Nonparametric Identification and Estimation of a First-price Auction Model With an Uncertain Number of Bidders

Unjy Song

University of British Columbia

Abstract

In this paper, I develop new methods for structural estimation of a first-price auction model with a stochastic number of bidders in which a bidder does not know how many competitors he faces when he submits a bid. Within the symmetric independent private values (IPV) model, I show that the first and second-highest bids nonparametrically identify bidders' value distribution and the distribution of the number of potential bidders. Unlike previous studies, knowledge of neither the number of potential bidders nor the number of potential bidders with valuations above the reserve price is necessary. The essence is that it is shown that the potential bidders' equilibrium mark-down is identified from the first and second-highest bids. Once the identification of bidders' equilibrium markdown is established, identification of bidders' value distribution and the distribution of the number of bidders is obtained by applying the results in Gurre, Perrigne, and Vuong (2000, Econometrica) and Song (2007, UBC). I then propose a consistent estimator of the bidders' value distribution for both cases: (1) a case in which all bids are available; and (2) a case in which only top two bids are available. I evaluate the estimator for the first case by Monte Carlo experiments and the results show that the estimator performs very well.

1 Introduction

In this paper, I develop a new method for structural estimation of a first-price auction model with a stochastic number of bidders in which a bidder does not know how many competitors he faces when he submits a bid. Recently structural analysis of auction data have attracted much interest mainly because structural analysis is essential to implementing an optimal auction mechanism. In spite of the extensive theory concerning mechanism design, it is impossible to determine the revenue-maximizing selling mechanism without knowledge of the potential bidders' value distribution. Furthermore, structural analysis often provides methods to test the predictions of game theory and/or decide which model fits a certain market better among the competing models.

The goal of the structural analysis is estimating the latent structure, typically the potential bidders' value distribution, from the observed bids. A natural strategy is computing the inverse bidding function which provides a map from the observed bids onto the latent structure. In the standard first-price auction model, however, the inverse bid function is not explicitly solved unless simple parametric specifications are imposed. Guerre, Perrigne, and Vuong (2000) resolve this difficulty within the symmetric independent private value (IPV) model. They develop an indirect two-step procedure for estimating the distribution of bidders' private values from observed bids, which requires neither parametric assumptions nor the computation of the inverse bidding function. Li, Perrigne, and Vuong (2000) and Li, Perrigne, and Vuong (2002) extend the results to the conditionally independent private information model and the affiliated private value model.

This paper extends the previous literature in two directions within the symmetric IPV model. First, I newly show that only top two bids identify the potential bidders' value distribution and the distribution of the number of potential bidders.[1] All existing results need observation of all bids above the reserve price or information regarding the number of potential bidders – either the number of potential bidders or the number of potential bidders with valuations above the reserve price. Even in sealed-bid auctions,

[1] Actually, in the model presented in Session 2, the number of potential bidders is constant and part of them become active bidders. Hence the number of active bidders is stochastic.

some potential bidders with valuations above the reserve price will not take part in auctions if there are positive bid preparation costs as in the model presented in Samuelson (1985). In that case, the researcher cannot recognize every potential bidder's existence. Moreover, the estimation strategy which requires only top two bids could be useful even to empirical analysis of auctions in which all bids are available. Generally, if we can estimate a model from part of observables, there is room to test the considered model by exploiting the observables which is not used for estimation of the model.

Second, I allow the number of potential bidders to be stochastic. The entry model proposed in Levin and Smith (1994) generates the stochastic number of potential bidders. In addition, the number of potential bidders is stochastic in some government auctions in which only selected contractors are invited to submit bids (McAfee and McMillan, 1987). If we assume that both bidders and researcher know the realization of the number of potential bidders, econometric methods for the constant number of potential bidders are easily extended to the case of the stochastic number of potential bidders. However, in a sealed-bid auction, it is not always true that bidders know the number of their competitors, since the bidders do not assemble together in one place. Actually, we often observe that a bidder submitted a bid greater than the reserve price even if he was only bidder; this implies that bidders are uncertain about the number of their competitors when they make a bid. The econometric methods in this paper is for the case where bidders know the distribution of the number of their competitors but do not know the actual number of their competitors. I consider two cases: (i) the case in which all bids are available; and (ii) the case in which only top two bids are available. Although there exists an obvious estimation method for the first case, I proposes an alternative, simple one.

The rest of the paper is organized as follows: The next session presents a first-price auction model with a stochastic number of bidders. In Session 2, the most crucial part of this paper, I provide an identification result. Consistent estimators are proposed in Session 4 and Monte Carlo simulation results are presented in Session 5. Session 6 mentions a tentative application very briefly. I make concluding remarks in Session 7.

2 Model within the Symmetric IPV Framework

I consider a first-price sealed-bid auction of a single indivisible good. Throughout the paper, I adopt the convention of denoting random variables in upper case, and their associated realizations in lower case. There are l risk-neutral *potential bidders*, and each potential bidder becomes an *active bidder* with the same probability. A is the set of active bidders and a random variable, N represents the number of active bidders. Let p_n denote $\Pr(N = n)$. Each active bidder i draws his valuation V^i independently from the absolutely continuous distribution $F(v)$, having support on $[\underline{v}, \overline{v}]$ ($\underline{v} > 0$). Since the presence of a reserve price does not change analysis much, I assume there is no reserve price for simplicity.[2] [3] Each active bidder knows only his valuation, the distribution $F(v)$, and the probabilities p_n. However, unlike the most empirical auction literature, an active bidder is assumed to be unaware of the realization of N. Every active bidder takes part in the auction.[4]

