# Estimating First-Price Auctions with Unknown Number of Bidders: A Misclassification Approach *

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

October 30, 2007
**Incomplete: Please do not quote**

**Abstract**

In this paper, we consider nonparametric identification and estimation of first-price auction models when $N^*$, the number of potential bidders, is unknown to the researcher. 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 $N^*$. Monte Carlo results illustrate that the procedure works well in practice. We present evidence from a dataset of procurement auctions, which shows that accounting for the unobservability of $N^*$ can lead to meaningful differences in the estimates of bidders' profit margins.

In many auction applications, researchers do not observe $N^*$, the number of bidders in the auction. (In the parlance of the literature, $N^*$ is the "number of potential bidders", a terminology we adopt in the remainder of the paper.) The most common scenario obtains under binding reserve prices. When reserve prices bind, the number of potential bidders $N^*$, which is observed by auction participants and influences their bidding behavior, differs from the observed number of bidders $A$ ($\leq N^*$), which is the number of auction participants whose bids exceed the reserve price. Other scenarios which would cause the true level of

*The authors can be reached at `yhu@jhu.edu` and `mshum@jhu.edu`. We thank Ken Hendricks, Harry Paarsch, Isabelle Perrigne, Jean-Marc Robin, Quang Vuong, and seminar participants at Brown, Caltech, UC-Irvine, Iowa, NC State, and SITE (Stanford) for helpful comments. Guofang Huang provided exceptional research assistance.

competition to be unobserved (and differ from the observed level of competition) include bidding or participation costs. In other cases, the number of auction participants may simply not be recorded in the researcher's dataset.

In this paper, we consider nonparametric identification and estimation of first-price auction models when $N^*$ is unobserved. Using recent results from the literature on misclassified regressors, we show how the equilibrium distribution of bids, given the unobserved $N^*$, can be identified and estimated. 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.

Accommodating the non-observability of $N^*$ is important for drawing valid policy implications from auction model estimates. Because $N^*$ is the level of competition in an auction, using a mismeasured value for $N^*$ can lead to wrong implications about the degree of competitiveness in the auction, and also the extent of bidders' markups and profit margins. This will be shown in the empirical illustration below.

Existing research 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 number of potential bidders $N^*$ is explicitly parameterized, so that not observing $N^*$ need not be a problem. For instance, Laffont, Ossard, and Vuong (1995) used a goodness-of-fit statistic to select the most plausible value of $N^*$ for French eggplant auctions. Paarsch (1997) treated $N^*$ essentially as a random effect and integrates it out over the 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 observing $N^*$ with error) raises difficulties. Within the independent private-values (IPV) 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) showed how to identify $N^*$ and the equilibrium bid distribution in the range of bids exceeding the reserve price. Hendricks, Pinkse, and Porter (2003) allowed $N^*$ to vary across auctions, and assume that $N^* = L$, where $L$ is a measure of the number of potential bidders which they construct.

Most closely related to our work is a paper by Song (2004). She solved the problem of the nonparametric estimation of ascending auction models in the IPV framework, when the number of potential bidders $N^*$ is unknown by the researcher. She showed that the distribution of valuations can be recovered from observation of any two valuations of which rankings from the top is known. Her methodology accommodates auctions where $N^*$ need not be fixed across auctions, which is also the case considered in this paper.[1]

In this paper, we consider a new approach to nonparametric identification and estimation of first-price auction models in which the number of bidders $N^*$ is observed by bidders, but unknown to the researcher. We use recent results from the recent econometric literature on models with misclassification error; e.g. Mahajan (2006), Hu (2006). Drawing an analogy between the misclassification problem and an auction model where the potential number of bidders $N^*$ is observed with error, 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. Our procedure requires two auxiliary variables: first, an imperfect noisy proxy for $N^*$; second, an instrument, which could be a second imperfect measure of $N^*$.

