Estimating Demand from eBay Prices

Christopher P. Adams*, FTC
Email: cadams@ftc.gov

April 28, 2006

Abstract

This paper presents results for identification and estimation of the value distribution from eBay auction prices. The paper presents results for eBay type auctions with independent private values and unobserved participation. It is first shown that the distribution of values is identified from observing the distribution of prices and knowing the distribution of potential bidders. The main result presents conditions for which the distribution of values and the distribution of potential bidders are simultaneously identified. Not surprisingly, the intuition is similar to the standard results for identifying demand from observed equilibrium prices. The estimation method suggested by the identification results is used to estimate the demand for the “C5” Chevrolet Corvette on eBay. The results suggest that a simple OLS model on prices will over estimate the mean value of the item. The order statistics approach also provides much more efficient estimates of the hedonic model.

*Thanks to Patrick Bajari, George Deltas, Luke Froeb, Dan Hosken, Laura Hosken, Marc Rysman, Mike Sandfort, Dave Schmidt, Joel Schrag, Unjy Song, Steve Tenn, Bill Vogt, Abe Wilkelgren and other FTC colleagues for helpful discussions. Thanks to eBay for providing the data. Thanks to Peter Newberry for excellent research assistance. Note that this paper does not necessarily represent the views of the Commission or any individual Commissioners. All remaining errors are our own.
1 Introduction

EBay and eBay type auctions are an economic phenomena. EBay is fast becoming a major distribution channel in everything from Beanie Babies to Humvees (Lucking-Reiley (2000); Cohen (2002)). EBay has also become a rich source of data for economists, other social scientists and even computer scientists (Bajari and Hortacsu (2004); Resnick et al. (2003); Zhang et al. (2002)). Recently, eBay began making data available to the general public through APIs. In order to use data from eBay or eBay type auctions to estimate the demand for cars, MP3 players, computer monitors or new digital cameras, it is necessary to identify the distribution from which each bidder’s value for the item is drawn. This paper presents non-parametric identification results for the case when the only information observed is the auction price and other auction characteristics such as the item description and the auction length. The method suggested by the results is used to estimate the demand for C5 Chevrolet Corvettes.

The first section of the paper describes the basic model of behavior on eBay (due to Song (2005)). The section presents a result due to Song (2005) that in a private value auction every Bayes Nash equilibrium each “potential” bidder in the auction will bid her value at her “last opportunity” to do so, unless she has already done so or unless she has been censored. Athey and Haile (2002) show that the value distribution is identified from the observed auction prices if the number of bidders in the auction is known and randomly determined. Song (2005) points out while the number of actual bidders on eBay is known it is not randomly determined. In particular the current posted price leads to selection bias with the observed bidders having values greater than the current posted price. Song (2005) suggests that there is a set of “potential” bidders which are randomly determined but for which the number is unknown. Song (2005) presents an identification result for the case

---

1 My colleague, Laura Hosken, even bought her wedding dress on eBay!
2 See Bajari and Hortacsu (2003) for analysis of common value eBay auctions. Such auctions are probably a better description of collectible auctions.
3 See Rezende (2002) for a straight forward estimation procedure for this case.
where the price and one other order statistic is observed but the number of potential bidders is unknown.\footnote{Song (2005) shows that if we observe a second order statistic then we can use the Athey and Haile (2002) result to identify the value distribution that is conditional on the value being above the second order statistic, and from that the whole value distribution is identified. For example if we observe the second and third highest order statistics then we can identify the value distribution conditional on the value being above the third highest order statistic as we know the number of bidders with that conditional valuation, ie 2.}

This paper generalizes a result in Athey and Haile (2002) by showing the value distribution is identified from observed prices when the distribution of the potential number of bidders is known. While this may be of some technical interest, censoring in the data may make observation of this distribution difficult. The paper presents conditions for which the distribution of the potential number of bidders and the value distribution are simultaneously identified. The intuition comes from the standard result for estimating the demand side of the market from observing equilibrium prices. In that case it is necessary to have observable characteristics which vary with demand and not supply and observable characteristics which vary with supply and not demand. Similarly, here we need observable characteristics which vary with the number of bidders in the auction but not the value of the item and observable characteristics which vary with the value of the item but not the number of bidders. The trick, as always, is finding such instruments! This basic identification result is generalized to the case where there is observed and unobserved item heterogeneity.