In a first-price auction, each bidder submits a sealed bid and the highest bidder buys the object at a price equal to his bid. I consider a symmetric, increasing, and differentiable Bayesian-Nash equilibrium strategy $\beta(v)$ which specifies the active bidder's bid amount as a function of his valuation. Define $q_n^i = \Pr(N = n | i \in A)$ as the probability that the number of active bidders is n *including* bidder i conditional on his being an active bidder. Clearly, $q_0^i = 0$. Since each potential bidder becomes an active bidder with the same probability, q_n^i is the same for all i. Hence define $q_n = q_n^i \ \forall \ i$. Krishna (2002) then

[2] If a seller sets a binding reserve price $r > 0$, only the active bidder with valuation above r takes part in the auction. In that case, we can identify only the truncated distribution $F(v|V \geq r)$ and p'_n $(n = 1, ..., l)$ when p'_n is defined as follows.

$$p'_n = \sum_{k=n}^{l} p_k \cdot \begin{pmatrix} k \\ n \end{pmatrix} (1 - F(r))^n F(r)^{k-n}$$

[3] Even though a seller does not set a reserve price, zero is a reserve price in practice. The followed analysis in the paper can be easily adapted by changing "0" to "reserve price" in auctions with a reserve price.

[4] A method in Sessions 4 will be applicable if top two active bidders take part in the auction.

shows that $\beta(V)$ is characterized as follows:

$$\beta(v) = \sum_{n=2}^{l} \left\{ \frac{q_n \cdot F(v)^{n-1}}{(q_1 \ + \ \sum_{n=2}^{l} q_n \cdot F(v)^{n-1})} \cdot \left(\ v - \frac{1}{F(v)^{n-1}} \int_{\underline{v}}^{v} F(u)^{n-1} du \ \right) \right\}. \tag{1}$$

For empirical purposes, as first suggested in Guerre, Perrigne, and Vuong (2000), it is more convenient to express each active bidder's valuation as a function of his bid and the distribution of all active bidders' bids. Hence I derive necessary conditions of $\beta(V)$ which are used for identification and estimation of the model in next two sessions. I first define necessary notations. Define $G(b) = F(\beta^{-1}(b))$. In other words, $G(b)$ is the distribution of $\beta(V)$. An associated density of $G(b)$ is denoted by $g(b)$. Let $G^{(i:n)}(b)$ denote the distribution of the ith order statistic from an i.i.d. sample of size n from the distribution $G(b)$ with having $g^{(i:n)}(b)$ as a density. Finally, M^i denote the maximum among the bids of an active bidder i's competitors. If there are no competitors, $M^i = 0$. Since I consider a symmetric model, the distribution of M^i is the same for all i. The same distribution of M^i is denoted by $G^M(b)$. The distribution $G^M(b)$ then can be written as follows:

$$G^M(b) = \begin{cases} q_1 + \sum_{n=2}^{l} q_n G^{(n-1:n-1)}(b), & b \geq 0 \\ 0, & b < 0 \end{cases} \tag{2}$$

Active bidder i's expected payoff is

$$(v^i - b^i) G^M(b^i)$$

where b^i is a bidder i's bid. Maximizing the expected payoff with respect to b^i yields the first-order condition:

$$v^i = b^i + \frac{G^M(b^i)}{g^M(b)} \tag{3}$$

where $g^M(b) = \frac{\partial}{\partial b} G^M(b) = \sum_{n=2}^{l} q_n g^{(n-1:n-1)}(b)$ where $b > 0$.[5] Every equilibrium bid satisfies Equation (3).

[5]Note that $g^M(b)$ is not a density function of $G^M(b)$ which is discontinuous at 0.

3 Nonparametric Identification

The model primitives are $F(v)$ and p_n $(n = 0, 1, ..., l)$. Identification depends on the data available. I derive identification results for the case where only the highest and second-highest bids are observed. I assume that the highest bid is recorded as zero if there are no active bidders and the second-highest bid is recorded as zero if there are less than two active bidders. Knowledge of the number of active bidders is not required. I start by Theorem 1 providing a new expression for the distribution of the highest competitor's bids conditional on a bidder's participation as a function of the distribution of top two bids.

Theorem 1 *Let*

$$Y1 = \begin{cases} \text{The highest bid,} & \text{if } N \geq 1 \\ 0 \quad , & \text{if } N = 0 \end{cases} \quad \text{and}$$

$$Y2 = \begin{cases} \text{The second-highest bid,} & \text{if } N \geq 2 \\ 0 \quad , & \text{if } N < 2 \end{cases} .$$

Let $H1(y1)$ denote the distribution of $Y1$, and $H2(y2)$ denote the distribution of $Y2$. Also define $h2(y2) = \frac{\partial}{\partial y2} H2(y2)$, $y2 > 0$.

Then,

$$\frac{G^M(b)}{g^M(b)} = \frac{H2(b) - H1(b)}{h2(b)}$$

Proof. Application of Lemma 1 of McAfee and McMillan (1987) yields:

$$q_n = \frac{np_n}{p_e} \tag{4}$$

From Equation (2) and (3),

$$\begin{aligned} G^M(b) &= q_1 + \sum_{n=2}^{l} q_n \cdot G^{(n-1:n-1)}(b) \\ &= \frac{p_1}{p_e} + \sum_{n=2}^{l} \frac{np_n}{p_e} \cdot G(b)^{n-1} \end{aligned}$$

$$\begin{aligned} g^M(b) &= \sum_{n=2}^{l} q_n g^{(n-1:n-1)}(b) \frac{\partial}{\partial b} G^M(b) \\ &= \sum_{n=2}^{l} \frac{np_n}{p_e} \cdot (n-1) \cdot G(b)^{n-2} \cdot g(b). \end{aligned}$$

Therefore,

$$\frac{G^M(b)}{g^M(b)} = \frac{p_1 + \sum_{n=2}^{l} np_n \cdot G(b)^{n-1}}{\sum_{n=2}^{l} np_n \cdot (n-1) \cdot G(b)^{n-2} \cdot g(b)} \tag{5}$$