For first-price auctions, allowing the unobserved $N^*$ to vary across auctions is important because $N^*$ is the level of competition perceived by the bidders, and affects their equilibrium bidding strategies. Hence, not observing $N^*$ implies that the observed bids are drawn from a mixture distribution, where the "mixing densities" $g(b|N^*)$ and the "mixing weights" $\Pr(A|N^*)$ are both unknown. This motivates the application of methods developed for models with a misclassified regressor, where (likewise) the observed outcomes are drawn from a mixture distribution. In contrast, when $N^*$ is unknown, but fixed across auctions, the observed bids are drawn from a homogeneous sample, so that methods from the misclassification literature would not be needed.

In a different context, Li, Perrigne, and Vuong (2000) applied deconvolution results from the (continuous) measurement error literature to identify and estimate conditionally independent auction models in which bidders' valuations have common and private (idiosyncratic) components. Krasnokutskaya (2005) applied these ideas to estimate auction models with unobserved heterogeneity. To our knowledge, however, our paper is the first application of (discrete) measurement error results to non-observability of the number of bidders, in auction models.

The issues considered in this paper are close to those considered in the literature on entry

[1] Song (2006) showed that the top two bids are also enough to identify first-price auctions where the number of active bidders is stochastic and uncertain from the bidders' perspective, under the assumption that the the number of potential bidders is fixed.

in auctions: eg. Li (2005), 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, their one commonality is to model more explicitly bidders' participation decisions in auctions, which can cause the number of observed bidders $A$ to differ from the number of potential bidders $N^*$. For instance, Haile, Hong, and Shum (2003) consider an endogenous participation model in which the number of potential bidders is observed by the researcher, and equal to the observed number of bidders (i.e., $N^* = A$), so that non-observability of $N^*$ is not a problem. However, $A$ is potentially endogenous, because it may be determined in part by auction-specific unobservables which also affect the bids. By contrast, in this paper we assume that $N^*$ is unobserved, and that $N^* \neq A$, but we do not consider the possible endogeneity of $N^*$.[2]

In section 2, we describe our auction framework. In section 3, we present the main identification results, and describe our estimation procedure. 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. In section 6, we consider extensions of the approach to scenarios where only the winning bid is observed. Section 7 concludes.

# 1 Model

In this paper, we consider the case of first-price auctions under the symmetric independent private values (IPV) paradigm, for which identification and estimation are most transparent. For a thorough discussion of identification and estimation of these models when the number of potential bidders $N^*$ is known, see Paarsch and Hong (2006, Chap. 4). For concreteness, we focus on the case where a binding reserve price is the reason why the number of potential bidders $N^*$ differs from the observed number of bidders, and is not known by the researcher.

There are $N^*$ bidders in the auction, with each bidder drawing a private valuation from the distribution $F(x)$ which has support $[\underline{x}, \bar{x}]$. $N^*$ can vary freely across the auctions. There is a reserve price $r$, assumed to be fixed across all auctions, where $r > \underline{x}$.[3] The equilibrium

[2] In principle, we recover the distribution of bids (and hence the distribution of valuations) separately for each value of $N^*$, which accommodates endogeneity in a general sense. However, because we do not model the entry process explicitly (as in the papers cited above), we do not deal with endogeneity in a direct manner.

[3] Our estimation methodology can potentially also be used to handle the case where $N^*$is fixed across all auctions, but $r$ varies freely across auctions.

bidding function for bidder $i$ with valuation $x_i$ is

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

Hence, the observed 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

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

where $\xi(b; N^*)$ 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), which does not depend on $A$, the observed number of bidders.

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

$$\xi(b, N^*) = b + \frac{1}{N^* - 1} \left[ \frac{G(b|N^*)}{g(b|N^*)} + \frac{F(r)}{1 - F(r)} \cdot \frac{1}{g(b|N^*)} \right]. \quad (3)$$

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.

## 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 unobserved number of potential bidders $N^*$, as well as the conditional distribution of $N|N^*$.

We require two auxiliary 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 three conditions. The first two conditions are given here:

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

This assumption implies that $N$ or $Z$ affects the equilibrium density of bids only through the unknown number of potential bidders. In the econometric literature, this is known as the "nondifferential" measurement error assumption.

In what follows, we only consider values of $b$ such that $g(b|N^*) > 0$, for $N^* = 2, \ldots, K$. This requires, implicitly, knowledge of the support of $g(b|N^*)$, which is typically unknown to the researcher. Below, when we discuss estimation, we present a two-step procedure which circumvents this problem.