The paper presents a parametric estimator based on the identification results. The paper assumes that the potential number of bidders is a simple linear function of the auction’s length. The estimator uses maximum likelihood techniques to estimate the parameters determining the relationship between the number of bidders and auction length as well as parameters of the value distribution. The value distribution is assumed to have a mean that is linear in observed item characteristics and have a log-normal distribution. The variance is also assumed to be a linear function of observable item characteristics.\footnote{See Giray et al. (2006) for alternative parametric specifications.} This model is estimated on eBay price data for new and used
“C5” Chevrolet Corvettes. Comparing the results from the order statistics model with a standard OLS model the paper shows that the OLS model overestimates the mean value of the item by more than 20% of the average price. The results also highlight the fact that the order statistic approach uses a lot more information about the distribution and so gives much more efficient estimates.

The paper proceeds as follows. Section 2 presents the basic model. Section 3 presents the main non-parametric identification results. Section 4 presents a generalization to auctions with item heterogeneity. Section 5 presents a parametric estimator based on the identification results. Section 6 presents demand estimation results for Chevrolet Corvettes. Section 7 concludes.

2 The Model and Notation

The model and notation closely follow Song (2005). It is a single eBay auction for a single item. There are $N$ “potential” bidders in the auction, with $p_n = \Pr(N = n)$, and $M$ observed bidders. The model is of a symmetric private information auction (Assumption 1), where each bidder knows the probability distribution over the number of bidders in the auction and the distribution from which the bidders draw their values (Assumption 2). Assumption 3 is made for simplicity.\footnote{This assumption is relaxed in the application. The paper presents an estimator that accounts for minimum bid requirements and other selection issues.}

**Assumption 1** Each potential bidder’s valuation $V^i$ is an independent draw from $F(\cdot)$, where $V^i \in [\underline{v}, \bar{v}]$.

**Assumption 2** Each potential bidder knows $p_n$, $F(\cdot)$ and their own value $V^i$.

**Assumption 3** There is no minimum bid and there is no minimum increment.
Finally, the auction lasts for the interval of time \([0, \tau]\) and each bidder is assumed to have a “last opportunity” to bid, although they don’t have to bid at that “last opportunity” (Assumption 4).

**Assumption 4** Each potential bidder \(i\) is assumed to have a “last opportunity” to bid, \(t^i \in [0, \tau]\), which is a random variable, such that the distribution of \(t^i\) is denoted \(G^i(\cdot)\).

The following lemma due to Song (2005) states that bidders will (have) bid their value at their last opportunity if they are not censored.

**Lemma 1** (Song (2005)) Let \(C_t\) be the “cut off” price at time \(t\), where \(C_t = B_{t}^{(M-1:M)}\). Given Assumptions (1 - 4), in every Bayes Nash equilibrium, every bidder whose value for the item is greater than \(C_t^i\) at their last opportunity \((t^i)\), \(B_{t}^i = V^i\) if they have not already done so.

As eBay is a second price auction, the current price or “cut off” price is equal to the current second highest bid. It is straightforward to see that it must be optimal for each bidder to bid her value at by her last opportunity. By assumption she has no chance to bid later than her last opportunity so it is a dominant strategy to bid her value.

In each auction, we assume that the amount of the lowest of the two highest bids, \(B_{\tau}^{(M-1:M)}\) or the price, is observed. Note that from above, the price in an eBay auction equals the value of the potential bidder with the second highest value.

In regards to entry into the auction, that decision is endogenous in that only bidders with a positive expected value of entering will enter the auction. This doesn’t really mean anything as the cost of entry for each bidder is either assumed to be 0 or infinity and is exogenously determined.\(^7\) So the probability distribution over the number of potential bidders \((p_n)\) is determined exogenously. Assumption 5 states that \(p_n\) is independent of the

\(^7\)Think of a bidder logging on to eBay at a particular date and time and either having an auction in which to bid (cost of entry is 0) or not having such an auction (cost of entry is infinity).
value distribution $F(\cdot)$. This assumption contrasts to the entry assumption in Bajari and Hortacsu (2003), who use endogenous entry and a zero-profit condition as part of their identification strategy.

**Assumption 5** Let $p_n$ be independent of $F(\cdot)$.

As we are only going to consider auctions with two or more bidders (potential or observed)$^8$ let $p_n = Pr(N = n|N \geq 2)$. The next lemma generalizes the result that the value distribution is identified when the number of bidders is known (Athey and Haile (2002)). Here the lemma shows that the value is distribution is identified when the probability distribution over the number of bidders is known.

**Lemma 2** Given Assumptions (1 - 5), if $p_n$ is known for all $n \in \{2, 3, \ldots\}$ and $\{V_2\}$ is observed then $F(\cdot)$ is identified.

