisms in terms of consumer benefits. Song [27] estimates a semi-nonparametric model using both the second and third highest bids in university yearbook auctions. She constructs an innovative methodology using the second and third highest bids and estimates the median consumer surplus in university yearbook auctions at $25.54. With a median price in her study of $22.50, the median consumers' share of the surplus is 53%. Our strategy is to search over various parametric models, which not only allows us to dispense with the need to use the third highest bid but it allows us to suggest a best parametric model which might be useful in other research. We also provide comparable semi nonparametric and nonparametric estimates of consumer surplus which utilize second and third highest bid data. Bapna, Jank, and Shmueli [6] also estimates consumer surplus, utilizing an innovative data collection technique that allows them to directly observe a bidder's stated value. With their rather heterogenous data, however, they cannot estimate a structural bidding function. They do, however, find that consumers capture at least 18.3% of the total surplus. Several other articles estimate consumer surplus in multi-unit auctions—Carare [10];

[2] See, for example, the website at http://www.baywotch.de/. We thank Rouwen Hahn from the University of Münster, Germany for this information.

Bapna, Paulo and Gupta [3], [4] and Bapna, Goes, Gupta, and Jin [5]—but these papers primarily focus on mechanism design issues and tend to use ad hoc techniques since the equilibrium bidding function in general multi-unit auctions is unknown.

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 original data set and the type of auctions on which we focus. When our data was collected bidding went from three to ten days and then stopped at a preset time.

Our estimation techniques are based on methods developed by Donald and Paarsch [13]. We do not need 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 also are able to estimate a full likelihood function since our data set includes all auctions where no one decided to bid.

There are a number of methodologies available in the literature. The non-parametric technique developed by Song [27] 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 it is unclear that compelling arguments can provide the same guarantee for the third highest bid. One must instead rely on the bidders planning not to update their bid and if this is not the case then results based on these methods could be biased. Adams [1] has recently significantly extended Song's technique and developed a methodology that non-parametrically identifies both the distribution of bidders' values and the probability of entry. This is a significant advancement because it is based only on the second highest bids. However, the current theory requires that there is some variable that affects entry but does not affect bidders' values. We do not find support for this assumption in our data. There are of course other generic problems with non-parametric estimators such as slow rates of convergence that are exacerbated by the need to estimate moments of sample order statistics. Another interesting technique is a Bayesian methodology developed in Bajari and Hortaçsu [2]. However these techniques require that the bidding functions are linearly scalable, a restriction unnecessary with our approach and violated by our structural form. Non-linear simulated least squares, developed by Laffont, Ossard, and Vuong [20] [17], is but another estimation methodology. This approach overcomes the complexity of calculating the likelihood function by simulating the auctions, and it is a flexible methodology that can be used for any bidding model where revenue equivalence holds.

Our paper is organized as follows. In section 2 we discuss our data set and collection techniques. In section 3 we outline the model we use in our estimation. Section 4 presents our estimates and briefly discusses them. In section 5 we present the core of our results, our estimates of consumer surplus. We also discuss a new measure of presenting consumer surplus, consumer share. Consumer share is the fraction of total available surplus that consumers capture. In section 6 we conduct a series of tests of which distribution best fits the data. Section 7provides our semi-nonparametric and nonparametric estimates. Section **??** concludes.

2 The Data Set and Our Collection Techniques.

At the time our data set was collected eBay saved all information about closed auctions on their website for a month after the auction closed. This allowed 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 stored database entries and creates a tab-delimited ASCII file.

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.

Running this program from February 23, 2000 to June 11, 2000 we were able to capture information on approximately 9000 English auctions of computer monitors, effectively all monitors auctioned during that time period. We deleted many of these monitors because they are clearly not in the same market as the one we are interested in—PC color computer monitors with a size between 14 and 21 inches. We also deleted some of our observations because they didn't contain enough information to construct measures of the competition each auction faced, resulting in 2934 auctions being used in our estimates. The monitors we analyze were in working order, were not touch screen monitors or LCD monitors, Apple monitors, or 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. We also deleted several auctions in which the auctioneer cancelled the auction early (usually within ten to fifteen minutes of the beginning of the auction.)

Since the level of competition a given auction faces is based on the number of auctions open at the same time we dropped auctions from our sample that took place within ten days of a break in our data collection. We then counted the number of auctions (including the given auction) that were open and had the same size of monitor, or open and were in the same category. This gave us three variables that measure the amount of competition an auction faces.

Descriptive variables except for monitor size were constructed using string searches. In Gonzalez, et al. [17] 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 close to a 10 percent market share while the smallest have close to a 3% market share. These 10 firms represent 57% 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. Descriptive statistics of the variables are presented in the Appendix.

3 The Model and Likelihood Functions.

In this section we use maximum likelihood to estimate bidders' values and an exogenous entry process in eBay auctions.

Bidding on eBay takes place by a proxy program. The bidder submits a reservation price and the computer raises the price until only one bidder remains. In such an auction the obvious action is to enter your reservation value (or simply "value") as your reservation price. However, bidders frequently do not do this, so to be certain that at least the second highest bidder does we will follow Haile and Tamer [18] by assuming that bidders follow two intuitive rules:

1. No bidder ever bids more than he is willing to pay.
2. No bidder allows opponents to win at a price he is willing to pay.

We assume that bidders' values are private, independent, and log-linear in a set of auction specific characteristics x_n (where n indicates the auction) and a private component ρ_j (where j indicates the person). If the winning price is b_n^w, r_n is the traditional open reservation price, and I is the number of potential bidders who bid in the auction the formula for the winning bid 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 private component of the second highest bidder in auction n. We allow for various models for the distribution of values and thus the distribution of b_n^w.

Let $F_n(z,\beta)$ be the cumulative distribution function (cdf) of the bidders' values at z and $f_n(z,\beta)$ be the probability density function (pdf)—where β may include some distribution specific coefficients. Let I_n^a be the number of active bidders in auction n—or the number who actually submitted bids, and for $i \in \{0,1\}$ $D_n^i = 1$ if $I_n^a = i$, $D_n^i = 0$ otherwise. If $I \geq I_n^a$ is the number of potential bidders in auction n the likelihood of auction n given I is:

$$\begin{aligned} l_n(\beta|I) &= \left(F_n(r_n,\beta)^I\right)^{D_n^0} * \\ &\quad \left(IF_n(r_n,\beta)^{I-1}(1-F_n(r_n,\beta))\right)^{D_n^1} * \\ &\quad \left(I(I-1)F_n(b_n^w,\beta)^{I-2}(1-F_n(b_n^w,\beta))f_n(b_n^w,\beta)\right)^{\left(1-D_n^0-D_n^1\right)} . \end{aligned} \tag{2}$$

We can not identify *inactive* bidders. Although we know there have been at least I_n^a who have bid but there might also be any number of bidders who thought about bidding and did not. Therefore I will be a stochastic variable that can range from I_n^a to $\overline{I}$—an arbitrary upper bound. One can view our treatment of identification and estimation of the number of bidders I as a direct treatment for what would otherwise be unobserved heterogeneity in each auction that potentially could be correlated with outcomes of the bidding process. We do not explicitly correlate the number of bidders with the other heterogeneity controls for differences in monitors types or the feed-back rating of the seller, but rather estimate this potential auction specific unobservable in our model. For an extensive treatment of unobserved heterogeneity in first-price auctions see Bierens and Song (2006a,b).

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:

$$\ln \lambda_n = z_n' \gamma \ . \tag{3}$$

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

$$l_n(\beta, \gamma) = \frac{\Sigma_{i=I_n^a+D_n^{sr}}^{\bar{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n} l_n(\beta|i)}{\Sigma_{i=D_n^{sr}}^{\bar{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n}} \ . \tag{4}$$

notice that the lower bound on i in both the denominator and the numerator are increased by one if there is a secret reservation price, thus we are following Bajari and Hortaçsu [2] 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. This is a rarity in auction data.

The choice of $\bar{I}$ is obviously arbitrary. In order to derive the estimates we chose $\bar{I} = 30$ and then tested the results with $\bar{I} = 50$. Changing $\bar{I}$ did not affect the coefficients. It appears our choice of $\bar{I} = 30$ is reasonable.

Since we cannot be certain *a-priori* of the true distribution of bidders' values, we utilize an array of distributions that have been proposed in the literature: half-logistic, gamma, Weibull, log-normal, and Pareto, all with positive support.

4 The Estimates

We first present estimated effects of the exogenous variables ($e^{x_n' \beta}$) on the winning bids. We then present estimates of the parameter of the entry process (λ). 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), and resolution (the size of picture that can be 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.

<Table 1 about here>

Results are in Table 1 in the Appendix. We first note the general stability of coefficients across estimates. The least stable coefficients are those for dot pitch and resolution. The coefficient on band name changes sign in one regression and is never significant, probably because what information brand name really conveys to the bidder is that the monitor is a common brand. While the differences in coefficients are generally small, the value of a computer monitor can be very different for a given auction using the different distributions. In Table 2 we report the descriptive statistics of the logs since the exponential function creates a skewed distribution and significantly biases the mean.