On the other hand,

$$\begin{aligned} H1(b) &= p_0 + \sum_{n=1}^{l} p_n G^{(n:n)}(b) \\ &= p_0 + \sum_{n=1}^{l} p_n G(b)^n \\ H2(b) &= p_0 + p_1 + \sum_{n=2}^{l} p_n G^{(n-1:n)}(b) \\ &= p_0 + p_1 + \sum_{n=2}^{l} p_n \left[nG(b)^{n-1} - (n-1)G(b)^n \right] \\ &= p_0 + p_1 + \sum_{n=2}^{l} p_n nG(b)^{n-1}(1-G(b)) + \sum_{n=2}^{l} p_n G(b)^n \\ &= H1(b) - p_1 G(b) + p_1 + \sum_{n=2}^{l} p_n nG(b)^{n-1}(1-G(b)) \\ &= (1-G(b)) \left[p_1 + \sum_{n=2}^{l} p_n nG(b)^{n-1} \right] + H1(b) \\ h2(b) &= \sum_{n=2}^{l} p_n n(n-1)G(b)^{n-2} g(b)(1-G(b)) \end{aligned}$$

Therefore,

$$\frac{H2(b) - H1(b)}{1 - G(b)} = p_1 + \sum_{n=2}^{l} p_n nG(b)^{n-1}$$

and

$$\frac{h2(b)}{1-G(b)} = \sum_{n=2}^{l} np_n(n-1)G(b)^{n-2}g(b).$$

This shows that

$$\frac{H2(b)-H1(b)}{h2(b)} = \frac{p_1 + \sum_{n=2}^{l} p_n nG(b)^{n-1}}{\sum_{n=2}^{l} np_n(n-1)G(b)^{n-2}g(b)} = \frac{G^M(b)}{g^M(b)}.$$

■

Corollary 2 *The highest and second-highest bid identify $F(v)$ and p_n $(n = 0, 1, ..., l)$.*

Proof. (i) Identification of $F(v)$

Theorem 1 implies that $Y1$ and $Y2$ identify $G^M(\cdot)/g^M(\cdot)$. Once $G^M(\cdot)/g^M(\cdot)$ is identified, the joint distribution of the highest and the second-highest bidders' valuation is identified by employing Equation (3). Lemma 1 in Song(2007)[6] then implies identification of $F(v)$.

(ii) Identification of p_n $(n = 0, 1, ..., l)$

Let denote by $F1(v)$ the distribution of the highest bidders' valuation. Then

$$F1(v) = p_0 + p_1F(v) + p_2F(v)^2 + \cdots\cdots + p_lF(v)^l.$$

Given identification of $F1(\cdot)$ and $F(\cdot)$, the coefficients of $F(y)$, $F(y)^2, ..., F(y)^l$ in the above equation is identified; therefore, identification of p_n $(n = 0, 1, ..., l)$ is established. ■

4 Nonparametric Estimation of $F(v)$

I consider a sample of T independent first-price auctions. Assume that the active bidders' value distribution, the distribution of the number of active bidders are fixed in all auctions in the sample. Let N_t denote the number of active bidders at auction t. At each auction t, $B_t^{(1:N_t)}, ..., B_t^{(N_t:N_t)}$ are the order statistics of active bidders' bids with $B_t^{(k:N_t)}$ denoting the kth-lowest among N_t. Bidder i's bid is denoted by B_t^i. $(i = 1, 2, ..., N_t)$.

[6] 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 is unknown and stochastic.

For a consistent estimator of $F(v)$, I will employ a two-step nonparametric estimation procedure which is similar to that in Guerre, Perrigne, and Vuong (2000) or Li, Perrigne, and Vuong (2000):

Step 1: Construct a sample of pseudo private values based on Equation (3) using a nonparametric estimate of $G^M(\cdot)/g^M(\cdot)$ from observed bids.

Step 2: Use the pseudo sample constructed in Step 1 to estimate $F(v)$ nonparametrically.

I propose a consistent estimator of $F(v)$ for two different cases. I start by the case in which all bids are available and then consider the case in which only top two bids are available.

4.1 Case 1: All bids are available

If all active bidders' bids are available, there is an obvious estimation procedure for $G^M(\cdot)/g^M(\cdot)$: we first estimate the distributions of N_t and B_t^i from the number of observed bidders and their bids. $G^M(\cdot)$ is a function of q_i and $G^{(n-1:n-1)}(b)$ $(i = 1, ..., l)$, which depend on the distributions of N_t and B_t^i respectively. Hence we can obtain a consistent estimator of $G^M(\cdot)/g^M(\cdot)$. In this session, I propose another, more simple estimation procedure which does not require separate estimation of the distributions of N_t and B_t^i.

For convenience, let $s = \sum_{t=1}^{T} n_t$. Construct a random variable, M_t^i: the maximum among bids of bidder i's competitors at auction t. If there are no competitors, $M_t^i = 0$. By construction, we can obtain s realizations of M_t^i. An important observation is that the distribution of M_t^i is $G^M(\cdot)$; this is shown in Theorem 3.

In the first step, noting that $M_t^i \backsim G^M(\cdot)$, I estimate $G^M(\cdot)/g^M(\cdot)$ by $\widehat{G^M}(\cdot)/\widehat{g^M}(\cdot)$, where

$$\begin{aligned}
\widehat{G^M}(b) &= \frac{1}{s}\sum_{t=1}^{T}\sum_{i=1}^{n_t} 1(m_t^i \leq b) \\
\widehat{g^M}(b) &= \frac{1}{s \cdot \lambda_{g^M}}\sum_{t=1}^{T}\sum_{i=1}^{n_t} K_{g^M}\left(\frac{b - m_t^i}{\lambda_{g^M}}\right) \cdot 1(m_t^i \neq 0).
\end{aligned}$$

where λ_{g^M} is a bandwidth and K_{g^M} is a kernel with ρ_{g^M}-length support. The term, $1(m_t^i \neq 0)$, is necessary because $G^M(b)$ has mass at 0. Note that $g^M(b)$ was not a density function of $G^M(b)$.

