Chapter 4 of *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*. RC Sickles and WC Horrace (eds.) Springer Science & Business Media, New York, NY, forthcoming, 2014.

Estimating Consumer Surplus in eBay Computer Monitor Auctions.*

Kevin Hasker
Economics Department
Bilkent University
Ankara, Turkey

Bibo Jiang
School of Economics
Fudan University
Shanghai, China

Robin Sickles
Economics department
Rice University
Houston, Texas

September 28, 2013

Abstract

Our paper utilizes semi-nonparametric and nonparametric methods to directly estimate consumer surplus in eBay computer monitor sales. We compare these with parametric estimates. We also develop a new measure of how competitive an auction is—the consumer share of surplus—and a new method to calculate these statistics. This reduced form methodology requires the assumption that the pool of potential bidders is constant across auctions, but it is not sensitive to the upper tail and does not require estimating the underlying distribution of values.

JEL classification: C22, C51

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

1

1 Introduction and brief discussion of the consumer surplus auction literature

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 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. It is still unclear, however, the degree to which 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. Our paper employs the technique proposed in Song (2004) with some adjustments to allow for the semi- nonparametric estimation of consumer surplus. Estimates of consumer surplus based on semi- nonparametric estimation of the bidding function also can be analyzed along with estimates generated from parametric analysis with the same data. Such parametric models of the bidding function, although not nonparametrically identified, can utilize a much larger set of our data and they provide a useful robustness check for our semi- nonparametric results. Fully nonparametric estimates of consumer surplus also can be derived and yield yet another method of comparison.

Our research also provides a new methodology for estimating consumer surplus. This methodology relies on a strong assumption of homogeneity in the pool of potential bidders as new auctions for the same generic product are conducted. This method is robust to tail probability properties of the underlying and nonparametrically specified distribution of private values. Essentially this methodology considers a counter factual wherein, if the price setting bidder in auction t' won auction t, then the consumer surplus would be the average over the t' that could have won the given auction. In reduced form estimations this would be the only feasible manner to estimate consumer surplus. In our paper it provides yet another set of robustness checks on our estimates.

We thus consider six different methods to calculate the consumer surplus and consumer share of the total surplus in our auction data. Although there is significant variation among these estimates, they do provide relatively tight bound for the consumer surplus for the auctioned good we analyze, computer monitors. According to our semi- nonparametric analysis, the median consumer surplus per computer monitor may be as high as \$51 or as low as \$17 with a median value of \$28. The median lower bound on the consumer share of surplus may be as high as 62.9% or as low as 9.5% with a median value of 19.0%.

Using a spider program we collected data on over 9000 computer monitors auctioned on eBay between February 23, 2000 and June 11, 2000 (Gonzalez,

2002, Gonzalez, et al. 2009). Lucking-Reiley, et al. (2007) utilized a spider program to collect eBay data on 461 "U.S. Cent" category auctions held at eBay over a 30-day period during July and August of 1999[1]. Recent methods for accessing data via "spider" programs have become commonplace[2] 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. Song (2004) 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 consumer share of the surplus is 53%. Bapna, Jank, and Shmueli (2008) 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 around 18.3% of the total surplus. Several other articles estimate consumer surplus in multi-unit auctions—Carare (2001); Bapna, Paulo and Gupta (2003a), (2003b) and Bapna, Goes, Gupta, and Jin (2004)—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 a modification of the methods developed in Song (2004). As discussed in Bulbul Toklu (2010), there have been several techniques identified for estimating bidders' values in online auctions. In order to nonparametrically identify his model, Adams (2007) assumes exogenous entry and that entry is not affected by any variables that affect bidders values. Parametric structural analysis of the bidding function and the entry rule can relax these assumptions, albeit at the cost of not being nonparametrically identified. Such structural analysis rejects these assumptions with our data but relies on the strong assumption that the pool of potential entrants is quite large. Nekipelov (2007) introduces an endogenous entry rule, wherein a rise in the auction price is assumed to decrease the probability of entry but increase the average bid conditional on entry. However, his model relies on the explicit calculation of the equilibrium bidding function at all parameter values and thus limits the computational appeal of his approach when there are a reasonable number of covariates. His model also requires the assumption that all bidders

---

[1] Specifically, Lucking-Reiley et al. (2007) focused on U.S. Indian Head pennies minted between 1859 and 1909, auctions in which only one coin was for sale, and the coin was in mint state (MS) with grades of between 60 and 66 on a 70-point scale.

