Appendix

for

"eBay in the Economic Literature: Analysis of an Auction Marketplace"

Kevin Hasker[1]
Bilkent University
Economics Department

Robin Sickles[2]
Rice University
Economics Department

[1] Bilkent University 06800 Bilkent Ankara, Turkey. Tel: +90 (312) 290 1875, Fax: +90 (312) 266 5140, hasker@bilkent.edu.tr
[2] Rice University 6100 South Main Street Houston, Texas 77005-1892, Tel: (713)-348-3322, Fax: (713)-348-5278, rsickles@rice.edu

Appendix A. Breakdown of Selling Formats

Table 1 Here

Appendix B. Methodology: Structural Parametric and Nonparametric Methods

In this section we outline standard approaches in the literature to estimating bidders' values and the entry process in eBay auctions using parametric and nonparametric methodologies. We assume throughout that the analyst is analyzing single item auctions, that the item has private values, entry is exogenous, and that the following insight from Haile and Tamer (2003) holds. This assumption is 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.

The second assumption must be applied with care in eBay auctions. Even if one considers exit as long as one assumes that the steady state hypothesis holds then this insight should be correct. Why would the second highest bidder leave "money on the table" in this auction to go to another auction where the competition will be just as strict?

As explained above, bidding 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, and the assumptions above makes sure that the second highest bidder will do this. For each bidder there will be a common 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 values are log linear then the formula for the winning bid is:

$$b_n^w = \max\left\{r_n, e^{x'_n\beta}\rho_n^{(2)} - c\right\} \tag{B-1}$$

where $\rho_n^{(2)}$ is the private component of the second highest bidder in auction n given there are I potential bidders, and c is the continuation value of bidders. We follow Sailer (2006) by assuming that bidders can not exit an auction until the auction is finished and simplify the analysis by assuming this constant is independent of the bidder.

Note that while ρ_n^i will be from a standard distribution, $\rho_n^{(2)}$, is not since it is the second order statistic from a sample of I bids. One method to estimate (B-1) is to use a Tobit model controlling for heteroskedasticity.However while this method is consistent it is a reduced form approach, leaving the analyst with no information about the true distribution of private values or how competitive the auction is. An auction is competitive if the number of potential bidders, I, is large, and this variable can not be estimated using reduced form techniques from this equation.

A straightforward methodology at this point would be to utilize formal maximum likelihood techniques, and extend them to allow for an exogenous entry process. Let $F_n(z, \beta)$ be the cumulative distribution function of the bidders' values at z and $f_n(z, \beta)$ be the probability density function—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:

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

Inactive bidders are somewhat problematic. Although there must have been at least I_n^a who have bid there might also be any number of bidders who thought about bidding and did not. Therefore I is often treated as a random variable that can range from I_n^a to $\overline{I}$—an arbitrary upper bound. One can view this 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. In this section we do not explicitly correlate the number of bidders with other potential heterogeneity controls or the feed-back rating of the seller, but rather estimate this potential auction specific unobservable. For an extensive treatment of unobserved heterogeneity in first-price auctions see Bierens and Song (2008).

One could estimate the number of potential bidders in each auction as a constant (perhaps varying with the length of auction or other discrete variables) or alternatively one could estimate an entry process. The number of bidders in an auction is often modeled as a Poisson entry process. The parameter of the entry process, λ_n, is often assumed to be log-linear in a set of auction specific characteristics z_n. Notice that some auction characteristics might affect entry but not private values and vice a versa, although it is difficult to understand what these non-overlapping variables might be a priori. However based on the analysis in the literature it seems clear that feedback—for example—has a much stronger effect on entry than it does on the sales price. A convenient functional form for entry is:

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

Let T_n be the length of the auction and D_n^{sr} be an indicator set to one if there is a secret reservation price and zero otherwise. Then the total likelihood for auction n with exogenous entry is:

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

The lower bound on i in both the denominator and the numerator is increased by one if there is a secret reservation price, following the treatment in Bajari and Hortaçsu (2003a) wherein the auctioneer is treated as another bidder if there is a secret reservation price. When analyzing eBay data sets based on spider programs that collect information on auctions that do not result in sales as well as those that do, full likelihood can be utilized, of course based on a parametric assumption for distribution of private values and the entry process.

Since one cannot be certain *a-priori* of the true distribution of bidders' values, an array of distributions can be used and results based on these different parametric distributions can be compared and analyzed, possibly with nonparametric diagnostic methods. Results can also be compared with nonparametric methods. Common one-sided distributions proposed in the literature are the folded logistic, gamma, Weibull, log-normal, and Pareto.