The idea of constructing M_t^i is an extension of Li, Perrigne, and Vuong (2000) where the number of active bidders is the same in all auctions. M_t^i is distributed according to $G^{(n_t-1:n_t-1)}(b)$. Hence, in the constant number of bidders case, M_t^i is distributed according to $G^M(\cdot)$ for all auction t. On the other hand, in the stochastic number of bidders case, the distribution of M_t^i is different according to n_t, and $G^M(\cdot)$ is a weighted sum of different $G^{(n_t-1:n_t-1)}(b)$. Therefore how to reflect M_t^i from different auctions to the estimation of $G^M(\cdot)$ is a critical issue. For reference, $\widetilde{G^M}(b)$ defined as follows might be tempting as an estimator of $G^M(b)$, but $\widetilde{G^M}(b)$ is inconsistent.

$$\widetilde{G^M}(b) = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{n_t}\sum_{i=1}^{n_t} 1(m_t^i \leq b)\right\} \tag{7}$$

Generally $\widetilde{G^M}(b)$ is different from $\widehat{G^M}(b)$, unless n_t is the same for all t.

$$\begin{aligned}
\widetilde{G^M}(b) &= \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{n_t}\sum_{i=1}^{n_t} 1(m_t^i \leq b)\right\} \\
&= \sum_{t=1}^{T}\left\{\sum_{i=1}^{n_t}\frac{1}{T}\frac{1}{n_t}\cdot 1(m_t^i \leq b)\right\} \\
&\neq \sum_{t=1}^{T}\sum_{i=1}^{n_t}\left\{\frac{1}{\sum_{t=1}^{T} n_t}\cdot 1(m_t^i \leq b)\right\} \\
&= \widehat{G^M}(b).
\end{aligned}$$

Even if $T \to \infty$, $\widetilde{G^M}(b) \neq \widehat{G^M}(b)$. Hence $\widetilde{G^M}(b)$ cannot be a consistent estimator if $\widehat{G^M}(b)$ is. The consistency of $\widehat{G^M}(b)$ is shown in proof of Theorem 3.

Since the kernel density estimator is biased at the boundaries, the same trimming as proposed in Guerre, Perrigne, and Vuong (2000) should be applied. Let $M_{\min}$ and $M_{\max}$ be the minimum and maximum of the all M_t^i except when $M_t^i = 0$. Define $C_S = [M_{\min} + \rho_{g^M}\lambda_{g^M}/2,\ M_{\max} - \rho_{g^M}\lambda_{g^M}/2]$. By applying Equation (3), I define the pseudo value $\widehat{V_t^i}$ corresponding to B_t^i as

$$\widehat{V_t^i} = \left\{ \begin{array}{lll} 0, & \text{if} & B_t^i < B_{\min} + \rho_{g^M}\lambda_{g^M}/2 \\ B_t^i + \widehat{G^M}(B_t^i)/\widehat{g^M}(B_t^i), & \text{if} & B_t^i \in C_S \\ \infty, & \text{if} & B_t^i > B_{\max} - \rho_{g^M}\lambda_{g^M}/2 \end{array} \right\}. \tag{8}$$

In the second step, I use the pseudo sample $\{v_t^i\}$ to estimate nonparametrically $F(v)$ by

$$\widehat{F}(v) = \frac{1}{s}\sum_{t=1}^{T}\sum_{i=1}^{n_t} 1(v_t^i \leq v). \tag{9}$$

Regarding λ_{g^M} and K_{g^M}, I make standard assumptions as follows.

(A1) $\lambda_{g^M} \rightarrow 0$ and $S \cdot (\lambda_{g^M})^2 \rightarrow \infty$, as $S \rightarrow \infty$.

(A2) $K_{g^M}(\cdot)$ is the class of all Borel measurable bounded real-valued functions such that

(i) $\int K_{g^M}(x)dx = 1$,

(ii) $\int |K_{g^M}(x)|dx < \infty$,

(iii) $|x||K_{g^M}| \rightarrow 0$, as $|x| \rightarrow \infty$,

(iv) $\sup |K_{g^M}| < \infty$,

(v) $\int |K_{g^M}^2(x)|dx < \infty$,

(vi) The characteristic function of K_{g^M} is absolutely integrable.

The next theorem establishes the uniform consistency of the above two-step nonparametric estimator.

Theorem 3 *Suppose (A1) and (A2) hold. Then $\widehat{F}(v)$ given in Equation (9) is uniformly consistent as $S \rightarrow \infty$.*

Proof. For the uniform consistency, I consider the supreme norm.

(i) $\widehat{G^M}(b)$ is uniformly consistent to $G^M(b)$.

Since the uniform consistency of the empirical distribution is well-known, it is enough if I show that $M_t^i \sim G^M(\cdot)$. Let C_t^i denote a random variable representing the number of bidder i's competitors in auction t. Then

$$\begin{aligned}
\Pr(C_t^i &= n) = \frac{(n+1)p_{n+1}}{\sum\limits_{n=0}^{l-1}(n+1)p_{n+1}} \\
&= \frac{(n+1)p_{n+1}}{p_e}.
\end{aligned}$$

Accordingly,

$$\begin{aligned}
\Pr(M_t^i &\leq b) = \Pr(C_t^i = 0) + \sum_{n=1}^{l-1}\Pr(C_t^i = n)\cdot G^{(n:n)}(b) \\
&= \frac{p_1}{p_e} + \sum_{n=1}^{l-1}\frac{(n+1)p_{n+1}}{p_e}\cdot G(b)^n \\
&= \frac{p_1}{p_e} + \sum_{n=2}^{l}\frac{np_n}{p_e}\cdot G(b)^{n-1} \\
&= q_1 + \sum_{n=2}^{l} q_n \cdot G(b)^{n-1} \\
&= G^M(b).
\end{aligned}$$

(ii) $\widehat{g^M}(b)$ is uniformly consistent to $g^M(b) = \sum\limits_{n=2}^{l} q_n g^{(n-1:n-1)}(b)$.