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

use the nonparametrically identified but rather opaque and complex equilibrium bidding functions. Nekipelov's model can explain both squat bidding (bid early to deter others from bidding: Ely, and Hossain, 2009) and snipe bidding (bid at the last second to deter counter bidding: Roth and Ockenfels, 2002). In contrast to the complex equilibrium bidding functions used by Nekipelov Song (2004) only assumes that the second and some of the third highest bidders bid their true value. It is known that not all the third bidders will bid their true value, but her methodology allows one test which bids should not be used. Her model also allows for exogenous entry, endogenous entry (for a range of models), and heterogeneous entry decisions and utilizes a relatively straightforward rule to determine how much to bid. One weakness of all these models is that they do not take into consideration the exit value of bidders. In an eBay auction this value may be significant. The only paper currently in the literature that takes this into consideration is Sailer (2006). Formally these models must rely on the *steady state hypothesis* (Hasker and Sickles, 2010) and the distribution is only identified up to a shift parameter.

There also several parametric techniques that deserve special mention, though currently none of these techniques are nonparametrically identified. Bajari and Hortaçsu (2003) develop a Bayesian methodology, but require that the bidding functions be linearly scalable. Non-linear simulated least squares, developed by Laffont, Ossard, and Vuong (1995) and used in Gonzeles, et al. (2007), is 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 in which revenue equivalence holds.

We organize our discussion of methods to analyze consumer surplus in eBay auction in the following way. Section 2 reviews our econometric methodology. Section 3 describes the data used in our estimation. Section 4 discusses our results. Sections 5 and 6 develop various measures of Consumer Surplus and Consumer Share of Surplus generated in our auctions and provides estimates of these measures. Section 7 concludes.

2   Econometric Methodology

Athey and Haile (2002, 2005) show that the underlying distribution of private values is uniquely determined if the distribution of any order statistic with a known number of bidders is identified. However, in eBay auctions, the number of potential bidders is generally not observable. Song (2004) addressed this issue by proving that, within the symmetric independent private values model, observations of any two valuations whose ranking is known can nonparametrically identify the bidders' underlying value distribution. Song goes on to point out that one 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 one should use data wherein either the third highest bidder had a good reason to believe she could win the auction or the higher bids were submitted late. She develops an econometric test to discover which third highest bids can be used. We will adopt her methodology to pursue our nonparametric estimation of consumer surplus from eBay auctions. We also follow Haile and Tamer (2003) by first assuming that bidders adhere to 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.

These rules are transparent and appealing, and guarantee that the second highest bidder will bid his value. Unfortunately the conditions are not sufficient for identification. As Haile and Tamer (2003) show, there are equilibria in second price auctions (and eBay auctions) in which these rules do not imply that the third highest bidder bid's his true value. Thus we make a third assumption

3. A finite number of bids from the third highest bidder reflect the bidders' true value.

We will thus need to examine which bids are the third highest bidder's true value. We should mention that this is not only a theoretical problem. Empirical evidence also shows that bidders do not bid their true value on eBay. To give an example to illustrate, assume that a bidder bids $50 for an item which he values at $80, and ten two other bidders immediately bid $100. After observing the higher bid of $100, this bidder will not update his bid and his final bid will be less than his value. On the other hand if only one bidder makes a bid higher than $80, and another bidder bids between $50 and $80, the given bidder will update his bid to his value $80. In this case the bidder bids more than once. This would not occur were bidders to bid their values. In other words, the existence of multiple bids is evidence that bidders are not bidding their value, and in fact there frequently are multiple bids per active bidder on eBay. Song (2004) points out that if the two highest bids are submitted right before the end of the auction (for example, within the last minute of the auction) then the third highest bid will almost certainly be that bidder's true value. In this case the third highest bidder must know that if he raises his bid then he might win the auction and thus his final bid must be his true value. This approach therefore requires that the third highest bidders who are outbid in the last minute are bidding their value. One then tests whether or not bidders outbid at earlier times are using the same bidding rule. We also make a relatively standard assumption about the bidder's values in our next condition.

4. Bidders' values are private, independent, and log-linear in a set of auction specific characteristics. Private values are given by:

$$\ln V_{m|i} = x_m' \beta + v_{m|i}, \tag{1}$$

5

with $m = 1, ..., M$, where M is the number of auctions and $i = 1, 2..., N_m$, where N_m is the number of potential bidders in auction m. For the estimation procedure we outline below, we require the potential number of bidders in any auction to be greater or equal to 3. We note that the standard assumption of private values is proscribed for the good under consideration. At the time of data collection computer monitors were subject to rapid technological development, thus anyone considering buying a monitor would know that the value of the monitor would decrease sharply in as little time as six months. It is also a relatively standard good, thus there would not be much information discovery from bids.