Another methodology proposed by Laffont, Ossard, and Vuong (1995) is to simulate the auction. Laffont et al. focused on first price auctions but the methodology is the same for eBay auctions. Imagine running S auctions with I bidders in each auction. In each simulation the second highest value is selected ($X_{sn}(\beta, I)$) and these values are averaged to form $\bar{X}_n(\beta, I) = \frac{1}{S} \Sigma_{s=1}^S X_{sn}(\beta, I)$. If S is large then the distance between $\bar{X}_n(\beta, I)$ and $E\left[v^{(2)}\right]$ will be small, and assuming one has the correct value for I then the distance between $\bar{X}_n(\beta, I)$ and $E[b_n^w]$ will be small. However, an unbiased methodology must take into account that in practice S is not large, and thus the objective function should compensate for the variance of the simulated estimator. This variance is $V_{Sn}(\beta, I) = \frac{1}{S(S-1)} \Sigma_{s=1}^S \left(X_{sn}(\beta, I) - \bar{X}_n(\beta, I)\right)^2$. Estimation of β and I are then given by:

$$\arg\min_{\beta,I} Q_{S,N}(\beta,I) = \frac{1}{N}\Sigma_{n=1}^{N}\left[\left(b_n^w - \bar{X}_n(\beta,I)\right)^2 - V_{Sn}(\beta,I)\right] \tag{B-5}$$

where N is the number of auctions. Note that the distribution of $v^{(2)}$ will be a non-degenerate function of β and I. This allows identification of I. On an intuitive level this is because I determines the amount of "skewness" in the observed prices. Theoretically one could allow for exogenous entry as outlined above, but this could prove computationally burdensome.

Another simulation based, parametric estimation approach comes from Bajari and Hortacsu (2003a). Bajari and Hortacsu (2003a) specifies a structural econometric model of eBay auctions, where Bayesian approach is taken to estimate the parameters of the model. They model the eBay bidding problem in a second price sealed bid auction setting with symmetric common value assumptions and stochastic entry, where the entry of the potential bidders depends on a zero profit condition and the number of actual bidders is determined by a Poisson process with mean λ_t. At each auction, bidders only observe a private signal x_{it} about the value of the auctioned good v_t, where $x_{it} = v_t + \varepsilon_{it}$, ε_i is distributed $i.i.d$, and $v \sim N(\mu_t, \sigma_t^2)$ and $x \sim N(\mu_t, k\sigma_t^2)$ (here t is the auction, and i is the bidder subscript). Being motivated by some empirical regularities in the observed characteristics of the auction data, they estimate reduced form relations between some observable auction variables and the structural parameters. Then, they specify the likelihood function of the observed bids using the data of the bidders who have not bid at all for each auction. By this way they assign a positive likelihood to auctions with no bidders, hence make a good use of all the available data. Showing the form of the likelihood function of the bids conditional on the auction specific data and the structural parameters to be estimated, they use this likelihood function to update their prior in the Markov-chain Monte Carlo method by which they simulate the posterior distribution of the model parameters. In the paper they emphasize the advantages to using a Bayesian approach when estimating parametric auction models. First, Bayesian methods are computationally simple and easier to implement than maximum likelihood. Since in many auction models, support of the distribution of the bids depends on the parameters, it is not possible to apply standard asymptotic theory straightforwardly. Moreover, confidence intervals in a classical framework require second order asymptotic approximations, whereas the results in this paper are correct in finite samples and do not require invoking the assumptions used in second order asymptotics. Lastly, in some parametric auction models, Bayesian models are asymptotically efficient while some commonly used classical methods are not efficient.[3]

Sailer (2006) is another interesting paper in terms of modeling and identification strategy, preferring parametric techniques for estimation. Sailer (2006) sets up an intertemporal optimization problem for an eBay bidder who faces an infinite sequence of Vickrey auctions for heterogenous products (She argues that only the bids at the last minutes of the auction, when bidders are no more able to observe the bids by their opponents, determine the sales price and the winner of the auction; that's why we can take into account only the last minutes of an eBay auction and evaluate it as a sealed bid second price auction.[4][5]) The intertemporal optimization setting is designed to reflect the tradeoff between bidding today and waiting for tomorrow, where the bidder optimally chooses a bid to maximize her expected benefit from bidding minus the cost of bidding, but still can choose to wait for the next auction without bidding in the current auction (hence not incurring the cost of bidding now) if the expected future payoff of waiting for the next auction is higher than the maximum of what she can expect to get from bidding in the current auction. The problem can be summarized as the following:

$$V_i(v_i,\mathbf{s}) = \left\{ \begin{array}{cc} \max\left\{ \max\limits_{b_i>r} E_{b_{(1)}}\left[\mathbf{1}\left\{b_{(1)}<b_i\right\}\left(v_i-b_{(1)}\right)-c_i+\mathbf{1}\left\{b_{(1)}\geq b_i\right\}V_i^e|\mathbf{s}\right], V_i^e \right\} & ,before\ win \\ 0 & ,after\ win \end{array} \right\}$$

where $V_i(v_i,\mathbf{s})$ is the value of having the opportunity of bidding in the current auction given the realization of the valuation of bidder i for the currently auctioned good (v_i), the realizations of the characteristics of the currently auctioned good and the current auction ($\mathbf{s}$). b_i is the bid of bidder i if she decides to participate in the current auction, $b_{(1)}$ is the maximum bid of the bidder i's opponents, v_i is the realization of bidder i's valuation of the auctioned good after seing the characteristics of the good and the auction, c_i is the cost of bidding for bidder i, and V_i^e is the continuation value of bidder i (the value she expects to get from the next auctions if she fails to win in the current auction or does not bid at all). The optimal bidding behaviour of bidder i requires the following equality:

$$c_i = E_{\mathbf{s},b_i^*}\left[E_{b_{(1)}}\left[\mathbf{1}\left\{b_{(1)}<b_i^*\right\}\left(b_i^*-b_{(1)}\right)|\mathbf{s}\right]|\delta_i^*=1\right] \quad \text{(B-6)}$$

where $\delta_i^*=1$ denotes that it is optimal for bidder i to participate in the current auction (the entry decision is affirmative),

---

3 see Porter and Hirano (2001).

4 Bajari and Hortacsu (2003) take a similar approach for the eBay auctions of some common value goods.

5 Nekipelov (2007) actually shows that early bidding can benefit the bidder by entry deterrence, hence the early bids may have an effect on the result of the auction.

and b_i^* is the optimal bid of her. Clearly, we could calculate the cost of bidder i if we could observe b_i^* when it is a winning bid. (We can only get biased estimates of the cost if we use the observed highest bid in place of the highest bid b_i^*). This is the point where Sailer(2006) suggests a stepwise procedure which allows her to show that both the distribution of valuations of the bidders and their costs are non-parametrically identified. Given that the model is set under the assumption of asymmetric bidders and that the available data are the bids of all loosing bidders together with the transactions prices and the identities of the bidders in each auction, the identification results from Athey and Haile (2007), Brendstrup and Paarsch (2006) and Song (2004) allow inference about the underlying parent bid distributions. Then, given these bid distributions, estimates of the unobserved winning bids can be built, which in turn are used to compute the costs from (B-6). The likelihood function of the problem is shown to be the following:

$$l = \sum_{t=1}^{T} \ln \left[\frac{f_{b_{l_t}}(b_{(2),t}|\mathbf{x}_t)(1 - F_{b_{m_t}}(b_{(2),t}|\mathbf{x}_t))}{(1 - F_{b_{l_t}}(b_{(3),t}|\mathbf{x}_t))(1 - F_{b_{m_t}}(b_{(3),t}|\mathbf{x}_t))} \right]$$

which is obtained from the summation of different auctions t due to the fact that the auctions are independent of each other. Since the parent bid distributions are bidder specific, the t dimension of the panel is not long enough to legitimize a non-parametric approach. Hence the author assumes a normal form for the bid distributions such that $f_{b_i}(b_i|x) = N\left[\mu_{b_i}, \sigma\right]$ with $\mu_{b_i} = cons + x\beta - V_i^0$, where $cons$ includes the part of the continuation value of bidder i that is common for everyone, and the individual specific part of the continuation value of i (V_i^0) is assumed to be a function of bidder i's cost c_i. Hence we can approximate the V_i^0 by a polynomial in c_i. However, the plan is to calculate the value of c_i from (B-6) after inferring the unobserved winning bids from the estimated bid distribution, for which we need to know the value of c_i. Therefore we start with a guess about the value of c_i, and this would be the beginning of an iterative process which goes on until convergence of the estimated bid distributions $\widehat{F}_{b_i|\mathbf{x}}$ and the bidding costs c_i.