Define a random variable, M': M_t^i conditional on that $M_t^i \neq 0$. Noting that $\Pr(M_t^i = 0) = q_1$, the density of M', $g^{M'}$is

$$\begin{aligned}
g^{M'}(b) &= \frac{1}{1-q_1}\cdot\sum_{n=2}^{l} q_n \cdot g(b)^{n-1}(b). \\
&= \frac{1}{1-q_1}\cdot g^M(b).
\end{aligned}$$

Accordingly,

$$\begin{aligned}
g^M(b) &= (1-q_1)\cdot g^{M'}(b) \\
&= (1-\Pr(m_t^i \neq 0))\cdot g^{M'}(b)
\end{aligned}$$

Noting that

$$\widehat{g^M}(b) = \frac{(S - \sum 1(m_t^i \neq 0))}{S}\cdot\frac{1}{(s - \sum 1(m_t^i \neq 0))\cdot\lambda_{g^M}}\sum_{t=1}^{T}\sum_{i=1}^{n_t} K_{g^M}(\frac{b-m_t}{\lambda_{h2}})1(m_t^i \neq 0).$$

$$\widehat{g^M}(b) = (1 - \widehat{\Pr}(m_t^i \neq 0)) \cdot \widehat{g^{M'}}(b)$$

where

$$\begin{aligned} \widehat{\Pr}(m_t^i &\neq 0) = \frac{\sum 1(m_t^i = 0)}{s} \qquad \text{and} \\ \widehat{g^{M'}}(b) &= \frac{1}{(s - \sum 1(m_t^i \neq 0)) \cdot \lambda_{g^M}} \sum_{t=1}^{T} \sum_{i=1}^{n_t} K_{g^M} \left(\frac{b - m_t}{\lambda_{h2}} \right) \cdot 1(m_t^i \neq 0). \end{aligned}$$

Since $\widehat{\Pr}(m_t^i \neq 0)$ is a consistent estimator of $\Pr(m_t^i \neq 0)$ and $\widehat{g^{M'}}(b)$ is uniformly consistent to $g^{M'}(b)$, the uniform consistency of $\widehat{g^M}(b)$ is established.

As $S \to \infty$, $M_{\min} \to \beta(\underline{v})$ and $M_{\max} \to \beta(\overline{v})$. Hence given the uniform consistency of $\widehat{G^M}(b)$ and $\widehat{g^M}(b)$, it is straightforward that $\widehat{V_t^i}$ converges uniformly to true value, V_t^i.

Let $F'(\cdot)$ be the true distribution of $\widehat{V_t^i}$. Using the triangle inequality,

$$\left\| \widehat{F}(v) - F(v) \right\| \leq \left\| \widehat{F}(v) - F'(v) \right\| + \left\| F'(v) - F(v) \right\|.$$

Since $\widehat{F}(v)$ is an empirical distribution of $\widehat{V_t^i}$, $\widehat{F}(v)$ converges uniformly to $F'(\cdot)$. Since $F'(v)$ and $F(v)$ are bounded, the uniform consistency of $\widehat{V_t^i}$ to V_t^i implies that $F'(\cdot)$ converges uniformly to $F(v)$. Accordingly, the uniform consistency of $\widehat{F}(v)$ is established. ■

4.2 Case 2: Only top two bids are available

Let $Y1_t$ be the highest bid at auction t and $Y2_t$ be the second-highest bid. By Theorem 1, $G^M(b)/g^M(b) = \{H2(b) - H1(b)\}/\{h2(b)\}$. Accordingly, in the first step, I estimate $G^M(\cdot)/g^M(\cdot)$ by $\{\widehat{H2}(\cdot) - \widehat{H1}(\cdot)\} \,/\, \{\widehat{h2}(\cdot)\}$, where

$$\begin{aligned} \widehat{H2}(b) &= \frac{1}{T} \sum_{t=1}^{T} 1(y2_t \leq b) \\ \widehat{H1}(b) &= \frac{1}{T} \sum_{t=1}^{T} 1(y1_t \leq b) \\ \widehat{h2}(b) &= \frac{1}{T \cdot \lambda_{h2}} \sum_{t=1}^{T} K_{h2} \left(\frac{b - y2_t}{\lambda_{h2}} \right) \cdot 1(y2_t \neq 0). \end{aligned}$$

where λ_{h2} is a bandwidth and K_{h2} is a kernel with ρ_{h2}-length support. Since the kernel density estimator is biased at the boundaries, the same trimming as in the previous subsession is applied. Let $Y2_{\min}$ and $Y2_{\max}$ be the minimum and maximum of the all $Y2_t$ except when $Y2_t = 0$. Define $C_T = [Y2_{\min} + \rho_{g^M}\lambda_{g^M}/2, Y2_{\max} - \rho_{g^M}\lambda_{g^M}/2]$. By applying Equation (3) I define the pseudo-values $\widehat{V_t^{(N_t:N_t)}}$ and $\widehat{V_t^{(N_t-1:N_t)}}$ corresponding to $Y1_t$ and $Y2_t$ as

$$\widehat{V_t^{(N_t:N_t)}} = \left\{ \begin{array}{cc} 0, & \text{if} \quad Y1_t < Y2_{\min} + \rho_{h2}\lambda_{h2}/2 \\ Y1_t + \{\widehat{H2}(Y1_t) - \widehat{H1}(Y1_t)\} \; / \; \widehat{h2}(Y1_t), & \text{if} \quad Y1_t \in C_T \\ \infty, & \text{if} \quad Y1_t > Y2_{\max} - \rho_{h2}\lambda_{h2}/2 \end{array} \right\}$$

and

$$\widehat{V_t^{(N_t-1:N_t)}} = \left\{ \begin{array}{cc} 0, & \text{if} \quad Y2_t < Y2_{\min} + \rho_{h2}\lambda_{h2}/2 \\ Y2_t + \{\widehat{H2}(Y2_t) - \widehat{H1}(Y2_t)\} \; / \; \widehat{h2}(Y2_t), & \text{if} \quad Y2_t \in C_T \\ \infty, & \text{if} \quad Y2_t > Y2_{\max} - \rho_{h2}\lambda_{h2}/2 \end{array} \right\}.$$