Table 2 – Log of Exogenous Values

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Average	3.63	4.58	2.38	1.74	3.33
Exponential of Ave.	\$37.71	\$97.51	\$10.80	\$5.70	\$27.94
Median	3.64	4.57	2.37	1.72	3.32
Exponential of Med.	\$38.09	\$96.54	\$10.70	\$5.58	\$27.66

The exponential of the average and median are in dollar terms Looking at either the median or the average one can see a wide variation in the estimates. The gamma distribution produces the highest estimate (\$98) the Pareto and half-logistic both predict medium values (\$28 or \$38 respectively), and the log-normal and Weibull distributions produce very low values (\$6 – \$11).

In our estimation of the entry process we include all variables that affect bidders' values and other variables that only affect entry. The first two additional variables are the log of the seller's feedback and a squared term for seller's feedback—allowing for a decreasing marginal benefit of experience. Feedback increases by one with every sale that results in a pleased customer, so it is an indicator of both the seller's experience and reputation. 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 monochrome sub-categories if he wishes. Notice that all monitors that are put in the monochrome sub-category are misplaced—all monitors in our data set are color monitors. Since our data collection process captured every monitor auctioned during our sampling period, we can construct three variables capturing the amount of competition a given auction faced. The first variable (Competing Auctions, all) is simply how many other auctions were open while this auction was running, divided by the length of the auction. The second variable (Competing Auctions, Same Size) is how many of those open auctions had a monitor of the same size normalized, again, by the length of the auction. The final one (Competing Auctions, Same Cat.) is the number of open auctions in the same category, and this variable is also normalized.

<Table 3 about here>

While these coefficient estimates are less stable than the estimates of bidders' exogenous values in general, the sign is consistent in all estimates and only one coefficient is surprising. This is the coefficient on the secret reservation price dummy. Theoretically we expect this coefficient to be negative but the estimate is positive—though generally statistically insignificant. We have tried using instruments for this variable and have also allowed for a distinct arrival probability for the first bidder but estimates of the impact of a secret reservation price remain relatively unchanged. We think the most likely reason for this positive and insignificant coefficient is because secret reservation prices are often used on items with very high values. On these items there will be more entry and the log-linear model of entry may not be able to fully capture this increase. Indeed, even if we control for the effect of the secret reservation price on entry, auctions with a secret reservation price generally can expect 50% more bidders than auctions without a secret reservation price.

It is also interesting to note that coefficients on the open reservation price are insignificant. The log-normal specification again has the wrong sign. It is well known that on eBay auctioneers are afraid to raise their reservation prices thinking they might drive bidders away. These coefficients indicate that they succeed in this goal and bidder behavior is essentially unaffected by the open reservation price.

The coefficients on the competitive variables have an interesting implication. Increasing the number of competing auctions increases the likelihood that a bidder will enter a given auction. This dramatically illustrates the economies of the marketplace that is the basis of eBay's market power. The primary reason buyers visit eBay instead of its competitors is because so many sellers use eBay. Likewise the sellers go to eBay because buyers use eBay. Thus we can not be certain that increasing the number of auctions will decrease the number of bidders per auction and these coefficients support this economic insight. Notice that increasing the number of competing auctions in an item's category decreases the number of bidders per auction. It is interesting to be able to find empirical support of the hypothesized reason for eBay's success, and would be exciting to know if these effects continue in the current market where there are a many more auctions each day.

The change in sign on the coefficient of the dummy for not reporting dot pitch merely indicates that in the log-normal specification not reporting dot pitch always decreases the number of bidders who enter. For other specifications it only sometimes decreases the number who enter. Again, except for this case, the ratio of the coefficients on dot pitch and resolution to the dummies for not reporting these variables are stable, around -0.85 for dot pitch and 0.14 for resolution. 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 or page of text. Of course this also means that the details of the picture or the text size are smaller, 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. This same effect is sometime seen with refurbished and brand name monitors.

The coefficients on the log of seller's feedback and its square deserve special attention. Notice that these coefficients are significant in every regression but the half-logistic, and the coefficient on the square indicates a diminishing marginal impact of increasing feedback. There are two reasons that we should expect a positive impact of feedback. First, more experienced auctioneers know more about setting up auctions. For example, their descriptions will be clearer or there will be pictures of the item for sale and this tends to encourage bidders to enter. Second, these coefficients could reflect bidders not trusting a seller who has a negative feedback. Indeed, Cabral and Hortaçsu [9] follow auctioneers over time and find that one negative feedback can decrease the growth rate of an auctioneer's sales from 7% to −7%.

Many papers have terms reflecting a seller's feedback in the sales price regression. However most of these papers do not estimate an entry model. If high feedback is actually affecting the number of bidders who enter the auction this would result in a higher sales price in a model where entry is not estimated. Perhaps regressions that do not explicitly model entry should include the number of bidders as a right hand side variable. In the few papers which do estimate both entry and sales price, the evidence is mixed, with much of it supporting our findings. Of these papers the only one that estimates a structural model of entry (Bajari and Hortaçsu [2]) does not include feedback in the entry equation. Jin and Kato [19] find that there is a trivial and negative effect of feedback on sales price but it has a significant and positive effect on the probability of sale. The methodology in Song [27] allows for any non-specified exogenous entry process, and she finds that feedback has an insignificant but positive impact on the bidders' values. Bajari and Hortaçsu [2] find that feedback has a positive and significant effect on bidders' values, but this regressor is missing from their entry equation. Livingston [21] found a significant and positive impact on both price and the probability of sale.

It may be that the literature has not paid enough attention to whether feedback affects entry or values. In our model trust is binary, either bidders trust a seller or they don't. If they trust the auctioneer they enter and bid. In a model where feedback affects values bidders who do not trust an auctioneer enter but shade their bids. We hope that in future research others will test our simple model of trust versus the more complicated models others have used.

We have examined non-homogeneous models of Poisson entry. Specifically we have allowed the probability of zero bidders to be unconstrained. Casual observation of the pdf of the number of bidders in Figure 1 indicates there are too many zero occurrences.

Figure 1: PDF of the Number of Bidders.

The average number of bidders is 3.92. One might expect that the pdf should be symmetric, increasing up to about 4 and then decreasing after that. Of course this does not take into consideration the covariance of our right hand side variables with the number of bidders. We found that allowing for a different probability of zero bidders did not meaningfully alter results.

We report in Table 4 the descriptive statistics of the log of λ since the exponential introduces a right skewness to the distribution.

Table 4 – Log of Entry Parameter

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Average	2.01	2.86	2.63	5.04	2.49
Exponential of Ave.	7.5	17.46	13.9	154.5	12.1
Median	1.28	1.93	1.82	2.32	1.7
Exponential of Med.	3.6	6.89	6.2	10.2	5.5

Even though we look at the log of λ there is right skewness in the distribution, which is illustrated by the differences between the average and the medians. For this reason, we prefer to look at the medians. For the gamma, Weibull, and Pareto the estimates of λ are essentially the same. The median number of bidders per day is between 6 and 7. For the half-logistic and the log-normal they vary significantly. The half-logistic it is surprisingly low and for the log-normal it is surprisingly high.

In our model λ is the expected number of bidders per time unit (one day). Thus in a three day auction the median number of bidders will be between 10 and 30, in a ten day auction between 35 and 100. Even our most conservative estimates suggest there are a large number of potential bidders for each computer monitor.

5 Consumer Surplus and Consumer Share.

Perhaps one of the most useful summary statistics for understanding the welfare impact of eBay is consumer surplus. The welfare impact can be calculated directly from our estimates because it is independent of the number of bidders in an auction. Ex-ante consumer surplus is derived in Appendix C for the different parametric distributions. *Ex-post* consumer surplus (the received welfare impact of the sale) in auction n is:

$$E\left(v_n^{(1:I)}|v_n^{(2:I)} = b_n^w|I \geq 1\right) - b_n^w \tag{5}$$

where $r_n^w = b_n^w$ when $I = 1$. For all $I \geq 1$ *ex-post* consumer surplus is:

$$\frac{\int_{b_n^w}^{\infty} z f_n(z, \beta)}{1 - F_n(b_n^w, \beta)} - b_n^w \tag{6}$$

The proof is nothing more than simplifying $E\left(v_n^{(1:I)}|v_n^{(2:I)} = b_n^w|I \geq 2\right)$ to show that it is equal to $E\left(v_n^{(1:I)}|v_n^{(2:I)} = b_n^w|I = 1\right)$, which is the above expression.

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

Proof. The explicit form of the expectation is:

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

if we let $b_n^w = r_n$ when $I = 1$. ∎

Using this insight we can estimate *ex-post* consumer surplus for the different parametric distributions we consider in this section. Summary statistics for estimates of consumer surplus for the various distributions are given in Table 5.

Table 5 – Expected Consumer Surplus

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Average	\$52.54	\$95.58	\$112.90	\$194.81	\$190.91
Median	\$39.69	\$72.26	\$87.55	\$142.95	\$139.72

There is wide variation in these estimates across distributions and in general the estimates are high given our median computer monitor sold for \$100. There is also significant right skewness. Because of this we will focus on median estimates in the remaining discussion. The half-logistic produces the lowest estimates, and the gamma the next lowest (their averages are actually the same). The distributions that fit the worst (the log-normal and the Pareto) also produce the highest estimates.