$V_{m|2}$ and $V_{m|3}$ represent the second and third highest bidders' valuations in auction m, respectively. We use the second and third highest bids as estimates of these two valuations. $v_{m|2}$, and $v_{m|3}$ are the corresponding error terms. x_m is the control variable including 7 auction specific characteristics that we specify below, $\beta = [\beta_1, \cdots, \beta_7]$ is the corresponding vector of coefficients. We consider the sample counterpart of conditional likelihood function $f\left(v_{m|2} \,|\, v_{m|3}\right)$ specified by Song (2004), since the full likelihood (the joint density of $\left(v_{m|2}, v_{m|3}\right)$) requires the unknown number of potential bidders. According to the basic theory of order statistics, the sample likelihood function can be written as:

$$L_M\left(\hat{f}\right) = \frac{1}{M} \sum_{m=1}^{M} \ln \frac{2\left[1 - \hat{F}\left(v_{m|2}\right)\right]\hat{f}\left(v_{m|2}\right)}{\left[1 - \hat{F}\left(v_{m|3}\right)\right]^2}, \tag{2}$$

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 $c = \min_{m=1,2,...M}\left(\log(V_{m|3})\right)$, since no information about $F(v)$ for $v < c$ can be observed. The reader should note that with our objective function this value does not affect our estimates. In order to estimate the unknown distribution of v we employ the method proposed by Coppejans and Gallant (2002) and use the hermite series to approximate the unknown distribution. Gallant and Nychka (1987), Fenton and Gallant (1996) and Coppejans and Gallant (2002) provide details on how to use this method to approximate the unknown distribution of private values. The optimal series length varies according to the data set under consideration. Coppejans and Gallant (2002) proposed a cross-validation strategy by employing the Integrated Squared Error (ISE) criterion to choose the optimal series length k^*. The ISE in their paper is defined as:

$$ISE\left(\hat{f}\right) = \int \hat{f}^2(y)\,dy - 2\int \hat{f}(y)f(y)\,dy + \int f^2(y)\,dy$$
$$= M_{(1)} - 2M_{(2)} + M_{(3)} \tag{3}$$

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

Here what we are interested is a conditional density. Along the line of Coppejans and Gallant (2002), we propose Weighted Integrated Squared Error

(WISE) which serves as our criteria in selecting the optimal series length. WISE is defined as follows:

$$
\begin{aligned}
ISE(\hat{f}) &= \int \int (\hat{f}(y|x) - f(y|x))^2 f(x) dy dx \\
&= \int \int \hat{f}(y|x)^2 dy f(x) dx - 2 \int \int \hat{f}(y|x) f(y|x) f(x) dy dx \\
&\quad + \int \int f(y|x)^2 dy f(x) dx \\
&= Q_1 - 2Q_2 + Q_3,
\end{aligned}
\tag{4}
$$

where $\hat{f}(y|x)$ is an estimator of true conditional density $f(y|x)$. In implementing the cross-validation strategy, first, we randomly partition the data set under consideration into 5 groups, denoted by χ_j, $j = 1, ..., 5$, making the sizes of these groups as close to equal as possible. Let $\hat{f}_{j,k}(\cdot)$ denote the semi- nonparametric estimate obtained from the sub-sample that remains after deletion of the j'th group when k is used as a series length. The cumulative distribution associated with $\hat{f}_{j,k}(\cdot)$ is denoted by $\hat{F}_{j,k}(\cdot)$. Since the third term only involves true densities, we only need to look at the first two terms. The estimates of the two terms are defined as below:

$$
\hat{Q}_1(k) = 1/M \sum_{j=1}^{5} \sum_{(x_m, y_m) \in \chi_j} \int [\hat{f}_{j,k}(y|x_m)]^2 dy
\tag{5}
$$

$$
\hat{Q}_2(k) = 1/M \sum_{j=1}^{5} \sum_{(x_m, y_m) \in \chi_j} \hat{f}_{j,k}(y_m|x_m).
\tag{6}
$$

We also define

$$
CVH(k) = \hat{Q}_1(k) - 2\hat{Q}_2(k).
\tag{7}
$$

It is worth mentioning that our criteria is WISE, and hence $\hat{Q}_1(k)$ and $\hat{Q}_2(k)$ are not the same as what were defined in Song (2004). 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 which brings the last abrupt drop of $CVH(k)$. Our result is listed in Table B.1 of Appendix B and shows that the abrupt drop of $CVH(k)$ occurs when k changes from 1 to 2. The small increase in $CVH(k)$ when k changes from 3 to 4 is due to small sample size. The second and fourth term in the series will be correlated in a small sample, and including the spurious fourth term causes identification problem.

