Nonparametric Estimation of Auctions with Unobserved Number of Bidders: a Misclassification Approach *

Yingyao Hu and Matthew Shum
Dept. of Economics, Johns Hopkins University

May 10, 2007

Preliminary and Incomplete

Please do not quote

Abstract

In this paper, we consider the nonparametric identification and estimation of auction models when N^*, the true number of bidders, is unobserved. Exploiting results from the recent econometric literature on models with misclassification error, we develop a nonparametric procedure for recovering the distribution of bids conditional on unobserved N^*. This procedure may be applied to both first- and second-price auction models. Monte Carlo results show that the procedure works well in practice, and we present some illustrative evidence from a dataset of procurement auctions.

In many auction applications, the true number of bidders N^* is not observed by the researcher. The most common scenario where this occurs is under binding reserve prices. When reserve prices bind, the true number of bidders N^*, which is observed by auction participants and influences their bidding behavior, differs from the actual number of bidders A ($\leq N^*$), which is the number of auction participants whose bid exceed the reserve price. Other scenarios which would cause the true level of competition to be unobserved (and differ from the observed level of competition) include bidding or participation costs. In some cases, the number of auction participants may simply not be noted in the researcher's dataset.

*The authors can be reached at yhu@jhu.edu and mshum@jhu.edu. Guofang Huang provided exceptional research assistance.

In this paper, we consider the nonparametric identification and estimation of auction models when N^* is unobserved. Using recent results from the literature on misclassified regressors (Hu (2006)), we show how the distribution of bids given the unobserved N^* can be identified, and estimated. In the case of second-price auctions, these conditional bid distributions are directly the parameters of interest, whereas in the case of first-price auctions, these bid distributions estimated using our procedure can be used as inputs into established nonparametric procedures (Guerre, Perrigne, and Vuong (2000), Li, Perrigne, and Vuong (2002)) to obtain estimates of bidders' valuations. Knowledge of bidder valuations can be useful for testing for common values (as in Haile, Hong, and Shum (2003)) or to determine the optimal design of auctions (Paarsch (1997),Athey, Levin, and Seira (2005)).

Previous work has dealt with the unobservability of N^* in several ways. In the parametric estimation of auction models, the functional relationship between the bids b and true number of bidders N^* is explicitly parameterized, so that not observing N^* need not be a problem. For instance, Laffont, Ossard, and Vuong (1995) use a goodness-of-fit statistic to select the most plausible value of N^* for French eggplant auctions. Paarsch (1997) treats N^* essentially as a random effect and integrates out over its assumed distribution in his analysis of timber auctions.

In a nonparametric approach to auctions, however, the relationship between the bids b and N^* must be inferred directly from the data, and not observing N^* (or observed N^* with error) raises difficulties. Within the independent private values framework, and under the additional assumption that the unobserved N^* is fixed across all auctions (or fixed across a known subset of the auctions), Guerre, Perrigne, and Vuong (2000) show the identification of N^* and the equilibrium bid distribution in the range of bids exceeding the reserve price. Hendricks, Pinkse, and Porter (2003) allow N^* to vary across auctions, and assume that $N^* = L$, where L is a measure of the potential number of bidders which they construct.

In this paper, we consider a different approach. Exploiting results from the recent econometric literature on models with misclassification error (Hu (2006)), we develop a nonparametric procedure for recovering the distribution of bids conditional on unobserved N^* which requires neither N^* to be fixed across auctions, nor for an (assumed) perfect measure of N^* to be available. We draw an analogy between the auction and misclassification problem, where the true number of bidders N^* is observed with error. Our procedure requires the availability of an imperfect measure of N^*, as well as an instrument (which could be a second imperfect measure of N^*).

In the next two sections, we present the model and estimation methodology, for the simple case of independent private value (IPV) auction models. In section 4, we provide Monte Carlo evidence of our estimation procedure, and discuss some practical implementation issues. In section 5, we present an empirical illustration to procurement auctions. In section 6, we consider a straightforward extension to auctions with affiliated values. Section 7 concludes.

1 Model

In the next few sections, we focus on auctions under the independent private values paradigm, for which identification and estimation are most transparent. After discussing this case fully, we consider extensions to the more general affiliated value case.

Consider the simplest case of first-price auctions under the symmetric independent private values (IPV) paradigm, with a binding reserve price. There are N^* bidders in the auction, with each bidder drawing a private valuation from the distribution $F(x)$ which has support $[\underline{x}, \bar{x}]$. Given the reserve price r (which is assumed to be fixed across all auctions), where $r > \underline{x}$, the equilibrium bidding function for bidder i with valuation x_i is

$$b^*(x_i; N^*) \begin{cases} = x_i - \frac{\int_r^{x_i} F(s)^{N^*-1} ds}{F(x_i)^{N^*-1}} & \text{for } x_i \geq r \\ 0 & \text{for } x_i < r. \end{cases} \tag{1}$$

Hence, the actual number of bidders $A \equiv \sum_{i=1}^{N^*} \mathbf{1}(x_i > r)$, the number of bidders whose valuations exceed the reserve price.

For this case, the equilibrium bids are i.i.d. and, using the change-of-variables formula, the density of interest $g(b|N^*, b > r)$ is equal to

$$g(b|N^*, b > r) = \frac{1}{b^{*\prime}(\beta^*(b); N^*))} \frac{f(\beta^*(b; N^*))}{1 - F(r)}, \text{ for } b > r \tag{2}$$

where $\beta^*(b; N^*)$ denotes the inverse of the equilibrium bid function $b^*(\cdot; N^*)$ evaluated at b. In equilibrium, each observed bid from an N^*-bidder auction is an i.i.d. draw from the distribution given in Eq. (2).

We propose a two-step estimation procedure. In the first step, the goal is to recover the density $g(b|N^*; b > r)$ of the equilibrium bids, for the truncated support $(r, +\infty)$. (For convenience, in what follows, we suppress the conditioning truncation event $b > r$.) To identify and estimate $g(b|N^*)$, we use the results from Hu (2006).