An alternative measure of consumer surplus is the consumers' share of the surplus instead of the consumer's absolute level of surplus. If v_b is the buyers's value and v_a the auctioneer's value then this statistic is $\frac{v_b - b_n^w}{v_b - v_a}$, or in terms of consumer surplus it is $\frac{CS_n}{CS_n + b_n^w - v_a}$. We can not accurately estimate v_a from our data and set it equal to its lower bound of zero. Using this assumption the estimates of the consumer share of surplus are presented in Table 6.

Table 6 – Expected Consumer Share of Surplus

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Average	0.333	0.451	0.489	0.624	0.573
Median	0.298	0.433	0.472	0.614	0.559

This gives a clearer picture of how much of the total surplus the bidders are capturing. Even though there are many bidders per monitor consumers are still receiving between one third and two thirds of the surplus.

These estimates also can be compared with those from other studies. Song [27] uses a non-parametric distribution and her median consumer share is 53% in yearbook auctions. Bapna et al. [?] elicit bidders' values via a sniping program and estimate the average consumer share to be 18%. Our results are closer to Song's estimates though her average number of bidders in auctions with three or more bidders is 3.6 while ours is 6.8. This implies our auctions are more competitive yet our consumer shares are similar, possibly because either the non-parametric techniques understate the tail probabilities or because parametric techniques overstate them.

The difference between our results and those of Bapna et al. [?] may also be due to the type of items their clients are buying and/or sample selection problems. In general those who use a sniping program are the most concerned about getting the lowest possible price. These bidders might both have low reservation values and might understate their values to the sniping program.

We can check the sensitivity of our estimates to the tail properties of our set of parametric distributions by constructing another estimate that is independent of the tails. This alternative estimate can be viewed as a "lower bound" for the true consumer surplus. If I were constant then assuming $v_n^{(1:I)} = v_n^{(2:I)} = b_n^w$ in every auction would provide us with a strict lower bound on the distribution of private values. Since I is stochastic in our auctions it is not a true lower bound, but we can still use this empirical distribution to construct estimates. The consumer's share of surplus using these lower bound estimates are in Table 7.

Table 7 – Lower Bound Estimates for Consumer Share

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Average	0.374	0.369	0.371	0.367	0.373
Median	0.318	0.314	0.315	0.311	0.319

These results are more comparable to Song's estimates. On average there are twice as many bidders in our auctions and the consumer share is 30% lower.

The fundamental point is clear. Consumers are capturing a large amount of surplus in these auctions. Even in the worst case approximately 30% of the surplus is being captured by the consumers.

6 Finding the best distribution.

We use two statistical approaches to test for the best parametric specification of private values. The first relies on information criteria tests ([1]). The second relies on goodness-of-fit tests between the empirical distribution of winning bids and their predicted distribution from the structural model based on differing parametric assumptions. We avoid pretest bias by using a portion of the original data set not used in estimation. We explain below two different methods to obtain these data. If the null hypothesis of a particular distribution for private values is true, then the probabilities have a uniform distribution on the interval from 0 to 1 (Fisher, 1948).

6.1 Tests based on Information Criteria.

The three information criteria tests, Akaike, Schwartz, and Browne-Cudek, favor the same parametric distributions since our sample size is relatively large and differences among the three test-statistics are $o(1)$. Results are in Table 8.

Table 8: Information Criteria Statistics.

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
AIC	3.806	3.729	3.770	3.852	3.856
BIC	3.840	3.765	3.806	3.889	3.892
BCC	3.806	3.720	3.770	3.852	3.856

The favored distribution is the gamma, followed by the Weibull, the half-logistic, log-normal, and Pareto, although differences in the test-statistics and corresponding p-values are not substantial.

6.2 Tests Against the Empirical Distribution.

A goodness-of-fit test between the empirical distribution of winning bids and the predicted distribution based on structural model estimates requires an out-of-sample forecast since we need data that was not used in estimation in order to avoid pre-test bias. Our data collection program sampled the entire population of monitors in eBay's records, but in order to know how much competition an auction faces we needed to know how many other auctions closed during a given auction we were examining. This led us to drop a significant subset of our data, a subset which we can now use. We can also employ a technique that Song [27] implemented in her estimation methodology and utilize the third highest bids to construct the predicted distributions of the winning bids. We thus construct two goodness-of-fit tests, one set using the data we dropped in order to measure the amount of competition an auction faced and another set using data for the third highest bids.

In our sample of dropped auctions we have 3608 data points. Our sample of third highest bids is drawn from our data for estimation so we have 2934 data points. We first drop all auctions where there are not two or three bidders (losing 1505 and 1408 data points respectively). We next drop auctions where there was a secret reservation price because this secret reservation price could be the true second highest or third highest bid and is unobserved, losing an additional 517 and 479 observations.

6.2.1 An Out of Sample Second Highest Values Test.

In a private value auction such as we consider in our paper, the winning bid is also the private value of the second highest bidder if there are at least two bidders who bid above the reservation price. After dropping the observations as discussed above, we are left with 1586 auctions out of the 3608 not used in estimation. In order to use these auctions we have to impute the values of three variables—competing auctions, all; competing auctions, same size; and competing auctions, same cat (category). Imputations for these missing observations are based on their average values over the auctions used in estimation (before normalizing by the auction length). This assumes that the market was not changing much over time and since our data was collected within a four month window this assumption is not unrealistic.

Conditional on I, the cdf of the of the second highest order statistic is:

$$G_n\left(b_n^w,\beta|I\right)=\int_0^{b_n^w}\frac{I!}{(I-2)!}F_n\left(z,\beta\right)^{I-2}\left(1-F_n\left(z,\beta\right)\right)f_n\left(z,\beta\right)dz\ . \quad (7)$$

If we then integrate over I:

$$G_n\left(b_n^w,\beta,\gamma\right)=\frac{\Sigma_{i=I_n^a}^{\overline{I}}\frac{(\lambda_n T_n)^i}{i!}e^{-\lambda_n T_n}G_n\left(b_n^w,\beta|i\right)}{\Sigma_{i=2}^{\overline{I}}\frac{(\lambda_n T_n)^i}{i!}e^{-\lambda_n T_n}} \quad (8)$$

this variable should be uniformly distributed. The resulting distributions based on structural estimates are compared to the uniform in Figure 2.

Figure 2 : Probabilities of Estimated Second Highest Bids

Probabilities based on the various structural estimates appear to have too many high values and all of them, except those based on the for the half-logistic, are approximately the same. The half-logistic is much closer to the uniform than any of the other distributions but none of the probabilities based on the structural estimates appear to be uniform. Test statistics reinforce this conclusion. The tests we use are Kolmogorov-Smirnov, the Kuiper, Cramer-von Mises, Watson, and the Anderson-Darling. Results are in Table 9. Although the half-logistic has the smallest test-statistic all tests are rejected at reasonable significance level.[3]

Table 9: Comparing Probabilities of Second Highest Bids to Uniform

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Kolmogorov	5.679	11.410	11.754	12.850	11.971
Kuiper	6.016	12.535	12.459	13.621	12.460
Cramer-von Mises	13.044	52.943	57.489	66.546	59.646
Watson	2.886	16.174	16.736	19.840	16.876
Anderson-Darling	63.822	243.293	263.040	303.727	273.823

Our model is apparently too parsimonious to be able to fit all of the vagaries of eBay auctions. Relaxing structural assumptions like the distribution of private values will always result in a better fit but often at the cost of less precisely estimated coefficients. Of course for all practical purposes any model will be rejected with enough data and our study uses a relatively large data set. However, one of our goals with this analysis is to find the best parsimonious structural model for bidders' private values and thus we are also interested in which of these distributions best fits the empirical distribution. The structural model based on the (folded) logistically distributed private values best fits the empirical distribution of private values using this second highest values test.

6.2.2 A Third Highest Values Test.

We can also use the third highest bid data to construct the same test. Unfortunately, however, the third highest bid may not be the true value of the third highest bidder. Second highest bids will always be the true private value of the second highest bidder. This is because if the second highest bidder raises their bid they might win the auction. On the other hand if both higher bidders bid before the third highest can update his bid then he might realize that he does not want to raise his bid. To give an example, assume that a bidder's true value is $100, but his first bid is $50. If two bidders then simultaneously bid $150 the price in the auction will rise to $150, and the first bidder will not find it worthwhile to raise his bid. We know that this sort of problem does occur

[3] For a full discussion of the strenghs and weakness of these tests see D'Agostino and Stephens [12].

because bidders do bid multiple times and thus we have evidence they did not initially bid their true value.