*Proof.* The proof has 4 steps. Step 1. Given $N = n$ where $n$ is greater than 2, the (large sample) probability of observing $V_2$ is

$$Pr(V_2|N = n) = \frac{n!f(V_2)(1 - F(V_2))F^{n-2}(V_2)}{(n - 2)!}$$

which is the probability that $V_2$ occurs times the probability that it is the 2nd highest of $n$ bids. Given Assumption 5 we have

$$Pr(V_2|N \geq 2) = f(V_2)(1 - F(V_2)) \sum_{n=2}^{\infty} \frac{p_n n! F^{n-2}(V_2)}{(n - 2)!}$$

Step 2. Let $[\underline{v}, \bar{v}]$ be segmented in to $K$ equal disjoint sets such that the union is equal to the original set. Let $v_k = \underline{v} + \frac{k-1}{K}(\bar{v} - \underline{v})$ and

$$f_K(v_k) = \int_{v=v_k}^{v_{k+1}} f(v)dv = F(v_{k+1}) - F(v_k)$$

and for $k > 1$,

$$F_K(v_k) = \sum_{h=1}^{k-1} f_K(v_h)$$

$^8$Note that the highest two potential bidders are never censored in this model.
Define $V_{2k}$ similarly, such that $V_{2k} \in \{v_1, v_2, \ldots, v_K\}$. Note that $v_1 = \underline{v}$. Note further that as $K \to \infty$, $F_K(.) \to F(.)$. Let $x_k$ denote the observed (large sample) probability of $V_{2k}$. So rewriting Equation (2) for the case of $V_{2k}$ and noting that the marginal probability of observing $v_k$ is $f_K(v_k)$ and the cumulative probability of observing $v_k$ is $F_K(v_k)$ we have the following equation.

$$x_k = f_K(V_{2k})(1 - F_K(V_{2k})) \sum_{n=2}^{\infty} \frac{p_n n! F_K^{n-2}(V_{2k})}{(n-2)!} \quad (5)$$

which can be rearranged to give

$$f_K(V_{2k}) = \frac{x_k}{(1 - F_K(V_{2k}))\sum_{n=2}^{\infty} \frac{p_n n! F_K^{n-2}(V_{2k})}{(n-2)!}} \quad (6)$$

Step 3. For the case where $V_{2k} = \underline{v}$ we have that $Pr(\underline{v}|N = 2) = 2p_2 f_K(\underline{v})$ and $Pr(\underline{v}|N > 2) = 0.$\(^9\) Therefore

$$f_K(\underline{v}) = f_K(v_1) = \frac{x_1}{2p_2} \quad (7)$$

Step 4. By Equations (4), (6) and (7), and induction we have that $f_K(.)$ is identified. Letting $K \to \infty$, $F(.)$ is identified. Q.E.D.

Lemma 2 shows that if we know the probability distribution over the number of bidders then the underlying value distribution is identified. This seems like a very useful result as it is much less restrictive than the result in Athey and Haile (2002). Unfortunately, it is not clear that it is possible to determine the distribution over the number of potential bidders given that bidders can have their existence censored. The next section shows that given certain data we can generalize this result to the case where the probability distribution is unknown.

### 3 Identification

The following assumption is critical to the main result of this section. Assume that auctions have some observable and numeric characteristic $t \in T$, where

\(^9\)Note that I’m assuming $F^0(v_k) = 1$ for all $v_k$.\)
$T$ is a countable set with $T$ elements and some observable characteristic $y \in Y$, where $Y$ is a countable set with $Y$ elements. An example of $T$ may be the set of auction lengths or the time of day the auction ends or an auction’s length conditional on the time of day the auction ends.\textsuperscript{10} An example of $Y$ may be the seller’s rating or the brand name of the product. The following assumption restricts $p_n$, $T$ and $Y$ to have certain properties.\textsuperscript{11}

**Assumption 6** Let $p_n$ and $f(\cdot|y)$ have the following properties:

1. $p_n(t|y) = p_n(t|y') = p_n(t)$ for all $y, y' \in Y$ and $t \in T$.
2. $p_n(t) \neq p_n(t')$ for all $t, t' \in T$
3. $f(\cdot|y) \neq f(\cdot|y')$ for all $y, y' \in Y$
4. As $t \to 0$, $p_2(t) \to 1$
5. As $t \to 0$, $p_n(t) \to 0$ for all $n > 2$.