In the second step, I use the pseudo sample $\{v_t^{(n_t:n_t)}, v_t^{(n_t-1:n_t)}\}$ to estimate $F(v)$ by the semi-nonparametric maximum likelihood (SNP) estimation method (Gallant and Nychka, 1987). From results in Song(2007), the density of $v_t^{(n_t:n_t)}$ conditional on $v_t^{(n_t-1:n_t)}$ is $f(v_t^{(n_t:n_t)})/(1 - F(v_t^{(n_t-1:n_t)}))$. Therefore, I obtain the sample likelihood function as follows.

$$L_{T'}(\widehat{f}) = \frac{1}{T'} \sum_{t=1}^{T'} \ln \frac{\widehat{f}(v_t^{(n_t:n_t)})}{1 - \widehat{F}(v_t^{(n_t-1:n_t)})} \qquad (\; \widehat{F}(x) = \int_c^x \widehat{f}(t)dt \;)$$

where c is the minimum of $v_t^{(n_t-1:n_t)}$. A pair in which one of $v_t^{(n_t:n_t)}$ and $v_t^{(n_t-1:n_t)}$ does not belong to C_T is not used in estimation, and T' is the number of pairs after that exclusion. The SNP estimator is denoted by $\widehat{f^{SNP}}$. Regarding λ_{h2} and K_{h2}, I make the same assumptions as made for λ_{g^M} and K_{g^M} in (A1) and (A2) in the previous subsession.

Theorem 4 *$\widehat{f^{SNP}}(v)$ converges uniformly to $f(v)$ as $s \to \infty$.*

Proof. For the uniform consistency, I consider the supreme norm.

(i) From the uniform consistency of the empirical distribution, $\widehat{H2}(b)$ and $\widehat{H1}(b)$ converges uniformly to $H2(b)$ and $H1(b)$ respectively.

(ii) $\widehat{h2}(b)$ is uniformly consistent to $h2(b) = \sum_{n=2}^{l} p_n n(n-1)G(b)^{n-2}g(b)(1-G(b))$.

Define a random variable, $Y2'$: $Y2_t^i$ conditional on that $Y2_t^i \neq 0$. Noting $\Pr(y2_t^i = 0) = p_0 + p1$, the density of $y2'$, $h2'(y2')$ is

$$\begin{aligned} h2'(b) &= \frac{1}{1-p_0-p_1} \cdot \sum_{n=2}^{l} p_n n(n-1)G(b)^{n-2}g(b)(1-G(b)) \\ &= \frac{1}{1-\Pr(y2_t^i = 0)} \cdot h2(b). \end{aligned}$$

Accordingly,

$$h2(b) = (1 - \Pr(y2_t^i \neq 0)) \cdot h2'(b).$$

Noting that

$$\begin{aligned} \widehat{h2}(b) &= \frac{1}{T \cdot \lambda_{h2}} \sum_{t=1}^{T} K_{h2}\left(\frac{b - y2_t}{\lambda_{h2}}\right) \cdot 1(y2_t \neq 0) \\ &= \frac{(T - \sum 1(y2_t \neq 0))}{T} \cdot \frac{1}{(T - \sum 1(y2_t \neq 0)) \cdot \lambda_{h2}} \sum_{t=1}^{T} K_{h2}\left(\frac{b - y2_t}{\lambda_{h2}}\right) \cdot 1(y2_t \neq 0), \end{aligned}$$

$$\widehat{h2}(b) = (1 - \widehat{\Pr}(y2_t \neq 0)) \cdot \widehat{h2'}(b)$$

where

$$\widehat{\Pr}(y2_t \neq 0) = \frac{\sum 1(y2_t = 0)}{T} \qquad \text{and}$$

$$\widehat{h2'}(b) = \frac{1}{(T - \sum 1(y2_t \neq 0)) \cdot \lambda_{h2}} \sum_{t=1}^{T} K_{h2}\left(\frac{b - y2_t}{\lambda_{h2}}\right) \cdot 1(y2_t \neq 0)$$

Since $\widehat{\Pr}(y2_t \neq 0)$ is a consistent estimator of $\Pr(y2_t^i \neq 0)$ and $\widehat{h2'}(b)$ is uniformly consistent to $h2'(b)$, the uniform consistency of $\widehat{h2}(b)$ is established. The uniform consistency of $\widehat{H2}(b)$, $\widehat{H1}(b)$, and $\widehat{h2}(b)$ implies the uniform consistency of $\{\widehat{H2}(b) - \widehat{H1}(b)\} \,/\, \widehat{h2}(b)$ to $\{H2(b) - H1(b)\}/\{h2(b)\} = G^M(\cdot)/g^M(\cdot)$. Moreover, $Y2_{\min} \to \beta(\underline{v})$ and $Y2_{\max} \to \beta(\overline{v})$ as $T \to \infty$. Hence it is straightforward that $\widehat{V_t^i}$ converges uniformly to true value, V_t^i.