A non-parametric methodology can be based on Song (2004). Much attention has been given to nonparametric alternatives to the structural models discussed above. Athey and Haile (2007) 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 non-parametrically 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.

In this technique one must assume not only that the second highest bidder bids his true value but also the third highest bidder does, thus we assume explicitly that

$$v_n^j = e^{x'_n \beta} \rho_n^j - c$$

for at least the second and third highest bidders. These bidders' values are denoted $v_n^{(2)}$ and $v_n^{(3)}$ with the corresponding error terms $\rho_n^{(2)}$, and $\rho_n^{(3)}$—note that I may vary by auction and is unknown throughout this analysis. Since we need both the second and third highest bids for estimation, all auctions with two or fewer bidders must be dropped. This methodology 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 details an econometric method to decide "how late" is proper. The interested reader should see Song (2004).

In this methodology we can only use the partial likelihood of $\rho_n^{(2)}$ given $\rho_n^{(3)}$ since the full likelihood requires the unknown number of potential bidders. According to the basic theory of order statistics, the sample likelihood function can be written as:

$$L_N\left(\hat{f}\right) = \frac{1}{N}\sum_{n=1}^{N} \ln \frac{2\left[1-\hat{F}\left(\rho_n^{(2)}\right)\right]\hat{f}\left(\rho_n^{(2)}\right)}{\left[1-\hat{F}\left(\rho_n^{(3)}\right)\right]^2},$$

where

$$\hat{F}(z) = \int_{\underline{v}}^{z} \hat{f}(t)\,dt.$$

Here and below, $\underline{v}$ is the lower bound of bidders' private value. Denote $m = \min\left(v_n^{(2)}\right)$. Note that no information about $F(v)$ for $v < m$ can be observed. If a starting price set by a seller is below m with positive probability, then m is a consistent estimate of $\underline{v}$. In order to estimate the unknown distribution one can employ the method proposed by Coppejans and Gallant (2002) and use the hermite series to approximate the unknown distribution:

$$\widehat{f}(z) = \frac{\left[1 + a_1\left(\frac{z-u}{\sigma}\right) + \cdots + a_k\left(\frac{z-u}{\sigma}\right)^k\right]^2 \phi(z; u, \sigma, \underline{v})}{\int_{\underline{v}}^{\infty}\left[1 + a_1\left(\frac{z-u}{\sigma}\right) + \cdots + a_k\left(\frac{z-u}{\sigma}\right)^k\right]^2 \phi(z; u, \sigma, \underline{v})\,dz}$$

where $\phi(z; u, \sigma, \underline{v})$ is the density of $N(u, \sigma)$ truncated at m .Then an estimator of $\widehat{f}(z)$, denoted as $\widehat{f}_N(z)$, is the maximizer of $L_N\left(\hat{f}\right)$, such that

$$(\widehat{a}_1,...,\widehat{a}_k,\widehat{u},\widehat{\sigma}) = \underset{a_1,...,a_k,u\in R,\sigma>0}{\arg\max} \; L_N(\widehat{f}) = \frac{1}{N}\sum_{n=1}^{N}\ln\frac{2\left[1-\hat{F}\left(\rho_n^{(2)}\right)\right]\hat{f}\left(\rho_n^{(2)}\right)}{\left[1-\hat{F}\left(\rho_n^{(3)}\right)\right]^2}$$

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. One can choose the optimal series length, k^*, using the cross-validation strategy employing the Integrated Squared Error (ISE) criterion (Coppejans and Gallant, 2002).

Another work that makes use of the SNP approach developed by Gallant and Nychka (1987) is by Brendstrup and Paarsch (2006). Brendstrup and Paarsch (2006) work under the assumption of asymmetric valuations and known number of potential bidders. They develop a theoretical model of bidder behaviour at single unit English auctions when valuations of the bidders are assumed to be independent draws from one of the J different classes of distributions. Then, in the light of Theorem 2 in Athey and Haile (2007), they demonstrate that the distribution of the different classes of latent valuations are nonparametrically identified when the winning bid, identity of the winner and the number of potential bidders n is observed. They propose a semi-nonparametric (SNP) estimation strategy. Finally, they extend the analysis to multi-unit auctions.

