Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders*

Unjy Song†

University of British Columbia

June 2004

Abstract

In this paper, I present new identification results and proposes an estimation method for an eBay auction model with an application. A key difficulty with data from eBay auctions is the fact that the number of potential bidders willing to pay the reserve price is not observable and the number of potential bidders varies auction by auction. While this precludes application of existing estimation methods, I show that this need not preclude structural analysis of the available bid data. In particular, I show that within the symmetric independent private values (IPV) model, observation of any two valuations of which rankings from the top is known (for example, the second- and third-highest valuations) nonparametrically identifies the bidders' underlying value distribution. In contrast to the results of previous studies, the researcher does not need to know the number of potential bidders willing to pay the reserve price nor assume that the number of potential bidders is fixed across auctions. I then propose a consistent estimator using the semi-nonparametric maximum likelihood estimation method developed by Gallant and his coauthors. Several Monte Carlo experiments are conducted to illustrate its performance. The simulation results show that the proposed estimator performs well. I apply the proposed method to university yearbook sales on eBay. Using my estimate of bidders' value distribution, I explore the effects of sellers' ratings on bidders' value distribution; compute consumers' surplus; and examine a regularity assumption that is often made in the mechanism design literature.

*I am indebted to Phil Haile for his warm encouragement and detailed comments. I am also grateful to Yeon-Koo Che, Bruce Hansen, Yuichi Kitamura, and Myunghwan Seo for their helpful comments and conversations, and thank to Salih Burak Gokturk, Ali Hortaçsu, and Pai-Ling Yin for their guidance in Perl programming. The financial support from the Wisconsin Alumni Research Foundation and the Social Science Research Council is appreciated.

† usong@interchange.ubc.ca, faculty.arts.ubc.ca/usong

1 Introduction

This paper develops and applies a new method for structural analysis of Internet auction data.[1] The rising popularity of Internet auctions has generated widespread interest in interpreting online bidding data. Nevertheless, structural analysis of Internet auction data has been severely limited due to a lack of adequate econometric methods. A key difficulty with data from Internet auctions is the fact that the number of potential bidders is unknown. Prior research has suggested solutions for the case in which the number of potential bidders is unavailable, but all of the studies require assumptions which is not plausible in Internet auction circumstances. Bajari and Hortaçsu (2002a), and Paarsch (1997) do not require knowledge of the number of potential bidders, but they assume that the observed bidders are all potential bidders willing to pay the reserve price. However that assumption is not plausible in Internet auctions, because Internet auctions use an ascending-bid format. In Internet auctions, the standing price starts at the reserve price (starting price) and rises as a new bid is submitted. Consequently one can never know of the number of bidders who intended to take part in auctions, but visited auction sites only to find that the standing prices were already raised by their competitors over their own willingness-to-pay. Although Laffont, Ossard, and Vuong (1995) do not assume that the observed bidders are all potential bidders willing to pay the reserve price, they instead assume that the number of potential bidders is the same across all auctions under consideration. Furthermore all of three papers make a parametric distributional assumption regarding the bidders' value distribution. Nonparametric approaches have been proposed by several recent papers, for instance, Athey and Haile (2002), Guerre, Perrigne, and Vuong (2000), and Haile and Tamer (2002). They all need knowledge of the number of potential bidders willing to pay the reserve price.

In order to overcome this difficulty with Internet auction data, I provide new identification results and an estimation strategy that can be used to recover the bidders' underlying value distribution without knowledge nor an assumption regarding the number of potential bidders. I show that the bidders' underlying value distribution is nonparametrically identified from observations of any two valuations of which rankings from the top are known (for example, the second- and third-highest valuations) within the standard symmetric independent private val-

[1] Actually, I focus only on eBay auctions. eBay is the dominant Internet auction site, and most auction sites have copied the eBay auction format.

ues (IPV) framework.[2] Within the eBay bidding model which I propose, the joint distribution of the second and third-highest valuations are identified from available eBay bidding data. I then propose a consistent estimator, using the semi-nonparametric maximum likelihood (SNP) estimation method developed by Gallant and his coauthors, and evaluate it by Monte Carlo experiments.[3]

The key insight into the identification result is revealed by Lemma 1 which shows that the distribution of two order statistics from an independent and identically distributed (i.i.d) sample uniquely determines the parent distribution, even if the size of each sample is unknown and varies sample by sample. To see the intuition, consider the special case in which we observe the joint distribution of the second-highest order statistic, Y, and the third-highest order statistic, X from an i.i.d. sample which is drawn from an absolutely continuous distribution $F(\cdot)$. Let Z denote the corresponding highest order statistic, which is not observed. If we condition on $X = x$, (Z, Y) is an ordered random sample of size *two* from the parent distribution, $[F(\cdot) - F(x)] \ / \ [1 - F(x)]$, whatever the size of the sample that (Z, Y, X) comes from. Hence, regardless of the sample size, the distribution of Y conditional on $X = x$ is the same as the distribution of the lower order statistic from a sample of size two from the distribution $[F(\cdot) - F(x)] \ / \ [1 - F(x)]$. The distribution of a single order statistic from an i.i.d. sample of a *known* sample size identifies the parent distribution. Hence, the truncated distribution $[F(\cdot) - F(x)] \ / \ [1 - F(x)]$ is identified for all x on the support of $F(\cdot)$. Letting x be the lower limit of the support of $F(\cdot)$ then gives the result.

Recently, research on Internet auctions become more and more important, because these auctions have become such a popular exchange mechanism. According to eBay, on any given day, more than 16 million items are listed on the site across 27,000 categories. In 2002, eBay members transacted \$14.87 billion.[4] In addition to the huge increase in transaction volume, professional eBay sellers have increased. Even big companies such as IBM and Xerox sell their products through eBay. Among the 923 auctions in my dataset, the seller has more than 1,000 ratings at 334 auctions. One seller has 18,271 ratings; this means that she has made at least 18,271

[2] For a definition of nonparametric identification, see Roehrig (1988).

[3] For the SNP estimator, see Gallant and Nychka (1987), Fenton and Gallant (1996 a,b), and Coppejans and Gallant (2002).

[4] http://pages.ebay.com/community/aboutebay/overview/index.html. For a broad survey of Internet auctions, see Lucking-Reiley (2000) or Bajari and Hortaçsu (2002b)

transactions through eBay. All of these imply that eBay is no longer an amateurs' casual market or a nationwide garage sale; it has become an alternative market through which many sellers make their living. Accordingly, we need to study this new enlarging market and this paper takes the first step toward an analysis of the issues involved with demand structure such as taxation or monopoly issues by providing a method to recover bidders' value distribution.[5] Moreover, in application I actually study the effect of sellers' ratings on bidders' value distribution; to compute consumers' surplus; and to check whether the assumption that the virtual value is increasing in valuation is satisfied.

Analysis of Internet auctions is important also because these auctions provide an extremely large amount of data and thus a strong basis for empirical study. A growing number of researchers use eBay auction data to study various economic issues.[6] The estimation strategy developed in this paper will be useful to future studies that require inference from eBay auction data.

Besides Internet auctions, there are many auctions in which the number of potential bidders willing to pay the reserve price is not available. First of all, in ascending auctions, even a bidder with valuation above a reserve price may not make a bid. Hence in order to know the number of potential bidders with valuations above a reserve price, complementary data is necessary in addition to each bidder's bid record. Such a data is not always available, for example British Colombian timber auctions (Paarsch, 1997). The free-structure of bidding in ascending auctions can lead to a debate regarding whether we can obtain two precise valuations with their top-down rankings, which are needed in my method of estimation. However, my method still improve the existing methods which assume that observed bidders are all potential bidders with valuations over the reserve price. Even in first-price sealed-bid auctions, the number of potential bidders willing to pay the reserve price may not be observed if there are bid participation costs. For example, in the participation model considered in Samuelson (1985), some of potential bidders with valuations above the reserve price do not take part in auctions. Song(2004) extends the results of this paper to a first-price auction model and so the method in this paper is useful to

[5] Sales tax is currently applied only to in-state transactions; how to impose sales tax on online transactions has been a debated issue. Monopoly issues may be raised because eBay seems to be the dominant auction site, even though it imposes relatively high posting fees.

[6] See Cabral and Hortaçsu (2003), Houser and Wooders (2001), Resnick and Zeckhauser (2002), Engelbrecht-Wiggans, List, and Lucking-Reiley (1999), Hossain and Morgan (2003), Yin (2002), etc.

structural analysis of first-price auctions in which the number of potential bidders willing to pay the reserve price is unknown.

The rest of the paper is organized as follows: The next section explains the eBay auction format, and Section 3 presents an eBay bidding model. In Section 4, the most crucial part of this paper, I provide the econometric model and its identification result. Section 5 proposes a consistent estimator, and then Section 6 evaluates it through Monte Carlo experiments. I apply the proposed method to a sample of eBay auctions in Section 7 and make concluding remarks in Section 8.

2 Background: eBay Auctions

This section briefly explains the eBay auction mechanism. Bajari and Hortaçsu (2002a,b) and Bryan et al. (2000) offer the interested reader richer descriptions. I consider only auctions in which a single item is sold. I exclude secret reserve price auctions as well as auctions ended by a bidder's use of a "buy it now" option. An eBay auction starts as soon as a seller registers it. An eBay seller has several options when she lists her item. She can set a starting price and also choose the time length of her auctions: three, five, seven or ten days. Potential buyers can find auctions of interest by browsing the categorized auction listings or by using a search engine. No advance announcement of an auction exists, therefore there is no reason to expect all potential bidders to become aware of the auction at the same time. Below, I will explicitly model the stochastic arrival times of bidders during an auction.

All auctions proceed according to the rules pre-announced by eBay. All eBay auctions use an open, ascending-bid format that is different from a more traditional auction's ascending bid format in two respects. First, there is a fixed ending time instead of a "going-going-gone" ending rule. Second, eBay uses the proxy bidding system. A bidder is asked to submit a *cutoff price*, a maximum bid, instead of his instant bid amount. The proxy bidding system then will issue a proxy bid equal only to the minimum increment over the next highest bid. If a competitor's bid is greater than the bidder's cutoff price, the proxy bidding system will not issue a bid for the bidder.

The maximum proxy bid is posted as the *standing price* next to a current winner's identity. For example, consider an auction in which a seller starts the bidding at \$5, and the *first-arrived* bidder submits \$25 as his cutoff price. The proxy server issues a proxy bid of \$5 on the first-arrived bidder's behalf and posts \$5 as the standing price. Suppose another bidder arrives and submits a cutoff price of \$20. The proxy server then bids \$20 plus the minimum increment for the first-arrived bidder and displays it as the standing price. As a result of the proxy bidding system, the standing price is the second-highest existing cutoff price plus the minimum increment. During the course of the auctions, if the cutoff price submitted by a new bidder is not high enough to lead the auction, or the current auction leader is outbid, eBay notifies the bidder via e-mail so that he may revise his cutoff price if he so desires. A bidder may keep or increase his previous cutoff price at any time, but may not decrease it.

Once an auction has concluded, the winner is notified by e-mail and pays the standing price posted at the closing time. Thus, a winner pays the second-highest bid plus the minimum increment. As soon as an auction ends, eBay discloses all bidders' cutoff prices except the highest one. (During the auction, only the bidders' identities and bidding times are available.) If a bidder submitted cutoff prices multiple times, every cutoff price and its corresponding bidding time is shown. All auction listings and their results remain publicly available on eBay for at least one month after the auction closes.

3 A Model of Bidding in eBay Auctions

3.1 The Model

Consider an eBay auction of a single object. The number of potential bidders, N, is a random variable, with $p_n = \Pr(N = n)$. Each potential bidder's valuation V^i is an independent draw from the absolutely continuous distribution $F(\cdot)$, having support on $[\underline{v}, \overline{v}]$. Each bidder knows only his valuation, the distribution $F(\cdot)$, and the probabilities p_n. The analysis would be identical were bidders to know the realization of N, as is usually assumed in the literature. For the sake of simplicity, I ignore the minimum increment.[7]

The auction is conducted over an interval of time $[0, \tau]$. In practice, this is 3, 5, 7 or 10 days, typically 7. I assume that bidders do not monitor the auction continuously. Rather, each

[7] The amount of the minimum increment is predetermined and posted by eBay. The minimum increment is, for instance, \$0.50 when the standing price is \$5.00 - \$24.99. Such a small, minimum increment seems unlikely to affect bidders' bidding behaviors significantly.

bidder has a finite set of monitoring opportunities $T^i = \{t_1^i, ..., t_{e^i}^i\}$ ($t_j^i \in [0, \tau]$). The number of monitoring opportunities and the time of each monitoring opportunity are random variables, the realizations of which are known at time t_1^i to bidder i but are unknown to i's opponents.[8] The number of monitoring opportunities, e^i, is drawn from a distribution defined on $\{1, 2, ..., \overline{e}\}$. As many monitoring times as e^i are drawn independently from a continuous distribution with support $[0, \tau]$. Without loss of generality, we order the monitoring times such that $t_1^i < ... < t_{e^i}^i$. At each monitoring opportunity in T^i, bidder i sees the standing price and is free to submit a new cutoff price.[9] At time 0, the standing price, denoted by h_t ($t \in [0, \tau]$), is initialized at the starting price set by the seller. As the auction proceeds, h_t is raised to the value of the second-highest cutoff price transmitted prior to t. If the number of existing bidders is less than two, h_t stays at the starting price. The auction ends at time τ, with the highest bidder declared the winner at a price equal to the second-highest cutoff price.