This assumption states that $p_n$ must not vary with the observable characteristic $y$ conditional on $t$. If $y$ is the product’s brand name then it may affect a potential bidder’s value for the item but not their entry decision. This is a fairly strict assumption and suggests some care is needed in choosing $Y$ and $T$. The assumption also states that $p_n$ must vary with $t$ at least for all $t$ in the set $T$ and similarly $f(\cdot|y)$ must vary with $y$. Finally, the assumption states that the probability of there being two potential bidders goes to 1 as $t$ gets small. When considering the reasonableness of this assumption remember that we have restricted ourselves to the set of auctions with 2 or more bidders. It seems reasonable to expect that as the length of the auction gets small the number of potential bidders will also get as small as possible. Also note Assumption 5 and remember that $T$ cannot be such

\textsuperscript{10}eBay auctions last for 3, 5, 7 or 10 days.
\textsuperscript{11}The assumptions (Assumptions (1 - 5)) are assumed to hold for $f(\cdot|y)$ as they do for $f(.)$. 

7
that the value distribution is likely to vary systematically with the number of bidders. The next assumption is a technical requirement on the probability distribution functions $p_n$. This assumption is sufficient for convergence of the Taylor Series.

**Assumption 7** Let $p_n$ have the following properties:

1. $p_n(t) \in C^\infty$ for all $t \in T$.
2. As $m \to \infty$, $p_n^{(m)}(t) \to 0$.

The notation $C^\infty$ refers to the fact that $p_n$ is infinitely continuous (all derivatives are continuous) on $T$. The second part of the assumption states that as the derivative index gets large the derivative gets small. There is no intuition for either condition, they are simply technical requirements for convergence of the Taylor Series approximation used in the proof. The following proposition shows that given these assumptions and the existence of the observed characteristics in $T$ the distribution over the number of bidders is identified and thus the value distribution is identified.

**Proposition 1** Given that Assumptions (1 - 7) hold and $T, Y$ and $\{V_2\}$ are observed, as $T \to \infty$ and $Y \to \infty$, $F(.)$ is identified.

**Proof.** The proof has two steps. Step 1. Following Step (2) of Lemma 2 we can approximated $f$ with $f_K$ and given this approximation we have, rearranging Equation (7) and condition on $y \in Y$,

$$x_1 = 2f_K(v_1|y)p_2$$  \hspace{1cm} (8)

Let $p_2$ be approximated by $p_{2J}(t) = 1 - \sum_{j=1}^{J} t^j q_{2j}$, so we have

$$x_{1|y} = 2f_K(v_1|y)(1 - t_1 q_{21} - ... - t_l q_{2J})$$  \hspace{1cm} (9)

and

$$x_{1my} = 2f_K(v_1|y)(1 - t_m q_{21} - ... - t^l_m q_{2J})$$  \hspace{1cm} (10)
where \( \{t_l, t_m\} \in T \). Dividing Equation (9) by Equation (10) we have

\[
\frac{x_{1ly}}{x_{1my}} = \frac{1 - t_l q_{21} - ... - t_l^j q_{2J}}{1 - t_m q_{21} - ... - t_m^J q_{2J}} \tag{11}
\]

Rearranging

\[
x_{1ly} - x_{1my} = q_{21}(t_m x_{1ly} - t_l x_{1my}) + ... + q_{2J}(t_m^J x_{1ly} - t_l^J x_{1my}) \tag{12}
\]

Let \( T \) be such that there are at least \( J \) separate pairs such that we can write out \( J \) simultaneous linear equations which can be represented by the following linear system.

\[
\begin{bmatrix}
  x_{11y} - x_{12y} \\
  ... \\
  x_{1ly} - x_{1my}
\end{bmatrix} =
\begin{bmatrix}
  t_2 x_{11y} - t_1 x_{12y} & ... & t_2^J x_{11y} - t_1^J x_{12y} \\
  ... \\
  t_m x_{11y} - t_l x_{1my} & ... & t_m^J x_{11y} - t_l^J x_{1my}
\end{bmatrix}
\begin{bmatrix}
  q_{21} \\
  ... \\
  q_{2J}
\end{bmatrix} \tag{13}
\]

or

\[
x_1 = A_1 q_2 \tag{14}
\]

where \( x_1 \) is a \( J \times 1 \) matrix, \( A_1 \) is \( J \times J \) and \( q_2 \) is \( J \times 1 \). As long as \( A_1 \) is full rank (Assumption 6) then

\[
q_2 = A_1^{-1} x_1 \tag{15}
\]

which means that we have identified \( p_{2J}(t) \) and given this approximation

\[
f_{KJ}(v_1|y) = \frac{x_{11y}}{2 p_{2J}(t_1)} \tag{16}
\]