Due to the hard stop time on eBay we know that this can not be true for all bidders. If, for example, the third highest bid is submitted with only fifteen seconds left in the auction a rational bidder should bid their true value—since the probability he can update his bid is zero. We can be fairly sure that the bids submitted in the last few minutes are almost certainly the bidders' true values. When we compare the distribution of these bids to earlier periods we find that we can not reject that bids as early as ten hours from the end of the auction are from the same distribution. There are 621 such cases and this constitutes the sample we use to analyze the third highest bid.[4]

Given I the cdf of the third highest bid is:

$$G_n\left(b_n^3, \beta | I\right) = \int_0^{b_n^3} \frac{I!}{(I-3)!2!} F_n\left(z, \beta\right)^{I-3}\left(1 - F_n\left(z, \beta\right)\right)^2 f_n\left(z, \beta\right) dz \qquad (9)$$

where b_n^3 is the third highest bid in auction n. We can then integrate over I:

$$G_n\left(b_n^w, \beta, \gamma\right) = \frac{\Sigma_{i=I_n^a}^{\overline{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n} G_n\left(b_n^w, \beta | i\right)}{\Sigma_{i=3}^{\overline{I}} \frac{(\lambda_n T_n)^i}{i!} e^{-\lambda_n T_n}} \qquad (10)$$

The resulting distribution of probabilities should be uniform. In Figure 3 we plot these probabilities against the uniform distribution.

Figure 3 : Probabilities of Third Highest Bids

Here we have two distributions that are very close to the Uniform, and which we accept depends on which test statistic we look at. The half-logistic dominates the gamma, Weibull, and log-normal, but the Pareto is not dominated. Probabilities based on structural estimates using Pareto distributed private values have changed substantially from those for the second highest bids. Nonparametric goodness-of-fit test results are in Table 10.

[4] We have constructed the same tests as described below for subsets of this data set and arrive at the same conclusions.

Table 10: Comparing Probabilities of Third Highest Bids to Uniform

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Kolmogorov	4.011	11.022	10.723	11.771	3.273
Kuiper	4.064	11.054	10.781	11.849	4.213
Cramer-von Mises	6.818	55.624	52.465	61.338	2.837
Watson	1.323	10.942	10.589	13.507	1.757
Anderson-Darling	33.402	289.968	274.861	320.447	33.841

For every test statistic either the Pareto or the half-logistic is always the lowest. In general the difference between the tests can be understood from their functional form. For example note that the Pareto crosses the uniform distribution while the half-logistic does not. This explains why the Kolmogorov-Smirnov (based on the largest absolute deviation) prefers the Pareto while the Kuiper (which punishes for both positive and negative deviations) prefers the half-logistic. However the clear overall statement is that based on this test both the Pareto and the half-logistic dominate the gamma, Weibull, and log-normal.

6.2.3 Comparing the distribution of Second and Third Highest Bids.

This difference in results between the probabilities of the Second and Third Highest bids for the Pareto raises an interesting issue. If our model is correct then clearly these two distributions should be the same. For this two-sample test we report only Kolmogorov-Smirnov statistic and results in Table 11 indicate the half-logistic dominates the other parametric distributions of private values in terms of this comparison.

Table 11—Comparing the CDF of the Second and Third Highest Bid

	Half-Logistic	Gamma	Weibull	Log-Normal	Pareto
Kolmogorov	0.961	4.982	4.718	6.225	6.960
Probability of Equality	0.314	0.000	0.000	0.000	0.000

7 Semi-nonparametric and Nonparametric Estimators of Consumer Surplus

Athey and Haile (2002, 2005) show that the parent distribution is uniquely determined if the distribution of any order statistic with a known sample size is identified. However, in eBay auctions, the number of potential bidders is generally not observable. Song (2004) addressed this issue by showing that within the symmetric independent private values model, observation of any two valuations of which ranking from the top is known nonparametrically identifies the bidders' underlying value distribution. Based on this theorem, Song argues that we can use the second and third highest bids to identify the distribution of bidders' private values. This approach is not without attendant problems, however, since whether or not the third highest bids reflect the third highest bidders' true private valuations can be questioned. To deal with this issue, Song suggests that we should "use data from auctions where the first or the second highest bidder submitted a cutoff price greater than the third highest valuation late in the auction". With this in mind, she suggested an econometric method to decide "how late" is proper. In this section, we follow Song's technique in estimating the distribution of bidders' valuations.

The following are the assumptions we make throughout this section.

1 No bidder ever bids more than he is willing to pay.

2 No bidder allows opponents to win at a price he is willing to pay.

3 Bidders' values are private, independent, and log-linear in a set of auction specific characteristics.

The assumptions 1 and 2 are the same as the ones used in parametric estimation in Section 3. Assumption 3 is also very standard in the research on auctions (c.f., Song, 2004). Private values are given by:

$$\ln V_t^{(i)} = x_t^{'}\beta + v_t^{(i)},$$

with $t = 1, ..., T$, where T is the number of auctions and $i = 1, 2..., N_t$, where N_t is the number of potential bidders in auction t. For the estimation procedure we outline below, we require the potential number of bidders in any auction to be greater or equal to 3. $V_t^{(2)}$ and $V_t^{(3)}$ represent the second and third highest bidders' valuations in auction t, respectively. We use the second and third highest bids as estimates of these two valuations. $v_t^{(2)}$, and $v_t^{(3)}$ are the corresponding error terms. x_t is the control variable including 7 auction specific characteristics that we specify below, $\beta = [\beta_1, \cdots, \beta_7]$ is the corresponding coefficient. We consider the partial likelihood specified by Song (2004) which is the sample counterpart of $p(v_t^{(2)}|v_t^{(3)})$, since the full likelihood (the joint density of $(v_t^{(2)}, v_t^{(3)})$) requires the unknown number of potential bidders. According to the basic theory of order statistics, the sample likelihood function can be written as:

$$L_T(\hat{f}) = \frac{1}{T}\sum_{t=1}^{T} \ln \frac{2[1-\hat{F}(v_t^{(2)})]\hat{f}(v_t^{(2)})}{[1-\hat{F}(v_t^{(3)})]^2},$$

where

$$\hat{F}(v) = \int_c^v \hat{f}(z)dz.$$

Here and below, c is the lower bound of bidders' private value. We choose $c = \min(v_t^{(2)})$, since no information about $F(v)$ for $v < c$ can be observed. In order to estimate the unknown distribution of v, as in Song (2004), we employ the method proposed by Coppejans and Gallant (2002) and use the hermite series to approximate the unknown distribution:

$$f_T(z) = \frac{[1+\hat{a}_1(\frac{z-u}{\sigma}) + \cdots + \hat{a}_k(\frac{z-u}{\sigma})^k]^2\phi(z;u,\sigma,c)}{\int_c^\infty [1+\hat{a}_1(\frac{z-u}{\sigma}) + \cdots + \hat{a}_k(\frac{z-u}{\sigma})^k]^2\phi(z;u,\sigma,c)dz}.$$

Gallant and Nychka (1987), Fenton and Gallant (1996) and Coppejans and Gallant (2002) provide details of this method to approximate the unknown distribution of private values. The optimal series length varies according to the data set under consideration. We choose the optimal series length, k^* using the cross-validation strategy employing the Integrated Squared Error (ISE) criterion (Coppejans and Gallant, 2002). The ISE criteria is defined as:

$$\begin{aligned} ISE(\hat{f}) &= \int \hat{f}^2(z)dz - 2\int \hat{f}(z)f(z)dz + \int f^2(z)dz \\ &= M_{(1)} - 2M_{(2)} + M_{(3)} \end{aligned} \tag{11}$$

Here, $\hat{f}(z)$ is an estimator of true density $f(z)$.

Since $M_{(3)}$ only depends on the unknown true density we focus on the first two terms. Following the steps suggested in Coppejans and Gallant (2002), we randomly partition the data into $J = 5$ sub data sets with similar size. Let $\hat{f}_{j,k}(\cdot)$ be the semi-nonparametric (SNP) estimate obtained from the data that remains after deletion of the jth group when k is used as the series length. The cumulative distribution associated with $\hat{f}_{j,k}(\cdot)$ is denoted by $\hat{F}_{j,k}(\cdot)$. Based on the formula given out in Song (2004), we calculate $\hat{M}_{(1)}(k)$ and $\hat{M}_{(2)}(k)$ as:

$$\begin{aligned} \hat{M}_{(1)}(k) &= 1/J\sum_{j=1}^{J}\int [\hat{p}_{j,k}(v^{(2)}|v^{(3)})]^2 dv^{(2)}dv^{(3)} \\ &= 1/J\sum_{j=1}^{J}\int \left(\frac{2[1-\hat{F}_{j,k}(v^{(2)})]\hat{f}_{j,k}(v^{(2)})}{[1-\hat{F}_{j,k}(v^{(3)})]^2}\right)^2 dv^{(2)}dv^{(3)} \end{aligned} \tag{12}$$

and,

$$
\begin{aligned}
\hat{M}_{(2)}(k) &= 1/T\sum_{j=1}^{J}\sum_{(v_t^{(3)},v_t^{(2)})\in\chi_j}\hat{p}_{j,k}(v_t^{(2)}|v_t^{(3)}) \\
&= 1/T\sum_{j=1}^{J}\sum_{(v_t^{(3)},v_t^{(2)})\in\chi_j}\frac{2[1-\hat{F}_{j,k}(v_t^{(2)})]\hat{f}_{j,k}(v_t^{(2)})}{[1-\hat{F}_{j,k}(v_t^{(3)})]^2}, \qquad (13)
\end{aligned}
$$