In second step, we use the methodology of Guerre, Perrigne, and Vuong (2000) to recover the valuations x from the joint density $g(b|N^*)$. For each b in the marginal support of $g(b|N^*)$, the corresponding valuation x is obtained by

$$\zeta(b; N^*) = b + \frac{G(b|N^*)}{(N^* - 1) \cdot g(b|N^*)}. \tag{3}$$

For most of this paper, we focus on the first step of this procedure, because the second step is a straightforward application of standard techniques.

1.1 Remarks

Thus far, we assume that N^* is observed and deterministic from bidders' point of view, but not observed by the researcher. However, our methodology could potentially also handle models with stochastic number of bidders. One such model is that two-hurdle participation model considered in the Hendricks, Pinkse, and Porter (2003) paper of bidding in oil and gas lease auctions. In their model, the true number of bidders differs from the actual number of bidders due to both (i) bidders' decisions to invest in a seismic survey of the tract, which is a necessary condition for submitting a bid; and (ii) a binding reserve price.

Hendricks, Pinkse, and Porter (2003) show that an equation analogous to Eq. (3) is still valid for generating valuations from observed bids in their model, so long as the number of potential bidders N^* is observed. Our methodology (suitably extended to the affiliated-value case, as explored in Section 6 below) would recover the equilibrium bid distribution $G(b|N^*)$ even when N^* is not perfectly observed.

In general, in cases when the number of bidders is stochastic from bidders' point of view, our procedure can still recover $g(b|N^*)$, where N^* is interpreted as a scalar variable which parameterizes the bidders' beliefs about the number of bidders. For example, Bajari and Hortacsu (2003) consider a model of internet auction bidding with a stochastic number of bidders, where the number of bidders in auction t is drawn from a Poisson distribution with mean μ_t. In our framework, we could interpret N^* as μ_t, and recover the distribution of b for different unobserved values of μ_t.

The issues considered in this paper are also related to the literature on entry in auctions (eg. Li (2000), Li and Zheng (2006), Athey, Levin, and Seira (2005), Krasnokutskaya and Seim (2005), Haile, Hong, and Shum (2003)). While the entry models considered in these papers differ, the econometric problem introduced by entry is that, while the true number of bidders is observed by the researcher, and equal to the actual number of bidders (ie. $N^* = A$), A is potentially endogenous, because it may be determined in part by auction-specific unobservables which also affect the bids. Contrastingly, in this paper, we assume that N^* is unobserved, and that $N^* \neq A$, but we do not allow N^* to be endogenous.

2 Nonparametric identification

In this section, we apply the results from Hu (2006) to show the identification of the conditional equilibrium bid distributions $g(b|N^*)$, conditioned on the true (but unobserved) number of bidders N^*, as well as the conditional distribution of $N|N^*$.

We require two variables:

1. a proxy N, which is a mismeasured version of N^*
2. an instrument Z, which could be a second corrupted measurement of N^*.

The variables (N, Z) must satisfy the following two conditions:

Condition 1 $g(b|N^*, N, Z) = g(b|N^*)$.

This assumption implies that N or Z affects the bids only through the true number of bidders.

Condition 2 $g(N|N^*, Z) = g(N|N^*)$.

This assumption implies that the instrument Z affects the mismeasured N only through the true number of bidders. Roughly, because N is a noisy measure of N^*, this condition requires that the noise is independent of the instrument Z, conditional on N^*.

One advantage to focusing on the IPV model is that A, the observed number of bidders, can be used in the role of N. Particularly, condition 1 is satisfied because, for a given N^*, the sampling density of any equilibrium bid exceeding the reserve price is equal to Eq. (2) above, which does not depend on A. A good candidate for the instrument Z could be a measure of the potential number of bidders, which we denote L.

We observe a random sample of $\left\{\vec{b}_t, N_t, Z_t\right\}$, where $\vec{b}_t$ denotes the observed bids $\{b_{1t}, b_{2t}, \ldots, b_{A_t t}\}$. We assume the variable N, Z, and N^* share the same support $\mathcal{N} = \{2..., K\}$. Here K can be interpreted as the maximum number of bidders, which is fixed across all auctions.[1]

By the law of total probability, the relationship between the observed distribution $g(b, N, Z)$ and the latent densities is as follows:

$$g(b, N|Z) = \sum_{N^*=2}^{K} g(b|N^*, N, Z)g(N|N^*, Z)g(N^*|Z). \tag{4}$$

Under conditions 1 and 2, Eq. (4) becomes

$$g(b, N|Z) = \sum_{N^*=2}^{K} g(b|N^*)g(N|N^*)g(N^*|Z). \tag{5}$$

In order to apply the identification technique in Hu (2006), we define matrices

$$\begin{aligned}
G_{b,N|Z} &= [g(b, N = i|Z = j)]_{i,j} , \\
G_{N|N^*} &= [g(N = i|N^* = k)]_{i,k} , \\
G_{N^*|Z} &= [g(N^* = k|Z = j)]_{k,j} , \\
G_{N|Z} &= [g(N = i|Z = j)]_{i,j} ,
\end{aligned}$$

and

$$G_{b|N^*} = \begin{pmatrix} g(b|N^* = 2) & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & g(b|N^* = K) \end{pmatrix}. \tag{6}$$

With this notation, Eq. (5) can be written as

$$G_{b,N|Z} = G_{N|N^*}G_{b|N^*}G_{N^*|Z}. \tag{7}$$

Condition 2 implies that

$$g(N|Z) = \sum_{N^*=2}^{K} g(N|N^*)g(N^*|Z), \tag{8}$$

which is equivalent to

$$G_{N|Z} = G_{N|N^*}G_{N^*|Z}. \tag{9}$$

[1]Our identification results still hold if Z has more possible values than N and N^*.