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

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

**Examples of $N$ and $Z$** Here we consider several examples of variables which could fulfill the roles of the special variables $N$ and $Z$.

1. 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, for a given $N^*$, the sampling density of any equilibrium bid exceeding the reserve price — as given in Eq. (2) above — does not depend on $A$, so that Condition 1 is satisfied.[4] A good candidate for the instrument $Z$ could be a noisy estimate of the potential number of bidders:

$$Z = h(N^*, \eta).$$

In order to satisfy conditions 1 and 2, we would require $b \perp \eta | N^*$, and also $A \perp \eta | N^*$. Because we are focused on the symmetric IPV model in this paper, we will consider this

[4]This is no longer true in affiliated value models.

example in the remainder of this section, and also in our Monte Carlo experiments and in the empirical illustration.

2. More generally, $N$ and $Z$ could be two noisy measures of $N^*$:

$$\begin{aligned} N &= f(N^*, \upsilon) \\ Z &= h(N^*, \eta). \end{aligned} \tag{4}$$

In order to satisfy conditions 1 and 2, we would require $b \perp (\upsilon, \eta)|N^*$, as well as $\eta \perp \upsilon|N^*$.

3. Another possibility is that $N$ is a noisy measure of $N^*$, as in example 2, but $Z$ is an exogenous variable which directly determines participation:

$$\begin{aligned} N &= f(N^*, \upsilon) \\ N^* &= k(Z, \nu). \end{aligned} \tag{5}$$

In order to satisfy conditions 1 and 2, we would require $b \perp (\upsilon, Z)|N^*$, as well as $\upsilon \perp Z|N^*$. This implies that $Z$ is excluded from the bidding strategy, and affects bids only through its effect on $N^*$.

Furthermore, in order for the second step of the estimation procedure (in which we recover bidders' valuations) to be valid, we also need to assume that $b \perp \nu|N^*$. Importantly, this rules out the case that the participation shock $\nu$ is a source of unobserved auction-specific heterogeneity.[5] Note that $\nu$ will generally be correlated with the bids $b$. ■

We observe a random sample of $\left\{\vec{b}_t, N_t, Z_t\right\}$, where $\vec{b}_t$ denotes the vector of observed bids $\{b_{1t}, b_{2t}, \ldots, b_{A_t t}\}$. (Note that we only observe $A_t$ bids for each auction $t$.) 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.[6]

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

$$g(b, N, Z) = \sum_{N^*=2}^{K} g(b|N^*, N, Z) g(N|N^*, Z) g(N^*, Z). \tag{6}$$

[5]In the case when $N^*$ is observed, correlation between bids and the participation shock $\nu$ can be accommodated, given additional restriction on the $k(\cdots)$ function. See Guerre, Perrigne, and Vuong (2005) and Haile, Hong, and Shum (2003) for details. However, when $N^*$ is unobserved, as is the case here, it is not clear how to generalize these results.

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

Under conditions 1 and 2, Eq. (6) becomes

$$g(b, N, Z) = \sum_{N^*=2}^{K} g(b|N^*)g(N|N^*)g(N^*, Z). \tag{7}$$

We define the 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

$$G_{b|N^*} = \begin{pmatrix} g(b|N^* = 2) & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & g(b|N^* = K) \end{pmatrix}. \tag{8}$$

All of these are $(K-1)$-dimensional square matrices. With this notation, Eq. (7) can be written as

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

Condition 2 implies that

$$g(N, Z) = \sum_{N^*=2}^{K} g(N|N^*)g(N^*, Z), \tag{10}$$

which, using the matrix notation above, is equivalent to

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

Equations (9) and (11) summarize the unknowns in the model, and the information in the data. The matrices on the left-hand sides of these equations are quantities which can be recovered from the data, whereas the matrices on the right-hand side are the unknown quantities of interest. As a counting exercise, we see that the matrices $G_{b,N,Z}$ and $G_{N,Z}$ contain $2(K-1)^2 - (K-1)$ known elements, while the $G_{N|N^*}$, $G_{N^*,Z}$ and $G_{b|N^*}$ matrices contain at most a total of also $2(K-1)^2 - (K-1)$ unknown elements. Hence, in principle, there is enough information in the data to identify the unknown matrices. The key part of the proof below is to characterize the solution and give conditions for uniqueness. Moreover, the proof is constructive in that it immediately suggests a way for estimation.