where $v^{(2)}$ and $v^{(3)}$ are random variables, satisfying the condition that $v^{(2)} > v^{(3)}$. Following the notation used in Song (2004), we let $CVH(k) = \hat{M}_{(1)}(k) - 2\hat{M}_{(2)}(k)$. According to Coppejans and Gallant (2002), a typical graph of $CVH(k)$ versus k is that $CVH(k)$ falls as k increases when k is small, periodically drops abruptly, and flattens right after the final abrupt drop. They recommend a choice of k based on which $CVH(k)$ has the last abrupt drop. Our results, however, indicate that $CVH(k)$ drops substantially when k changes from 0 to 1, increases from 1 to 2, and drops again gradually after 2. These results are in Appendix B.2. One possible reason for this is that the model with $k = 2$ oversmoothes, but the effect is partially offset when we include higher order polynomial into the distribution function. Based on our results, we choose $k = 1$ as the optimal series length. The density function of v_t follows immediately as:

$$
f_T(v_t) = \frac{[1+a(\frac{v_t-u}{\sigma})]^2\phi(v_t;u,\sigma,c)}{\int_c^{\infty}[1+a(\frac{v_t-u}{\sigma})]^2\phi(v_t;u,\sigma,c)dv_t} \qquad (14)
$$

The nonparametric maximum likelihood estimator is then defined as:

$$
(\beta_1,\cdots,\beta_9,\hat{a},\hat{u},\hat{\sigma}) = \text{argmax}_{\beta_1,\cdots,\beta_9,a,u\in R,\sigma>0}L_T(\hat{f}) = \frac{1}{T}\sum_{t=1}^{T}\ln\frac{2[1-\hat{F}(v_t^{(2)})]\hat{f}(v_t^{(2)})}{[1-\hat{F}(v_t^{(3)})]^2}.
$$

A major criticism of this method is that the third highest bids usually do not reflect the bidders' private values. Remember that we use the second and third highest bids as estimates of the second the third highest bidders' private values. Song (2004) argues that "by looking at auctions where the first- or second-highest bidder submitted a cutoff price grater than the third-highest bid late in the auction, we can increase the probability of obtaining the actual third-highest valuation". To determine 'how late' is sufficient, Song (2004) provides a method which also employs the ISE criterion. Following her procedure, we consider a sequence of 5 sub data sets, $A_{w1},\cdots,A_{w5}$, with different window sizes. Song

(2004) does not provide out a general rule on how to choose the window sizes for the sub data sets, although in her paper she chooses window sizes which make the size differences between successive sets similar. We will follow the same rule. We choose window sizes as $w1 = 1$ minute, $w2 = 5$ minutes, $w3 = 40$ minutes, $w4 = 3.5$ hours and $w5 =$all. A_{w1} represents the auction set in which the first or second highest bidder submits a bid greater than the third highest bid no earlier than 1 minute before the auction ends. Other sub data sets are defined in the similar way. Obviously, we have $A_{w1} \subset A_{w2} \subset \cdots, \subset A_{w5}$. It is intuitive that the third highest bids are more likely to reflect the third highest valuations for auctions in set A_{w1} than auctions in other sub sets. However, A_{w1} has the least number of observations and thus a potentially larger sample variance. Song's approach considers this trade-off by applying the same cross-validation strategy that is used for choosing the optimal series length. For each auction set A_{wi}, she computes $CVH_{wi}(k^*)$ using data from auctions in A_{wi} in the same way it is computed in choosing the optimal series length, except that she uses $\hat{M}_{(2)}$ defined in equation (15) instead of the one in equation (13).

$$\begin{aligned}
\hat{M}_{(2)}(k^*) &= 1/T_1 \sum_{j=1}^{J} \sum_{(v_t^{(3)}, v_t^{(2)}) \in \chi_j \cap A_{w1}} \hat{p}_{j,k^*}(v_t^{(2)} | v_t^{(3)}) \\
&= 1/T_1 \sum_{j=1}^{J} \sum_{(v_t^{(3)}, v_t^{(2)}) \in \chi_j \cap A_{w1}} \frac{2[1 - \hat{F}_{j,k^*}(v_t^{(2)})] \hat{f}_{j,k^*}(v_t^{(2)})}{[1 - \hat{F}_{j,k^*}(v_t^{(3)})]^2}, \quad (15)
\end{aligned}$$

where T_1 is the sample size of A_{w1} and $J = 5$ as before. Note that $\hat{p}_{j,k*}(\cdot|\cdot)$ is evaluated only at sample points in A_{w1} implying that "CVH_{wi} measures how well the estimate obtained by using data from an auction set A_{wi} fits the data from auctions in A_{w1}" (Song (2004)). We calculate CVH_{wi} based on her method, and present the result in Appendix B.3. CVH_{wi} decreases from window size 1 to window size 4, and then increases. Therefore, we choose $w4 = 3.5$ hours as our optimal window size.

7.1 Data

In selecting the sample for the semi-nonparametric and nonparametric analysis we needed to alter the criteria used for selecting the sample used in our parametric modeling above. This is in part because we need the second and third highest bids which are not available for some observations in the data which was used in parametric estimation since the parametric method does not need that information. Moreover, selecting a relatively homogeneous data set is important in conducting and interpreting results from nonparametric analysis. Since we needed the information for both the second and third highest bids in order to estimate the models, we dropped the auctions that had less than 3 bidders. In order to get a relatively homogeneous sample to make nonparametric estimation both feasible and meaningful, we picked the auctions for the 17 inch colored PC monitor only, which gave us 476 observations, subvs. Monitor size has the most pronounced and significant effect on bidders' private values. To make the data set even more more homogeneous, we also dropped 12 auctions in which warranties were offered on the auctioned monitors. Our final data set used for estimating consumer surplus using nonparametric and semi-nonparametric methods has 464 observations.

In estimating the distribution of bidders' private values with the semi-nonparametric approach, we used the following 9 control variables: monitor dot pitch (sample average of dot pitch is used when no dot pitch is reported), dummy for the cases when no dot pitch is reported, monitor resolution (sample average of resolution is used when no resolution is reported), dummy for the cases when no resolution is reported, condition of auctioned items (2 for new, 1 for like new or refurbished, 0 for no condition report), dummy for flat screen and dummy for "Brand name". We use 1 for both "like new" and "refurbished" because we did not see significant sample mean difference for these two categories and there are only 17 observations with condition specified as "like new" in our reduced sample. Descriptive statistics of the variables for the second sample are presented in Table 4 in Appendix A.

7.2 Estimation Results and Consumer Surplus

For the results that follow we choose the optimal hermite series number as $k^* = 1$ and the optimal window size as $w4 = 3.5$ hours, i.e., we choose the auctions where the first or second highest bidder submits a bid greater than the third highest bid no earlier than 3.5 hours before the auction ends. This yields a sample of 376 observations on which to base our semi-nonparametric estimates of consumer surplus.

The estimation results of the semi-nonparametric approach is presented in Table B.3 in Appendix B and the statistics of bidders' private values is calculated accordingly. Because the data we use for the SNP analysis is relatively homogeneous, we also present nonparametric results as comparison. In the nonparametric estimation, we use Song's method without considering the control variables. The estimated expectation and standard deviation of bidders' private valuation are in Table 12. SNP and NP denote semi-parametric and nonparametric methods respectively.

Table 12

Statistics of Estimated Distribution		
	Mean	Std
SNP	\$31.39	\$3.48
NP	\$25.47	\$26.63

The mean and standard deviation are computed with the median values of $x_1, x_2, \cdots$, and, x_7

In order to investigate the welfare impact of eBay, we also calculate the consumer surplus and consumer share of surplus as well. A consumer's surplus at auction t is calculated as:

$$CS_t = V_t^{(1)} - p_t,$$

where p_t denotes the price the winner paid, which equals to the second highest bid in eBay auctions. $V_t^{(1)}$ denotes the valuation of the winner. Since we do not observe $V_t^{(1)}$, we estimate the expected consumer surplus as:

$$E(CS_t|V_t^{(2)}) = \int_{\hat{v}_t^{(2)}}^{\infty} \frac{f(v)}{1 - F(\hat{v}_t^{(2)})} \cdot v\, dv + x_m' \hat{\beta} - p_t$$

Again, $\hat{v}_t^{(2)}$ is the estimator of $v_t^{(2)}$ calculated with estimated coefficient $\hat{\beta}$ and x_m, which is the vector of median values of control variables. We include the results from the preferred parametric likelihood (PL) model above, which assumes that private values are distributed as half-logistic. The descriptive statistics of expected consumer surplus from our SNP, NP methods, and PL approaches are presented in the following Table 13.