Step 2. From Equation (5)

\[
x_2 = f_K(v_2|y)(1 - F_K(v_2|y)) \sum_{n=2}^{\infty} \frac{p_n n! F_K^{n-2}(v_2|y)}{(n - 2)!} \tag{17}
\]

and from Equations (4), \( F_K(v_2|y) = f_K(v_1|y) \) which can be approximated by \( \frac{x_{11y}}{2 p_{2J}(t_1)} \) (Equation (16)). Let there be an \( N \) such that \( p_n = 0 \) for all \( n > N \). With this approximation we have

\[
x_2 = f_K(v_2|y)(1 - F_K(v_2|y)) \times \left( 2p_2 + 6 p_3 F_K(v_2|y) + ... + p_N \frac{N!}{(N-2)!} F_K^{N-2}(v_2|y) \right) \tag{18}
\]
Let $p_n$ be approximated by $p_n(t) = \sum_{j=1}^{J} t^j q_{nj}$ for all $n$ between 2 and $N$. Given these approximations, for $t_i$ and $t_m$ we have

$$x_{2i,y} = f_K(v_2|y)(1 - F_K(v_2|y))(2p_{2,j}(t_i) + (t_i q_{31} + ... + t_i q_{3, j}) 6F_K(v_2|y) + ... + (t_i q_{N1} + ... + t_i q_{Nj}) \frac{N!}{(N-2)!} F^{N-2}_K(v_2|y))$$

(19)

and

$$x_{2my} = f_K(v_2|y)(1 - F_K(v_2|y))(2p_{2,j}(t_m) + (t_m q_{31} + ... + t_m q_{3, j}) 6F_K(v_2|y) + ... + (t_m q_{N1} + ... + t_m q_{Nj}) \frac{N!}{(N-2)!} F^{N-2}_K(v_2|y))$$

(20)

Dividing Equation (19) by Equation (20) and rearranging

$$2x_{2i,y}p_{2,j}(t_m) - 2x_{2my}p_{2,j}(t_i) = q_{31}(t_i x_{2m} - t_m x_{2i}) 6F_K(v_1|y) + ... + q_{Nj}(t_i^j x_{2m} - t_m^j x_{2i}) \frac{N!}{(N-2)!} F^{N-2}_K(v_1|y)$$

(21)

Similar to Step (1) we can represent this as a set of simultaneous equations if $T$ has at least $J$ separate pairs and $Y$ has at least $N - 2$ elements, where

$$x_2 = A_2 q$$

(22)

and $A_2$ is

$$\begin{bmatrix}
(t_2 x_{21y} - t_1 x_{22y}) 6F_K(v_2|y) & ... & (t_2^j x_{21y} - t_1^j x_{22y}) N(N-1) F^{N-2}_K(v_2|y) \\
... & ... & ... \\
(t_m x_{21y} - t_1 x_{22y}) 6F_K(v_2|y') & ... & (t_m^j x_{21y} - t_1^j x_{22y}) N(N-1) F^{N-2}_K(v_2|y')
\end{bmatrix}$$

(23)

and $x_2$ is a $(N-2)J \times 1$ matrix, $A_2$ is $(N-2)J \times (N-2)J$ and $q$ is $(N-2)J \times 1$. If $A_2$ has full rank (Assumptions 6) then

$$q = A_2^{-1} x_2$$

(24)

and so letting $K \to \infty$, $q_{nj}$ is identified for $n$ between 3 and $N$ and all $j$ between 1 and $J$. Letting $T \to \infty$, $q_{nj}$ is identified for any arbitrary $j$ where $j$ is strictly greater than 0. Letting $Y \to \infty$, $q_{nj}$ is identified for any arbitrary $n$ greater or equal to 2. By Theorem 9.28\textsuperscript{12} (Apostol (1974)) given

\textsuperscript{12}This theorem presents sufficient conditions for Taylor’s Series to converge.
Assumption 7 and as $p_n(t)$ is bounded by 0 and 1 then

$$p_2(t) = 1 - \sum_{j=0}^{\infty} t^j q_{2j}$$

(25)

and

$$p_n(t) = \sum_{j=0}^{\infty} t^j q_{nj} \forall n > 2$$

(26)

and by Assumption 6, $q_{n0} = 0$. By Lemma 2, $F(.)$ is identified. Q.E.D.

The proposition states that as the number of states ($t$ and $y$) get arbitrarily large then the probability distribution over the number of bidders can be approximated arbitrarily closely and thus (from Lemma 2) the value distribution is identified. The proof is based on the intuition that varying auction length, the probability of that there are two bidders (or $N$ bidders) varies but the value distribution does not. So any observed changes in price is coming from changes in the number of bidders. As the distribution is non-parametric the richer the variation the more detailed the description of the distribution. Unfortunately, auction length (at least on eBay) is not continuous and in fact only has 4 states. However if we also believe that the number of bidders varies exogenously over the time of day that the auction ends then we can condition auction length on the time of day and generate a much larger number of states.\textsuperscript{13} We also need the value distribution to vary over some other observable characteristic of the auction. Without this the matrix $A_2$ is not of full rank and any $p_n$ greater than 2 is not identified. The researcher needs to be careful to choose a set of observable states for which the number of bidders varies exogenously and the probability distribution doesn’t vary with $y$.