I consider a strategy for bidder i that specifies the cutoff price he will submit at each time $t \in T^i$ as a function of his valuation, the history of the standing prices that he has monitored, and his most recent cutoff price.[10] The history of the standing prices at i's monitoring opportunities is denoted by a set $H_{t_k^i}^i = \{h_0, h_{t_1^i}, ..., h_{t_k^i}\}(k = 1, ..., e^i)$. C_t^i represents bidder i's most recent cutoff price at time t, and $C_t^i = 0$ indicates that bidder i has not transmitted a cutoff price by time t. Obviously $C_t^i = 0$ for $t < t_1^i$. Those who submit a cutoff price are called *actual bidders.* Potential bidder i's strategy can be described by functions $s_{t_k^i}(v^i | H_{t_k^i}, c_{t_{(k-1)}}^i)$ $(k = 1, ..., e^i)$ which specify his cutoff price at each time t_k^i.

Theorem 1.1 provides a characterization of all symmetric Bayesian-Nash equilibria of this game and show that all satisfy two key conditions: (a) No bidder ever submits a cutoff price greater than his valuation, and (b) At his final submission time ($t_{e^i}^i$), bidder i submits a cutoff

[8] One might consider a model in which T^i were endogenous rather than exogenous, as assumed here. I model T^i as exogenous, because T^i is likely to be more influenced by bidder i's personal schedule than by the proceedings of the auction. Further, since the model allows bidder i to choose his actual bidding time(s) from the set T^i, little may be lost by making this simplifying assumption.

[9] An eBay bidder can submit a cutoff price as many times as he wants. I will call only the finally submitted cutoff price his "bid."

[10] I do not include previous bidders' identities in bidding functions. eBay does not post each bidder's bid amount during auctions, so previous bidders' identities play the same role as the number of previous bidders in symmetric equilibria, on which I focus. Including the number of previous bidders in bidding functions would have no effect on Theorem 1.

price equal to his valuation if he has not yet done so, and his valuation is greater than the standing price at that time ($h_{t^i_{e^i}}$). Note that many patterns of bidding are possible in equilibrium. For example, bidder i may submit a cutoff price equal to his valuation v^i immediately after he finds the auction, he may postpone his submission until time $t^i_{e^i}$, or he may submit a cutoff price lower than v^i and update his cutoff price over time.

Theorem 1 *The strategies* $S^o = (s^o_{t^i_1}(v^i|H_{t^i_1}, c_{t^i_0}), ...s^o_{t^i_{e^i}}(v^i|H_{t^i_{e^i}}, c_{t^i_{e^i-1}}))$ *constitute a symmetric Bayesian-Nash equilibrium if and only if they induce:*

(a) $c^i_t \leq v^i \quad \forall\, t$ *and* *(b)* $c^i_{t^i_{e^i}} = v^i$ *if* $v^i > h_{t^i_{e^i}}$.

Proof. (Sufficiency) Let $m = \arg\max_k v^k$. Suppose all bidders other than i use a strategy inducing (a) and (b). Regardless of $i's$ strategy, the standing price cannot have risen above v^m by time t_{e^m}, so that by part (b), $c^m_\tau = v^m = \max_{k \neq i} c^k_\tau$. Hence, if bidder i wins, he pays c^m_τ. Thus, any strategy inducing bidder i (i) to lose when $v^i < c^m_\tau$, and (ii) to win when $v^i > c^m_\tau$, is a best response. Condition (a) guarantees (i), and condition (b) guarantees (ii).

(Necessity) *The necessity of (b)* $c^i_{t^i_{e^i}} = v^i$ *if* $v^i > h_{t^i_{e^i}}$: A bidder might affect rivals' cutoff prices through the effect on the sequences of standing prices (the second-highest existing cutoff price) posted during the auction. Let me denote by $h(c^i_{t^i_{e^i}})$ the (final) second-highest cutoff price as a function of $c^i_{t^i_{e^i}}$.[11]

(1) A strategy inducing $c^i_{t^i_{e^i}} = v^i$ is strictly preferable to a strategy inducing $c^i_{t^i_{e^i}} = \underline{c} < v^i$.

(i) $c^i_{t^i_{e^i}} = v^i$ is strictly preferable to $c^i_{t^i_{e^i}} = \underline{c}$ if $\Pr(\underline{c} < B^j < v^i) > 0$ where $B^j = C^j_{t^j_{e^j}}(v^j)$ $(j \neq i)$.

Consider the case in which $h(\underline{c})$ is no greater than $\underline{c}$. Noting that the standing price is the second-highest cutoff price, any c which is greater than $\underline{c}$ would result in the same sequence of standing prices as $\underline{c}$ would result; therefore, $h(c) = h(\underline{c})$ for any $c \geq \underline{c}$. Then $c^i_{t^i_{e^i}} = v^i$ yield the same payoff as $c^i_{t^i_{e^i}} = \underline{c}$. Next, consider the case in which $h(\underline{c})$ is greater than $\underline{c}$. Bidder i then loses an auction to obtain zero payoff if he uses a strategy inducing $c^i_{t^i_{e^i}} = \underline{c}$. Noting that $c^i_{t^i_{e^i}} = v^i$ always guarantees at least zero payoff, we can conclude that a strategy inducing $c^i_{t^i_{e^i}} = v^i$ always yields as much as a strategy inducing $c^i_{t^i_{e^i}} = \underline{c}$. Now if we find one case in which a strategy inducing $c^i_{t^i_{e^i}} = v^i$ yields more payoff than a strategy inducing $c^i_{t^i_{e^i}} = \underline{c}$ does, we can

[11] The (final) second-highest cutoff price depends on other cutoff prices as well. When I use $h(c^i_{t^i_{e^i}})$ below, I assume that only $c^i_{t^i_{e^i}}$ is changed.

say that $c^i_{t^i_{e^i}} = v^i$ is strictly preferable to $c^i_{t^i_{e^i}} = \underline{c}$. Consider Case A: $\underline{c} < h(\underline{c}) < v^i$ and $h(\underline{c})$ is only cutoff price greater than $\underline{c}$. In that case, if the second-highest cutoff price is h_1 when $h(\underline{c})$ is submitted, under a strategy inducing $c^i_{t^i_{e^i}} = \underline{c}$, the second-highest cutoff price will be also h_1 under a strategy inducing $c^i_{t^i_{e^i}} = v^i$. Hence, $h(v^i) = h(\underline{c})$. Therefore a strategy inducing $c^i_{t^i_{e^i}} = v^i$ yields more payoff $(v^i - h(\underline{c}))$ than a strategy inducing $c^i_{t^i_{e^i}} = v^i$ yields (zero). Noting the symmetry assumption, it is straightforward that Case A happens with positive probability if $\Pr(\underline{c} < B^j < v^i) > 0$ where $B^j = C^j_{t^j_{e^j}}(v^j)(j \neq i)$.

(ii) In any symmetric Bayesian equilibrium, $\Pr(b_1 < B < b_2) > 0$ for $\underline{v} \leq b_1 < b_2 \leq \overline{v}$ where B represents $C^k_{t^k_{e^k}}(v^k)$ of an arbitrary bidder k. Suppose that there exists (b_1, b_2) such that $\Pr(b_1 < B < b_2) = 0$. Let $\underline{b} = \sup_b\{b| \Pr(B \in [b, b_1) = 0\}$ and $\overline{b} = \inf_b\{b| \Pr(B \in (b_2, b]) = 0\}$. Then $\Pr(\underline{b} < B < \overline{b}) = 0$; this implies that every bidder uses a strategy inducing $b \leq \underline{b}$ or $b \geq \overline{b}$ in the symmetric equilibrium. Below I show that there is a bidder j who has a strictly preferable strategy to a strategy inducing $b^j \leq \underline{b}$ or $b^j \geq \overline{b}$. Consider a bidder j whose valuation, v^j is between $\underline{b}$ and $\overline{b}$. First, suppose that bidder j uses a strategy inducing $b^j \leq \underline{b}$ with positive probability. By construction of $\underline{b}$, $\Pr(\underline{c} < B^j < v^j) > 0$ for any $\underline{c} < \underline{b}$. Then by (i), a strategy inducing $b^j = v^j$ is a strictly preferable strategy to a strategy inducing $b^j < \underline{b}$. Accordingly, bidder j must use a strategy inducing $b^j = \underline{b}$. Since this is the case for all bidders whose valuation is between $\underline{b}$ and $\overline{b}$, $\Pr(B = \underline{b}) > 0$. However, if $\Pr(B = \underline{b})$ is positive, a strategy inducing $b^j = v^j$ is strictly preferable to a strategy inducing $b^j = \underline{b}$. To see this, note that $\Pr(\underline{b} < B < v^j) = 0$. Hence the deviation from $b^j = \underline{b}$ to $b^j = v^j$ increases the winning probability without increasing the payment conditional on winning. Accordingly no bidder whose valuation is between $\underline{b}$ and $\overline{b}$ use a strategy inducing $b^j \leq \underline{b}$ in a symmetric Bayesian equilibrium.

Next, suppose that bidder j uses a strategy inducing $b^j \geq \overline{b}$ with positive probability. By construction of $\overline{b}$, $\Pr(v^j < B^j < \overline{c}) > 0$ for any $\overline{c} > \overline{b}$. Hence any strategy inducing $b^j > \overline{b}$ obtains a negative payoff with positive probability. Therefore any strategy inducing $b^j > \overline{b}$ should not be played with positive probability. Accordingly, bidder j must use a strategy inducing $b^j = \overline{b}$. Since this is the case for all bidders whose valuation is between $\underline{b}$ and $\overline{b}$, $\Pr(B = \overline{b}) > 0$. However, if $\Pr(B = \overline{b})$ is positive, a strategy inducing $b^j = v^j$ is strictly preferable to a strategy inducing $b^j = \overline{b}$, because a strategy inducing $b^j = \overline{b}$ results in a negative payoff with positive probability. Accordingly no bidder whose valuation is between $\underline{b}$ and $\overline{b}$ use a strategy inducing $b^j \geq \overline{b}$ in a symmetric Bayesian equilibrium.

Therefore no bidder whose valuation is between $\underline{b}$ and $\overline{b}$ use a strategy inducing $b^j \leq \underline{b}$ or $b^j \geq \overline{b}$; this contradicts there exists (b_1, b_2) such that $\Pr(b_1 < B < b_2) = 0$.

From (i) and (ii), we can conclude that a strategy inducing $c^i_{t^i_{e^i}} = v^i$ is strictly preferable to a strategy inducing $c^i_{t^i_{e^i}} = \underline{c} < v^i$.

(2) A strategy inducing $c^i_{t^i_{e^i}} = v^i$ is strictly preferable to a strategy inducing $c^i_{t^i_{e^i}} = \overline{c} > v^i$.

Consider the case in which $h(\overline{c})$ is no greater than v^i. Noting that the standing price is the second-highest cutoff price, any c which is greater than $h(\overline{c})$ would result in the same sequence of standing prices as $\overline{c}$ would result; therefore, $h(c) = h(\overline{c})$ for any $c \geq h(\overline{c})$. Then $c^i_{t^i_{e^i}} = v^i$ yield the same payoff as $c^i_{t^i_{e^i}} = \overline{c}$. Next, consider the case in which $h(\overline{c})$ is greater than $\overline{c}$. Bidder i then loses an auction to obtain zero payoff by using a strategy inducing $c^i_{t^i_{e^i}} = \overline{c}$. Noting that $c^i_{t^i_{e^i}} = v^i$ always guarantees at least zero payoff, a strategy inducing $c^i_{t^i_{e^i}} = v^i$ always yields as much as a strategy inducing $c^i_{t^i_{e^i}} = \overline{c}$. Finally consider the case in which $h(\overline{c})$ is between v^i and $\overline{c}$. In this case, bidder i then obtains a negative payoff by using a strategy inducing $c^i_{t^i_{e^i}} = \overline{c}$. Again, since a strategy inducing $c^i_{t^i_{e^i}} = v^i$ always guarantees at least zero payoff, a strategy inducing $c^i_{t^i_{e^i}} = v^i$ is more profitable. Moreover by (ii) of (1) above, it happens with positive probability that $h(\overline{c})$ is between v^i and $\overline{c}$.