The density function of v_m follows immediately as:

$$f\left(v_m\right) = \frac{\left[1 + a_1\left(\frac{v_m - u}{\sigma}\right) + a_2\left(\frac{v_m - u}{\sigma}\right)^2\right]^2 \phi\left(v_m; u, \sigma, c\right)}{\int_c^\infty \left[1 + a_1\left(\frac{z - u}{\sigma}\right) + a_2\left(\frac{z - u}{\sigma}\right)^2\right]^2 \phi\left(z; u, \sigma, c\right) dz} \quad . \tag{8}$$

The nonparametric maximum likelihood estimator is then defined as:

$$\left(\hat{\beta}, \hat{a}, \hat{u}, \hat{\sigma}\right) = \arg\max_{(\beta, a_1, a_2, u) \in \mathbb{R}^{10}, \sigma \in \mathbb{R}_{++}} L_M\left(\hat{f}\right)$$

$$= \arg\max_{(\beta, a_1, a_2, u) \in \mathbb{R}^{10}, \sigma \in \mathbb{R}_{++}} \frac{1}{M} \sum_{m=1}^M \ln \frac{2\left[1 - \hat{F}\left(v_{m|2}\right)\right] \hat{f}\left(v_{m|2}\right)}{\left[1 - \hat{F}\left(v_{m|3}\right)\right]^2} \quad . \tag{9}$$

One criticism of this method is that the third highest bids may not reflect the bidders' private values since they utilize the second and third highest bids as estimates of the second and the third highest bidders' private values. We follow Song (2004) by using data in which the first or second highest bidder submitted a cutoff price greater than the third highest bid late in the auction. To determine how late is sufficient, Song (2004) provides a method based on the CVH using a modified formula for $\hat{Q}_1$ and $\hat{Q}_2$. Following her method, we consider a sequence of 6 sub data sets, $A_{w1}, \cdots, A_{w6}$, with different window sizes, arbitrarily chosen so that the size difference between two adjacent subsets is similar. We choose window sizes as $w1 = 5$ minute, $w2 = 15$ minutes, $w3 = 40$ minutes, $w4 = 2$ hours, $w5 = 3.5$ hours and 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 5 minute before the auction ends. Other sub data sets are defined in the similar way. Obviously, we have $A_{w1} \subset A_{w2} \subset \cdots \subset A_{w6}$. 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 and suggests choosing the window size which has the smallest CVH_{wi}. For each auction set A_{wi}, we calculate CVH_{wi}. Corresponding to equation (7), CVH_{wi} is defined as

$$CVH_{wi}(k^*) = \hat{Q}_{1_{wi}}\left(k^*\right) - 2\hat{Q}_{2_{wi}}\left(k^*\right),$$

where

$$\hat{Q}_{1_{wi}}(k^*) = 1/M_{wi} \sum_{j=1}^5 \sum_{(x_m, y_m) \in \chi_j \cap A_{wi}} \int [\hat{f}_{j,k^*}(y|x_m)]^2 dy$$

$$\hat{Q}_{2_{wi}}(k^*) = 1/M_{wi} \sum_{j=1}^5 \sum_{(x_m, y_m) \in \chi_j \cap A_{wi}} \hat{f}_{j,k^*}(y_m|x_m),$$

where M_{wi} is the sample size of subset A_{wi}. We present the results in Table B.2 in Appendix B. It is clear that CVH_{wi} decreases when window size increases from $w1$ to $w4$. The values of CVH_{w4} and CVH_{w5} are almost identical to each other. The change of window size from $w5$ to $w6$ follows by a dramatic increase of CVH_{wi}. Since our analysis is based on semi- nonparametric and nonparametric methods, we want to keep as much data as possible. We choose $w5 = 3.5$ hours instead of $w4 = 2$ hours as our optimal window size since the difference between CVH_{w4} and CVH_{w5} is rather small.

3 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 excluded any non-working, touch screen, 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% 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.)

Descriptive variables except for monitor size were constructed using string searches. In Gonzalez, et al. (2009) 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% market share while the smallest have close to a 3% market share. These 10 firms represent 59% of the market. Dot pitch and resolution are not reported in all of the auctions. Dot Pitch is reported in 42% of the auctions, resolution in 64%.

Since selecting a relatively homogeneous data set is important in conducting and interpreting results from nonparametric analysis we dropped all auctions that were not clearly for 17" color PC monitors. Monitor size has the most pronounced and significant effect on bidders' private values. Since we need 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. This gives us 476 observations. To make the data set even more homogeneous, we also drop 12 auctions in which warranties were offered on the auctioned monitors. Our final data set has 464 observations.

In estimating the distribution of bidders' private values with the semi- nonparametric approach, we used the following 7 control variables: monitor dot pitch (0 is used when no dot pitch is reported), dummy for the cases when no dot pitch is reported, monitor resolution (0 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 sample are presented in Table A.1 in Appendix A.

Notice that we do not use auctioneer's feedback rating – a reputation system on eBay – in our estimates. While this variable may affect entry under the private value assumption, it cannot affect bids conditional on entry.