4 Auction Heterogeneity

Traditionally the demand for differentiated products is estimated by assuming that product choices can be mapped into observed and unobserved prod-

\textsuperscript{13}eBay provides the time that the auction ends to the second.
uct characteristics and using hedonic regression models (Berry et al. (1995)). The major argument for doing this is that there are often so many different products that it is not possible to identify the demand for each product without using information about the demand for similar products. As discussed below part of the identification strategy in this literature is to use this variation in product characteristics. The following results suggest that a similar model can be used with eBay type data.

**Assumption 8** Let $V^j_i$ be distributed $F(., X^j)$ where $X^j$ is a $J$ dimensional vector of observed item characteristics.

Assumption 8 states that the value distribution is some general function of the set of observed characteristics of the item $X^j$. An example would be a random coefficients model (Berry et al. (1995); Nevo (2000)). Note that in Berry et al. (1995) and more recently Bajari and Benkard (2006) there is a unobserved component of the product characteristics, this issue is discussed below.

**Corollary 1** If Assumptions (1 - 7) and Assumption 8 hold, then if $\{V^j_2\}$, $T$ and $X^j$ are observed for all items $j$, then $F(., X^j)$ is identified.

**Proof.** By Proposition 1, given $X^j$, $F(.|X^j)$ is identified. Therefore, for all $X^j$, $F(., X^j)$ is identified. Q.E.D.

Corollary 1 shows that it is straight forward to generalize Proposition 1 to a hedonic model. Note, that for simplicity it is assumed that the distribution of bidders is not a function of the item characteristics. It seems reasonable that this assumption can be relaxed although one needs to remain aware of exclusion restrictions necessary for identification. The rest of the section considers a model with unobserved item heterogeneity.

**Assumption 9** Let $V^j_i = v^j_i + \xi_j$ where $v^j_i$ is distributed $F(., X^j)$ and $\xi_j \in [\underline{\xi}, \overline{\xi}]$, $X^j$ is a $J$ dimensional vector of observable characteristics of the item, and $\xi_j$ is a characteristic of the item observed by the bidder and unobserved by the researcher.

12
Note that $\xi_j$ is constant across auctions for the same item. That is it represents some unobserved characteristic about the car or MP3 player rather than some unobserved characteristic that is auction specific. An important simplification in Assumption 9 is that unobserved item heterogeneity enters the value function additively. This assumption makes identification straightforward, but one may be concerned that it is unnecessarily simple (Bajari and Benkard (2006)).

**Proposition 2** If Assumptions (1 - 7) and Assumption 9 hold, then if $\{V_2\}$, $T$ and $X_j$ are observed for all items $j$, then $F(., X_j)$ and $\xi_j$ are identified.

**Proof.** Step 1. Let $G(., X_j)$ be the distribution of $v_{ij} + \xi_j$. For a given item $j$ by Corollary 1, $G_j(., X_j)$ is identified. Step 2. Let there be two items $j$ and $k$ such that $X_j = X_k$, then for some $a > b$ such that $G_j(a, X_j) = G_k(b, X_j)$, $a - b = \xi_j - \xi_k$. When this difference is equal to $\bar{\xi} - \bar{\xi}$, $\xi_j = \bar{\xi}$ and $\xi_k = \bar{\xi}$, and so $F(., X_j)$ and $\xi_j$ are identified. Q.E.D.

Proposition 2 shows that Corollary 1 can be generalized to the case of unobserved item heterogeneity. Identification comes from comparing the distributions for items with similar observed characteristics.

### 5 Estimation Model

This section presents a parametric maximum likelihood estimator based on the results presented in the previous section. The main distributional assumption is that the error structure on the item’s value is log-normal, although both the mean and variance are allowed to vary with observed item characteristics. The number of potential bidders is assumed to be a linear function of auction length. The structural assumption is that the bids of the two highest bidders are equal to each bidder’s value for the item. Further, Adams (2004) suggests that it is necessary to observe individual bidders across auctions in order to account for unobserved auction heterogeneity.
for bidders that bid below the final price, their value for the item is less than
the final price.