Accordingly, we can conclude that a strategy inducing $c^i_{t^i_{e^i}} = v^i$ is strictly preferable to a strategy inducing $c^i_{t^i_{e^i}} = \overline{c} > v^i$.

Next, the necessity of (a) $c^i_t \leq v^i$: It is straightforward from the necessity of (b). Since an eBay bidder may not decrease his previous cutoff price, any strategy inducing $c^i_t > v^i$ cannot equalize $c^i_{t^i_{e^i}}$ with v^i. ■

Theorem 1 shows that, although there may be multiple equilibria, in every equilibrium, the two highest-valued potential bidders, whose valuations cannot be lower than the standing price at any time, always bid their valuations before the auction ends. On the other hand, lower-than-second-highest-valued bidders will bid their valuations only if they choose to do so before the standing price rises above their valuations. Therefore, some potential bidders may not make a bid at all or may not update their early cutoff prices, even though these were lower than their valuations.

The proof of Theorem 1 demonstrates that every equilibrium is an *ex-post* equilibrium:[12] even if the actual number of potential bidders n and all potential bidders' private information

[12] For a definition of ex post equilibrium, see Appendix F of Krishna (2002) or references therein.

$\{v^i, T^i\}_{i=1}^n$ were known to a particular bidder i, his equilibrium strategy would still be optimal. This suggests robustness to changes in assumptions that a bidder knows distribution $F(\cdot)$ and the probabilities p_n, which might be valuable for an Internet auctions where the used common knowledge assumptions concerning $F(\cdot)$ and p_n may be implausible.

3.2 Discussion

At this point, I will briefly discuss the relationship of this model to some stylized facts about eBay auctions, and to other existing models of eBay auctions. Previous research concerning eBay auctions such as Bajari and Hortaçsu (2002a), Ockenfels and Roth (2002), and Roth and Ockenfels (2002) has pointed out that late-bidding is prevalent. The above result does not contradict late-bidding. For example, in one equilibrium, all bidders wait until their own, last monitoring times to submit any cutoff prices. However, this model does not explain why late-bidding is observed more frequently than early-bidding: here, bidders have no reason to bid late, but also no reason not to. Equilibrium selection is a difficult issue that I need not address here.

Bajari and Hortasçu (2002a) study eBay auctions within the common value paradigm. They show that on an equilibrium, bidders will bid at the end of the auction in order not to reveal their private information to other bidders in a common value environment. Ockenfels and Roth (2002) construct a model which, like mine, has multiple equilibria, including one involving last-minute bidding in private value environments. However, in their model, on an equilibrium path in which the last-minute bidding happens, every bid in an auction should be submitted in the auction's last seconds; that is hardly ever observed in practice. The strength of Theorem 1 is that it characterizes all symmetric Bayesian-Nash equilibria. In Bajari and Hortasçu (2002a) and Ockenfels and Roth (2002), they show only that their equilibria is an equilibrium.

4 Econometric Model

Throughout this paper I represent random variables in upper case and their realizations in lower case. I consider a sample of independent eBay auctions. Assume that potential bidders' value distribution, $F(\cdot)$, is fixed in all auctions under consideration.[13] Let N_t denote the number of po-

[13]Standard arguments extend the IPV model to the model in which bidders' valuations are i.i.d. conditional on auction-specific observables. To avoid unnecessary complexity, the model maintains the IPV assumption. In

tential bidders at each auction t. Unlike the potential bidders' value distribution, the distribution of N_t may vary auction by auction in an arbitrary way. At each auction t, $V_t^{(1:N_t)}, ..., V_t^{(N_t:N_t)}$ are order statistics of potential bidders' valuations, with $V_t^{(k:N_t)}$ denoting the kth-lowest among N_t. Each $V_t^{(i:N_t)}$ is distributed according to $F^{(i:N_t)}$. Let $I_t^{(k:N_t)}$ denote the identity of the potential bidder whose valuation is $V_t^{(k:N_t)}$.

The primary model primitive of interest is $F(\cdot)$. While $F(\cdot)$ is sufficient for answering many policy questions, in some cases we may also be interested in the distribution of N_t. I focus on a structure in which the distribution of N_t varies freely across auctions, and is therefore not identified. However, in Appendix B, I discuss identification of the distribution of N_t under the restrictive assumption (but the minimum requirement for identification) that the distribution of N_t is fixed across auctions. To make the key idea conspicuous, I derive identification results in this section by assuming that starting prices are below $\underline{v}$. Appendix C extends the results to auctions in which a seller sets a binding starting price, $s > \underline{v}$, in a way similar to that in Athey and Haile (2002) and Haile and Tamer (2003).

For each auction, a researcher can observe each actual bidder's history of cutoff prices: identity, the amounts of all submitted cutoff prices (except for the winning bid), and their submission times (including winning bid's submission time). The amount of the winning bid is not revealed, even after an auction ends. On top of this information concerning actual bidders, various auction details, such as starting prices and sellers' ratings, are available as well. I explain those observables when it becomes necessary to this study. Actual bidder i's final cutoff price, $C^i_{t^i_{e^i}} (= C^i_\tau)$, is called i's *bid* and denoted by B^i_t. Let M_t denote the number of actual bidders, and $B_t^{(1:M_t)}, ..., B_t^{(M_t-1:M_t)}$ denote the ordered set of the observed bids. $\widetilde{I}_t^{(k:M_t)}$ is the identity of the actual bidder who submitted the kth-lowest among M_t bids.

4.1 Two Order Statistics Identify the Parent Distribution

A key feature of eBay auctions, captured by the theoretical model, is the fact that the number of potential bidders is not observable. As you will see this identification issue eventually reduces to a statistical question of whether a parent distribution is uniquely determined by the distribution of its order statistics from a sample, of which the size is unknown. It is not obvious whether or not one can discriminate between changes in the parent distribution and changes in the sample

the application, however, bidders' value distribution is allowed to vary with auction characteristics.

size, since the joint distribution of order statistics depends on both the parent distribution and the sample size.

The ith order statistic and jth order statistic $(n \geq j > i \geq 1)$ from an i.i.d sample of size n from distribution $F(\cdot)$ have a joint probability density function (PDF)

$$g^{(i,j:n)}(x,y) = \begin{cases} \frac{n![F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y)}{(i-1)!(j-i-1)!(n-j)!}, & y > x \\ 0, & otherwise. \end{cases} \tag{1}$$

The *ith* order statistic has a PDF

$$g^{(i:n)}(x) = \frac{n!}{(i-1)!(n-i)!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{n-i} f(x) \tag{2}$$

and CDF

$$G^{(i:n)}(x) = \frac{n!}{(i-1)!(n-i)!}\int_0^{F(x)} t^{i-1}(1-t)^{n-i}dt \tag{3}$$

(See, for example Arnold et al., 1992).

Lemma 1 *An arbitrary absolutely-continuous distribution $F(\cdot)$ is nonparametrically identified from observations of any two order statistics from an i.i.d sample, even when the sample size, n, is unknown and stochastic.*

Proof. There are two possibilities concerning available order statistics: (1) The distance from the top is known; for instance, there is a pair consisting of the second- and third-highest order statistic, and (2) The distance from the bottom is known; for instance there is a pair consisting of the lowest and third-lowest order statistic. The first case is relevant to eBay auctions, so I present its proof here, while the proof for (2) is provided in Appendix A. The second case may be relevant to procurement auctions.[14]

Let Y denote the k_1th-highest value, which is the $(n-k_1+1)$th order statistic, and X denote the k_2th-highest value, which is the $(n-k_2+1)$th order statistic $(1 \leq k_1 < k_2 \leq n)$. The lower limit of support of $F(\cdot)$ is denoted by $\underline{v}$, and $f(\cdot)$ is an associated density of $F(\cdot)$. It is

[14] One could imagine a possibility that for one, the ranking from the top is known and for the other, the ranking from the bottom is known. Lemma 1 is not applicable to that case.

possible that $\underline{v} = -\infty$. For the sake of notational convenience, let $F(\cdot|x)$ denote the truncated distribution of $F(\cdot)$ truncated at x, and $f(\cdot|x)$ denote an associated density.

The density of Y conditional on X, $p_{(k_2,k_1)}(y|x)$, is computed by employing equations (1) and (2):

$$\begin{aligned} p_{(k_2,k_1)}(y|x) &= \frac{(k_2-1)!}{(k_2-k_1-1)!(k_1-1)!} \\ &\quad \times \frac{[F(y)-F(x)]^{k_2-k_1-1}[1-F(y)]^{k_1-1}f(y)}{(1-F(x))^{k_2-1}} \cdot I_{\{y\geq x\}} \\ &= \frac{(k_2-1)!}{(k_2-k_1-1)!(k_1-1)!} \times [(1-F(x))F(y|x)]^{k_2-k_1-1} \\ &\quad \times \frac{[(1-F(x))(1-F(y|x))]^{k_1-1}f(y|x)(1-F(x))}{(1-F(r))^{k_2-1}} \cdot I_{\{y\geq x\}} \\ &= \frac{(k_2-1)!}{(k_2-k_1-1)!(k_2-1-(k_2-k_1))!} \\ &\quad \times F(y|x)^{k_2-k_1-1}(1-F(y|x))^{k_1-1}f(y|x) \cdot I_{\{y\geq x\}} \\ &= f^{(k_2-k_1:k_2-1)}(y|x). \end{aligned} \tag{4}$$

Now note that $\lim_{x\to\underline{v}} p_{(k_2,k_1)}(y|x) = f^{(k_2-k_1:k_2-1)}(y)$. This implies that the k_1th- and k_2th-highest order statistics identify the density of the (k_2-k_1)th order statistic from a sample of which the size is (k_2-1) from $F(\cdot)$. The proof of Theorem 1 in Athey and Haile (2002) shows that the parent distribution is identified whenever the distribution of any order statistic with a known sample size is identified. Thus, the result follows. ■

4.2 Identification in the eBay Model

Theorem 1 implies that $B_t^{(M_t-1:M_t)} = V_t^{(N_t-1:N_t)}$, $\widetilde{I}_t^{(M_t:M_t)} = I_t^{(N_t:N_t)}$, and $\widetilde{I}_t^{(M_t-1:M_t)} = I_t^{(N_t-1:N_t)}$ in all auctions in which $M_t \geq 2$.[15] In other words, one observes the second-highest valuation and the identities of the highest and second-highest potential bidders. Generally, $N_t \neq M_t$, so a researcher knows that the observed $B_t^{(M_t-1:M_t)}$ is the second-highest valuation, but she/he does not know the number of potential bidders among whose valuations,

[15] More precisely, $V_t^{(N_t-1:N_t)}$ is the shipping and handling charge over the second-highest bid. In estimation, I use the second-highest bid plus the shipping and handling charge as the observation of the second-highest valuation. I adjust other observables related to the shipping and handling charge in the same way.

$B_t^{(M_t-1:M_t)}$ is the second-highest. However, observation of the second-highest valuation alone is not enough to identify $F(\cdot)$, when N_t is unknown. To see this, consider two underlying structures: (i) $F(x) = x^{1/2}$ defined on $[0,1]$ with $\Pr(N_t = 3) = 2/3$, and $\Pr(N_t = 4) = 1/3$, (ii) $F(x) = x$ defined on $[0,1]$ with $\Pr(N_t = 2) = 1$. Both structures generate the same distribution of $B_t^{(M_t-1:M_t)}$ which is $F^{(N_{t-1}:N_t)} = 2x - x^2$.

Lemma 1 implies that observations of two order statistics of valuations identify $F(x)$, even if N_t is unknown. Theorem 1 guarantees only that two highest-valued potential bidders always submit their true valuations, so a researcher can conclude only that $B_t^{(k:M_t)} \leq V_t^{(k:N_t)}$ for $1 \leq k \leq M_t - 2$. However, observations of the two highest-valued potential bidders' submission times of all cutoff prices enable us to know in which auctions $B_t^{(M_t-2:M_t)} = V_t^{(N_t-2:N_t)}$ is more likely. Indeed, I shall be able to construct a sequence of auction sets, $\{A_{window}\}$, such that $\Pr(B_a^{(M_a-2:M_a)} = V_a^{(N_a-2:N_a)}) \to 1$ for an auction $a \in A_{window}$, as $window \to 0$.