4   Estimates

For the results that follow we choose the optimal hermite series number as $k^* = 2$ and the optimal window size as $w5 = 3.5$ hours, i.e., we choose the auctions where the first or second highest bidder submitted a bid greater than the third highest bid no earlier than 3.5 hours before the auction ended. This yields a sample of 376 observations on which to base our semi- nonparametric estimates of consumer surplus.

Table 1: Semi- Nonparametric Estimates

Dot Pitch	-7.9842***
	(0.1310)
Dummy, No Dot Pitch	-1.9606***
	(0.0347)
Resolution	0.0486***
	(0.0051)
Dummy, No Resolution	-0.1024***
	(0.0060)
Dummy, Brand Name	-0.0465***
	(0.0005)
Dummy, Flat Screen	-0.0769***
	(0.0004)
Status	0.0568***
	(0.0001)
Number of Auctions	376

***Significant at the 1% confidence level

The coefficients except for the ones on "Brand Name" and "Flat Screen" dummies have the expected sign and are highly significant. A small dot pitch is better so we expect the coefficient on the log of dot pitch to be negative, likewise a larger resolution and better condition are both good so those coefficients should be positive. When no dot pitch or resolution are reported the bidders expect a worse than average value for these variables. "Brand Name" means popular brand,which includes 10 brands in our data set. There is no consensus on what the sign should be for the variable "Brand Name". Both positive and negative results have been seen in literature. In our study, we find negative effect of "Brand Name" on bidders' private value. The coefficient on "Flat Screen" is negative and significant which does not suite our intuition. However, since the magnitude of the estimate is tiny, we have a reason to believe that the "Flat dummy" does not play an important role in determining the bidder's private. Of course, another possible reason for the incorrect sign on "Flat dummy" might be because we use semi- nonparametric estimation method. Parametric methods shall give more precise estimates.

Our consumer surplus estimates are based on these coefficients. Because the data we use for the 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 2. SNP and NP denote semi-nonparametric and nonparametric methods respectively.

Table 2

Statistics of Estimated Distribution		
	Mean	Standard Deviation
SNP	$30.18	$26.77
NP	$28.26	$26.63

In the semi- nonparametric analysis, the mean and standard deviation are computed with the median values of $x_1, x_2, \cdots,$ and, x_7.

5 Structural Consumer Surplus and Consumer Share of Surplus

In order to investigate the welfare impact of eBay, we calculate the consumer surplus and consumer share of surplus. Consumer surplus in auction m is calculated as:

$$CS_m = V_{m|1} - p_m, \tag{10}$$

where p_m denotes the price the winner paid, which equals to the second highest bid in eBay auctions. $V_{m|1}$ denotes the valuation of the winner. Since we do not observe $V_{m|1}$, we estimate the expected consumer surplus as:

$$E\left(CS_m | \hat{V}_{m|2}\right) = e^{x'_m \hat{\beta}} \int_{\hat{v}_{m|2}}^{\infty} \frac{f(v)}{1 - F\left(\hat{v}_{m|2}\right)} e^v dv - p_m \tag{11}$$

Again, $\hat{v}_{m|2}$ is the estimator of $v_{m|2}$ calculated based on model (1) with estimated coefficient $\hat{\beta}$ and x_m, which is the vector of the values of control variables in auction m; and p_m is the price. For comparison we include the parametric estimates(PL) wherein private values are distributed as half-logistic. These preferred estimates are based on a battery of nonparametric good-of-fit tests of a number of parametric distributions for private values. The descriptive statistics of expected consumer surplus from our SNP, NP methods, and PL approaches are presented in Table 3.

Table 3

Structural Consumer Surplus					
	Mean	Median	Std. Dev.	Min	Max
SNP	$28.32	$28.31	$3.85	$16.94	$51.06
NP	$28.57	$28.31	$1.57	$19.28	$39.28
PL	$52.54	$39.69	$38.23	$13.51	$210.09

Notice that the distribution of consumer surplus based on the parametric method is highly skewed. For this reason here and below we focus on the median values, not the averages, for comparison among the three methodologies, since for the SNP and NP estimates the median and mean are quite similar. There could be at least two reasons for the significant divergence of parametric and semi-nonparametric and nonparametric estimates. Of course one reason could be the more flexible distribution of private values in the semi- nonparametric and nonparametric methodologies. However, it could be because the auctions used for the semi- nonparametric and nonparametric auctions are more competitive. Since in order to estimate the semi- nonparametric and nonparametric models, all auctions with less than three bidders were dropped and thus the average number of bidders in the semi- nonparametric and nonparametric data subset

is 8.1, versus 6.8 for the parametric data set. One way to find which effect is more pronounced is to examine the parametric estimates of consumer surplus using the semi-nonparametric data set based on a methodology which does not depend on the distribution function. We do this below.