This equation implies that

$$Rank\left(G_{N|Z}\right) \leq \min\left\{Rank\left(G_{N|N^*}\right), Rank\left(G_{N^*|Z}\right)\right\}. \tag{10}$$

We assume

Condition 3 *$Rank\left(G_{N|Z}\right) = K - 1$.*

Note that this condition is directly testable from the sample. This assumption implies that $Rank\left(G_{N|N^*}\right) = K - 1$ and $Rank\left(G_{N^*|Z}\right) = K - 1$. In other words, the matrices $G_{N|Z}$, $G_{N|N^*}$, and $G_{N^*|Z}$ are all invertible. Therefore, we have

$$G_{b,N|Z}G_{N|Z}^{-1} = G_{N|N^*}G_{b|N^*}G_{N|N^*}^{-1}. \tag{11}$$

Because $G_{b|N^*}$ is diagonal (cf. Eq. (6)), this expression implies an eigenvalue-eigenvector decomposition of the observed matrix on the left-hand side. In order to make the decomposition unique, we assume that

Condition 4 *for any $i, j \in \mathcal{N}$, the set $\{(b) : g(b|N^* = i) \neq g(b|N^* = j)\}$ has nonzero Lebesgue measure whenever $i \neq j$.*

This assumption guarantees that eigenvalues in $G_{b|N^*}$ are distinctive for some bid b. It also implies that the true number of bidders N^* is relevant to the bids b.

Eq. (11) shows that a matrix decomposition of the observed $G_{b,N|Z}G_{N|Z}^{-1}$ matrix identifies the unknown $G_{N|N^*}$ and $G_{b|N^*}$ matrices, up to the ordering of the columns of $G_{N|N^*}$ and $G_{b|N^*}$. In order to complete the identification, we need an additional assumption which pins down the ordering of these columns. One such assumption is:

Condition 5 *$N \leq N^*$.*

The condition $N \leq N^*$ is natural, and automatically satisfied, when $N = A$, the actual number of bidders. This condition implies that for any $i, j \in \mathcal{N}$

$$g\left(N = j|N^* = i\right) = 0 \text{ for } j > i. \tag{12}$$

In other words, $G_{N|N^*}$ is an upper-triangular matrix. Since the triangular matrix $G_{N|N^*}$ is invertible, its diagonal entries are all nonzero, i.e.,

$$g\left(N = i|N^* = i\right) > 0 \text{ for all } i \in \mathcal{N}. \tag{13}$$

Equations 12 and 13 imply that n^* is the 100th percentile of the discrete distribution $g(N|N^* = n^*)$, i.e.,

$$n^* = \inf \left\{ \widetilde{n}^* : \sum_{i=2}^{\widetilde{n}^*} g(N = i|N^* = n^*) \geq 1 \right\}.$$

In other words, condition 5 implies that, once we have the columns of $G_{N|N^*}$ obtained as the eigenvectors from the matrix decomposition (11), the right ordering can be obtained by re-arranging these columns fo that they form an upper-triangular matrix.

Hence, under the assumptions 1-5, both $G_{b|N^*}$ and $G_{N|N^*}$ are identified (the former point-wise in b).

3 Estimation

In this section, we give details on the estimation of $(b|N^*)$ given observations of (b, N, Z), for the symmetric independent private values model. As shown in the previous section, the distributions $g(b|N^*)$, $g(N|N^*)$ and $g(N^*|Z)$ are nonparametrically identified from the observed distribution $g(b, N|Z)$ as follows:

$$g(b, N|Z) = \sum_{N^*=2}^{K} g(b|N^*)g(N|N^*)g(N^*|Z). \tag{14}$$

Note that the bid b may have a different unknown support for different N^*. We assume

$$g(b|N^*) = \begin{cases} > 0 & \text{for } b \in [r, u_{N^*}] \\ = 0 & \text{otherwise} \end{cases},$$

where u_{N^*} is unknown. This fact makes the direct estimation of $g(b|N^*)$ difficult. Therefore, we propose a two-step estimation procedure.

Step 1 is to estimate the eigenvector matrix $G_{N|N^*}$. We consider the conditional expectation of the bid b to avoiding directly estimating the unknown density $g(b|N^*)$ with unknown support. We define

$$G_{Eb,N|Z} = [E(b|N = i, Z = j)\, g(N = i|Z = j)]_{i,j}\,, \tag{15}$$

and

$$G_{Eb|N^*} = \begin{pmatrix} E[b|N^* = 2] & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & E[b|N^* = K] \end{pmatrix}.$$

From equation 5, we have

$$E(b|N,Z)g(N|Z)=\sum_{N^*=2}^{K}E(b|N^*)g(N|N^*)g(N^*|Z). \tag{16}$$

Therefore, the same diagonalization result in equation 11 holds for equation 16 with $G_{b,N|Z}$ and $G_{b|N^*}$ replaced with $G_{Eb,N|Z}$ and $G_{Eb|N^*}$ as follows:

$$G_{Eb,N|Z}G_{N|Z}^{-1}=G_{N|N^*}G_{Eb|N^*}G_{N|N^*}^{-1}.$$

That means we may have

$$G_{N|N^*}=\psi\left(G_{Eb,N|Z}G_{N|Z}^{-1}\right),$$

where $\psi(\cdot)$ denotes the mapping from a square matrix to its eigenvector matrix following the identification procedure in the previous section. Therefore,we may estimate $G_{N|N^*}$ as follows:

$$\widehat{G}_{N|N^*}:=\psi\left(\widehat{G}_{Eb,N|Z}\widehat{G}_{N|Z}^{-1}\right), \tag{17}$$

where $\widehat{G}_{Eb,N|Z}$ and $\widehat{G}_{N|Z}$ may be constructed directly from the sample.

Since we don't have a covariate in the simulation, all the entries in the matrices in the equation 16 are constants. We may then use the eigenvalue/vector decomposition of the left-hand side $\widehat{G}_{Eb,N|Z}\widehat{G}_{N|Z}^{-1}$ to directly estimate $\widehat{G}_{N|N^*}$. When there are covariates w, we may also use a semi-nonparametric method (Ai and Chen (2003)) to estimate $g(N|N^*,w)$, which we will discuss later.