Before providing the formal result, I explain how the two highest-valued potential bidders' submission times are related to finding auctions in which $B_t^{(M_t-2:M_t)} = V_t^{(N_t-2:N_t)}$. The model shows that the only reason for a potential bidder not to have bid his valuation is that the standing price became higher than his valuation before he had a chance to submit that valuation. The standing price cannot rise to higher than the third-highest valuation until *both* the first- and second-highest bidders have submitted cutoff prices greater than the third-highest valuation. Thus, by looking at auctions where the first- *or* the second-highest bidder submitted a cutoff price greater than the third-highest bid late in the auction, we can increase the probability of obtaining the actual third-highest valuation. The following example illustrates this. Consider an item that has four potential bidders: Bidder A's valuation is \$25; Bidder B's is \$20; Bidder C's is \$15; and Bidder D's is \$10. Examine the following two situations:

Situation 1:

(1) An auction starts at \$5. (Standing price: \$5)

(2) Bidder A submits a \$25 cutoff price 24 hours before the auction ends. (Standing price: \$5)

(3) Bidder D submits a \$10 cutoff price 12 hours before the auction ends. (Standing price: \$10)

(4) Bidder B submits a \$20 cutoff price 2 hours before the auction ends. (Standing price: \$20)

Situation 2:

(1) An auction starts at \$5. (Standing price: \$5)

(2) Bidder A submits a \$12 cutoff price 24 hours before the auction ends. (Standing price: \$5)

(3) Bidder D submits a \$10 cutoff price 12 hours before the auction ends. (Standing price: \$10)

(4) Bidder B submits a \$20 cutoff price 2 hours before the auction ends. (Standing price: \$12)

(5) Bidder A submits a \$25 cutoff price 10 seconds before the auction ends. (Standing price: \$20)

Suppose that in Situation 1, Bidder C tries to submit a cutoff price of his valuation 10 minutes before the auction ends. He would be unsuccessful, because the standing price is already higher than his valuation. The observed third-highest bid would then be the fourth-highest valuation. In Situation 2, on the other hand, he would successfully submit a cutoff price equal to his valuation, so the observed third-highest bid would, in fact, be the third-highest valuation.

In order to obtain the formal result, I start by defining an auction set that consists of auctions in which the observed, third-highest bids are in fact the third-highest valuations.

Definition 1 *(The set of auctions $A^*(\tau^*)$): Let τ^* be the (unobserved) third-highest potential bidder's final monitoring opportunity. An auction a is in the set $A^*(\tau^*)$ if and only if*

$$\min\{C_{\tau^*}^{\widetilde{I}^{N_a-1:N_a}},\ C_{\tau^*}^{\widetilde{I}^{N_a:N_a}}\} < B_a^{(M_a-2:M_a)}. \tag{5}$$

In other words, auctions in the set $A^*(\tau^*)$ are those in which, at time τ^*, at least one of the first- and second-highest-valued bidders' cutoff prices is less than $B_a^{(N_a-2:N_a)}$. Recall that the top two observed bidders are the potential bidders with the highest valuations.

Corollary 1 *In every auction in the set $A^*(\tau^*)$, $B_a^{(M_a-2:M_a)} = V_a^{(N_a-2:N_a)}$.*

Proof. This is obvious in the case where the third-highest potential bidder submitted his valuation before his final opportunity. For any other cases, note that only the two highest-valued bidders would ever submit cutoff prices greater than $V_a^{(N_a-2:N_a)}$. Hence, the standing price at time τ^* must be less than $V_a^{(N_a-2:N_a)}$, when condition (5) holds. Condition (b) in Theorem 1 then yields that $C_{\tau^*}^{\widetilde{I}^{N_a-2:N_a}} = V_a^{(N_a-2:N_a)} = B_a^{(M_a-2:M_a)}$. ■

Definition 2 *(The set of auction A_{window}): Let τ_a be the ending time of auction a. An auction $a \in A_{window}$ if and only if*

$$\min\{C^{\widetilde{I}^{N_a-1:N_a}}_{\tau_a-window},\ C^{\widetilde{I}^{N_a:N_a}}_{\tau_a-window}\} < B_a^{(M_a-2:M_a)}.$$

In other words, auctions in the A_{window} are those in which the first- *or* the second-highest bidder submitted cutoff prices greater than the third-highest bid no earlier than *window* (time) before the auction ends. Since $\lim_{window \to 0} \Pr(\tau^* \leq \tau_a - window) = 1$,
$\lim_{window \to 0} \Pr(B_a^{(M_a-2:M_a)} = V_a^{(N_a-2:N_a)} \mid a \in A_{window}) = 1$; i.e., as *window* goes to zero, $B_a^{(M_a-2:M_a)} = V_a^{(N_a-2:N_a)}$ in every auction a in the set A_{window}. In any auction t, $B_t^{(M_t-1:M_t)} = V_t^{(N_t-1:N_t)}$.

Consider the following assumption:

Assumption 1:

$$\begin{aligned}\Pr(V_a^{(N_a-1:N_a)} &\leq v_2 | V_a^{(N_a-2:N_a)} = v_3, a \in A_{window}) \\ &= \Pr(V_a^{(N_a-1:N_a)} \leq v_2 | V_a^{(N_a-2:N_a)} = v_3) \ \ \forall \ window\end{aligned}$$

Under Assumption 1, the distribution of the second-highest valuation conditional on the third-highest valuation is identified from: (a) the second- and third-highest bids; and (b) the first- and second-highest bidders' cutoff price submission times. Assumption 1 is required because the sample selection problem may arise if one uses only that part of a dataset chosen on the basis of cutoff submission times. For example, if a highly-valued bidder is apt to bid late, data from only auctions in which bids are made late will be different from data from all auctions. If the number of monitoring opportunities, e^i, is one for all i, Assumption 1 is satisfied if the bid submission time and bid amount are independent. If $e^i > 1$, it is not clear what assumptions on model primitives and equilibrium selection ensure Assumption 1.[16] By applying Lemma 1, I find the following.

Theorem 2 *Under Assumption 1.1, $F(\cdot)$ is identified from the following observables:*

(a) The second- and third-highest bids

(b) The first- and second-highest bidders' cutoff price submission times.

It is necessary at least to compare descriptive statistics across auction sets with different *window* to see if there are fundamental differences. In my dataset, this does not appear to be

[16] Multiple bidding complicates the analysis. Not only the bid submission time and final bid amount but also the intermediate bid amount is critical.

a problem as I shall demonstrate later. The above theorem shows that potential bidders' value distribution is identified from the data available in eBay auctions. The idea behind the above identification result also suggests an estimation strategy: I will 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. An econometric method to decide "how late" is appropriate is proposed in the following section.

5 Estimation of the Distribution of Valuations

For a consistent estimate of $F(\cdot)$, I employ the semi-nonparametric (SNP) method developed by Gallant and his coauthors.[17] They showed that simply replacing the unknown density with a Hermite series and applying the standard, finite dimensional maximum likelihood methods yields consistent estimators of model parameters, if there are any, and nearly all aspects of the unknown density itself, provided that the length of the series increases with the sample size. The rule for increasing series length can be data-dependent.

Since my estimation technique requires knowledge of the second- and third-highest bids in each auction, I consider only auctions in which observed bidders are no fewer than three for estimation of $F(\cdot)$. Let (Y_t, X_t) denote the second- and third-highest bids, including the shipping and handling charges at auctions $t = 1, 2, ..., T$. Let $c = \min_t \ x_t$. Since no information about $F(v)$ for $v < c$ can be discovered from this dataset, I treat $F^*(\cdot) = F(\cdot|c)$ as the model primitive of interest. An associated density is denoted by $f^*(\cdot)$. If a starting price set by a seller is below $\underline{v}$ with positive probability, c is a consistent estimate of $\underline{v}$; in that case, a consistent estimate of $F^*(\cdot)$ is a consistent estimate of $F(\cdot)$.[18]

The density of Y_t conditional on X_t, $p^*(y_t|X_t = x_t)$, is calculated by substituting 3 for k_2, and 2 for k_1 in Equation (4):

$$p^*(y_t|X_t \quad = \quad x_t) = \frac{2[1 - F^*(y_t)]f^*(y_t)}{[1 - F^*(x_t)]^2} \quad \text{for } y_t \geq x_t \geq c.$$

[17]See Gallant and Nychka (1987), Fenton and Gallant (1996a,b), and Coppejans and Gallant (2002).

[18]In practice, the property of c is not important unless c is far above $\underline{v}$. For most economic issues, including three issues analyzed in my application, $F^*(\cdot)$ is sufficient.

Hence, I consider the following sample likelihood function:

$$L_T(\widehat{f}) = \frac{1}{T}\sum_{t=1}^{T} \ln \frac{2[1-\widehat{F}(y_t)]\widehat{f}(y_t)}{[1-\widehat{F}(x_t)]^2} \qquad (\ \widehat{F}(x) = \int_c^x \ \widehat{f}(t)dt \ ).$$

I consider a partial likelihood which is the sample counterpart of $p^*(y_t|X_t = x_t)$, because the full likelihood which is joint density of (Y_t, X_t) includes the unknown number of potential bidders. The proof of Lemma 1 implies that $p^*(y|x)$ characterizes $F^*(v)$, when the lower limit of the support of $F^*(v)$ is known. By construction, c is the lower limit of the support of $F^*(v)$, so my partial likelihood is uniquely maximized at $\widehat{F}(v) = F^*(v)$.

I use the following specification of $\widehat{f}(x)$:

$$\widehat{f}(x) = \frac{[1 + a_1(\frac{x-\mu}{\sigma}) + ... + a_k(\frac{x-\mu}{\sigma})^k]^2\phi(x;\mu,\sigma,c)}{\int_c^\infty [1 + a_1(\frac{x-\mu}{\sigma}) + ... + a_k(\frac{x-\mu}{\sigma})^k]^2\phi(x;\mu,\sigma,c)dx}$$

where $\phi(x;\mu,\sigma,c)$ is the density of $N(\mu,\sigma)$ truncated at c. An estimator, $\widehat{f_T}$, is the maximizer of $L_T(\widehat{f})$. So,

$$\widehat{f_T}(x) = \frac{[1 + \widehat{a}_1(\frac{x-\widehat{\mu}}{\widehat{\sigma}}) + ... + \ \widehat{a}_k(\frac{x-\widehat{\mu}}{\widehat{\sigma}})^k]^2\phi(x;\widehat{\mu},\widehat{\sigma},c)}{\int_c^\infty [1 + \widehat{a}_1(\frac{x-\widehat{\mu}}{\widehat{\sigma}}) + ... + \ \widehat{a}_k(\frac{x-\widehat{\mu}}{\widehat{\sigma}})^k]^2\phi(x;\widehat{\mu},\widehat{\sigma},c)dx}$$

such that

$$(\widehat{a}_1, ..., \ \widehat{a}_k, \widehat{\mu}, \widehat{\sigma}) = \underset{a_1,..., \ a_k, \mu\in R, \ \sigma>0}{\arg\max} \ L_T(\widehat{f}) = \frac{1}{T}\sum_{t=1}^{T} \ln \frac{2[1-\widehat{F}(y_t)]\widehat{f}(y_t)}{[1-\widehat{F}(x_t)]^2}.$$

I choose the optimal series length, k^*, following the method proposed in Coppejans and Gallant (2002). They consider a cross-validation strategy, which employs the *ISE*[Integrated Squared Error] criteria. When $\widehat{h}(x)$ is a density estimate of $h(x)$, the ISE criterion is defined as follows:

$$\begin{aligned} \text{ISE}(\widehat{h}) &= \int \widehat{h}^2(x)dx - 2\int \widehat{h}(x) \ h(x)dx + \int h^2(x)dx \\ &= M_{(1)} - 2M_{(2)} + M_{(3)}. \end{aligned}$$

First, I randomly partition a dataset under consideration into J groups, denoted by χ_j, $j = 1, ..., J$, making the sizes of these groups as close to equal as possible. Let $\widehat{f}_{j,k}(\cdot)$ denote the SNP estimate obtained from the sample points that remain after deletion of the jth group when k is used as a series length. The cumulative distribution associated with $\widehat{f}_{j,k}(\cdot)$ is denoted by $\widehat{F}_{j,k}(\cdot)$.