In order to show the identification of the J different classes of distributions, Brendstrup and Paarsch (2006) start from the probability density function of the second highest order statistics of n independent draws each from one of the J different types of distributions, due to Balakrishnan and Rao (1998):

$$g_{(2:n)}(y|\mathbf{F}) = \frac{1}{(n-2)!}Perm\begin{pmatrix} F_{type(1)}(y) & . \;\; . \;\; . & F_{type(n)}(y) \\ . & . \quad\quad & . \\ . & \quad . \quad & . \\ . & \quad\quad . & . \\ F_{type(1)}(y) & . \;\; . \;\; . & F_{type(n)}(y) \\ f_{type(1)}(y) & . \;\; . \;\; . & f_{type(n)}(y) \\ \left[1-F_{type(1)}(y)\right] & . \;\; . \;\; . & \left[1-F_{type(n)}(y)\right] \end{pmatrix}$$

where $\mathbf{F}$ is the vector of the cumulative distribution functions of the J parent classes of distributions, $F_{type(i)}(y)$ and $f_{type(i)}(y)$ stand for the CDF and the PDF of the i'th bidder's valuation respectively, and $Perm$ is the permanent operator. Then they develop the following system of Pfaffian integral equations:

$$\mathbf{F}^0_{type}(y) = \exp\left\{\mathbf{A}^{-1}\log\left[diag\left(\int_0^y \exp\left\{-\log\left[\boldsymbol{\iota}_n - \mathbf{F}^0_{type}(u)\right]\right\} d\mathbf{G}^0(u)^T\right)\right]\right\}$$

where $\mathbf{F}^0_{type}$ and $d\mathbf{G}^0$ are $(n \times 1)$ column vectors whose ith rows equal $F^0_{type(i)}(y)$ (the true population cumulative distribution function for the class $type(i)$) and $d\mathbf{G}^0_{(2:n)}(y,i)$ (the derivative of the true population cumulative distribution function of the winning bid at an auction won by a bidder whose identity is i), respectively. From Meilijson (1981), this system of Pfaffian integral equations have a unique solution, which in turn leads the authors to conclude that the distributions of the valuations are identified from the winning bids and the identities of the winners.

Having shown the identification of the latent distributions of the valuations, the authors propose an estimator using the SNP estimator developed by Gallant and Nychka (1987) in a way to introduce covariates into the system. They define the draw of bidder i who is in class j in the tth auction as follows:

$$\log V_t^{ij} = \mathbf{x}_t \boldsymbol{\beta}_j + U_t^{ij}$$

They approximate $f_j(u)$ by

$$f_j^{p_T}(u) = \left[\sum_{k=0}^{p_T} \gamma_{jk} H_k(u)\right]^2 \exp\left(-u^2/2\right) + \varepsilon \exp\left(-u^2/2\right)$$

where $H_k(u)$ denotes an Hermite polynomial of order k. Then they implement this finite order approximation into the method of quasi-maximum likelihood , defining the estimator

$$\left\{\widehat{f_{jT}}\right\}_{j=1}^{J} = \arg\max_{f_j \in \mathcal{F}_{jT}} \frac{1}{T}\sum_{t=1}^{T} \log g_{(2:n)}(y_t, i_t|\mathbf{F})$$

where

$$\mathcal{F}_{jT} = \left\{f_{jT} \in \mathcal{F}_j : f_{jT}(u|\boldsymbol{\gamma}_j) = \left[\sum_{k=0}^{p_T} \gamma_{jk} H_k(u)\right]^2 \exp\left(-u^2/2\right) + \varepsilon \exp(-u^2/2), \boldsymbol{\gamma}_j \in \boldsymbol{\Theta}_{jT}\right\}$$

and

$$\boldsymbol{\Theta}_{jT} = \left\{\boldsymbol{\gamma}_j = (\gamma_{j0}, ..., \gamma_{jp_T}) : \int_0^\infty f_{jT}(u|\boldsymbol{\gamma}_j) du = 1\right\}$$

The approximation will converge to the truth by letting p_T increase at a rate that is slower than the rate at which the sample size T increases.