Let bidder $i$’s value function for item $j$ be represented by the following
function.

$$\ln(v_{ij}) = \beta \ln(X_j) + \epsilon_{ij}$$

(27)

where $v_{ij}$ is $i$’s value for item $j$, $X_j$ is a set of observable characteristics of item
$j$, and $\epsilon_{ij}$ are unobservable characteristics of the bidder and the item. Let $\epsilon_{ij}$
be distributed $N(0, \gamma Z_j \sigma)$, where $Z_j$ is a set of observable characteristics of
item $j$. Let the distribution of the number of potential bidders be given by
the following function.

$$\hat{N} = 2 + \tau \text{Auction Length}$$

(28)

That is, the probability $N = \hat{N}$ is one if $\hat{N} = 2 + \tau \text{Auction Length}$ and zero
otherwise.

Given these distributional and structural assumptions. The likelihood
function for observing a set of prices in $K$ auctions is

$$\ln L = \sum_{k=1}^{K} (-0.5 \ln(\hat{\sigma}) + \ln(\phi(z_k)) + \ln(1 - \Phi(z_k)) + \ln(\hat{N}(\hat{N} - 1)\Phi(z_k)^{\hat{N} - 2})
+ \ln(1 - \hat{N}!(\Phi(z_{ck})^{\hat{N}} + \Phi(z_{ck})^{\hat{N} - 1})))$$

(29)

where $\hat{\sigma} = \gamma Z_j \sigma$, $z_k = \frac{V_2 - \beta \ln(X_j)}{\hat{\sigma}}$ and $z_{ck} = \frac{S - \beta \ln(X_j)}{\hat{\sigma}}$, $V_2$ is the auction
ending price and $S$ is the auction starting price.

This likelihood function is derived from the probability distribution of
prices given $N$ bidders (see Equation (1)).

$$Pr(V_2|N) = f(V_2)(1 - F(V_2))\frac{N!F^{N-2}(V_2)}{(N - 2)!}$$

(30)

This distribution is then multiplied by the probability of observing a price
above the starting price (or minimum bid) of the auction (the last term). We
also replace $N$ with $\hat{N}$ which is simultaneously estimated (Equation (28)).

As suggested above, identification relies on exclusion restrictions. The
number of bidders is assumed to be a function of the observed auction length
but not a function of other item characteristics. The mean and the variance
of the distribution are assumed to be a function of item characteristics but not a function of auction length. It is also assumed that there is an item characteristic that affects the variance of the distribution but not the mean of the distribution. The first set of restrictions seem intuitive given the identification results presented above. The last restriction is not obvious but seems to be necessary.\footnote{While not presented, montecarlo analysis was conducted on the model. In this analysis, when the second exclusion restriction was not made, $\sigma$ and $\tau$ were poorly estimated.}

6 Estimation Results

The model presented above is estimated on a subset of cars sold in eBay auctions in 2003. The cars are new and used “C5” (or version 5) Chevrolet Corvettes. The ‘Vette data is described in detail in Adams et al. (2006). For the purposes of this analysis, the data includes the auction price, the starting price, auction length, auction month, and car characteristics such as color, condition and mileage. The data includes all C5 ‘Vettes sold at auction in 2003 in which the final price was above the minimum price or the reserve price.

The data includes only “successful” auctions which are auctions in which there was at least one bidder above the starting price or the reserve price (if there was one). I further selected the data to include just auctions where there were at least two bidders above the starting price or the reserve price. Auctions in which the car sold with the Buy-It-Now option were also dropped. The model estimated below accounts for selection of auctions above the starting price (or reserve price) but does not account for the use of the Buy-It-Now option.

Table 1 presents results from OLS on the price data and results from the Order Statistic approach with an instrument for the number of bidders. The columns headed with $\beta$ are the point estimates for the model parameters associated with the observed characteristics. The standard errors are represented by the 5th and 95th percentiles for the estimated parameter. Note
that the order statistic percentiles are from bootstrapping the estimate (100 reps).

The table highlights two important differences with estimates from the two approaches. First, the order statistics approach gives an estimate for the mean of the value distribution that is substantially lower than the OLS estimate suggests. In this case the difference from the point estimate \( (e^{10.95} - e^{10.83}) \) is about $6,500, which is more than 20% of the average price of the cars in the data. Note however, the 90% confidence intervals around both point estimates overlap. Second, the order statistic approach provides a substantially more efficient estimate. Efficiency comes from the fact that an order statistic provides a lot of information about the distribution. Not only does it give an estimate of a single point, but it also provides information about the weight of the distribution lying both above and below that point. Another difference to note is that “Age” is used in the order statistics approach as a variable that affects the variance but not the mean. In the estimated model \( \sigma_j = 0.976\text{Age} \), where Age is measured in years (these cars are between 0 and 5 years old). Finally, the number of bidders is found to increase linearly at one new bidder per extra day of the auction. That is, \( \hat{N} = 2 + 1.003\text{AuctionLength} \).