Adjustment of the Coppejans and Gallant (2002) estimators to my case gives the following estimators:

$$\widehat{M}_{(1)}(k) = \frac{1}{s}\sum_{j=1}^{J}\int [\widehat{p^*}_{j,k}(y|x)\ ]^2 dydx \tag{6}$$
$$= \frac{1}{s}\sum_{j=1}^{J}\int \{\frac{2[1-\widehat{F}_{j,k}(y)]\widehat{f}_{j,k}(y)}{[1-\widehat{F}_{j,k}(x)]^2}\}^{\ 2}\ \ dydx,$$
$$\widehat{M}_{(2)}(k) = \frac{1}{s}\sum_{j=1}^{J}\sum_{(x_t,y_t)\in\chi_j}\widehat{p^*}_{j,k}(y_t|x_t)$$
$$= \frac{1}{s}\sum_{j=1}^{J}\sum_{(x_t,y_t)\in\chi_j}\frac{2[1-\widehat{F}_{j,k}(y_t)]\widehat{f}_{j,k}(y_t)}{[1-\widehat{F}_{j,k}(x_t)]^2}$$

where s is the sample size of a considered dataset. Let $CVH(k) = \widehat{M}_{(1)}(k) - 2\widehat{M}_{(2)}(k)$. Since $M_{(3)}$ does not depend on k, $M_{(3)}$ does not need to be considered here. In a typical graph of CVH(k) *versus* k, CVH(k) falls as k increases when k is small, periodically drops abruptly, and flattens right after the final abrupt drop. Coppejans and Gallant (2002) recommend use of the final abrupt drop point in the graph of CVH(k) *versus* k as the series length, k^*. For my estimation, I will use J of ten.

Finally, I propose a method for the selection of an auction set to be used in estimation of $F^*(\cdot)$: a method for how to choose the size of *window*. The choice of *window* has a similar trade-off to the one that occurred in the bandwidth selection of the Kernel density estimation. Two properties of the set of auction A_{window} are important to make this clear. First, $\Pr(a \in A^*(\tau^*)|a \in A_{window})$ is decreasing in *window*; as *window* decreases, it is more likely that the third-highest bid equals the third-highest valuation. Second, $\Pr(a \in A_{window})$ is increasing in *window*; as *window* decreases, the number of available auctions becomes smaller. Both properties are straightforward to show, given the fact that $\min\{C_t^{\widetilde{I}^{N_a-1:N_a}}, C_t^{\widetilde{I}^{N_a:N_a}}\}$ is increasing in t. The first property of A_{window} implies that as *window* decreases, an estimate obtained by using data from auctions in A_{window} would be less biased. On the other hand, the second property implies that as *window* decreases, an estimate would have greater variance. I construct a method for choosing the optimal *window* which clearly considers this trade-off by applying the same cross-validation strategy that is used for the choice of series length. Detailed procedures follow.

I consider a sequence of auction sets, $A_{w_1} \supset A_{w_2} \supset ... \supset A_{w_I}$, classified by the later one between the times at which the first- and second-highest bidders submitted cutoff prices greater than the third-highest bid. For example, a sequence of auction sets can be constructed in the following way: A_{w_1} is a set of all auctions; w_2 =2 days (in other words, A_{w_2} is a set of auctions in which the highest or second-highest bidder submits a cutoff price greater than the third-highest bid no earlier than 2 days before an auction ends.),..., and w_I =10 seconds.[19] Clearly A_{w_I} will have the least number of incorrect, third-highest valuations. For each auction set A_{w_i}, I compute $\text{CVH}_{w_i}(k^*)$ using data from auctions in A_{w_i} in the same way it is computed in the choice of the optimal series length, except that I use of the following new $\widehat{M}_{(2)}(k^*)$, instead of that in Equation (6):

$$\begin{aligned} \widehat{M}_{(2)}(k^*) &= \frac{1}{s_2} \sum_{j=1}^{J} \sum_{(x_t, y_t) \in \chi_j \cap A_{w_I}} \widehat{p}_{j,k^*}(y_t | x_t) \\ &= \frac{1}{s_2} \sum_{j=1}^{J} \sum_{(x_t, y_t) \in \chi_j \cap A_{w_I}} \frac{2[1 - \widehat{F}_{j,k^*}(y_t)] \widehat{f}_{j,k^*}(y_t)}{[1 - \widehat{F}_{j,k^*}(x_t)]^2} \end{aligned}$$

where s_2 is the size of A_{w_I}. Note that $\widehat{p}_{j,k^*}(\cdot|\cdot)$ is evaluated only at sample points in A_{w_I}; therefore, CVH_{w_i} measures how well the estimate obtained by using data from an auction set A_{w_i} fits the data from auctions in A_{w_I}. Although I postpone theoretical justification of the above method for future work, my use of it here is based on the standard cross-validation strategy.

An obvious complementary method to CVH_{w_i} comparison is examining descriptive statistics regarding the third-highest bids among A_{w_i}. For example, suppose that there is i_0 such that the distribution of the third-highest bids does not change much among A_{w_i} for $i \geq i_0$. Then the bias caused by the third-highest bids being smaller than the third-highest valuations also cannot be critical among A_{w_i} for $i \geq i_0$. It is not clear, however, when two distributions are not much different from each other.

[19] I do not suggest a general way of choosing $w_1, ..., w_I$. In my application, a sequence of auction sets is constructed such that size differences between successive sets are similar.

6 Monte Carlo Experiments

To illustrate the performance of my estimation method, I conduct two sets of Monte Carlo experiments. The first set of experiments is a benchmark. It illustrates the performance when the true second- and third-highest order statistics are available. The second set of experiments simulates eBay auction environments. A potential bidder monitors an auction at a randomly generated time to make a bid according to eBay rules. The simulated data set may have third-highest bids that are lower than the true third-highest bidders' valuations.

6.1 Benchmark Experiments

For each experiment, artificial data of 600 auctions are generated.[20] The number of potential bidders, N_t $(t = 1, ..., 600)$, was first drawn from a Binomial distribution with trial number 50 and success probability 0.1. N_t potential bidders' valuations were then generated according to the equation:

$$\ln V_t^i = \alpha_1 SR_{1t} + \alpha_2 SR_{2t} + \nu_t^i \tag{7}$$

where $\alpha_1 = 1$, $\alpha_2 = -1$, $SR_{1t} \backsim N(0,1)$, $SR_{2t} \backsim Exp(1)$, and $\nu_t^i \backsim f(\cdot) = Gamma(9,3)$. For the sake of reference, $E(\nu_t^i) = 3$, and $Var(\nu_t^i) = 1$. The variables SR_{1t} and SR_{2t} represent observable item characteristics, and ν_t^i is bidder i's private information, the distribution of which is to be estimated.

To reflect the fact that not all potential bidders make bids, reserve prices, R_t, were set as follows:

$$\ln R_t = \alpha_1 SR_{1t} + \alpha_2 SR_{2t} + \omega_t$$

where $\omega_t = Y_t - 2$, $Y_t \backsim Gamma(9,3)$. Both Y_t and ν_t^i follow the same distribution, but they are independent. Since both V_t^i and R_t depend on SR_{1t} and SR_{2t}, V_t^i and R_t are positively correlated, but independent conditional on SR_{1t} and SR_{2t}. *Actual bidders' bids* are those V_t^i greater than R_t. Each experiment works as if the researcher does not know the presence of potential bidders with valuations below R_t. Thus, the distribution of observed bids biased upward relative to that of V_t^i, and the number of observed bidders is smaller than the number of potential bidders.

[20] Six hundred may seem like a large number. I do not use data from all 600 auctions, as will become clear later on. Moreover, a benefit of using eBay auction data is that many observations are available.

In each experiment, a dataset consists of SR_{1t}, SR_{2t}, and the second- and third-highest among actual bidders' bids. I do not use the highest bidder's bid. Auctions with fewer than three actual bidders are dropped. Hence, the number of the auctions used for estimation is less than 600, actually 412, on average, across 100 repetitions. I estimate α_1, α_2, and $f(\cdot)$ by varying the series lengths of the SNP estimator, k from 0 to 8. A researcher does not make a parametric distributional assumption on ν_t^i, though the specification in (7) is assumed to be known. In principle, the functional form of the effects of item characteristics is also identified up to location (Athey and Haile, 2002), but its estimation requires a huge data set.

Table 1 documents the averages of various statistics for 100 repetitions. The Best SNP means the SNP estimate obtained by using the series length that minimizes a true ISE between 0 and 8 in each experiment. It would be ideal to obtain the performance of the SNP estimates by using the value of k^* chosen through the method described in the estimation section. But this would take a great deal of time, and furthermore, as Table 1.1 shows, the SNP estimates are robust in terms of the choice of k in this simulation.

Table 1: Benchmark Experiment Results

	$\widehat{E(\nu_t^i)}$	$\widehat{STD(\nu_t^i)}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{SE(\widehat{\alpha_1})}$	$\widehat{SE(\widehat{\alpha_2})}$
True	3	1	1	-1		
Best SNP	3.057	1.021	1.015	-1.003	.091	.047
$k=0$	2.950	1.058	1.015	-1.002	.101	.052
$k=1$	2.989	1.040	1.014	-1.002	.100	.051
$k=2$	3.081	1.013	1.015	-1.004	.091	.047
$k=3$	3.115	1.007	1.016	-1.010	.090	.047
$k=4$	3.027	1.014	1.012	-1.014	.088	.046
$k=5$	3.099	1.000	1.018	-1.012	.081	.042
$k=6$	3.057	1.009	1.010	-1.007	.075	.038
$k=7$	3.078	1.001	1.067	-1.005	.068	.034
$k=8$	3.054	1.003	1.009	-.996	.061	.031

- SE($\widehat{\alpha_1}$), SE($\widehat{\alpha_2}$) : BHHH estimators.

The results in Table 1 show that the estimators perform very well. I calculate the estimates of $E(\nu_t^i)$ and $STD(\nu_t^i)$ to examine the performance of the estimates of $f(\cdot)$. Figure 1a-1c in

Appendix D shows the estimate of $f(\cdot)$ and the density estimates of V_t^i evaluated at the median of (SR_1, SR_2), along with the corresponding true densities. The graphs also illustrates that the estimators perform very well. The explanation of how to draw Figure 1a-1c is given in Appendix D.

6.2 Experiments under eBay Circumstances

For each experiment, artificial data of 1,200 auctions are generated. The item characteristics SR_{1t} and SR_{2t} are first generated such that $SR_{1t} \backsim N(0,1)$, and $SR_{2t} \backsim [Exp(2) - 1]$ $(t = 1, ..., 1200)$. The number of potential bidders, N_t, was then drawn from $B(30, p)$ where $p = 0.2 \cdot \Phi(SR_{1t}) + 0.1$. Φ is the cumulative distribution function of the standard normal. Thus, the distribution of N_t varies, auction by auction. Each potential bidder draws a *bidding time*, T_t^i from $U[0, 100]$, which represents the *remaining auction time*. Also, each potential bidder obtains V_t^i according to the following equation:

$$\ln V_t^i = \alpha_1 SR_{1t} + \alpha_2 SR_{2t} + \nu_t^i$$

where $\alpha_1 = 1$, $\alpha_2 = -1$, and $\nu_t^i \sim f(\cdot) = Gamma(9, 3)$. Each potential bidder i monitors the auction at time T_t^i and submits a bid equal to V_t^i, if the standing price is less than V_t^i. A standing price is zero until two bidders submit bids, and it is raised to the second-highest existing bid when a new bid is submitted. With the introduction of a random bidding time, it happens that the third-highest bidder does not submit a bid when the first two highest bidders' bidding times are earlier than the third-highest bidder's time. Furthermore, if a lower–than-third-highest bidder's bidding time is the earliest, as in the example in Section 4.2, the third-highest bid is not the third-highest bidder's valuation. The third-highest bid can be the third-highest valuation, the fourth-highest one, and so on.

A dataset in each experiment includes SR_{1t}, SR_{2t}, the second- and third-highest among actual bidders' bids, and the highest and second-highest bidders' bidding times. I consider four subsets of the artificial dataset: (1) one including all auction data; (2) one including data from auctions in which the highest or the second-highest bidder's bidding time is no earlier than 50 (time) before an auction ends; (3) no earlier than 25 (time) before the end; and (4) no earlier than 10 (time) before the end. The SNP estimates of α_1, α_2, and $f(\cdot)$ are obtained in the same way as in the previous simulation, by using each subset of the dataset. I replicate the described experiment 50 times. Table 2 documents the averages of various statistics obtained through application of the series length that minimizes a true ISE.

Table 2: eBay Experiment Results

	# of obs	Wrong third *	$\widehat{E(\nu_t^i)}$	$\widehat{STD(\nu_t^i)}$
True			3	1
(1) All	979	26.03%	3.281	1.057
(2) 50**	785	19.23%	3.176	1.054
(3) 25	467	10.48%	3.066	1.039
(4) 10	204	4.44%	2.999	1.011

	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{SE1(\widehat{\alpha_1})}$	$\widehat{SE1(\widehat{\alpha_2})}$	$\widehat{SE2(\widehat{\alpha_1})}$	$\widehat{SE2(\widehat{\alpha_2})}$
True	1	-1				
(1) All	1.144	-.997	.049	.024	.049	.024
(2) 50	1.055	-.995	.062	.029	.062	.030
(3) 25	1.055	-.999	.095	.044	.094	.045
(4) 10	1.006	-1.018	.158	.075	.162	.082

* Percentage of third-highest bids that are not third-highest valuations

- SE1($\widehat{\alpha_1}$), SE1($\widehat{\alpha_2}$) : Estimated standard error computed by evaluating the Hessian matrix.
- SE2($\widehat{\alpha_1}$), SE2($\widehat{\alpha_2}$) : BHHH estimators.