While the amount of consumer surplus in an auction is a significant statistic it reveals only one dimension of the surplus being generated. If the value of the average monitor is high relative to the surplus even if the size of the surplus is substantial the fraction of the surplus captured by consumers might be small. A low share of surplus indicates that auctions under consideration were highly competitive, and that auctioneers were not earning large profits on eBay. Hence, the consumer share of surplus is another important measure to understand the eBay auction market. This measure is the fraction of total surplus that is captured by the consumers and is defined as:

$$CSS_m = \frac{CS_m}{V_{m|1} - V_{m|a}} = \frac{V_{m|1} - p_m}{V_{m|1} - V_{m|a}}. \tag{12}$$

where $V_{m|1}$ is the winning bidder's private valuation in auction m, p_m is the price, and $V_{m|a}$ is the value the auctioneer places on the item auctioned. Although we do not have a direct measure of $V_{m|a}$ we do know that its lower bound is 0. One could use the auctioneer's reservation price to produce a tighter bound. Theoretically this should be equal to the auctioneer's reservation value, but when one looks at the data one realizes that if that were true almost all auctioneer's do not value their good. Thus using this information would not change much and we obviously do not have a correct theory for the relationship between reservation prices and reservation values. Thus it is better to ignore this value, and we can estimate a lower bound for the consumer share of surplus as:

$$CSS_m = \frac{V_{m|1} - p_m}{V_{m|1}} = 1 - \frac{p_m}{V_{m|1}} \in [0, 1] . \tag{13}$$

We also note that CSS_m is less sensitive to outliers than CS_m. Although we do not directly observe CSS_m we can derive the expected consumer share of surplus in auction m:

$$E\left(CSS_m | \hat{V}_{m|2}\right) = 1 - p_m e^{-x'_m \hat{\beta}} \int_{\hat{v}_{m|2}}^{\infty} \frac{f(v)}{1 - F\left(\hat{v}_{m|2}\right)} e^{-v} dv . \tag{14}$$

Estimates of this expectation are in Table 4.

Table 4

Structural Lower Bound of Consumer Share of Surplus				
	Mean	Median	Min	Max
SNP	20.63%	19.03%	9.52%	62.84%
NP	20.83%	18.98%	9.96%	64.91%
PL	33.90%	30.30%	8.70%	99.90%

In Table 4, PL again 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. Again this could be due to differences in methodology or the fact that there were more bidders on average in the semi- nonparametric and non-parametric data set.

For comparison with other analyses it is useful to substitute and examine the median values of the lower bounds of the consumer share of surplus. In this analysis we use the median price and median consumer's value. In general this is easily computed from the reported coefficients and descriptive statistics. To construct the median consumer's value (if it is not immediately given) one uses the sales price and the coefficients of the regression. If the median price is p_m and x_m the median values of the right hand side variables then with a log linear specification this is $p_m e^{-x_m'\beta}$. These results are shown in Table 5. 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 5

Medians Lower Bound of Consumer Share of Surplus			
	SNP	NP	PL
MCSS	18.93%	18.98%	28.41%

We can see that the results are very close, although the 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 using the same methodology. Song's result is significantly higher than all of the results above. The difference can be explained with competition levels involved. The average number of bidders is 3.6 in Song (2004), which is significantly lower than either subset of monitor auctions. More competition on the bidders' side would appear to result in lower consumer share of surplus.

6 Distribution Free Consumer Surplus and Consumer Share of Surplus

A problem with both parametric and semi- nonparametric and nonparametric estimation is upper tail sensitivity. The parameters determining the weight on the upper tail are determined by observations at the center of the distribution, thus the upper tail can be easily too thick or too thin. For extreme value statistics like consumer surplus this can cause significant problems. It would be desirable to find an alternative method that is not as sensitive to the underlying distribution.

A secondary problem is that there is no simple method to estimate consumer surplus if one does not use structural estimation. Thus this important statistic is often overlooked in empirical analysis. It is possible to estimate consumer

surplus without performing structural estimation but it requires additional assumptions. One that we pursue in this section is that the set of potential bidders is constant for all auctions, which is not equivalent to assuming a constant set of active bidders. Randomizing the entry order over the set of potential bidders would produce a large variation in the number of active bidders. However, we do require this number to be nonstochastic and that it does not vary from auction to auction, which is often implicit in interpreting results from many reduced form auction studies.