Step 2 is to estimate $g(b|N^*)$. With $G_{N|N^*}$ estimated by $\widehat{G}_{N|N^*}$ in step 1, we may estimate $g(b|N^*)$ directly even without knowing its support. From equation 11, we have for any b

$$G_{N|N^*}^{-1}\left(G_{b,N|Z}G_{N|Z}^{-1}\right)G_{N|N^*}=G_{b|N^*}. \tag{18}$$

Define $e_{N^*}=(0,...0,1,0,...,0)^T$, where 1 is at the N^*-th position in the vector. We have

$$g(b|N^*)=e_{N^*}^T\left[G_{N|N^*}^{-1}\left(G_{b,N|Z}G_{N|Z}^{-1}\right)G_{N|N^*}\right]e_{N^*}. \tag{19}$$

which holds for all $b\in(-\infty,\infty)$. This equation also implies that we may identify the upper bound u_{N^*} as follows:

$$u_{N^*}=\sup\left\{b:g(b|N^*)>0\right\}.$$

Finally, we may estimate $g(b|N^*)$ as follows:

$$\widehat{g}(b|N^*) := e_{N^*}^T \left[\widehat{G}_{N|N^*}^{-1} \left(\widehat{G}_{b,N|Z}\widehat{G}_{N|Z}^{-1}\right) \widehat{G}_{N|N^*}\right] e_{N^*},$$

where $\widehat{G}_{N|N^*}$ is estimated in step 1 and $\widehat{G}_{b,N|Z}$ may be constructed directly from the sample. In our empirical work, we use a kernel estimate for $\widehat{G_{b,N|Z}}$:

$$\begin{aligned}\widehat{G}_{b,N|Z} &= \widehat{g}(b|N,Z) \cdot \widehat{g}(N|Z) \\ &= \left[\frac{1}{Th}\sum_t \frac{1}{A_t}\sum_{i=1}^{N_t} K\left(\frac{b-b_{it}}{h}\right)\mathbf{1}(N_t = N, Z_t = Z)\cdot\right]\widehat{g}(N|Z).\end{aligned} \tag{20}$$

We have not yet derived the asymptotics in full detail, but here we comment briefly. Given the discreteness of N, Z, and the use of sample average of $b|N,Z$ to construct $\widehat{G}_{b,N|Z}$ (via. Eq. (15)), the estimates of $\widehat{G}_{N|N^*}$ (obtained using Eq. (17)) and $\widehat{G}_{N|Z}$ should converge at a $\sqrt{T}$-rate (where T denotes the total number of auctions).

Hence, pointwise in b, the convergence properties of $\widehat{g}(b|N^*)$ to $g(b|N^*)$, where $\widehat{g}(b|N^*)$ is estimated using Eq. (19), will be determined by the convergence properties of the kernel estimate of $g(b|N,Z)$, which converges slower than $\sqrt{T}$.

3.1 An alternative approach

In step 1 above, we may also use a semi-nonparametric method to estimate $g(N|N^*)$. This method is particularly useful when there are covariates w in the model with unknowns $g(b|N^*,w)$ and $g(N|N^*,w)$. As suggested in Ai and Chen (2003), we consider the following moment condition

$$\begin{aligned}E\left[b|N,Z,w\right] &= \sum_{N^*=2}^{K} E\left[b|N^*,w\right] g(N|N^*,w)g(N^*|Z,w)\frac{1}{g(N|Z,w)} \\ &\equiv m\left(N,Z,w,\alpha_0\right).\end{aligned}$$

where

$$\begin{aligned}\alpha_0 &= (g_{10},g_{20},g_{30}) \\ g_{10} &= E\left[b|N^*,w\right] \\ g_{20} &= g(N|N^*,w) \\ g_{30} &= g(N^*|Z,w).\end{aligned}$$

Let $p_i(\cdot)$ be a series of known basis functions, such as power series, splines, Fourier series, etc. For example, the Hermite polynomial series $\{H_k : k = 1, 2, ...\}$ is an orthonormal basis of $\mathcal{L}^2(\mathbb{R}, \exp\{-x^2\})$. It can be obtained by applying the Gram-Schmidt procedure to the polynomial series $\{x^{k-1} : k = 1, 2, ...\}$ under the inner product $\langle f, g \rangle_\omega = \int_{\mathcal{R}} f(x)g(x) \exp\{-x^2\}dx$. That is, $H_1(x) = 1/\sqrt{\int_{\mathbb{R}} \exp\{-x^2\}dx} = \pi^{-1/4}$, and for all $k \geq 2$,

$$H_k(x) = \frac{x^{k-1} - \sum_{j=1}^{k-1} \langle x^{k-1}, H_j \rangle_\omega H_j(x)}{\sqrt{\int_{\mathbb{R}} [x^{k-1} - \sum_{j=1}^{k-1} \langle x^{k-1}, H_j \rangle_\omega H_j(x)]^2 \exp\{-x^2\}dx}}. \tag{21}$$

We then consider the sieve expression corresponding to g_{10} as follows:

$$g_1(N^*, w) = \sum_{i=1}^{\infty} \sum_{j=2}^{K} \beta_{i,j} \times I(N^* = j) \times p_i(w). \tag{22}$$

In the estimation, we shall use finite-dimensional sieve spaces since they are easier to implement as follows:

$$g_{1n}(N^*, w) = \sum_{i=1}^{I_{1n}} \sum_{j=2}^{K} \beta_{i,j} \times I(N^* = j) \times p_i(w). \tag{23}$$

We let $I_{1n} \to \infty$ as $T \to \infty$.