Table 2 shows that the estimates obtained by using dataset (3) and (4) are very close to the true value even if dataset (3) and (4) have 4.44% and 10.48% inaccurate third-highest valuations. The simulation results demonstrate the trade-off between a less biased dataset and a large dataset. For example, as a less biased dataset is used, the bias of $\widehat{\alpha_1}$ becomes smaller, but its standard error becomes larger. Although the change in the biases of $\widehat{\alpha_2}$ moves in an ambiguous direction, the standard error of $\widehat{\alpha_2}$ becomes larger as a less biased dataset is used, because of the decrease in the sample size. Figure 2a-2d in Appendix E shows the performance of estimates obtained using dataset (1)-(4). When the estimates which accomplish median performance are compared, as expected, estimates obtained by using dataset (4) are closest to the true. Explanation of how to produce Figure 2a-2d is provided in Appendix E.

7 Application to eBay University Yearbook Auctions

7.1 Data

I illustrate my method using data from auctions of university yearbooks from the years 1930-1997 held on eBay from November 17, 2001 to November 16, 2002. The university yearbook auction was chosen as an application because I want to illustrate my method using data from auctions that match my maintained assumptions as closely as possible. The university yearbook is an item hardly traded except in eBay auctions. In most auction models as well as in my bidding model in Section 3, a bidder does not have an outside option, the choice of buying the same item from other markets. Therefore, items transacted seldom outside of eBay suit the model best. Furthermore, the items transacted mainly through eBay are particularly interesting because their demand structure can be recovered only through analysis of eBay auction data.

I did a small number of surveys to ask winners why they bought or would buy a yearbook. The winners bought yearbooks as presents for their relatives, to donate to libraries, out of interest in a particular region or time period, etc. Some bought yearbooks to obtain celebrities' pictures. Since yearbooks containing celebrities' pictures may have a common value element, and their item heterogeneity is difficult to control for with observables, I excluded all auctions in which a seller advertised a particular person's picture, whether or not the advertised person actually seemed famous. Given the thin market of university yearbooks, resale opportunity cannot be an incentive for buying unless the books feature celebrities. Thus the IPV assumption is plausible in this market.

In Appendix F, Table 3 provides the summary statistics of all auctions, and Table 4 gives more detailed summary statistics of auctions in which there were more than two active bidders.

7.2 Estimation Results

I calculated descriptive statistics regarding various auction/item characteristics (year, the presence or absence of advertising imagery, and payment methods) to see if those characteristics have effects on bidders' valuations. The descriptive statistics given in Appendix G suggest that none of these characteristics affect bidders' valuations. Motivated by this observation, I consider the following specification of a bidder's valuation[21]:

$$\ln V_t^i = \alpha_1 SR_{1t} + \alpha_2 SR_{2t} + \nu_t^i$$

where SR_{1t} is the number of seller's positive ratings at auction t, and SR_{2t} is (the number of seller's neutral and negative ratings) / (SR_{1t} +1).

I consider six sets of auctions: (1) all auctions in the dataset; (2) auctions in which the highest or second-highest bidder submits a bid greater than the third-highest bid no earlier than 2 days before an auction ends; (3) no earlier than 12 hours before the end; (4) no earlier than 2 hours before the end; (5) no earlier than 5 minutes before the end; and (6) no earlier than 10 seconds before the end. Descriptive statistics about the second- and third-highest bids in each set are given in Table 5.

Table 5: Descriptive Statistics

		All	2 Days	12 Hours	2 Hours	5 Minutes	10 Seconds
	# of obs	915	806	685	503	338	162
2nd-highest bid	Mean	\$31.71	\$31.82	\$32.00	\$32.19	\$31.92	\$31.33
	Median	\$25.50	\$25.23	\$25.12	\$25.46	\$26.03	\$27.33
	Std	\$25.26	\$26.23	\$26.91	\$28.65	\$23.52	\$15.86
	Min	\$4.87	\$4.87	\$4.87	\$4.87	\$5.50	\$6.99
	Max	\$304.23	\$304.23	\$304.23	\$304.23	\$247.00	\$92.23
3rd-highest bid	Mean	\$21.82	\$22.11	\$22.58	\$23.40	\$23.67	\$22.26
	Median	\$17.61	\$18.12	\$18.68	\$19.07	\$19.26	\$19.38
	Std	\$15.59	\$16.07	\$16.40	\$17.55	\$16.53	\$11.47
	Min	\$3.76	\$3.76	\$3.76	\$3.76	\$5.22	\$5.22
	Max	\$152.54	\$152.54	\$152.54	\$152.54	\$140.00	\$73.00

Table 5 documents that the patterns of second-highest bids, which are supposed to be the same among auction sets, appear to be very similar across the six auction sets. The difference in the standard deviations between 5 minute set and 10 second set is relatively large. However, if

[21] As I mentioned in the section that presents Monte Carlo experiments, although I consider a particular functional form here, the functional form of auction characteristics is also identified up to location.

they are compared after excluding outliers, the difference is small.[22] The medians of the third-highest bids show a very gentle increasing trend, but neither mean nor median is significantly different within any pair of sets. Estimation results in each set are as follows:

Table 6: Estimation Results

	k^*	$\widehat{E(V)}$	$\widehat{STD(V)}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$
(1) All	2	\$25.85	\$70.00	.00 (.00005)	.95 (4.029)
(2) 2 Days	2	\$24.71	\$51.14	.00 (.00005)	-3.48 (6.646)
(3) 12 hours	2	\$22.63	\$61.54	.00 (.00005)	-25.18 (11.590)
(4) 2 hours	2	\$21.17	\$39.92	.00 (.00006)	-15.31 (15.203)
(5) 5 minutes	0	\$19.97	\$18.85	.00 (0.00009)	-14.13 (13.266)
(6) 10 seconds	0	\$23.95	\$20.99	.00 (0.00008)	-19.540 (13.350)

* $\widehat{E(V)}$ and $\widehat{STD(V)}$ are computed at $SR_{1t} = 633$ and $SR_{2t} = .0032$, which are median values.

Applying the method described in the previous section, I chose the fourth dataset (2 hours). Figure 3 in Appendix H illustrates the density estimate of ν_t^i and the density estimate of bidders' valuations conditional on $SR_{1t} = 633$ and $SR_{2t} = .0032$.

7.3 Issues on Economic Variables

7.3.1 The Effects of Auction/Item Characteristics on Valuations: Sellers' Ratings

Often one is interested in the effects of auction/item characteristics on potential bidders' value distribution. In eBay auctions, the effects of sellers' ratings on valuations have attracted a great deal of attention.[23] eBay encourages buyers and sellers to rate each other at the close of a transaction. They can give positive, neutral or negative ratings and comments, which are publicly available to every eBay customer. Thus, a natural question is whether or not sellers' ratings can affect bidders' valuations of auctioned items. Although some prior research shows a relationship between transaction prices and sellers' ratings, there is none showing a relationship

[22] When the lower and upper 1% is excluded, the standard deviations are 17.26 and 15.09.

[23] See Cabral and Hortaçsu (2003), Houser and Wooders (2001), Resnick and Zeckhauser (2002), Resnick, Zeckhauser, Swanson, and Lockwood (2002), etc.

between bidders' latent value distribution and sellers' ratings; exploring this matter is possible only through structural estimation. A seller has options to control trade-off between a sale price and the probability of sale. For example, if a seller has set a higher starting price, she can extract a higher sale price at the expense of a lower probability of sale. It would be misleading if we were to interpret the relationship between transaction prices and sellers' ratings as a relationship between bidders' valuations and ratings.

The results, in particular $\widehat{\alpha_1}$ and $\widehat{\alpha_2}$, in Table 6 imply that sellers' positive ratings do not affect bidders' valuations. Although the ratio of nonpositive to positive ratings has a negative effect on bidders' valuations, the negative effect is not statistically significant. The results of prior studies exploring the relationship between transaction prices and sellers' ratings are slightly mixed, but most models show a significant effect of negative ratings reducing transaction price, and a trend toward a positive effect of positive ratings on transaction price (Resnick and Zeckhauser, 2002). However, even when the effect of negative ratings is statistically significant, typically the effect is small. Furthermore, Resnick and Zeckhauser (2002) show that more positive ratings and fewer negatives and neutrals do appear to affect the probability of sale. Accordingly, a significant effect of negative ratings on transaction prices does not necessarily mean a significant effect on bidders' valuations.

7.3.2 Consumers' Surplus

A consumer's surplus at an auction t is

$$CS_t = v_t^{(n_t:n_t)} - p_t$$

where p_t denotes the price that a buyer actually paid, including the shipping and handling charges. Since $v_t^{(n_t:n_t)}$ is not observed, I estimate an *expected* consumer's surplus as follows:

$$E[CS_t | V_t^{(N_t-1:N_t)} = v_t^{(n_t-1:n_t)}] = \int_{v_t^{(n_t-1:n_t)}}^{\infty} \frac{f(x)}{1 - F(v_t^{(n_t-1:n_t)})} \cdot x \, dx \; - \; p_t \, .$$

The descriptive statistics of the estimate of the expected consumer's surplus at each auction are as follows. The estimate of the sum of consumer's surplus (923 auctions) is \$33435.41 with standard error of \$660.

Mean	Std	Median	Min	Max
\$36.26	\$34.07	\$25.54	\$9.88	\$306.28

This estimate of consumers' surplus may indicate how much eBay customers can benefit from an introduction to the eBay auction site, because yearbook auctions are seldom transacted outside of eBay. However, there are possible sources of bias; consumers' surplus underestimates the benefit because it ignores suppliers' surplus, and it overestimates because it ignores bidders' entry costs, if any.

7.3.3 Examination of Increasing Virtual Valuation Assumption

The virtual valuation, $v - [1 - F(v)]/f(v)$, is often assumed to be an increasing function of v in the mechanism design literature, e.g., Myerson (1981). This paper examines whether this assumption is satisfied in the university yearbook market. In Figure 4 in Appendix I, I present the graph of $\widehat{f}(v)$ and $v-[1-\widehat{F}(v)]/\widehat{f}(v)$, conditional on the median of (SR_{1t}, SR_{2t}). The graph shows that virtual values increase where the density is greater than zero.

8 Concluding Remarks

In this paper, I develop and apply new methods for analyzing auctions in which the number of potential bidders is unknown. This research was motivated by Internet auctions, but is certainly useful for analysis of other ascending auctions as well. Furthermore, the method in this paper is extended to a first-price auction model, although we need to clarify assumptions concerning potential bidders' participation decisions. Song(2004) shows that the highest and second-highest bids per auction nonparametrically identify potential bidders' value distribution with an uncertain number of bidders within the symmetric IPV framework. Again, unlike previous studies, a researcher does not need to know the number of potential bidders.

A few open questions related to interpreting online bidding data exist: First, similar items are often sold side-by-side in Internet auctions, and many auctioned objects are easily available outside online auction markets. My application, university yearbooks, does not have these characteristics; however, a substantial number of online auction items do. If bidders consider participating in more than one eBay auction featuring the same item and consider buying outside of eBay auctions, their behavior would probably be different.

Second, a seller has a number of options when she lists her object for an eBay auction. In this paper, I regard sellers' choices as exogenous. It would be an interesting extension to set up a model of sellers' behavior and estimate it.

References

Arnold, Barry C., N. Balakrishnan, and H. N. Nagaraja (1992), *A first course in order statistics*, New York: John Wiley & Sons.

Athey, Susan and Philip A. Haile (2002), "Identification of Standard Auction Models", *Econometrica*, 70(6), November 2002, 2107-2140.

Bajari, Patrick and Ali Hortaçsu (2002a), "Winner's curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auction, " *RAND Journal of Economics*, Summer 2003, 329-355.

Bajari, Patrick and Ali Hortaçsu (2002b), "Cyberspace Auctions and Pricing Issues: A Revies of Empirical Findings," *New Economic Handbook*, Edited by Derek C. Jones.

Bryan, Doug, David Lucking-Reiley, Naghi Prasad, and Daniel Reeve (2000), "Pennies from eBay: the Determinants of Price in Online Auctions", Working Paper, University of Arizona.

Cabral, Luis and Ali Hortaçsu (2003), "The Dynamics of Seller Reputation: Theory and Evidence from eBay", Working paper, University of Chicago.

Coppejans, Mark and A. Ronald Gallant (2002), "Cross-validated SNP density estimates, " *Journal of Econometrics*, 110, 27-65.

Fenton, Victor M. and A. Ronald Gallant (1996a), "Convergence Rates of SNP Density Estimators," *Econometrica*, 64(3), 719-27.