Let $f'(\cdot)$ be the true density of $\widehat{V_t^i}$. Using the triangle inequality,

$$\left\| \widehat{f^{SNP}}(v) - f(v) \right\| \leq \left\| \widehat{f^{SNP}}(v) - f'(v) \right\| + \left\| f'(v) - f(v) \right\|.$$

As $T \to \infty$, $\left\| \widehat{f^{SNP}}(v) - f'(v) \right\| \to 0$ (Gallant and Nychka, 1987). Since $F'(v)$ and $F(v)$ are bounded and $\widehat{V_t^i}$ converges uniformly to V_t^i, $\|f'(v) - f(v)\| \to 0$, as $T \to \infty$. Accordingly, the uniform consistency of $\widehat{F}(v)$ is established. ■

5 Monte Carlo Experiments

This session presents the results of Monte Carlo experiments to illustrate performance of the proposed estimators. For the case in which only top two bids are available, I conducts only the first step. I conducted both step in the case in which all bids are available.

My Monte Carlo experiment consists of 10,000 replications. For each experiment, artificial data of 200 auctions are generated. The number of active bidders, N_t ($t = 1, ..., 200$), was first drawn from a Binomial distribution with trial number 12 and success probability 0.25. Active bidders' valuations, V_t^i were then drawn from $Gamma(9, .5)$. For the sake of reference, $E(V_t^i) = 18$, and $Var(V_t^i) = 36$. I compute numerically the corresponding bids B_t^i using Equation (1).

5.1 Case 1: All bids are available

I choose the kernel and bandwidths according to Guerre, Perrigne, and Vuong (2000): $K_{g^M} = (35/32)(1 - u^2)^3 1(|u| \leq 1)$ and $h_{g^M} = 1.06\widehat{\sigma}(S)^{1/5}$ where $\widehat{\sigma}$ is the estimated standard deviations of observed M'. I first report the results only from Step 1 to examine the performance of $\widehat{G^M}(b)/\widehat{g^M}(b)$. In Figure 1, I display the 10 % percentile, median, and 90th-percentile of $b + \widehat{G^M}(b)/\widehat{g^M}(b)$ among 10,000 repetitions (three dotted lines) , along with $b + G^M(b)/g^M(b)$ (solid line). For reference, the means of boundaries of $C_S = [M_{\min} + \rho_{g^M}\lambda_{g^M}/2,\ M_{\max} - \rho_{g^M}\lambda_{g^M}/2]$ which appears in Equation (9) is 1.57 and 19.99. Figure 1 shows that the first step estimator performs very well. Within the average boundaries, the true curve falls within the 90% confidence band and the median of the estimates perfectly matches the true curve. Table 1 gives descriptive statistics of the mean and standard deviation of (after trimming) pseudo-values and true values at each experiments.

Figure 1

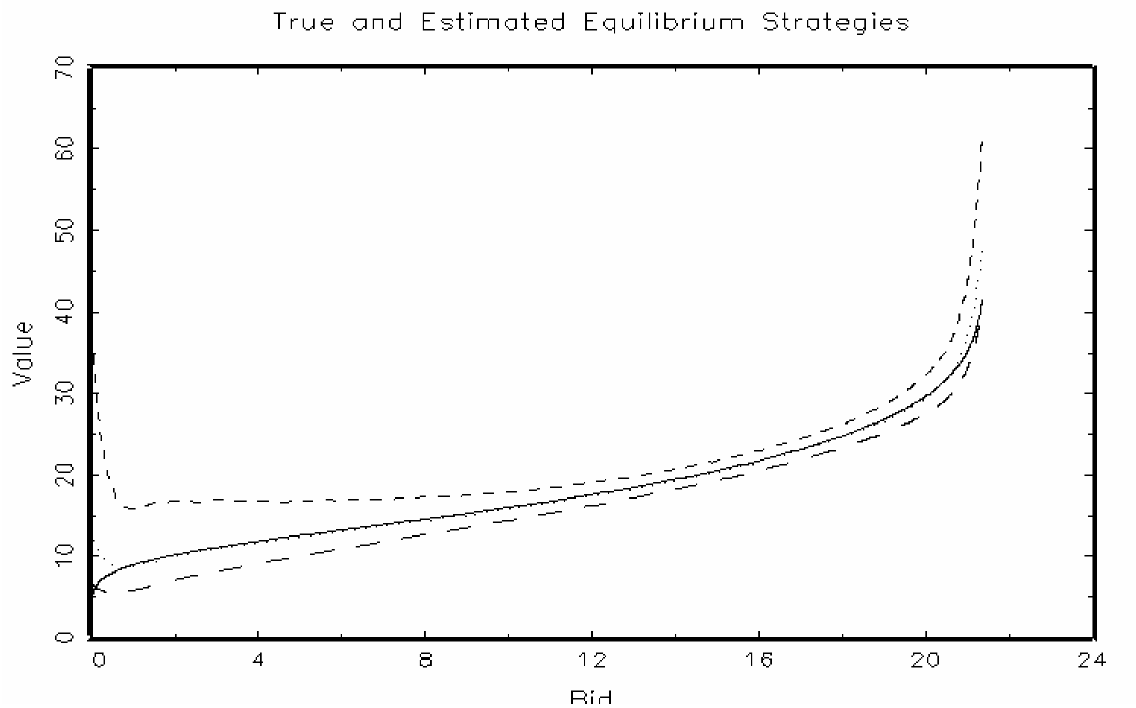
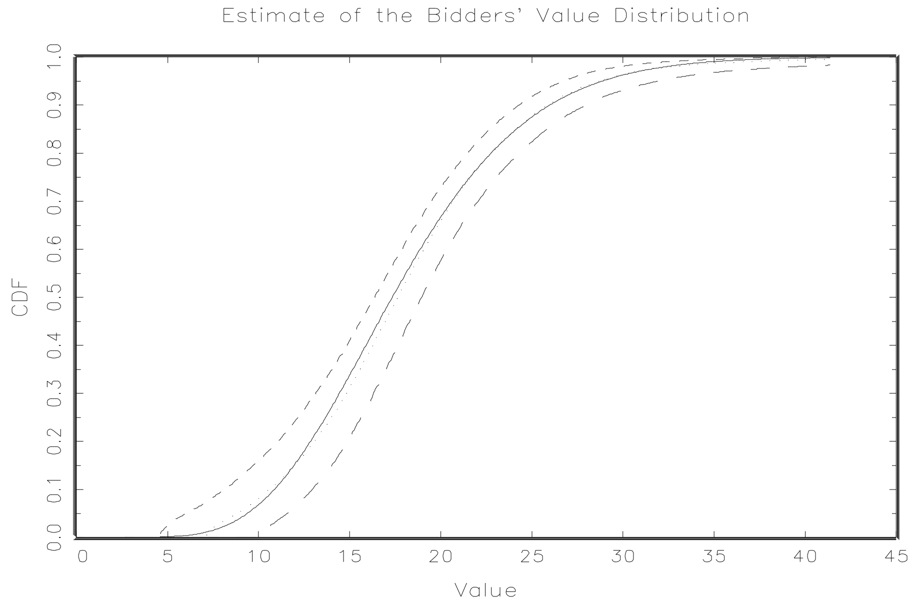