The sieve expressions corresponding to g_{20} and g_{30} are as follows:

$$g_2(N|N^*, w) = \sum_{i=1}^{\infty} \sum_{j=2}^{K} \sum_{k=1}^{K} \gamma_{i,j,k} \times I(N = j) \times I(N^* = k) \times p_i(w),$$

$$g_3(N^*|Z, w) = \sum_{i=1}^{\infty} \sum_{j=2}^{K} \sum_{k=1}^{K} \delta_{i,j,k} \times I(N^* = j) \times I(Z = k) \times p_i(w),$$

with their finite-dimensional counterparts

$$g_{2n}(N|N^*, w) = \sum_{i=1}^{I_{2n}} \sum_{j=2}^{K} \sum_{k=1}^{K} \gamma_{i,j,k} \times I(N = j) \times I(N^* = k) \times p_i(w),$$

$$g_{3n}(N^*|Z, w) = \sum_{i=1}^{I_{3n}} \sum_{j=2}^{K} \sum_{k=1}^{K} \delta_{i,j,k} \times I(N^* = j) \times I(Z = k) \times p_i(w).$$

The coefficients $\gamma_{i,j,k}$ and $\delta_{i,j,k}$ satisfies

$$\sum_{N=1}^{K} g_{2n}(N|N^*, w) = 1 \text{ for any } N^* \text{ and } w \tag{24}$$

and

$$\sum_{N^*=2}^{K} g_{3n}(N^*|Z, w) = 1 \text{ for any } Z \text{ and } w \tag{25}$$

These two conditions implies linear restrictions on the coefficients $\gamma_{i,j,k}$ and $\delta_{i,j,k}$.

We define an alternative values of α_0 as follows:

$$\alpha = (g_1, g_2, g_3).$$

We also define the data as $D_t = (b_{kt}, k = 1, \ldots, A_t; A_t, L_t)$. We then have

$$\alpha_0 = \underset{\alpha=(g_1,g_2,g_3)}{\arg\sup} E\left[b - m(D_t, \alpha)\right]^2$$

The semi-nonparametric estimator $\widehat{\alpha}_n = (\widehat{g}_1, \widehat{g}_2, \widehat{g}_3)$ for α_0 is defined as:

$$\widehat{\alpha}_n = \underset{\alpha=(g_{1n},g_{2n},g_{3n})}{\arg\max} \sum_{t=1}^{T} \left[b_t - m(D_t, \alpha)\right]^2.$$

To be specific, we have

$$\begin{aligned}\left(\widehat{\beta}_{i,j}, \widehat{\gamma}_{i,j,k}, \widehat{\delta}_{i,j,k}\right) &= \underset{\left(\beta_{i,j},\gamma_{i,j,k},\delta_{i,j,k}\right) \text{ in } \alpha=(g_{1n},g_{2n},g_{3n})}{\arg\max} \sum_{t=1}^{T} \left[b_t - m(D_t, \alpha)\right]^2, \\ &\quad \text{such that } \left(\beta_{i,j}, \gamma_{i,j,k}, \delta_{i,j,k}\right) \text{ satisfies conditions 24,25.}\end{aligned}$$

Our estimate of the distribution of $(N|N^*, w)$ is

$$\widehat{g}_2(N|N^*, w) = \sum_{i=1}^{I_{2n}} \sum_{j=1}^{K} \sum_{k=1}^{K} \widehat{\gamma}_{i,j,k} \times I(N = j) \times I(N^* = k) \times p_i(w).$$

We may them use the procedure in Step 2 to estimation the distribution of interest $g(b|N^*, w)$ as follows:

$$\widehat{g}(b|N^*, w) = e_{N^*}^T \left[\widehat{G}_{N|N^*,w}^{-1} \left(\widehat{G}_{b,N|Z,w} \widehat{G}_{N|Z,w}^{-1}\right) \widehat{G}_{N|N^*,w}\right] e_{N^*},$$

where $\widehat{G}_{b,N|Z,w}$ and $\widehat{G}_{N|Z,w}$ may be constructed directly from the sample.

4 Monte Carlo Evidence

In this section, we present some Monte Carlo evidence for the independent private value model. We focus only on the first part of the estimation procedure, in which the conditional bid distributions $g(b|N^*)$, for each N^*, are recovered. We also consider the case of no covariates, and estimate these bid distributions using the direct matrix decomposition method presented in section 3 above.

We consider first price auctions where bidders' valuations $x_i \sim U[0,1]$, independently across bidders i. With a reserve price $r > 0$, the equilibrium bidding strategy with N^* bidders is:

$$b^*(x; N^*) = \begin{cases} \left(\frac{N^*-1}{N^*}\right) x + \frac{1}{N^*} \left(\frac{r}{x}\right)^{N^*-1} r & \text{if } x \geq r \\ \text{some } c < r & \text{if } x < r. \end{cases} \tag{26}$$

For each auction t, we need to generate the equilibrium bids b_{jt}, for $j = 1, \ldots N_t^*$, as well as (N_t^*, N_t, Z_t). In this exercise, N_t is taken to be the number of actual bidders A_t, and Z_t is a second corrupted measure of N_t^*.

For each auction t, the true number of bidders N_t^* is generated uniformly on $\{2, 3, \ldots, K\}$, where K is the maximum number of bidders. Subsequently, the corrupted measure Z_t is generated as:

$$Z_t = \begin{cases} N_t^* & \text{with prob. } q \\ \text{unif.} \{2, 3, \ldots, J\} & \text{with prob. } 1 - q. \end{cases} \tag{27}$$

For each auction t, and each participating bidder $j = 1, \ldots, N_t^*$, draw $x_j \sim U[0,1]$. Subsequently, the number of actual bidders is determined as the number of bidders whose valuations exceed the reserve price:

$$A_t = \sum_{j \in \mathcal{N}_t^*} \mathbf{1}(x_j \geq r) \tag{28}$$

Finally, for each auction t, and each actual bidder $j \in \mathcal{A}_t$, we can calculate the equilibrium bid using Eq. (26).

Note that the estimation procedure in section 3 above requires the matrix $G_{A|N^*}$ to be square, but in generating the variables here, the support of A is $\{1, 2, .., K\}$ while the support of N^* is $\{2, .., K\}$. To accommodate this, we define

$$N = \begin{cases} A & \text{if } A \geq 3 \\ 2 & \text{if } A \leq 2 \end{cases}.$$

Therefore, N has the same support as N^*. And all the identification conditions hold.

4.1 Results

In this section, we present results from $S = 200$ replications of a simulation experiment. In these experiments, we focus on the first step of the estimation procedure, where the equilibrium bid density $g(b|N^*)$ is recovered, for different values of N^*. In estimation, we use the first approach, where each density $g(b|N^*)$ is obtained by a matrix decomposition operation, pointwise for every value of b (cf. Eq. (19).