The third condition which the special variables $N$ and $Z$ must satisfy is a rank condition:

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

Note that this condition is directly testable from the sample. It essentially ensures that the instrument $Z$ affects the distribution of the proxy variable $N$ (resembling the standard instrumental relevance assumption in usual IV models).

Because Eq. (11) implies that

$$Rank\left(G_{N,Z}\right) \leq \min\left\{Rank\left(G_{N|N^*}\right), Rank\left(G_{N^*,Z}\right)\right\}, \tag{12}$$

it follows from Condition 3 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 the key equation

$$G_{b,N,Z}G_{N,Z}^{-1} = G_{N|N^*}G_{b|N^*}G_{N|N^*}^{-1}. \tag{13}$$

The matrix on the left-hand side can be formed from the data. For the expression on the right-hand side, note that because $G_{b|N^*}$ is diagonal (cf. Eq. (8)), the RHS matrix represents an eigenvalue-eigenvector decomposition of the LHS matrix. This is the key representation which will facilitate estimation of the unknown matrices $G_{N|N^*}$ and $_{b|N^*}$.

In order to make the eigenvalue-eigenvector decomposition in Eq. (13) 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 the eigenvalues in $G_{b|N^*}$ are distinctive for some bid $b$, which ensures that the eigenvalue decomposition in Eq. (13) exists and is unique, for some bid $b$. This assumption guarantees that all the linearly independent eigenvectors are identified from the decomposition in Eq. (13). Suppose that for some value $\widetilde{b}$, $g(\widetilde{b}|N^* = i) = g(\widetilde{b}|N^* = j)$, which implies that the two eigenvalues corresponding to $N^* = i$ and $N^* = j$ are the same, so that the two corresponding eigenvectors cannot be uniquely identified. This is because any linear combination of the two eigenvectors is still an eigenvector. Assumption 4 guarantees that there exists another value $\overline{b}$ such that $g(\overline{b}|N^* = i) \neq g(\overline{b}|N^* = j)$. Notice that Eq. (13) holds for every $b$, implying that $g(\widetilde{b}|N^* = i)$ and $g(\overline{b}|N^* = i)$ correspond to the same eigenvector, as do $g(\widetilde{b}|N^* = j)$ and $g(\overline{b}|N^* = j)$. Therefore, although we cannot use $\widetilde{b}$ to uniquely identify the two eigenvectors corresponding to $N^* = i$ and $N^* = j$, we can use the value $\overline{b}$ to identify them.

Eq. (13) shows that an eigenvalue decomposition of the observed $G_{b,N,Z}G_{N,Z}^{-1}$ matrix recovers the unknown $G_{N|N^*}$ and $G_{b|N^*}$ matrices, with $G_{b|N^*}$ being the diagonal matrix of eigenvalues, and $G_{N|N^*}$ being the corresponding matrix of eigenvectors. This identifies $G_{b|N^*}$ and $G_{N|N^*}$ up to a normalization and ordering of the columns of the eigenvector matrix $G_{N|N^*}$.

There is a clear appropriate choice for the normalization of the eigenvectors; because each column of $G_{N|N^*}$ should add up to one, we can multiply each element $G_{N|N^*}(i,j)$ by the reciprocal of the column sum $\sum_i G_{N|N^*}(i,j)$, as long as $G_{N|N^*}(i,j)$ is non-negative.

The appropriate ordering of the columns of $G_{N|N^*}$ is less clear, and 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 observed number of bidders. This condition implies that for any $i, j \in \mathcal{N}$

$$g\left(N = j | N^* = i\right) = 0 \text{ for } j > i. \tag{14}$$

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.,

$$g\left(N = i | N^* = i\right) > 0 \text{ for all } i \in \mathcal{N}. \tag{15}$$

Equations 14 and 15 imply that $n^*$ is the 100th percentile of the discrete distribution $g\left(N | N^* = n^*\right)$, i.e.,

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

In other words, Condition 5 implies that, once we have the columns of $G_{N|N^*}$ obtained as the eigenvectors from the matrix decomposition (13), the right ordering can be obtained by re-arranging these columns so that they form an upper-triangular matrix.