Fenton, Victor M. and A. Ronald Gallant (1996b), "Qualitative and asymptotic performance of SNP density estimators," *Journal of Econometrics,* 74, 77-118.

Engelbrecht-Wiggans, Richard, John A. List, and David Lucking-Reiley (1999), "Demand Reduction in Multi-unit Auctions with Varying Numbers of Bidders: Theory and Field Experiments", Working Paper, University of Arizona.

Gallant, A. Ronald and Douglas Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong (2000), “Optimal nonparametric estimation of first-price auctions,” *Econometrica*, 68, 525-574.

Haile, Philip A. and Elie Tamer (2003), “Inference with an Incomplete Model of English Auctions,” *Journal of Political Economy,* 111 (1), 1-51.

Hossain, Tanjim and John Morgan (2003), “A Test of the Revenue Equivalence Theorem using Field Experiments on eBay," Working Paper, UC-Berkeley.

Houser, Daniel and John Wooders (2001), "Reputation in Auctions: Theory, and Evidence from eBay," Working Paper, University of Arizona.

Krishna, Vijay (2002), *Auction Theory*, San Diego: Academic Press.

Laffont, Jean-Jacques (1999), “Game Theory and Empirical Economics: The Case of Auction Data,” *European Economic Review*, 41, 1-35.

Laffont, Jean-Jacques, Herve Ossard, and Quang Vuong (1995), “Econometrics of First price auctions, ” *Econometrica*, 63, 953-980.

Lucking-Reiley, David (2000), “Auctions on the Internet: What’s Being Auctioned, and How?”, *Journal of Industrial Economics*, 48, 227-252.

Myerson, Roger B. (1981), "Optimal Auction Design, " *Mathmatics of Operations Research*, 6, 58-73.

Ockenfels, Axel and Alvin E. Roth (2002), "Late Bidding in Second Price Internet Auctions: Theory and Evidence Concerning Different Rules for Ending an Auction," Working Paper, Harvard University.

Paarsch, Harry J. (1997), “Deriving an Estimate of the Optimal Reserve Price: An application to British Columbian Timber Sales,” *Journal of Econometrics*, 51, 191-215.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood (2002), "The Value of Reputation on eBay: A Controlled Experiment," Working Paper, School of Information, University of Michigan.

Resnick, Paul and Richard Zeckhauser (2002), "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay’s Reputation System," *The Economics of the Internet and E-Commerce*, Michael R. Baye, editor, Volume 11 of *Advances in Applied Microeconomics.*

Roth, Alvin E. and Axel Ockenfels (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 (4), 1093-1103.

Roehrig, Charles (1988), "Conditions for Identification in Nonparametric and Parametric Models, " *Econometrica*, 56, 433-477.

Song (2004), "Nonparametric Identification and Estimation of a First-price Auction Model with an Uncertain Number of Bidders", Working paper, University of British Columbia.

Yin, Pai-Ling (2002), "Information Dispersion and Auction Prices," Working Paper, Harvard University.

Appendix

Appendix A: Identification from the Lowest Two Bids

Let $i = k_1$ and $j = k_2$ where $1 \leq k_1 < k_2 \leq n$, for notational convenience, and let $F(\cdot|r) = \frac{F(x)}{F(r)}$ and $f(\cdot|r) = \frac{f(x)}{F(r)}$.

From equations (1) and (2),

$$p_{(k_1,k_2)}(x|y) = \begin{cases} \frac{(k_2-1)!}{(k_1-1)!(k_2-k_1-1)!} \frac{F(x)^{k_1-1}[F(y)-F(x)]^{k_2-k_1-1}f(x)}{F(y)^{k_2-1}}, & y \geqq x \\ 0, & otherwise \end{cases}$$

where y : k_2th order statistic, x : k_1th order statistic.

For $x \leq y$,

$$\begin{aligned} p_{(k_1,k_2)}(x|y) &= \frac{(k_2-1)!}{(k_1-1)!(k_2-k_1-1)!} \times \\ &\quad \frac{[F(y)F(x|y)]^{k_1-1}[F(y)(1-F(y|r))]^{k_2-k_1-1}F(y)f(x|y)}{F(y)^{k_2-1}} \\ &= \frac{(k_2-1)!}{(k_1-1)!(k_2-k_1-1)!}F(x|y)^{k_1-1}[1-F(x|y)]^{k_2-k_1-1}f(x|y) \\ &= \frac{(k_2-1)!}{(k_1-1)!(k_2-1-k_1)!}F(x|y)^{k_1-1}[1-F(x|y)]^{k_2-1-k_1}f(x|y) \\ &= f^{(k_1:k_2-1)}(x|y). \end{aligned}$$

Appendix B: Identification of the Distribution of the Number of Potential Bidders

A random variable, Y_t, represents the second-highest bid. In contrast to my use of only auctions in which $M_t \geq 3$ for identification of $F(\cdot)$, I here include auctions in which $M_t < 3$ as well. If $M_t \leq 1$, $Y_t = 0$. Let $p_i = \Pr(N_t = i)$ where $i = 1, ..., l$, and $G(y)$ denote the cumulative distribution of Y_t. I start with the case in which starting prices are not binding and Appendix C below discusses the case in which there is a binding starting price. Applying Equation (3) and rearranging terms lead to:

$$
\begin{aligned}
G(y) &= p_0 + p_1 + \sum_{i=2}^{l} p_i F^{(i-1:i)}(y) \qquad (8) \\
&= p_0 + p_1 + \sum_{i=2}^{l} p_i [iF(y)^{i-1} - (i-1)F(y)^i] \\
&= p_0 + p_1 + 2p_2 F(y) + (3p_3 - p_2)F(y)^2 + \\
&\quad \cdots\cdots + [lp_l - (l-2)p_{l-1}]F(y)^{l-1} - (l-1)p_l F(y)^l.
\end{aligned}
$$

Even if a starting price is not binding, generally $N_t \neq M_t$, because a standing price is raised to the second-highest cutoff price as an auction proceeds. However, $N_t = M_t$ when $M_t \leq 1$; therefore, the identification of p_0 and p_1 is obvious. The previous section proves identification of $F(\cdot)$, and $G(\cdot)$ is observed. Hence the coefficients of $F(y)$, $F(y)^2, ..., F(y)^l$ in Equation (8) are identified; this implies identification of $\{p_i\}_{i=2}^{l}$.

Binding Starting Prices

If a seller sets a binding starting price, $s > \underline{v}$, no auctions can reveal information about $F(v)$ for $v < s$ or potential bidders with valuations below s. However, the previous identification results extend to identify the truncated distribution, $F(\cdot|s) = \frac{F(\cdot)-F(s)}{1-F(s)}$ and distribution of N_t' where N_t' represents the number of potential bidders with valuations above s.

Corollary 2 *Suppose we observe a pair of order statistics from an arbitrary absolutely continuous distribution $F(\cdot)$, where each order statistic is greater than s. Then $F(v|s) = \frac{F(v)-F(s)}{1-F(s)}$ is identified where $v \geq s$.*

Proof. The order statistics are observed on the condition that they are greater than s. So the (joint) distribution of the observed order statistics is equal to the (joint) distribution of order statistics from $F(v|s)$. The result is then immediate by applying Lemma 1. ■

The identification of the distribution of N_t' is easily established from the previous section result. Let $q_i = \Pr(N_t' = i)$ and $Y_t = s$ if $N_t' \leq 1$. Equation (8) is still satisfied with q_i substituted for p_i and with $F(y|s)$ substituted for $F(y)$. So the result follows.

Appendix C: Graph of Benchmark Monte Carlo Experiments

The 100 Best SNP are obtained through application of the series length that minimizes a true ISE in each experiment. Figure 1a presents the graph of the density estimate of which true ISE is the smallest among 100 Best SNP estimates. Figure 1b graphs the density estimate of which the true ISE is the 50th. Finally, Figure 1c graphs the density estimate of which the true ISE is the biggest. Throughout these three figures, the left graph presents a density estimate of $f(\cdot)$, and the right graph presents a density estimate of V_t^i conditional on the median of (SR_1, SR_2). It is important to remember that potential bidders' valuations were generated according to the equation:

$$\ln V_t^i = \alpha_1 SR_{1t} + \alpha_2 SR_{2t} + \nu_t^i \tag{7}$$

where $\alpha_1 = 1$, $\alpha_2 = -1$, $SR_{1t} \backsim N(0,1)$, $SR_{2t} \backsim Exp(1)$, and $\nu_t^i \backsim f(\cdot) = Gamma(9,3)$. The solid line represents the true density, and the dotted line represents an estimate.

Figure 1a: The Best Performance

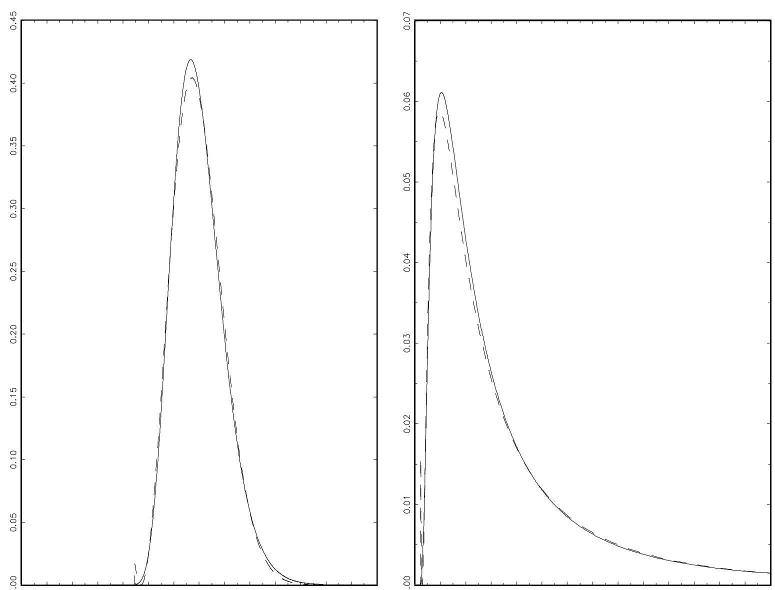


Figure 1b: The Median Performance

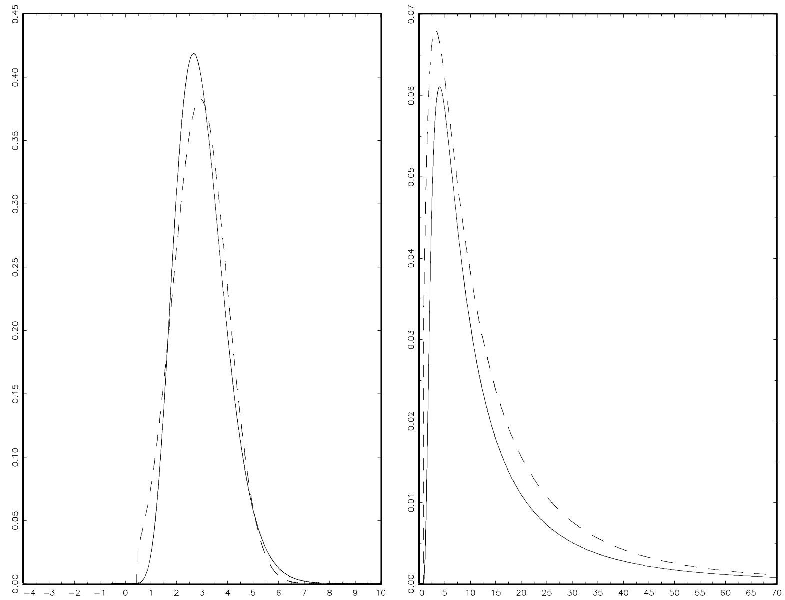
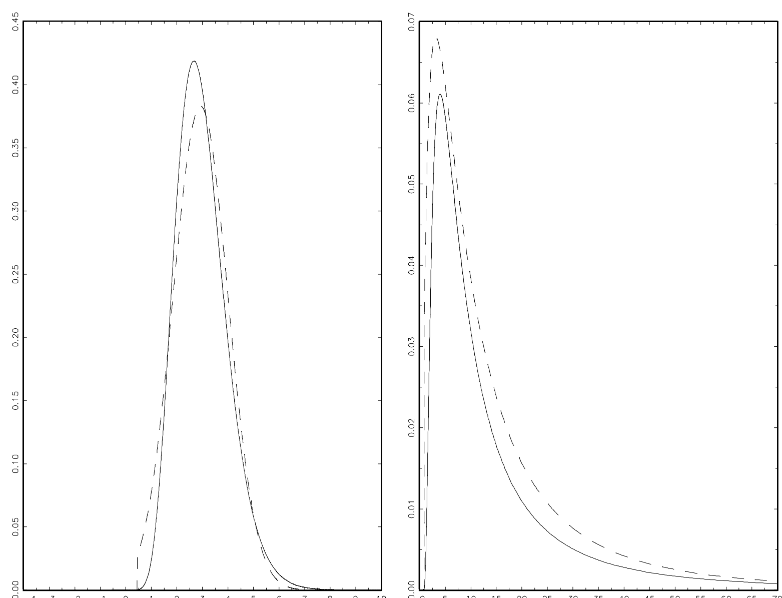


Figure 1c: The Worst Performance



Appendix D: Graph of Monte Carlo Experiments under eBay Circumstances

The 50 (experiments) $\times$ 4 (subsets) Best SNP are computed in the same way as in the benchmark Monte Carlo Experiments. Each figure below presents the graph of the density estimate of which the true ISE is the 25th among 50. As before, the left graph presents a density estimate of $f(\cdot)$, and the right graph presents a density estimate of V_t^i conditional on the median of (SR_1, SR_2). The solid line represents the true density, and the dotted line represents an estimate.