First, we consider the case where K (the maximum number of bidders) is equal to 4. The performance of our estimation procedure is illustrated in Figure 1. The estimator perform well for all values of $N^* = 2, 3, 4$, and for a modest-sized dataset of $T = 302$ auctions. Across the Monte Carlo replications, the estimated density functions track the actual densities quite closely. In these graphs, we also plot $g(b|A = n)$, the bid density conditioned on the actual number of bidders, for $n = \{1, 2\}, 3, 4$, which could be one "naive" estimator for $g(b|N^* = n)$. For $N^* = 2, 3$, our estimator outperforms the naive estimator, especially for the case of $N^* = 2$.

[Figure 1 about here.]

The good performance of our estimator is confirmed in Table 1, which lists the percentiles (across the $S = 200$ replications) of the L^2-norm of the errors between the estimate $g(b\hat{|}N^*)$ and the true $g(b|N^*)$. We see that the errors become smaller as N^* increases. We conjecture this may be due to the increased number of bids used to construct the estimate of the bid density, at larger values of N^*. For comparison, in the bottom panel of Table 1, we also list the percentiles of the L^2-norm of the errors between the "naive" estimator $g(b|A = n)$ and the true $g(b|N^* = n)$. Across all N^* and all percentiles, the errors from this naive approach are larger than the errors from using our estimation procedure.

[Table 1 about here.]

In a second set of experiments, we consider the case where the maximum number of bidders is $K = 6$. In these experiments, we increased the number of auctions to be $T = 1000$. Graphs summarizing these simulations are presented in Figure 2. Clearly, our estimator continues to perform well. The percentiles of the errors are presented in Table 2, and confirm the good performance of our estimator. However, for larger values of $N^* = 5, 6$, we see that the errors in our estimation procedure can be larger than the errors from the "naive" estimator $g(b|A)$. This may not be surprising, that the naive approach does less badly as the number of bidders grows: as N^* increases, the bidding strategies are less distinguishable for different values for N^* and, in the limit, as $N^* \to \infty$, the equilibrium bid density will approach the distribution of the valuations x. Hence, the error in using $g(b|A = n)$ as the estimator for $g(b|N^* = n)$ for larger n will be less severe.

[Figure 2 about here.]

[Table 2 about here.]

5 Empirical illustration: Procurement auctions

To be added.

6 Affiliated Values Case

In this section, we consider the extension of our estimation procedure to the case of auction within the general affiliated values paradigm (Milgrom and Weber (1982)). Within the affiliated values setting, the goal of the identification and estimation is the joint density $g(b, m|N^*; b > r, m > r)$, where b denotes a bid and m the maximum of rivals' bids, subject to both b and m being above the reserve price. (As above, we will suppress the truncation conditions $b > r, m > r$ for convenience.)

After estimating $g(b, m|N^*)$, a first-order condition analogous to Eq. (1) can be utilized to obtain estimates of the "pseudovalues":

$$\xi(b; N^*) \equiv b - \frac{G_{M|B}(b|b; N^*)}{g_{M|B}(b|b; N^*)}. \tag{29}$$

As is well-known, in the affiliated private values case (cf. Li, Perrigne, and Vuong (2002)), the pseudovalue $\xi(b; N^*)$ is an estimate of the valuation $x \equiv {b^*}^{-1}(b; N^*)$ corresponding to each bid b. In the common value case, $\xi(b; N^*)$ is estimate of $E\left[V_i|X_i = x, Y_i = x\right]$, where $x = {b^*}^{-1}(b_i; N^*)$ and $Y_i \equiv \max_{j \neq i} X_j$ (cf. Hendricks, Pinkse, and Porter (2003), Haile, Hong, and Shum (2003)).

The conditions that must be satisfied by the proxy N and instrument Z in this case are (analogous to conditions 1 and 2 above):

Condition 6 $g(b, m|N^*, N, Z) = g(b, m|N^*)$.

Condition 7 $g(N|N^*, Z) = g(N|N^*)$.

Unlike in the IPV case, the number of actual bidders A no longer satisfies condition 6, and should not be used as the proxy variable N. For a given value of N^*, when bids are affiliated in equilibrium, a larger value of A will imply that the bids exceeding the reserve price should, on average, have higher values than under a lower value of A. Thus the appropriate choice of N and Z is an important practical consideration in applying this procedure to affiliated auction environments. We will consider several possibilities below.

For a given pair of variables (N, Z), most of the identification argument in Section 2 still goes through, with $g(b, m|N^*)$ replacing $g(b|N^*)$. One important change to the argument is that condition 5, which allows one to "order" the columns of the $G_{N|N^*}$ matrix by the values of N^*, may no longer be valid when $N \neq A$. An alternative condition must be used. Hu (2006) gives two alternative conditions, either of which may be substituted for condition 5 to achieve identification. Both of these are restrictions on the matrix $G_{N|N^*}$, the matrix of conditional probabilities $Pr(N = i|N^* = j)$:

Condition 8 $g(N = n|N^* = n) > g(N = n|N^* = m),\ m \neq n.$

This condition implies that the probability that our corrupted measurement N is correct is higher than the probability of taking any other values. This implies that the diagonal element of each column of the matrix $G_{N|N^*}$ is the largest element in that column; subsequently, given the columns of $G_{N|N^*}$, we can reorder them such that the maximum in each column occurs on the diagonal of the matrix.

A second alternative assumption is:

Condition 9 *There exists some row j, of $G_{N|N^*}$, such that* $g(N = j|N^* = 2) > g(N = j|N^* = 3) > \cdots > g(N = j|N^* = J)$.

That is, we know *a priori* the ordering of the elements within a given row of $G_{N|N^*}$. With this information, if we were given the columns of $G_{N|N^*}$, we would be able to reorder them such that the elements in the j-th row were monotonic.

Clearly, the plausibility of either of these conditions would depend on the specific choice of the N variable.[2]

7 Conclusions

To be added.