Hence, under Conditions 1-5, $G_{b|N^*}$, $G_{N|N^*}$ and also $G_{N^*,Z}$ are identified (the former pointwise in $b$).

Before proceeding, we note that in the key equation (13), the matrix $G_{N|N^*}$ is identical for all $b$, and because this equation holds for all $b$, there is a large degree of overidentification in this model. This suggests the possibility of achieving identification with weaker assumptions. In particular, it may be possible to relax the non-differentiability condition 1 so that we require $g(b|N^*, N, Z) = g(b|N^*)$ only at one particular value of $b$. We are exploring the usefulness of such possibilities in ongoing work.

## 3 Estimation of bid densities $\mathbf{g(b|N^*)}$: two-step procedure

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:

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

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 One** In Step 1, we 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

$$G_{Eb,N,Z} = \left[E\left(b|N = i, Z = j\right) g(N = i, Z = j)\right]_{i,j}, \tag{17}$$

and

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

From equation 7, we have

$$E(b|N,Z)g(N,Z) = \sum_{N^*=2}^{K} E(b|N^*)g(N|N^*)g(N^*,Z). \tag{18}$$

Therefore, the same diagonalization result in equation 13 holds for equation 18 with $G_{b,N,Z}$ and $G_{b|N^*}$ replaced with $G_{Eb,N,Z}$ and $G_{Eb,N^*}$ as follows:

$$G_{Eb,N,Z}G_{N,Z}^{-1} = G_{N|N^*}G_{Eb|N^*}G_{N|N^*}^{-1}. \tag{19}$$

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. As shown in Hu (2006), the function $\psi(\cdot)$ is a nonstochastic analytic function. Therefore,we may estimate $G_{N|N^*}$ as follows:

$$\widehat{G}_{N|N^*} := \psi\left(\widehat{G}_{Eb,N,Z}\widehat{G}_{N,Z}^{-1}\right), \tag{20}$$

where $\widehat{G}_{Eb,N,Z}$ and $\widehat{G}_{N,Z}$ may be constructed directly from the sample. In our empirical example, we use an estimate for $\widehat{G}_{Eb,N,Z}$ as follows:

$$\widehat{G}_{Eb,N,Z} = \left[\frac{1}{T}\sum_{t}\frac{1}{N_j}\sum_{i=1}^{N_j} b_{it}\mathbf{1}(N_t = N_j, Z_t = Z_k)\right]_{j,k}.$$

Since we don't have a covariate in the simulation, all the entries in the matrices in the equation 18 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 $g(N|N^*,w)$. This alternative is presented in the Appendix.

**Step Two** In Step 2, we 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 13, we have for any $b$

$$G_{N|N^*}^{-1}\left(G_{b,N,Z}G_{N,Z}^{-1}\right)G_{N|N^*} = G_{b|N^*}. \tag{21}$$

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

$$g(b|N^*) = e_{N^*}^T \left[ G_{N|N^*}^{-1} \left( G_{b,N,Z} G_{N,Z}^{-1} \right) G_{N|N^*} \right] e_{N^*}. \tag{22}$$

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

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

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 $\widehat{G_{b,N,Z}} = \left[ \widehat{g}_{b,N,Z} (b, N_j, Z_k) \right]_{j,k}$:

$$\widehat{g}_{b,N,Z} (b, N_j, Z_k) = \left[ \frac{1}{Th} \sum_t \frac{1}{N_t} \sum_{i=1}^{N_t} K \left( \frac{b - b_{it}}{h} \right) \mathbf{1}(N_t = N_j, Z_t = Z_k) \right].$$

We analyze the asymptotic properties of our estimator in detail in the appendix. Here we provide a brief summary. 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. (17)), the estimates of $\widehat{G}_{N|N^*}$ (obtained using Eq. (20)) 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. (22), will be determined by the convergence properties of the kernel estimate of $g(b, N, Z)$, which converges slower than $\sqrt{T}$. We also present the uniform convergence rate of $\widehat{g}(b|N^*)$ in the appendix. Similarly, the pointwise asymptotic distribution of $\widehat{g}(b|N^*)$ is also determined by that of the kernel estimator of $g(b, N, Z)$. In fact, we show in the appendix that $(Th)^{1/2} [\widehat{g}(b|N^*) - g(b|N^*)]$ for a given $b$ converges to a normal distribution.