One of the important papers in terms of identification strategy is Adams (2007). In a setting similar to Song (2004), Adams (2007) generalizes the result that the value distribution is identified from the observed auction prices when the number of bidders is known or randomly determined (Athey and Haile, 2007). Although the number of actual bidders in an eBay auction is observable, we cannot say that it is randomly determined, because of the selection bias resulting from the format of the eBay auctions. It can be argued that there is a set of potential bidders which are randomly determined, but for which the number is not known (Song, 2004). Adams (2007) shows that the value distribution is identified when the auction prices are observed and the probability distribution over the number of potential bidders is known, under the assumption that the potential number of bidders is independent from the values of bidders in an auction. Nevertheless, it is not clear if it is possible to determine this distribution, since the existence of potential bidders in an auction can be censored. The author argues and shows in a formal proof, that some auction characteristics that affect the number of bidders but not the distribution of bidder valuations in an auction will allow us approximate this distribution arbitrarily closely under some assumptions.The auction length can be given as an example of such a characteristic: with varying auction length, the probability of having n bidders in an auction varies, but the value distribution does not. Hence, for the estimation of the structural parameters, the length of the auction is included in the function defining the probability of having n people in the auction, but not in the value function of the bidder. The author applies OLS on the price data, and uses the order statistics approach in a maximum likelihood estimation, preferring a log-normal assumption for the distribution of the value. Results of the two estimations show that OLS may substantially overestimate the average value.

Another non-parametric estimation approach comes from Haile and Tamer (2003). As mentiod at the beginning of the current section, the authors have two weak assumptions as to how to interprete the observed bids in an English auction. But it is clear that the distribution of the bidder valuations is not identified having assumed these rules which imply neither (i) a unique distribution of bids given a distribution of valuations nor (ii) a unique distribution of valuations given a distribution of bids. However, they argue that informative bounds on the distribution functions of bidder valuations are possible to identify out of this incomplete model of bidding. The identification of the upper bound for the distribution of the valuations $F(v)$ is provided by the property of the i.i.d random variables that the distribution of the i'th order statistic $F_{i:n}(.)$ is related to the parent distribution $F(.)$ by

$$F(v) = \phi(F_{i:n}(.); i, n)^6 \qquad \text{(B-7)}$$

together with the implication of the first assumption that $b_{i:n} \leq v_{i:n}$ for all i. Therefore

$$F_{i:n}(v) \leq G_{i:n}(v) \qquad \forall i, n, v \qquad \text{(B-8)}$$

where i denotes the bidder, n is the total number of bids, $G_{i:n}(.)$ denotes the distribution of the i'th lowest bid, and $\phi(.; i, n)$ is a strictly increasing differentiable function. Applying the monotone transformation $\phi(.; i, n)$ on both sides of (B-8) implies the identification of the upper bound $F_U(v)$ as the following:

$$F(v) \leq F_U(v) \equiv \min_{n \in \{2,\ldots,\overline{M}\}, i \in \{1,\ldots,n\}} \phi(G_{i:n}(v); i, n) \qquad \text{(B-9)}$$

for all v in the relevant range and higher than the reservation price. Note that there is a different bound on $F(v)$ for each pair of indices (i, n), thus the minimization operator chooses the tightest one among them. For the identification of the lower bound, the following implication of the second assumption (bidders do not allow an opponent to win at a price they are willing to beat) gains importance: The second highest valuation has to be smaller than or equal to the highest bid plus the minimum bid increment:

$$v_{n-1:n} \leq b_{n:n} + \triangle \qquad \text{(B-10)}$$

Obviously this provides an upper bound on the realization of only one order statistic of the valuation at each auction, but (B-7) enables us to construct a lower bound for the latent distribution. From (B-10) we can easily write

$$F_{n-1:n(v)} \geq G^{\triangle}_{n:n}(v) \quad \forall n, v \qquad \text{(B-11)}$$

where $G^{\triangle}_{n:n}(.)$ is the distribution of $B_{n:n} + \triangle$. Then, analogous to the derivation of the upper bound, we apply the monotone transformation to both sides of (B-11) and recall (B-7) to finally get the identification of the lower bound $F_L(v)$:

$$F(v) \geq F_L(v) \equiv \max_n \phi(G^{\triangle}_{n:n}(v); n-1, n) \qquad \text{(B-12)}$$

for all v in the relevant range and greater than the reservation price. Note that whenever $v_{n-1:n} = b_{n:n}$ for some n, the lower bound implied by (B-10) is the true distribution; and whenever $b_{i:n} = v_{i:n}$ for some (i, n), the upper bound obtained from above is the true distribution. Therefore we can say that there is no cost to taking this bound approach