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Athey, S., J. Levin, and E. Seira (2005): "Comparing Open and Sealed Bid Auctions: Theory and Evidence from Timber Auctions," working paper, Harvard University.

Bajari, P., and A. Hortacsu (2003): "Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions," *RAND Journal of Economics*, 34, 329–355.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–74.

Haile, P., H. Hong, and M. Shum (2003): "Nonparametric Tests for Common Values in First-Price Auctions," NBER working paper #10105.

Hendricks, K., J. Pinkse, and R. Porter (2003): "Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common-Value Auctions," *Review of Economic Studies*, 70, 115–145.

Hu, Y. (2006): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables," Manuscript, University of Texas.

Krasnokutskaya, E., and K. Seim (2005): "Bid Preference Programs and Participation in Highway Procurement Auctions," working paper, University of Pennsylvania.

[2] Hu (2006) also gives two additional alternative conditions, which are monotonicity restrictions on the conditional distribution of $(b, m|N^*)$. In the affiliated value case, however, it is well-known that the equilibrium bids may not be monotonic in N^* (cf. Pinkse and Tan (2005)) so that these conditions will generally not hold in affiliated-value auction environments.

LAFFONT, J. J., H. OSSARD, AND Q. VUONG (1995): "Econometrics of First-Price Auctions," *Econometrica*, 63, 953–980.

LI, T. (2000): "Econometrics of First-Price Auctions with Binding Reservation Prices," Manuscript, Indiana University.

LI, T., I. PERRIGNE, AND Q. VUONG (2002): "Structural Estimation of the Affiliated Private Value Acution Model," *RAND Journal of Economics*, 33, 171–193.

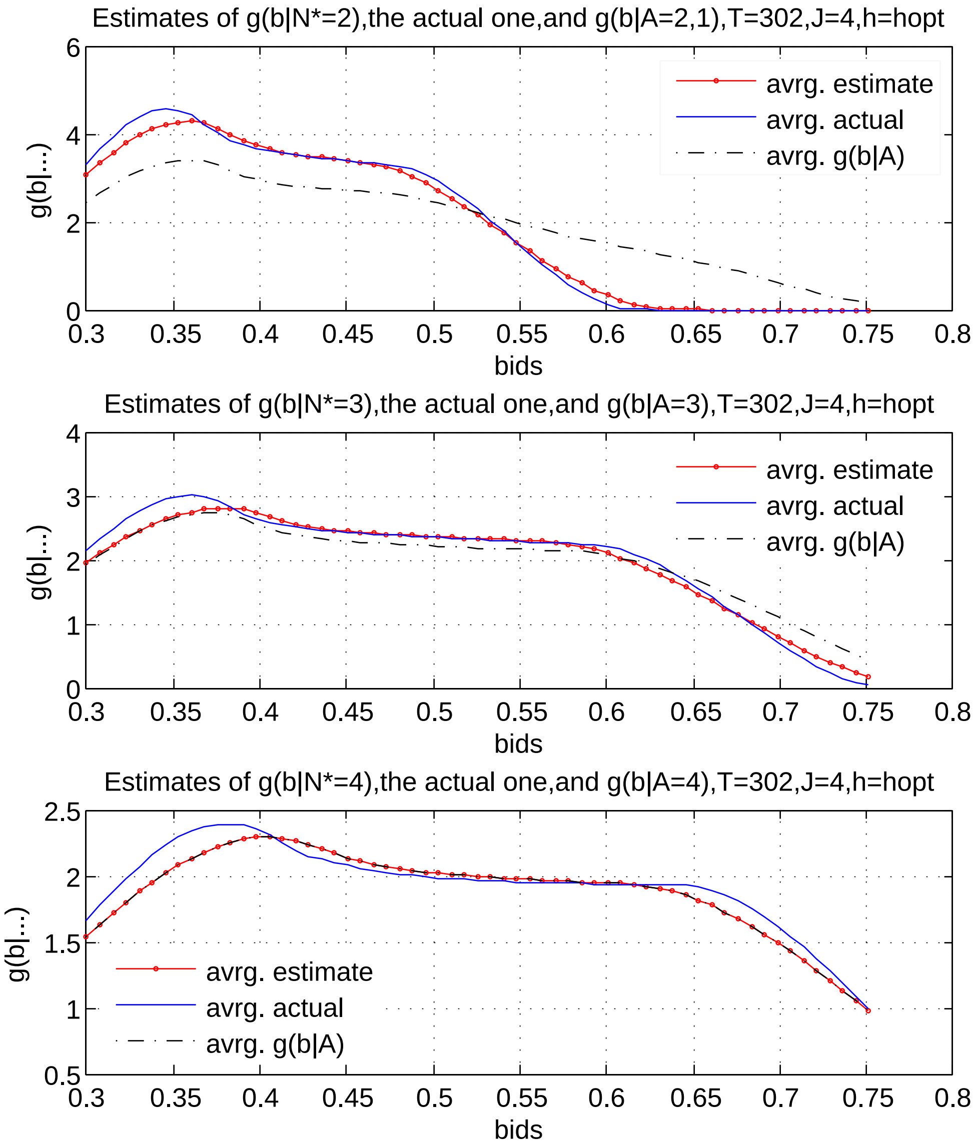
LI, T., AND X. ZHENG (2006): "Entry and Competition Effects in First-Price Auctions: Theory and evidence from Procurement Auctions," working paper, Vanderbilt University.

MILGROM, P., AND R. WEBER (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1089–1122.

PAARSCH, H. (1997): "Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbian Timber Sales," *Journal of Econometrics*, 78, 333–357.

PINKSE, J., AND G. TAN (2005): "The affiliation effect in first-price auctions," *Econometrica*, pp. 263–277.