In structural estimations the following methodology produces an alternative way to measure consumer surplus and provides a potentially more robust picture of the size of the consumer surplus. Let N be the number of potential bidders, it is clear that the distribution of the first order statistic—$H^1(V|N)$—first order stochastically dominates the distribution of the second order statistic—$H^{(2)}(V|N)$. Utilizing the fact that it is a lower bound for $H^{(1)}(V|N)$ one can produce a lower bound for consumer surplus. Under reasonable assumptions on $H^{(2)}(V|N)$ we know that:

$$H^{(2)}(V|N) = \lim_{M \to \infty} \frac{\#\left(m' \in M | p_{m'} e^{-x'_{m'}\hat{\beta}} \leq V\right)}{M} .$$

(15)

For finite M of course the right hand side is only an alternative estimator for $H^{(2)}(V|N)$. Obviously if the potential number of bidders is stochastic or due to simple bad draws we can have $H^{(2)}(V|N) < \frac{\#\left(m' \in M | p_{m'} e^{-x'_{m'}\hat{\beta}} \leq V\right)}{M}$ for some V. As we will show with our data, it is possible that this estimator has a fatter upper tail than the structural estimates. We essentially construct the estimator by setting up a counter factual wherein the price setter in auction m' wins auction m instead. Averaging this over the m' that could have won auction m we derive an estimate of the consumer surplus in auction m. This statistic can easily be calculated using only the estimated coefficients and the data. Let 1_x be the indicator function which is 1 if x is true, 0 otherwise. Based on Model (1) and this estimating approach, consumer surplus can be derived as:

$$\widehat{CS}_m = e^{x'_m \hat{\beta}} \frac{\sum_{m'=1}^{T} p_{m'} e^{-x'_{m'}\hat{\beta}} 1_{p_{m'} e^{-x'_{m'}\hat{\beta}} \geq p_m e^{-x'_m\hat{\beta}}}}{\#\left(m' \in T | p_{m'} e^{-x'_{m'}\hat{\beta}} \geq p_m e^{-x'_m\hat{\beta}}\right)} - p_m .$$

(16)

We refer to this as the **distribution free consumer surplus** because it does not require nor make use of any estimates of the distribution of the error term.

The descriptive statistics for this estimate of consumer surplus are in Table 6.

<div align="center">Table 6</div>

	Mean	Median	Std. Dev.	Min	Max
Distribution Free Consumer Surplus					
SNP	$44.47	$39.64	$17.36	$0	$119.89
NP	$41.48	$37.02	$16.06	$0	$112.50
PL, All Data	$65.47	$45.69	$53.25	$0	$1185.82
PL, SNP Data	$40.88	$36.65	$16.60	$0	$130.63

The minimum consumer surplus is zero by construction. Interestingly, median estimates of this measure of consumer surplus are higher than those based on the structural estimates we presented earlier, due to the presence of outliers. These outliers could either be due to bad draws from the underlying distribution or due to the number of potential bidders being stochastic. Either problem could cause a given auction to be quite competitive and result in a relatively high value for the price setting bidder resulting in larger reduced form estimates of consumer surplus in every auction. To take account of the outliers we trim both the top and bottom varying percentages to see how much trimming is necessary to stabilize the estimated consumer surplus. For the parametric and nonparametric methodologies we estimates stabilize with 2% total trimming . For the semi- nonparametric methodology 8% total trimming was necessary. The statistics generated without any trimming are significantly larger with the difference in medians about $10 for the semi- nonparametric and nonparametric and $6 for the parametric estimates.

Recall our findings above wherein the parametric structural estimates of consumer surplus where larger than those based on the semi-nonparametric and nonparametric estimates. These differences have at least two causes. One is that the parametric methodology is less flexible. Another is that we utilize a more competitive data set for the semi- and nonparametric estimations. The last row in Table 7 below points to the latter rationale. When we estimate the distribution free consumer surplus using the data set where matching allowed us to utilize the semi- and nonparametric methods we find that the estimates are a bit lower than the semi- and nonparametric estimates, but comparable. Indeed, when one compares these estimates of consumer surplus this those in the row above it is clear that a major explanation for differences in estimates of consumer surplus are due to selecting a more competitive data set for the semi- and nonparametric estimates.

The new measure of consumer share of surplus also will be less sensitive to these outliers and would provide a potentially more robust picture of how much surplus is being generated. In the consumer share of surplus the value of the counter factual is always between zero and one and this normalization also makes the statistic less sensitive to outliers. The statistic is:

$$\widehat{CSS}_m = 1 - p_m e^{-x'_m \hat{\beta}} \frac{\sum\limits_{m'=1}^{M} \frac{1}{p_{m'} e^{-x'_{m'} \hat{\beta}}} 1_{p_{m'} e^{-x'_{m'} \hat{\beta}} \geq p_m e^{-x'_m \hat{\beta}}}}{\# \left(m' \in M | p_{m'} e^{-x'_{m'} \hat{\beta}} \geq p_m e^{-x'_m \hat{\beta}} \right)} , \qquad (17)$$

Estimates of this new measure of the share of consumer surplus are given in Table 7.