Table 1 Estimation of Valuations

	pseudo values		true values	
	mean	std	mean	std
10% percentile	17.33	4.378	17.65	4.509
median	17.88	4.939	17.93	4.738
90% percentile	18.52	5.523	18.23	4.900
mean	17.95	5.409	17.93	4.736

Figure 2 displays the 10 % percentile, median, and 90th-percentile of $\widehat{F}(v)$ among 10,000 repetitions, along with the true bidders' value distribution solid line). The performance of $\widehat{F}(v)$ is very strong. The average boundary values outside of which are trimmed are 8.34 and 39.32.

Figure 2



5.2 Case 2: Only top two bids are available

In these experiments, I estimate $G^M(b)/g^M(b)$ using only top two bids. I choose the kernel and bandwidths in the same way: $K_{h2}(u) = (35/32)(1-u^2)^3 1(|u| \leq 1)$ and $\lambda_{h2} = 1.06\widehat{\sigma}(S)^{1/5}$.

Table 3 and 4 give descriptive statistics of the mean and standard deviation of (after trimming) pseudo-values and true values at each experiments.

Table 3: Estimation of the Highest Value

	pseudo values		true values	
	mean	std	mean	std
10% percentile	21.02	3.14	21.92	4.61
median	22.01	3.91	22.54	5.08
90% percentile	23.20	5.03	23.14	5.57
mean	22.14	4.66	22.53	5.09

Table 4: Estimation of the Second-Highest Value

	pseudo values		true values	
	mean	std	mean	std
10% percentile	16.12	2.92	16.10	3.27
median	16.83	3.62	16.63	3.54
90% percentile	17.57	4.27	17.16	3.79
mean	16.83	3.61	16.63	3.53

6 Tentative Applications: BC Timber Auctions

The British Columbia timber sales (BCTS)[7] develops and sells blocks of publicly-owned timber across the province. One of their major goals is to provide a credible reference point for costs and pricing of timber harvested from public land in BC. To achieve that goal, BCTS uses the market pricing system which is based on the results of auction by BCTS: They auction part of the provincial allowable annual cut through hundreds of timber sales and then use the auction results with auction characteristics to determine the price of the remaining timber. Currently, they use simple regressions.

My future work is to see if structural analysis can do better than the current market pricing system. The estimation methods proposed in this paper will be useful specifically because the estimation methods do not require knowledge regarding the number of potential bidders. Most structural estimation methods have proposed an estimator of the potential bidders' value distribution, given knowledge regarding the number of potential bidders. As information of the number of potential bidder is hard to obtain except in auctions already held, it is difficult to apply the methods requiring knowledge regarding the number of potential bidders to this application because we need to predict the "auction price" of the timber which has not been sold through the auction. In addition, BCTS uses first-price sealed-bid auctions and protects information of how many bidders have

[7] See http://www.for.gov.bc.ca/bcts/.

submitted a bid until bid opening. Hence the model with an uncertain number of bidders is relevant.

7 Concluding Remarks

This paper develops econometric methods for structural analysis of first-price auction data in which bidders do not know how much competition they face with allowing the stochastic number of bidders. In particular, I newly show that only top two bids identify the bidders' value distribution and the distribution of the number of potential bidders. I propose an estimation strategy for the bidders' value distribution in both cases: (i) case in which all bids are available; and (ii) case in which only top two bids are available. It is future work to propose an estimation strategy for the distribution of the number of potential bidders and apply the proposed methods to BC timber auctions.

References

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.

Hendricks, Kenneth and Harry Paarsch (1995), "A Survey of Recent Empirical Work Concerning Auctions, " Canadian Journal of Economics, 28, 403-426.

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.

Levin, Dan and James L. Smith (1994), "Equilibrium in Auctions with Entry," American Economic Review, 84 (3), 585-599.

Li, Tong, Isabelle Perrigne, and Quong Vuong (2000), "Conditionally Independent Private Information in OLS Wildcat Auctions," Journal of Econometrics, 98, 129-161.

Li, Tong, Isabelle Perrigne, and Quong Vuong (2002), "Structural Estimation of the Affiliated Private Value Auction Model, " RAND Journal of Economics, 33 (2), 171-193.

McAfee, R. Preston and John McMillan (1987), "Auctions with a Stochastic Number of Bidders," Journal of Economic Theory, 43, 1-19.

Samuelson, William F. (1985), "Competitive Bidding with Entry Costs," Economics Letters, 17, 53-57.

Song, Unjy (2007), "Estimation of Auction Models with an Unknown Number of Bidders", working paper.