Figure 1: Monte Carlo Evidence: $K = 4$



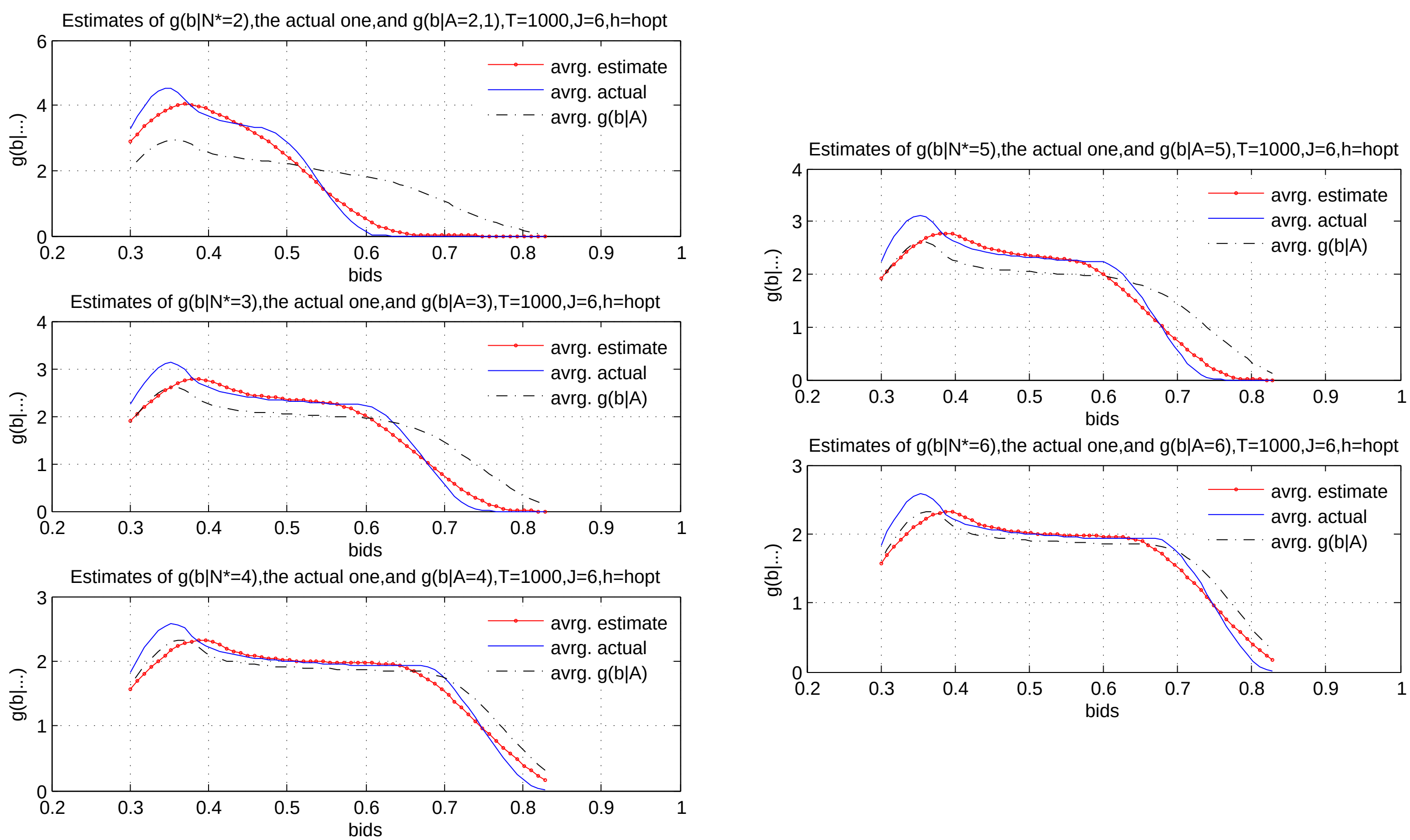


Figure 2: Monte Carlo Evidence: $K = 6$

Table 1: Monte Carlo Evidence: $K = 4$

Percentiles of the L^2 norms of the error of $\widehat{g(b|N^*)}$, i.e. $\left(\int \left(\widehat{g(b|N^*)} - g(b|N^*)\right)^2 db\right)^{0.5}$

Percentiles	$N^* = 2$	$N^* = 3$	$N^* = 4$
10%	1.6264	1.2695	0.9626
25%	2.0083	1.5394	1.2424
50%	2.5363	1.8347	1.6311
75%	3.458	2.3425	2.2104
90%	4.3469	2.8247	2.8988

Percentiles of the L^2 norms of the error of $g(b|A)$, i.e.

$$\left(\int \left(g(b|A = n) - g(b|N^* = n)\right)^2 db\right)^{0.5}$$

Percentiles	$A = 1$ or 2	$A = 3$	$A = 4$
10%	5.9286	1.4979	0.9626
25%	6.3439	1.7112	1.2424
50%	6.9409	2.0866	1.6311
75%	7.619	2.4846	2.2104
90%	8.2022	2.9435	2.8988

Table 2: Monte Carlo Evidence: $K = 6$

Percentiles of the L^2 norms of the error of $\widehat{g(b|N^*)}$, i.e. $\left(\int \left(\widehat{g(b|N^*)} - g(b|N^*)\right)^2 db\right)^{0.5}$

Percentiles	$N^* = 2$	$N^* = 3$	$N^* = 4$	$N^* = 5$	$N^* = 6$
10%	2.6992	1.7579	1.3615	1.1119	1.0304
25%	3.6732	2.1221	1.5779	1.2877	1.3606
50%	4.6263	2.6744	1.9319	1.5360	1.7903
75%	6.4446	3.2925	2.2107	1.8711	2.3820
90%	8.6287	4.1864	2.5614	2.3099	3.0962

Percentiles of the L^2 norms of the error of $g(b|A)$, i.e.

$$\left(\int \left(g(b|A = n) - g(b|N^* = n)\right)^2 db\right)^{0.5}$$

Percentiles	$A = 1$ or 2	$A = 3$	$A = 4$	$A = 5$	$A = 6$
10%	7.8520	3.3873	1.4608	0.8810	1.0304
25%	8.2920	3.6284	1.6421	1.0258	1.3606
50%	8.8599	3.8466	1.8236	1.2006	1.7903
75%	9.3901	4.1615	2.0642	1.4200	2.3820
90%	10.0023	4.4825	2.2262	1.7103	3.0962