As shown in the appendix, the convergence properties of our estimator relies on the smoothness of $g(b|N, Z)$. This conditional distribution is a mixture of distributions $g(b|N^*)$ with a different support for a different $N^*$. Therefore, the smoothness of $g(b, N, Z)$ requires that of $g(b|N^*)$ on not just its support $[r, u_{N^*}]$ but the largest support $[r, u_K]$. When the supports of $g(b|N^*)$ are known, we only require the smoothness of $g(b|N^*)$ on its own support $[r, u_{N^*}]$ because the distribution $g(b|N, Z)$ can be estimated piecewise on $[r, u_2], [u_2, u_3], ..., [u_{K-1}, u_K]$. When the supports of $g(b|N^*)$ are unknown, we require the density $g(b|N^*)$ for each value of $N^*$ to be smooth at the upper boundary.

As pointed out in GPV (2000), the density $g(b|N^*)$ may not be bounded at the reserve price $r$. As they suggest, we may use the transformation $\sqrt{b-r}$. Notice that our identification and estimation remain the same if $b$ replaced by $\sqrt{b-r}$.

The asymptotic properties of the private values may also be derived given the direct relationship between the density $g(b|N^*)$ and the private values, as suggested in GPV (2000). We also shown in the appendix that the private values converge uniformly and has a normal distribution in the limit.

The matrix $G_{N|N^*}$, which is a by-product of the estimation procedure, can be useful for specification testing, when $N = A$, the observed number of bidders. Under the assumption that the difference between the observed number of bidders $A$ and the number of potential bidders $N^*$ arises from a binding reserve price, and that the reserve price $r$ is fixed across all the auctions with the same $N^*$ in the dataset, it is well-known (cf. Paarsch (1997)) that

$$A|N^* \sim \text{Binomial}(N^*, 1 - F_v(r)) \tag{23}$$

where $F_v(r)$ denotes the CDF of bidders' valuations, evaluated at the reserve price. This suggests that the recovered matrix $G_{A|N^*}$ can be useful in two respects. First, using Eq. (23), the truncation probability $F_v(r)$ could be estimated. This is useful when we use the first-order condition (3) to recover bidders' valuations. Alternatively, we could also test whether the columns of $G_{A|N^*}$, which correspond to the probabilities $\Pr(A|N^*)$ fir a fixed $N^*$, are consistent with the binomial distribution in Eq. (23).

## 4 Monte Carlo Evidence

In this section, we present some Monte Carlo evidence for the IPV model. We 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:

$$b^*(x;N^*) = \begin{cases} \left(\frac{N^*-1}{N^*}\right)x + \frac{1}{N^*}\left(\frac{r}{x}\right)^{N^*-1} r & \text{if } x \geq r \\ 0 & \text{if } x < r. \end{cases} \tag{24}$$

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 observed bidders $A_t$, and $Z_t$ is a second corrupted measure of $N_t^*$.

For each auction $t$, the number of potential 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:

$$Z_t = \begin{cases} N_t^* & \text{with probability } q \\ \text{unif.} \{2, 3, \ldots, J\} & \text{with probability } 1 - q. \end{cases} \tag{25}$$

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

$$A_t = \sum_{j \in \mathcal{N}_t^*} \mathbf{1}(x_j \geq r) \tag{26}$$

Finally, for each auction $t$, and each observed bidder $j \in \mathcal{A}_t$, we can calculate the equilibrium bid using Eq. (24).

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 $N^*$. All the identification arguments above continue to hold.