Table 13

Consumer Surplus					
	Mean	Median	Std	Min	Max
SNP	\$34.91	\$34.44	\$7.74	\$14.96	\$72.38
NP	\$33.39	\$32.88	\$6.07	\$22.96	\$68.09
PL	\$52.54	\$39.69	\$38.23	\$13.51	\$210.09

The alternative measure of consumer surplus is the consumers' share of surplus, which is defined as:

$$CSS_t = \frac{CS_t}{V_t^{(1)} - V_t^{(a)}}.$$

where $V_t^{(1)}$ is the winning bidder's private valuation in auction t, which we do not observe. However, we can estimate its expected value with

$$\hat{V}_t^{(1)} = p_t + \hat{CS}_t$$

$V_t^{(a)}$ is the auctioneer's valuation in auction t which we do not observe either. However, as above, we assume $V_t^{(a)} = 0$, which gives an underestimated result for CSS_t that provides us the lower bound of the consumers' share of surplus. The descriptive statistics of expected lower bound of the consumers' share of surplus is presented as follows:

Table 14

Lower Bound of Consumer Share of Surplus				
	Mean	Median	Min	Max
SNP	22.54%	22.30%	11.09%	37.62%
NP	21.77%	21.51%	16.06%	36.20%
PL	33.90%	30.30%	8.70%	99.90%

Again, PL represents the parametric method with the assumption of half-logistic distributed private valuations. The results from SNP and NP are comparable, however, obviously lower than those from PL except for the minimum.

If we simply substitute CS_t and p_t with their median values, we can have results as shown in Table 15. The median winning price p_m is \$120 for the data used in semi-nonparametric and nonparametric methods and 100 for the data used in parametric estimation.

Table 15

Expected Consumer Share of Surplus			
	SNP	NP	PL
ECSS	22.30%	21.59%	28.41%

We can see that the results are very close, although the expected consumer share of surplus from SNP and NP are smaller than that from PL. In Song (2004), the consumer share of surplus for yearbook auctions is 53% if calculated in the same way. This result is higher than the values in Table 13. The difference can be explained with competition levels involved. The average number of bidders is 3.6 in Song (2004), 6.8 for the data used in the PL estimates and 8.1 for the data used in estimating the SNP and NP estimates. More competition on the bidders' side would appear to result in lower consumers' share of surplus.

8 Conclusion

In this paper, we estimate consumer surplus for eBay computer monitor auctions with parametric, semi-nonparametric and parametric methods. In the parametric estimation, 5 distributions proposed in the auction literature for constructing bidders' private value are considered separately. More specifically, Half-Logistic, Gamma, Weibull, Log-Normal and Pareto distributions are considered, estimated and compared with *Out of Sample Second Highest Value Test* and *Third Highest Value Test*. The half-logistic distribution is shown to be the most favorable one among the class of distributions. As comparison, we also provide nonparametric and semi-nonparametric estimation in line with the approach introduced in Song (2004). Consumer surplus and share of consumer surplus are calculated and compared among the nonparametric, semi-nonparametric and the favorable parametric models. Comparable results from these three models indicate that the nonparametrically chosen distributional assumption is reasonable and that there is substantial agreement on the magnitude of the consumer surplus generated in this particular eBay auction.

References

[1] Adams, C. P., 2006. Estimating demand from eBay prices, mimeo.

[2] Adams, R. M., Berger, A., and Sickles, R. C., 1999. Semiparametric approaches to stochastic panel frontiers with applications to the banking industry, *Journal of Business and Economic Statistics* 17, 349-358.

[3] Adams, R. M., and Sickles, R. C, 2007. Semi-parametric efficient distribution free estimation of panel models, *Communication in Statistics: Theory and Methods* 36, 2425–2442.

[4] Athey, S., and Haile, P. A., 2002, Identification of standard auction models, *Econometrica* 65, 1097-1128.

[5] Athey, S., and Haile, P.A., 2005. Nonparametric approaches to auctions, in *Handbook of Econometrics,*

Volume 6, edited by J.J. Heckman and E. Leamer. Elsevier: Amsterdam.

References

[1] Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, In Petrove B.N. and Csaki F. (Eds), *Second International Symposium on Information Theory*, Budapest, Akademiai Kiado, 67-281.

[2] Bajari, P. and Hortaçsu, A., 2003. Winner's curse, reserve prices and endogenous entry: empirical insights from eBay auctions, *RAND Journal of Economics* 34, 329-355.

[3] Bapna, R., Goes P., Gupta, A., 2003. Analysis and design of business-to-consumer online auctions, *Management Science* 49, 85-101.

[4] Bapna, R., Goes P., Gupta, A., 2003. Replicating online yankee auctions to analyze auctioneers' and bidders' strategies, *Information Systems Research* 14, 244-268.

[5] Bapna, R., Goes, P., Gupta, A., and Jin Y., 2004. User heterogeneity and its impact on electronic auction market design: an empirical exploration, *MIS Quarterly* 28, 21-43.

[6] Bapna, R., Jank, W., Shmueli, G., 2008. Consumer surplus in online auctions, forthcoming in *Information Systems Research*.

[7] Bierens, H., and Song, H., 2006a. Nonparametric identification of first-price auction models with unbounded values and observed auction-specific heterogeneity, mimeo, Pennsylvania State University.

[8] Bierens, H., and Song, H., 2006b. Semi-nonparametric estimation of first-price auctions with auction-specific heterogeneity using simulated method of moments, mimeo, Pennsylvania State University.

[9] Cabral, L., and Hortaçsu, A., 2004. Dynamics of seller reputation: theory and evidence from eBay, NBER Working Paper 10363.

[10] Carare, O., 2001. Need for speed: demand estimation ssing auction data, Rutgers University, Economics Working Paper, available at: http://econweb.rutgers.edu/carare/ns.pdf

[11] Coppejans, M., Gallant, A.R., 2002. Cross-validated snp density estimates, *Journal of Econometrics* 110, 27-65.

[12] D'Agostino, R.B. and Stephens, M.A., (Eds.). 1986, *Goodness-of-Fit Techniques*, Marcel Dekker, Inc.

[13] Donald, S., and Paarsch, H., 1993, Piecewise maximum likelihood estimation in empirical models of auctions, *International Economic Review* 34, 121-148.

[14] Fisher, R. A., 1948. Combining independent tests of significance, *American Statistician* 2, 32-33.

[15] Gallant, A.R., Nychka, D.W., 1987. Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 363-390.

[16] Gonzalez, R., 2002, *Essays in Structural Parametric and Semi-Parametric Estimation of Auction and Agency Models*, unpublished dissertation, Rice University.

[17] Gonzalez, R, Hasker, K., and Sickles, R.C., 2007. An analysis of strategic behavior in eBay auctions, mimeo, Rice University.

[18] Haile, P.A., and Tamer, E.T., 2003. Inference with an incomplete model of English auctions. *Journal of Political Economy* 111, 1-51.

[19] Jin, G.Z., and Kato, A., 2005. Price, quality and reputation: evidence from an online field experiment, forthcoming in *Rand Journal of Economics*.

[20] Laffont, J.J., Ossard, H., and Vuong, Q., 1995. Econometrics of first-price auctions, *Econometrica* 63, 953-980.

[21] Livingston, J.A., 2005. How valuable is a good reputation? a sample selection model of internet auctions, *Review of Economics and Statistics* 87, 453 - 465.

[22] Park, B.U. and Simar, L., 1994. Efficient semiparametric estimation in stochastic frontier models, *Journal of the American Statistical Association* 89, 929–936.

[23] Park, B.U., Sickles, R.C., and Simar, L., 1998. Stochastic frontiers: a semiparametric approach, *Journal of Econometrics*, 84, 273–301.

[24] Park, B.U., Sickles, R.C. and Simar, L., 2003. Semiparametric efficient estimation of ar(1) panel data models, *Journal of Econometrics* 117, 279-309.

[25] Park, B.U., Sickles, R.C. and Simar, L., 2007. Semiparametric Efficient Estimation of Dynamic Panel Data Models, *Journal of Econometrics* 136, 281-301.

[26] Silverman, B.W. 1986. *Density estimation for statistics and data analysis,* 1-140.

[27] Song, U., 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

Appendix A: Tables and Descriptive Statistics

Table 1: Estimates of the Exogenous Value.