Figure 2a: Using simulated data from auctions in which the highest or the second-highest bidder's bidding time is no earlier than 10 (time) before an auction ends

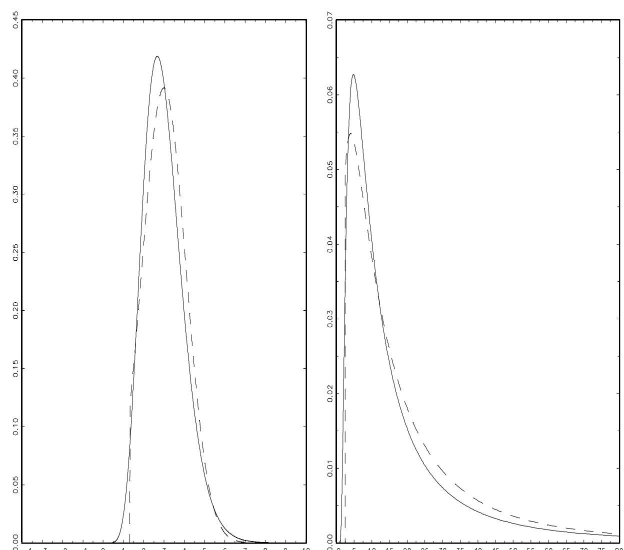


Figure 2b: No earlier than 25 (time) before the end

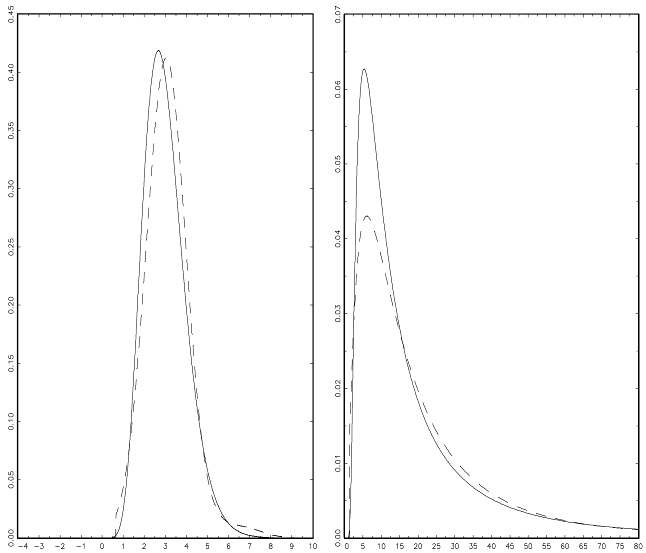


Figure 2c: No earlier than 50 (time) before the end

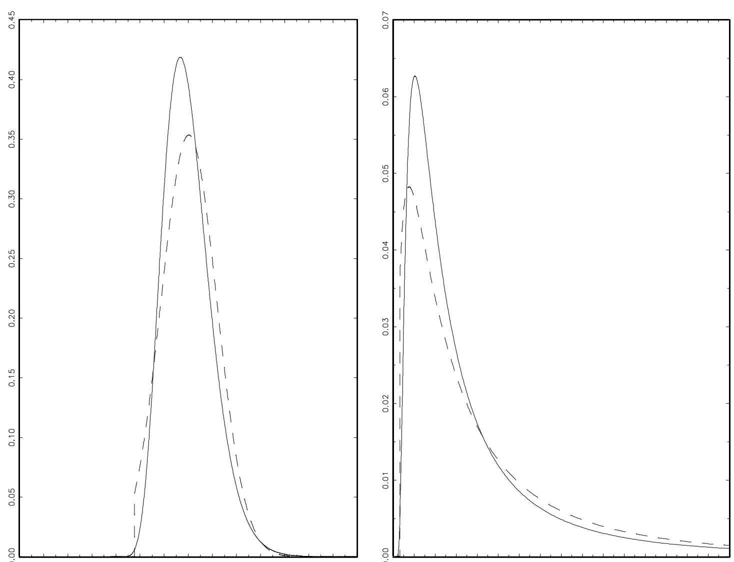
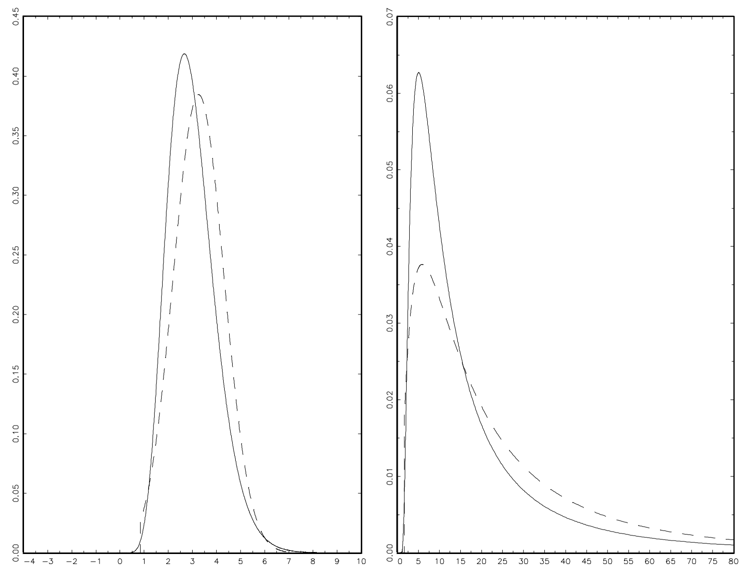


Figure 2d: Using simulated data from all auctions



Appendix E: Descriptive Statistics

Table 3 presents summary statistics on the dataset that includes all auctions: averages and standard deviations in parentheses. While no auctions in which sellers advertise yearbooks on the basis of any particular person are used for estimation of bidders' value distribution, the summary statistics below are for the dataset including some auctions in which a seller advertises a particular person. Actually, I first dropped only auctions in which sellers advertise "famous" persons (i.e., I did not exclude auctions of which sellers advertise persons, who did not seem to be famous), and later on I dropped any auctions if a seller advertised a particular person in that auction. I made the second drop only for auctions used for estimation of bidders' value distribution, because it would take too long to examine all the auctions in the dataset. Here, for the sake of comparison, all statistics in Table 3 and 4 are for the dataset before the second drop.

Table 3: Summary Statistics of All Auctions: Averages and Standard Deviations in Parentheses

	Sellers' Ratings			Sale
# obs in parentheses	Positive	Neutral	Negative	Price
All auctions (14,151)	1176(1587)	5.65(15.04)	3.63(12.19)	
All with Sale (6,612)	1130(1485)	5.13(13.62)	3.09(10.43)	$14.97(16.49)
Zero Bid (7,496)	1215(1672)	6.10(16.19)	4.09(13.56)	
End by "Buy It Now"	975(1016)	3.78(9.20)	2.21(6.41)	*
One Bidder (3,494)	1181(1611)	5.51(14.83)	3.41(11.5)	$9.54(6.31)
Two Bidders (1,674)	1121(1352)	5.04(11.39)	2.85(8.70)	$15.81(14.03)
Over Two (1,160)	1039(1377)	4.59(13.71)	2.79(10.15)	$30.13 (27.59)

* "Buy It Now" price does not remain on the auction site after an auction is ended by a bidder's use of the "Buy It Now" option. We can know only whether or not an auction is ended by the "Buy It Now" option.

* The number of auctions ended by a bidder's use of the "Buy It Now" option is 327.

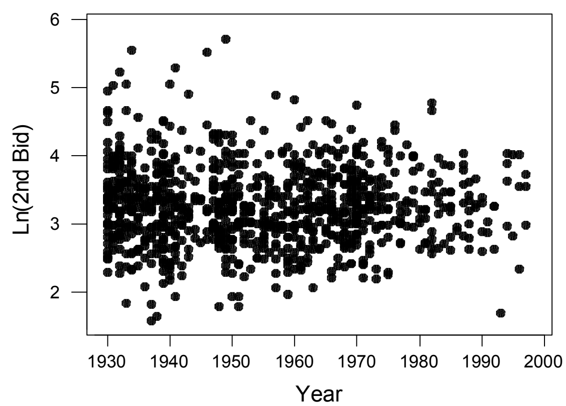
Table 4: Detailed Descriptive Stat. of Auctions in Which Bidders > 2

	Mean	Median	STD	Min	Max
Price	$30.13	$22.50	$27.59	$1.37	$305
2nd highest bid	$29.32	$22.00	$27.16	$1.12	$300
3rd highest bid	$18.71	$14.5	$17.31	$.01	$162
Positive ratings	1,039	633	1,377	0	17,865
Neutral ratings	4.59	1	13.71	0	226
Negative ratings	2.79	1	10.15	0	180
Starting price	$8.35	$8.99	$4.99	$.01	$39.99
Number of bidders	3.61	3	.99	3	10
Number of bids	5.84	5	3.08	3	26
Number of auction days	7.21	7	1.25	3	10

Appendix F: Auction Characteristics

I examined descriptive statistics to see if there are auction characteristics (other than sellers' ratings) which have effects on a bidder's willingness-to-pay. The descriptive statistics strongly suggest that none of auction characteristics examined below affect a bidder's willingness-to-pay.

(1) Year – Plot of ln*(the second-highest bids) vs. year of yearbook:*



(2) Picture

– Descriptive statistics of the second-highest bids under two different categories:

	With advertising image (834 obs)	Without advertising image (89 obs)
Mean	$31.82	$30.63
Median	$25.49	$25.5
STD	$25.47	$23.25

(3) Payment methods

– Descriptive statistics of the second-highest bid under two different categories:

	With online payment (724 obs)*	Without online payment (199 obs)
Mean	$31.93	$30.89
Median	$25.50	$25.00
STD	$25.43	$24.64

* Auctions in which a seller provides a payment method that enables a bidder to pay online, such as PayPal, credit card, and so on.

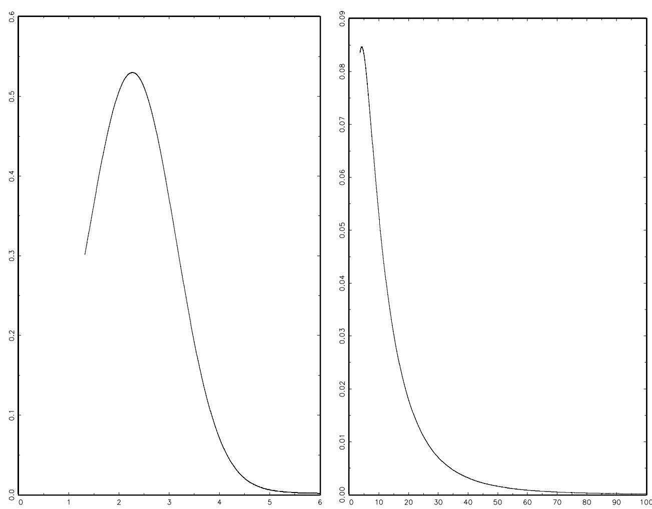
Appendix G: Graph of the Density Estimate

The left graph presents a density estimate of ν_t^i, and the right graph presents a density estimate of V_t^i conditional on the median of (SR_{1t}, SR_{2t}). It is useful to recall that the specification of a bidder's valuation is:

$$\ln V_t^i = \alpha_1 SR_{1t} + \alpha_2 SR_{2t} + \nu_t^i$$

where SR_{1t} is the number of seller's positive ratings at auction t, and SR_{2t} is (the number of seller's neutral and negative ratings) / (SR_{1t} +1).

Figure 3



Appendix H: Examination of an Assumption that the Virtual Value Is an Increasing Function in the Valuation

The left graph presents a density estimate of V_t^i, and the right graph presents $(1-\widehat{F}(v))/\widehat{f}(v)$, conditional on the median of (SR_{1t}, SR_{2t}).

Figure 4