## 4.1 Results

We present results from $S = 200$ replications of a simulation experiment. 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 observed number of bidders, for $n = \{1, 2\}, 3, 4$, which we consider a "naïve" estimator for $g(b|N^* = n)$. For $N^* = 2, 3$, our estimator outperforms the naïve estimator, especially for the case of $N^* = 2$.

[Figure 1 about here.]

[Figure 2 about here.]

In Figure 2, we present estimates of bidders' valuations. In each graph, we graph the bids against three measures of the corresponding valuation: (i) the actual valuation, computed from Eq. (3) using the actual bid densities $g(b|N^*)$, and labeled "values"; (ii) the estimated valuations using our estimates of $g(b|N^*)$, labeled "Pseudovalue"[7]; and (iii) naïve estimates of the valuations, computed using $g(b|A)$, the observed bid densities conditional on the observed number of bidders.[8]

The graphs show that the most notable differences between the valuation estimates arise for larger bids. This is not surprising, because as the estimated density graphs above showed, the biggest differences in the densities were also at the upper tail. For the $N^* = 2$ case, where the differences between the estimated and naïve valuations are most apparent, we see that the naïve approach underestimates the valuations, which implies that bidders' profit margins $(v - b)$ will also be underestimated. This makes sense, because $N^* \geq A$, so that the set of auctions with a given value of $A$ actually have a true level of competition larger than $A$. Hence, the naïve approach overstates the true level of competition for each value of $A$ (except the highest value), which leads to underestimation of the profit margin.

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 3. Clearly, our estimator continues to perform well. In both the $K = 4$ as well as the $K = 6$ case, we see that the differences between our estimator and the naïve estimator diminish. This may not be surprising, because 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 3 about here.]

[Figure 4 about here.]

[7] In computing these valuations, the truncation probability $F(r)$ in Eq. (3) is obtained from the first-step estimates of the misclassification probability matrix $G_{N|N^*}$ as $\hat{F}(r) = 1 - \left[\hat{G}(N^*|N^*)\right]^{1/N^*}$.

[8] In computing these valuations, we use the first-order condition $\xi(b; A) = b + \frac{G(b|A)}{(A-1)\cdot g(b|A)}$, which ignores the possibility of a binding reserve price.

The valuations implied by our estimates of the bid densities, for the $K = 6$ case, are presented in Figure 4. Qualitatively, the results are very similar to the $K = 4$ results presented earlier, with the largest differences between the valuations estimated using our approach, and using a naïve approach coming at the upper tail of the bids.

## 5 Empirical illustration

In this section, we consider an application of the methodology presented above to a dataset of low-bid construction procurement auctions held by the New Jersey Department of Transportation (NJDOT) in the years 1989–1997. This dataset was previously analyzed in Hong and Shum (2002), and a full description of it is given there.

For the purposes of applying our estimation methodology, we assume that bidders' valuations are drawn in an IPV framework. The earlier analysis in Hong and Shum (2002) allowed for affiliated values, but assumed a parametric family (joint log-normality) for bidders' private information. Here, we estimate nonparametrically, and allow for the number of bidders $N^*$ to be unobserved by the researcher, but at the cost of the more restrictive IPV information structure.

Among all the auctions in our dataset, we focus on two categories of construction projects, for which the number of auctions is the largest: highway work and road grading/paving. In Table 1, we present some summary statistics on the auctions used in the analysis. Note that for both project categories, there are auctions with just one bidder, in which non-infinite bids were submitted. If the observed number of bidders is equal to $N^*$, the number of potential bidders perceived by bidders when they bid, then the non-infinite bids observed in these one-bidder auctions is difficult to explain from a competitive bidding point of view.[9] However, occurrences of one-bidder auctions can be a sign that the observed number of bidders is not indicative of the true extent of competition, and the methodology developed in this paper allows for this possibility.

[Table 1 about here.]

For the two special variables, we used $A$, the number of observed bidders, in the role of the

[9] Indeed, Li and Zheng (2006, pg. 9) point out that even when bidders are uncertain about the number of competitors they are facing, finite bids are difficult to explain when bidders face a non-zero probability that they could be the only bidder.

noisy measure $N$. In the role of the instrument $Z$, we constructed a measure of the average number of observed bidders in the five previous auctions of the same project category which took place before a given auction.