.	Logistic	Gamma	Weibull	Log-Normal	Pareto
Constant	-11.3159***	-11.1785***	-13.7639***	-14.3522***	-12.1096***
.	(0.000)	(0.000)	(0.000)	(0.001)	(0.006)
Log, Size	4.6806***	4.8012	4.7326***	4.806***	4.6192***
.	(0.004)	(0.894)	(0.000)	(0.000)	(0.000)
Log, Dot Pitch	-0.5418	-0.7289	-0.5844	-1.1047	-0.2101
.	(0.793)	(0.500)	(0.425)	(0.563)	(0.912)
Dummy, No Dot Pitch	0.6132	0.8606	0.6666	1.366*	0.1142
.	(0.487)	(0.521)	(0.815)	(0.055)	(0.872)
Log, Resolution	0.1409	0.1695	0.2811	0.1448	0.307
.	(0.911)	(0.992)	(0.168)	(0.717)	(0.441)
Dummy, No Resolution	1.0803	1.309	2.0814**	1.0922	2.2261***
.	(0.415)	(0.786)	(0.025)	(0.107)	(0.001)
Dummy, New	0.2539	0.3345	0.3242	0.3939	0.3203
.	(0.864)	(0.907)	(0.945)	(0.149)	(0.241)
Dummy, Like-new	0.0877	0.2247	0.2351	0.3396	0.2425
.	(0.960)	(0.923)	(0.802)	(0.188)	(0.347)
Dummy, Refurbished	0.0594	0.0619	0.0482	0.04	0.0204
.	(0.998)	(0.987)	(0.998)	(0.965)	(0.982)
Dummy, Warranty	0.0779	0.0976	0.1033	0.189	0.1154
.	(0.935)	(0.952)	(0.959)	(0.798)	(0.876)
Dummy, Brand Name	0.016	0.0061	0.0055	0.009	-0.001
.	(0.861)	(0.996)	(0.997)	(0.990)	(0.999)
Dummy, Flat Screen	0.1979	0.246	0.2246	0.2213	0.2047
.	(0.733)	(0.618)	(0.645)	(0.592)	(0.620)
Distribution Variable^{+}	NA	-1.695	-0.6841	1.6563***	0.7099**
.	.	(0.762)	(0.527)	(0.000)	(0.022)
Number of Auctions	2934	2934	2934	2934	2934
-Log Likelihood/Number of Auctions	3.7939	3.7175	3.7583	3.8402	3.8441

$^{+}$ For the Log-Normal this is the Standard Deviation. For the Weibull, Gamma, and Pareto this is the log of the shape parameter.

P-values are reported below the coefficients in Parentheses.

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

Table 2: Estimates of the Entry Parameter.

.	Logistic	Gamma	Weibull	Log-Normal	Pareto
Constant	-15.5587***	-13.033***	-14.5406***	-5.1905***	-16.1312***
.	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log, Size	3.3723***	3.2725***	3.4683	4.2382***	3.8408***
.	(0.000)	(0.000)	(0.463)	(0.000)	(0.000)
Log, Dot Pitch	-5.4067	-7.2483*	-8.4491***	-3.4813	-7.9655
.	(0.637)	(0.080)	(0.006)	(0.992)	(0.981)
Dummy, No Dot Pitch	6.5705***	8.5044***	10.0708***	-3.3655***	9.6205***
.	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log, Resolution	-0.8514	-1.4036*	-1.5012***	-1.2527	-1.3859
.	(0.968)	(0.097)	(0.000)	(0.316)	(0.267)
Dummy, No Resolution	-6.0989***	-10.0171***	-10.7186	-8.966	-9.8957
.	(0.000)	(0.000)	(0.130)	(0.554)	(0.514)
Dummy, New	7.3845	7.4000	7.5104	5.5506	7.5072
.	(0.983)	(0.999)	(0.998)	(0.990)	(0.987)
Dummy, Like-new	6.5463	6.5000	6.0767	3.0008	6.0748
.	(0.979)	(1.000)	(0.996)	(0.916)	(0.831)
Dummy, Refurbished	-0.0011	-0.0205	0.0039	0.0161	0.0366
.	(1.000)	(0.991)	(0.999)	(0.986)	(0.968)
Dummy, Warranty	0.4447	0.8402	0.971	0.7463	1.0831
.	(0.573)	(0.964)	(0.741)	(0.488)	(0.315)
Dummy, Brand Name	-0.0075	0.0128	0.0092	-0.006	0.0147
.	(0.998)	(0.999)	(0.989)	(0.966)	(0.917)
Dummy, Flat Screen	-0.0549	-0.2003	-0.183	-0.1187	-0.2295
.	(0.967)	(0.964)	(0.797)	(0.974)	(0.951)
Log, Seller's Feedback +1	0.715	1.5201***	1.4805***	2.2702***	1.7373***
.	(0.992)	(0.002)	(0.000)	(0.000)	(0.004)
Log, (Seller's Feedback $+1)^2+1$	-0.3473	-0.7443***	-0.7198***	-1.1072**	-0.8432*
.	(0.955)	(0.006)	(0.002)	(0.021)	(0.079)
Category Dummy, $\leq 17''$ Screen	0.8153*	1.1007***	0.9541	0.8032	0.7102
.	(0.094)	(0.001)	(0.114)	(0.205)	(0.262)
Category Dummy, $\geq 19''$ Screen	0.1085	0.2498	0.1388	-0.2512	-0.0794
.	(0.951)	(0.717)	(0.897)	(0.931)	(0.978)
Category Dummy, Monochrome	-1.5328***	-1.6876	-1.411	-0.7847	-1.1656*
.	(0.002)	(0.283)	(0.989)	(0.227)	(0.073)
Competing Auctions, all	1.2191***	1.1813**	1.1274**	1.0966**	1.0732**
.	(0.007)	(0.014)	(0.014)	(0.021)	(0.024)
Competing Auctions, Same Size	0.2364	0.241	0.2373	0.1552	0.2319
.	(0.738)	(0.629)	(0.914)	(0.797)	(0.701)
Competing Auctions, Same Cat.	-0.2957	-0.3409	-0.2529	-0.0674	-0.1192
.	(0.886)	(0.341)	(0.317)	(0.877)	(0.785)
Open Reservation Price	-0.0754	-0.0253	-0.0199	0.005	-0.0206
.	(0.997)	(0.983)	(0.980)	(0.994)	(0.976)
Dummy, Secret Reserve	0.5861	2.2273	1.1178	1.3458	1.0856
.	(0.916)	(0.952)	(0.152)	(0.916)	(0.932)
Number of Auctions	2934	2934	2934	2934	2934
-Log Likelihood/Number of Auctions	3.7939	3.7175	3.7583	3.8402	3.8441

P-values are reported below the coefficients in Parentheses.

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

Table 3: Descriptive Statistics of Key Variables-First Sample

	Average	Median	Standard deviation	Skewness	Maximum	Minimum
Sales Price	135.7	100	132.85	2.01	1430	0.01
Open Reservation Price	75.65	35	106.37	2.56	1100	0.01
Number of Bidders	3.92	3	4.06	1.06	22	0
The Length of the Auction	5.08	5	2.16	0.65	10	3
Size	16.92	17	2.39	0.42	21	14
Dot Pitch^{+}	0.26	0.26	0.02	-0.7	0.31	0.2
Dummy, Dot Pitch Not Reported	0.64	1	0.48	-0.56	1	0
Resolution^{+}	1116.24	1024	271.05	0.32	1600	640
Dummy, Resolution Not Reported	0.38	0	0.49	0.49	1	0
Dummy, New Monitor	0.07	0	0.26	3.31	1	0
Dummy, Like-New Monitor	0.03	0	0.17	5.38	1	0
Dummy, Refurbished Monitor	0.13	0	0.33	2.23	1	0
Dummy, Warranty on Monitor	0.03	0	0.17	5.51	1	0
Dummy, Brand Name Monitor	0.59	1	0.49	-0.36	1	0
Dummy, Flat Screen Monitor	0.18	0	0.38	1.71	1	0
Seller's Feedback	207.99	108	382.49	4.35	4343	1
Dummy, Item sold in <=17" size Screen Category	0.62	1	0.49	-0.5	1	0
Dummy, Item sold in >=19" size Screen Category	0.27	0	0.45	1.02	1	0
Dummy, Item sold in Monochrome Category	0.00	0	0.05	22.06	1	0
Competing Auctions, all	924.33	906	116.55	0.9	1322	705
Competing Auctions, Same Size	182.01	191	68.77	-0.06	357	3
Competing Auctions, Same Category	435.06	496	189.87	-0.35	903	3
Dummy, if there was a Secret Reservation Price	0.18	0	0.38	1.68	1	0

$^{+}$Statistics for these variables are only for items where a value was reported

Table 4: Descriptive Statistics of Key Variables-Second Sample

	Average	Median	Standard deviation	Skewness	Maximum	Minimum
Sales Price	124.27	120	2.12	0.82	355	10.5
Number of Bidders	7.82	7	0.15	0.57	20	3
The Length of the Auction	5.30	5	0.10	0.40	10	3
Size	17	17	1	2.73	17	17
Dot Pitch$^+$	0.57	1	1.03	0.72	1	0.2
Dummy, Dot Pitch Not Reported	0.58	1	0.02	-0.32	1	0
Resolution$^+$	86.74	800	1.17	0.57	1600	1
Dummy, Resolution Not Reported	0.36	0	0.02	0.58	1	0
Dummy, New Monitor	0.08	0	0.010	3.23	1	0
Dummy, Like-New Monitor	0.04	0	0.01	4.79	1	0
Dummy, Refurbished Monitor	0.13	0	0.02	2.19	1	0
Dummy, Warranty on Monitor	0	0	0	0	0	0
Dummy, Brand Name Monitor	0.59	1	0.02	-0.38	1	0
Dummy, Flat Screen Monitor	0.28	0	0.02	1.01	1	0
Seller's Feedback	42.87	57	1.09	0.76	4344	1
Dummy, Item sold in <=17" size Screen Category	0.96	1	0.01	-0.51	1	0
Dummy, Item sold in >=19" size Screen Category	0	0	0	0	0	0
Dummy, Item sold in Monochrome Category	0	0	0	0	0	0
Competing Auctions, all	929.67	907	5.13	0.63	1322	707
Competing Auctions, Same Size	246.08	248	1.54	0.41	357	172
Competing Auctions, Same Category	558.14	558	5.71	-1.81	903	76
Dummy, if there was a Secret Reservation Price	0.33	0	0.02	0.71	1	0