[6] see for example Arnold,Balakrishnan and Nagaraja (1992)

rather than assuming the full structure of the standard model, since only when the data is inconsistant with the standard model we identify bounds on $F(.)$ rather than identify $F(.)$ itself. Consistent non-paarmetric estimators for the upper and lower bounds $F_U(v)$ and $F_L(v)$ are obtained by substituting the relevant empirical distribution functions for their population analogs in (B-9) and (B-12) respectively. The estimators are

$$\begin{aligned}\widehat{F}_U(v) &\equiv \min_{n\in\{2,\ldots,\overline{M}\},i\in\{1,\ldots,n\}} \phi(\widehat{G}_{i:n}(v);i,n),\\ \widehat{F}_L(v) &\equiv \max_{n\in\{2,\ldots,\overline{M}\}} \phi(\widehat{G}^{\triangle}_{n:n}(v);n-1,n)\end{aligned}$$

where

$$\begin{aligned}\widehat{G}_{i:n}(v) &= \frac{1}{T_n}\sum_{t=1}^{T}\mathbf{1}\left[n_t=n, b_{i:n_t}\leq v\right],\\ \widehat{G}^{\triangle}_{n:n}(v) &= \frac{1}{T_n}\sum_{t=1}^{T}\mathbf{1}\left[n_t=n, b_{n_t:n_t}+\triangle_t\leq v\right]\end{aligned}$$

and

$$T_n=\sum_{t=1}^{T}\mathbf{1}\left[n_t=n\right]$$

A negative property of these fairly simple non-parametric estimators is that, although they are uniformly consistent[7], these estimators can be badly biased in small samples. The authors propose a smoothing operation in order to improve the small sample performance of the estimators.

A final non-parametric methodology comes from Nekipelov (2007). In his paper, he explains two types of agressive bidder behavior induced by the multi-auction structure of eBay in a continuous time stochastic auction model with endogenous entry, in which bidder types are differentiated by their initial information regarding the entry process. It is the only continuous time auction model that we are aware of in the literature, though there are more aspects that are specific to this paper. One is the visibility parameter θ of a given auction, which brings a stochastic component into the model: even if the bidder is completely certain about the quality of the object itself, she can be uncertain about the group of potential rivals. Bidders have prior beliefs in the form of probability distributions over the value of visibility, and they can update their beliefs over the course of the auction. The other is that the bidder who arrives in the auction gives a best response to the entire path of price process rather than to actions of particular rival bidders.The model is composed of four key structural functions: the instantaneous demand function (this is the frequency of the poisson

[7] The asymptotic distribution of the estimators and the consistency of the bootstrap bands are shown in Haile and Tamer (2002).

process that specifies the entry rate to an auction) $\lambda(t, p_t, \theta)$, the price jump size function $h(t, p_t, \eta)$ (where η is the bid increment), the distribution of valuations of the bidders $F(v)$, and the distribution of bidders' beliefs about the visibility parameter (the mean and the variance of it) $G_\mu(\mu_0)$ and $G_\sigma(\sigma_0)$. These structural functions of the model are defined to be non-parametrically identified from the data if no two different sets of characteristics of the model produces the same distribution of simulated prices. This definition requires two conditions to hold. First, the observable distribution (this is the joint distribution of timing and the sizes of price jumps for the data set used in the paper) should contain sufficient information about the structural functions of the model. Second, the structural functions of the model can be recovered by using some method of inversion of the observed joint distribution. The author lists the assumptions that allow the unique recovery from the data the set of structural functions of the model. The distribution of bidders' valuation is identified from the distribution of the number of active bidders across auctions given the price, if the bidding function is assumed to be monotone with respect to the valuation in each moment of time and that the beliefs are independent from valuations.After the distribution of the valuations is identified, it is possible to identify the mean beliefs of the bidders by the distance between their observed bidding patterns given their infomation and the pattern computed for an auction with given visibility and given structural functions. Lastly, sorting the bidders according to the relative number of their bids will identify the variance of the beliefs of bidders, as the model predicts that bidders with more diffuse priors bid more frequently. Finally, given the distributions of valuations and beliefs, the path of the second highest bid can be simulated for any given instantaneous demand and price jump size function, and minimizing the distance between the observed price path and the simulated price path gives us the estimate for the parameters of the Poisson process. The multi-step estimation procedure starts with non-parametric estimation of the distribution of the observed price and timing of jump data $f(p_t, N_t, t, \gamma_0)$, which is characterized by the structural parameter γ_0, using a kernel:

$$\widehat{f}(p,t) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{h_p h_t}\sum_{i=1}^{I_k}\kappa\left(\frac{p_i^{(k)}-p}{h_p}\right)\kappa\left(\frac{t_i^{(k)}-t}{h_t}\right)$$