The results suggest that the simple OLS model captures most of the change in in the mean value due to observable characteristics. However, the OLS model provides a biased estimate of the overall mean value of the item. The model presented above may provide a method for reducing the bias and increasing the efficiency of the estimate. Note however, the order statistics model provides relatively imprecise estimates for the parameters on the constant, the variance and the estimated number of bidders. These imprecise estimates suggest that the model is misspecified. In particular, the arbitrariness of choosing “Age” to vary with the variance but not the mean is problematic. Still, these results provide some flavor for how to estimate demand when only auction prices and auction characteristics are observed.
<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>Order IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 pctl</td>
<td>95 pctl</td>
</tr>
<tr>
<td>Constant</td>
<td>10.951</td>
<td>10.771</td>
</tr>
<tr>
<td>Log Mileage</td>
<td>0.030</td>
<td>-0.050</td>
</tr>
<tr>
<td>Age</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Log Age</td>
<td>-0.367</td>
<td>-0.428</td>
</tr>
<tr>
<td>New</td>
<td>-0.455</td>
<td>-0.590</td>
</tr>
<tr>
<td>Automatic</td>
<td>-0.060</td>
<td>-0.103</td>
</tr>
<tr>
<td>Convertible</td>
<td>0.142</td>
<td>0.099</td>
</tr>
<tr>
<td>Black</td>
<td>-0.109</td>
<td>-0.165</td>
</tr>
<tr>
<td>Burgandy</td>
<td>-0.263</td>
<td>-0.377</td>
</tr>
<tr>
<td>Red</td>
<td>-0.041</td>
<td>-0.100</td>
</tr>
<tr>
<td>Silver</td>
<td>-0.055</td>
<td>-0.118</td>
</tr>
<tr>
<td>White</td>
<td>-0.038</td>
<td>-0.114</td>
</tr>
<tr>
<td>Negative Feedback</td>
<td>-0.063</td>
<td>-0.168</td>
</tr>
<tr>
<td>Log Feedback</td>
<td>-0.010</td>
<td>-0.021</td>
</tr>
<tr>
<td>February</td>
<td>-0.013</td>
<td>-0.133</td>
</tr>
<tr>
<td>March</td>
<td>0.156</td>
<td>0.043</td>
</tr>
<tr>
<td>April</td>
<td>0.160</td>
<td>0.054</td>
</tr>
<tr>
<td>May</td>
<td>0.133</td>
<td>0.031</td>
</tr>
<tr>
<td>June</td>
<td>0.112</td>
<td>0.006</td>
</tr>
<tr>
<td>July</td>
<td>0.212</td>
<td>0.110</td>
</tr>
<tr>
<td>August</td>
<td>0.157</td>
<td>0.051</td>
</tr>
<tr>
<td>September</td>
<td>0.160</td>
<td>0.064</td>
</tr>
<tr>
<td>October</td>
<td>0.149</td>
<td>0.045</td>
</tr>
<tr>
<td>November</td>
<td>0.072</td>
<td>-0.034</td>
</tr>
<tr>
<td>December</td>
<td>0.055</td>
<td>-0.054</td>
</tr>
<tr>
<td>Auction Length</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Number of Auctions: 768

Table 1: “C5” Chevrolet Corvette Demand Estimates
7 Conclusion

This paper develops on ideas presented in Song (2005) and Athey and Haile (2002), and suggests an alternative method for identifying demand in single eBay auctions. Athey and Haile (2002) shows that in certain auctions demand can be identified from observing the price and the number of bidders. Unfortunately, in general eBay auctions it is not possible to observe the number of bidders. Above I show that the value distribution is identified if the probability distribution over the number of bidders is known. However, given censoring in eBay type auctions it is not clear that it is possible identify that distribution. Song (2005) shows that for a certain set of eBay auctions it is possible to identify demand even when the number of bidders is unknown if the distribution of two order statistics are observed. However, in many cases it is not possible to observe two order statistics in eBay auctions. The paper presents the necessary conditions to simultaneously identify both the value distribution and the distribution over the number of potential bidders. The intuition is similar to solving the “simultaneity problem.” These distributions are identified if there are observable characteristics of the auctions that vary with one (the value distribution) but not the other (the distribution of potential bidders) and vice versa. I show that the basic identification results can be generalized to the case where there is observed and unobserved item characteristics.

A parametric estimator based on the identification results is estimated on “C5” Chevrolet Corvettes. Comparing the OLS results to the results from the order statistics approach, the latter provides for a much lower estimate of the mean value of the car and much more precise estimates of the hedonic model.

References