Table 7

Reduced Form, Lower Bound of Consumer Share of Surplus				
	Mean	Median	Min	Max
SNP	28.34%	25.16%	0%	91.05%
NP	27.20%	23.58%	0%	91.43%
PL, All Data	32.42%	26.35%	0%	99.90%
PL, SNP Data	23.62%	19.83%	0%	90.44%

These results are consistent with those in Table 6 and indicate that the median distribution-free semi-nonparametric and nonparametric estimates are somewhat higher than their structural counterparts, with a difference of about 6% for SNP and 5% for NP. The corresponding consumer share measure based on the parametric estimates is lower than that its structural counterpart by about 4%. When we use the parametric model to estimate consumer surplus using the data set based on the matching needed to employ our semi-nonparametric estimators it is now the lowest estimate of all, as in Table 6, which we would expect as estimated average values are lower for both consumer surplus and consumer's share of surplus. This would suggest that estimates of consumer surplus fall significantly when these statistics are high and have little impact on them when they are low. Thus if we estimated the parametric model on the restricted data set it is likely that the consumer's surplus and consumer's share of surplus will be higher.

The advantages of this particular distribution free methodology are twofold. First it is easy and immediate to calculate after any estimation. Second it appears to be relatively robust. Its disadvantage is the assumption that the pool of potential bidders is the same for each auction.

7   Conclusion

In this paper we estimate consumer surplus for eBay computer monitor auctions with semi- nonparametric and nonparametric methods. We compare our results with estimates based on parametric assumptions on the distributions of private values. We also develop a new technique, a reduced form technique, to estimate these important statistics. This new technique provides robustness checks for our estimates. We also develop a new method that places a lower bound on the consumers' benefit from these auctions, the consumer share of surplus. This provides more insight into the degree of competition in these auctions.

The general conclusions from our empirical study is that the market for computer monitors on eBay was competitive, but not at the extreme, from February 23, 2000 to June 11, 2000. It seems that the median consumer was capturing around $28 in consumer surplus or 19% of the total surplus available. This suggests that the auctioneers were capturing at most 81% of the total surplus.

While this is a hefty share this does not take into account the unknown value that auctioneers place on their computer monitors. It would be interesting to know what share of the surplus consumers are capturing in a similar market today. Since eBay is an auction marketplace, high profits will draw more auctioneers to the market and high consumer surplus will draw more bidders. However, it is much more costly to become an auctioneer and thus the number of auctioneers per bidder is likely to have increased over time.

We would like to encourage more analysts to estimate the consumer surplus and consumer share of surplus generated in online auctions. It would be worthwhile to develop a more general picture of how much eBay is benefitting our economy. In this vein we point out that our reduced form methodology does not require the standard structural assumptions necessary to estimate consumer surplus and seems to produce estimates that are close to structural estimates, especially if the data is trimmed by reasonable trimming factors.

References

[1] Adams, C. P., 2007. Estimating demand from eBay prices, *International Journal of Industrial Organization.* 25, 1213-1232

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

[3] 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.

[4] 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.

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

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

[7] 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.

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

[9] Bulbul Toklu, S., 2010. Methodology: Structural Parametric and Nonparametric Methods for Online Auctions. *mimeo.* available at http://www.owlnet.rice.edu/~seda.bulbul/.

[10] Carare, O., 2001. Need for speed: demand estimation using 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] Ely, J.C., and Hossain, T., 2009. Sniping and Squatting in Auction Markets. *American Economic Journal: Microeconomics*, 1, 68-94.

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

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

[15] Gonzalez, R, Hasker, K., and Sickles, R.C. 2009. An Analysis of Strategic Behavior in Ebay Auctions *The Singapore Economic Review* 54, 441-472.

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

[17] Hasker, K., Sickles, R.C. 2010. eBay in the Economic Literature: Analysis of an Auction Marketplace. *Review of Industrial Organization.* forthcoming.

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

[19] Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D., 2007. Pennies from eBay: the determinants of price in online auctions, *Journal of Industrial Economics*, 55, 223-233.

[20] Nekipelov, D., 2007. Entry Deterrence and Learning Prevention on eBay. Working paper.

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

[22] Sailer, K. 2006. Searching the eBay Marketplace. *University of Munich, Department of Economics Discussion Papers,* Number 1234.

[23] 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 A.1: Descriptive Statistics of Key Variables-Second Sample

	Average	Median	Std. Dev.	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

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

Appendix B: Tables of Semi- Nonparametric Estimation

Table B.1

Relations Between CVH and k					
	k=0	k=1	k=2	k=3	k=4
CVH	-3.83	-3.83	-3.89	-3.89	-3.88

Table B.2

Relations Between CVH and Window Size, k*=2					
	w1 ≤ 5 min	w2 ≤ 15 min	w3 ≤ 40 min	w4 ≤ 2 hours	w5 ≤ 3.5 hours	w5 =all
CVH	-4.65	-4.69	-4.71	-4.72	-4.72	-4.62