where N_t is the total number of bidders who have entered the auction up to time t, n is the number of observed auctions, k is the index of an auction, I_k is the number of price jumps in the auction k, $\kappa(.)$ is a kernel function and h_p and h_t are bandwith parameters. This density estimate is consistent and asymptotically normal. In the second step the entry of the bidders N_t is simulated given the parameter vector γ, then optimal bidding problem of each bidder is solved to calculate the second highest bid in the auction at any given instant. This simulated price data is named as *the response of the structural model to the data.* Then the same non-parametric estimation procedure is applied to the simulated price data to get the estimated distribution $\widehat{f_\gamma}(p,t)$, which is also a consistent estimate. Then, Kullback-Leibler Information Criterion (KLIC) is used to compare the joint distribution of the observed and the simulated price and time of the price jump data:

$$\widehat{KLIC} = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{I_k}\log\left[\frac{\widehat{f}(p_i^{(k)}, t_i^{(k)})}{\widehat{f_\gamma}(p_i^{(k)}, t_i^{(k)})}\right] \tag{B-13}$$

The idea here is to compare the empirical model with the structural model. Note that minimizing the KLIC is equivalent to maximizing

$$L_n(\gamma) = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{I_k}\log\left[\widehat{f_\gamma}(p_i^{(k)}, t_i^{(k)})\right]$$

The author shows the estimate of the parameter γ obtained by minimizing (B-13) is asymptotically normal. Marcov-Chain Monte Carlo method is used to minimize (B-13) with a substantially reduced computational burden.This also gives an asymptotically normally distributed estimator, though the asymptotic variance of the estimate needs a small correction.

Hong and Nekipelov (2009) develop an efficient local instrumental variable estimation of an empirical auction model. In particular, they derive semiparametric efficiency bounds and efficient estimators for the conditional monotone local instrumental variable model studied in Abadie, Angrist and Imbens (2002). They apply this semiparametrically efficient estimation method to analyze the relation between bid dispersion and early bidding in the dataset which is collected from a natural experiment conducted in Nekipelov (2007). One of the implications of the theoretical model developed in Nekipelov (2007) is that early bidding has competing effects on the dispersion of bids. On the one hand, early bidding deters entry and decreases the bid dispersion. On the other hand, early bidding provides more information and potentially increases both learning and bid dispersion. In an attempt to test which effect is dominant, Hong and Nekipelov (2009) apply the local instrumental variable approach trying to walk around the endogeneity problem that can arise with a simple regression of bid variation on an early bidding indicator because of the correlation of some unobserved characteristics of auction (visibility) with both early jump bidding and bid dispersion. A local instrumental variable is given by an exogenous change of supply in the natural experiment conducted in Nekipelov (2007), who auctioned off additional supply of Robbie Williams' CD on eBay. Such an exogenous increase in the supply weakly increases the incentive for early jump bidding. The set of compliers are defined as the auctions which have no early jump bid prior to the supply increase but have early jump bid after the supply increases. In the first step of the estimation, the basic result in Abadie, Angrist and Imbens (2002) is used to find the initial consistent but inefficient estimate of the parameter vector.In the second step, an estimate of the weighting matrix is formed, and efficient estimates of the parameter vector is obtained. Results of the estimation actually reveal the inefficiency of the first step estimates; although the same parameter values were obtained from the first step and at the end of second step, the standard errors from the second step are much smaller than the ones obtained from the first step estimation. Moreover, the estimates from OLS and 2SLS are compared. They find that the coefficient of the early bid indicator is positive in both, but significantly greater in 2SLS than in OLS : a result consistent with the prediction of omitted variable bias.

The field of econometrics has been providing methods for investigation of bidder behavior in eBay auctions that are well grounded in economic theory. Today, a researcher has a number of commonly used alternative structural methods like maximum likelihood (Adams 2007), non parametric methods (Song, 2004; Brendstrup and Paarsch, 2006; Nekipelov, 2007), simulation based methods (Bajari and Hortacsu, 2003) and bounds estimation of incomplete models (Haile and Tamer, 2003), as well as methods for identifying empirical relationship between various characteristics of online auctions (Lucking-Reiley et al. 2007; Hong and Nekipelov, 2009). As long as one has enough sample size for standard asymptotic results to hold, using non-parametric methods have advantages in terms of robustness to distributional misspecification, although parametric methods are more convenient with small sample sizes and high number of covariates included in the analyses, besides having a higher convergence rate. Another issue that gains importance with non-parametric estimation of structural models is the identification problem. Although we have some standard results about identification under different assumptions (Athey and Haile, 2007), development of new structural models will require new identification results (Nekipelov, 2007).