$^+$Statistics for these variables are only for items where a value was reported

Appendix B: Tables of Semi-nonparametric Estimation

B.1

Relations Between CVH and k					
	k=0	k=1	k=2	k=3	k=4
CVH	36.35	26.78	57.52	52.63	29.35

B.2

Relations Between CVH and Window Size, k*=1					
	w1=1 min	w2=5 min	w3=40 min	w4=3.5 hour	w5=all
CVH	91.43	22.39	17.75	17.30	23.89

B.3

Table of Estimates from SNP Estimation			
Constant*	-0.343(0.0849)	Status*	0.3222(0.1564)
Log, Dot Pitch*	-1.6813(0.0624)	Dummy, Brand Name	0.0213(0.1155)
Dummy, No Dot Pitch*	-0.2271(0.1146)	Dummy, Flat Screen	0.0458(0.2075)
Log, Resolution*	0.0216(0.0115)	Dummy, No Resolution*	-0.3045 (0.1696)

Note: * coefficient is significant at 2.5% level.

Appendix D: Derivations of Consumer Surplus

We provide derivations of the formulas for consumer surplus using our five competing parametric distributions. We use the notation $e^{x'_n\beta} = |b|$ for notational clarity.

Half-Logistic:

$$
\begin{aligned}
f(z,b) &= 2\left(1+e^{-\frac{z}{|b|}}\right)^{-1}\left(1-\left(1+e^{-\frac{z}{|b|}}\right)^{-1}\right)\frac{1}{|b|} \\
F(y) &= \frac{e^{\frac{y}{|b|}}-1}{e^{\frac{y}{|b|}}+1} \\
E(x|x\geq y) &= \frac{1}{1-\frac{e^{\frac{y}{|b|}}-1}{e^{\frac{y}{|b|}}+1}}\int_y^\infty 2z\left(1+e^{-\frac{z}{|b|}}\right)^{-1}\left(1-\left(1+e^{-\frac{z}{|b|}}\right)^{-1}\right)\frac{1}{|b|}dz \\
&= \left(e^{\frac{y}{|b|}}+1\right)|b|\ln\left(1+e^{-\frac{y}{|b|}}\right)+y \\
CS &= \left(e^{\frac{y}{|b|}}+1\right)|b|\ln\left(1+e^{-\frac{y}{|b|}}\right). \qquad (16)
\end{aligned}
$$

Gamma:

Let $F(\cdot,\alpha)$ be the gamma distribution function with parameter α.

$$
\begin{aligned}
f_n(z,\alpha) &= \frac{1}{\Gamma(\alpha)}\frac{z^{\alpha-1}}{(|b|)^\alpha}e^{-\frac{z}{|b|}} \\
F(y,\alpha) &= \int_0^y \frac{1}{\Gamma(\alpha)}\frac{z^{\alpha-1}}{(|b|)^\alpha}e^{-\frac{z}{|b|}}dz \\
E(x|x\geq y) &= \frac{1}{1-F(y,\alpha)}\int_y^\infty \frac{1}{\Gamma(\alpha)}z\frac{z^{\alpha-1}}{(|b|)^\alpha}e^{-\frac{z}{|b|}}dz \\
&= \frac{1}{1-F(y,\alpha)}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}|b|\int_y^\infty \frac{1}{\Gamma(\alpha+1)}\frac{z^{\alpha}}{(|b|)^{\alpha+1}}e^{-\frac{z}{|b|}}dz \\
&= \frac{1-F(y,\alpha+1)}{1-F(y,\alpha)}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}|b| \\
CS &= \frac{1-F(y,\alpha+1)}{1-F(y,\alpha)}\alpha|b|-y.
\end{aligned}
$$

Weibull:

$$
\begin{aligned}
f_n(z,\alpha) &= \alpha\frac{z^{\alpha-1}}{(|b|)^\alpha}e^{-\left(\frac{z}{|b|}\right)^\alpha} \\
F(y,\alpha) &= 1-e^{-y^\alpha|b|^{-\alpha}} \\
E(x|x\geq y) &= \frac{1}{1-F(y,\alpha)}\int_y^\infty z\alpha\frac{z^{\alpha-1}}{(|b|)^\alpha}e^{-\left(\frac{z}{|b|}\right)^\alpha}dz \\
&= \frac{1}{e^{-y^\alpha|b|^{-\alpha}}}\int_y^\infty z\alpha\frac{z^{\alpha-1}}{(|b|)^\alpha}e^{-\left(\frac{z}{|b|}\right)^\alpha}dz. \qquad (17)
\end{aligned}
$$

Note that with a change of variables, we can reform the integration as:

$$\int_y^\infty z\alpha \frac{z^{\alpha-1}}{(|b|)^\alpha} e^{-\left(\frac{z}{|b|}\right)^\alpha} dz = \int_{\left(\frac{y}{|b|}\right)^\alpha}^\infty |b| x^{\frac{1}{\alpha}} e^{-x} dx$$

with $x = \frac{z}{|b|}^\alpha$. The result is derived according to fundamental calculus.

Moreover, the pdf for gamma distribution is:

$$gamma\,(x, \rho) = \frac{1}{\Gamma(\rho)} x^{\rho-1} e^{-x}.$$

The conditional expectation then can be expressed as

$$\begin{aligned} E\,(x|x \geq y) &= \frac{|b|\Gamma\left(\frac{\alpha+1}{\alpha}\right)}{e^{-\left(\frac{y}{|b|}\right)^\alpha}} \int_{\left(\frac{y}{|b|}\right)^\alpha}^\infty \frac{1}{\Gamma\left(\frac{\alpha+1}{\alpha}\right)} z^{\frac{1}{\alpha}} e^{-z} dz \\ &= \frac{|b|\Gamma\left(\frac{\alpha+1}{\alpha}\right)}{e^{-\left(\frac{y}{|b|}\right)^\alpha}} \left(1 - F_{gamma}\left(\left(\frac{y}{|b|}\right)^\alpha, \frac{\alpha+1}{\alpha}, 1\right)\right) \end{aligned}$$

Therefore

$$CS = \frac{|b|\Gamma\left(\frac{\alpha+1}{\alpha}\right)}{e^{-\left(\frac{y}{|b|}\right)^\alpha}} \left(1 - F_{gamma}\left(\left(\frac{y}{|b|}\right)^\alpha, \frac{\alpha+1}{\alpha}, 1\right)\right) - y. \tag{18}$$

Log-normal:

$$\begin{aligned} f_n\,(z, \beta) &= \frac{1}{\sigma z \sqrt{2\pi}} e^{-\left(\ln z - x_n' \beta\right)^2 / 2\sigma^2} \\ F\,(y) &= \int_0^y \frac{1}{\sigma z \sqrt{2\pi}} e^{-\left(\ln z - x_n' \beta\right)^2 / 2\sigma^2} \\ E\,(x|x \geq y) &= \frac{1}{1 - F_{norm}\,(\ln y)} \int_y^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\ln z - |b|)^2}{2\sigma^2}} dz \\ &= \frac{\frac{1}{2} e^{|b| + \frac{1}{2}\sigma^2}}{1 - F_{norm}\,(\ln y)} \left(1 + erf\left(\frac{1}{2}\frac{\sqrt{2}}{\sigma}\left(-\ln y + |b| + \sigma^2\right)\right)\right) \end{aligned}$$

where erf is the error function and $|b|$ represents $x_n'\beta$ instead of $e^{x_n'\beta}$ like in other distributions. Therefore,

$$CS = \frac{\frac{1}{2} e^{|b| + \frac{1}{2}\sigma^2}}{1 - F_{norm}\,(\ln y)} \left(1 + erf\left(\frac{1}{\sqrt{2}\sigma}\left(-\ln y + |b| + \sigma^2\right)\right)\right) - y \tag{19}$$

Pareto:

$$\begin{aligned} f_n\,(z, \alpha) &= \frac{\alpha}{|b|}\left(1 + \frac{z}{|b|}\right)^{-(\alpha+1)} \\ F\,(y, \alpha) &= -|b|^\alpha \left(\frac{1}{|b| + y}\right)^\alpha + 1 \\ E\,(x|x \geq y) &= \frac{1}{1 - F\,(y, \alpha)} \int_y^\infty \frac{z}{|b|}\alpha \left(1 + \frac{z}{|b|}\right)^{-(\alpha+1)} dz \\ &= \frac{1}{\alpha - 1}\left(\alpha y + |b|\right) \end{aligned}$$

Therefore, the consumer surplus is

$$CS = \frac{1}{\alpha - 1}\left(\alpha y + |b|\right) - y \tag{20}$$

Note that for this to be true it is necessary that $\alpha > 1$.