In order to satisfy condition 3, which requires that the matrix $G_{N|Z}$ be full rank, we divided the values of $A$ and $Z$ into three categories: $\{(1, 2, 3), 4, 5+\}$, and correspondingly consider only three distinct values for $N^* \in \{3, 4, 5\}$. Furthermore, the ordering assumption that we make is that $A \leq N^*$, which is consistent with the story that bidders decide not to submit a bid due to an (implicit) reserve price.[10]

Because we model these auctions in a simplified setting, we do not attempt a full analysis of these auctions. Rather, this application also highlights some practical issues in implementing the estimation methodology, which did not arise in the Monte Carlo studies. There are two important issues. First, the assumption that $A \leq N^*$ implies that the matrix on the right-hand side of the key equation (19) should be upper triangular, and hence that the matrix on the left-hand side, $G_{Eb,N|Z}G_{N|Z}^{-1}$, which is observed from the data, should also be upper-triangular. However, in practice, the left-hand side matrix is not upper-triangular. We consider two remedies for this. For most of this section, we focus on results obtained by imposing upper-triangularity on the left-hand side matrix, by setting all lower-triangular elements of the matrix to zero.[11] Later, however, we also consider results obtained without imposing upper-triangularity, which requires an alternative ordering condition to identify the column order of the eigenvector matrix $G_{N|N^*}$.

Second, even after imposing upper-triangularity, it is still possible that the eigenvectors and eigenvalues could have negative elements, which is inconsistent with the interpretation of them as densities and probabilities.[12] When our estimate of the densities $g(b|N^*)$ took on negative values, our remedy was to set the density equal to zero, but normalize our density estimate so that the resulting density integrated to one.[13]

**Results: Highway work auctions** This procedure is apparent in Figure 5, which contains the graphs of the estimated densities $g(b|N^*)$ for $N^* = 3, 4, 5$, for the highway work auctions. In each graph, we present three estimates of each $g(b|N^*)$: (i) the naïve estimate,

[10] See Hong and Shum (2002, Appendix B.1) for more discussion of a model with implicit reserve prices, for this dataset.

[11] If we assume that the non-upper triangularity is just due to small sample noise, then this is a valid procedure.

[12] This issue also arose in our Monte Carlo studies, but went away when we increased the sample size.

[13] Here we follow the recommendation of Efromovich (1999, pg. 63).

given by $g(b|A)$; (ii) the raw un-normalized estimate, which includes the negative values for the density, labeled "eig est"; and (iii) the normalized estimate with the negative portions removed, labeled "trunc est".

[Figure 5 about here.]

Figure 5 shows that the naïve bid density estimates, using $A$ in place of $N^*$, overweights small bids, which is reminiscent of the Monte Carlo results. As above, the reason for this seems to be that the number of potential bidders $N^*$ exceeds the observed number of bidders $A$. In the IPV framework, more competition drives down bids, implying that using $A$ to proxy for the unobserved level of competition $N^*$ may overstate the effects of competition. Because in this empirical application we do not know and control the data-generating process, so these economically sensible differences between the naïve estimates (using $g(b|A)$) and our estimates (using $g(b|N^*)$) serve as a reality check on the assumptions underlying our estimator.

For these estimates, the estimated $G_{A|N^*}$ matrix was

| | $N^* = 3$ | $N^* = 4$ | $N^* = 5$ |
|---|---|---|---|
| $A = (1, 2, 3)$ | 1.0000 | 0.1490 | 0.2138 |
| $A = 4$ | 0 | 0.8510 | 0.4237 |
| $A \geq 5$ | 0 | 0 | 0.3625 |

Furthermore, for the normalized estimates of the bid densities with the negative portions removed, the implied values for $E[b|N^*]$, the average equilibrium bids conditional on $N^*$, were 7.984, 7.694, 4.162 for, respectively, $N^* = 3, 4, 5$ (in millions of dollars).

The corresponding valuation estimates, obtained by solving Eq. (3) pointwise in $b$ using our bid density estimates, are graphed in Figure 6. We present the valuations estimated using our approach, as well as a naïve approach using $g(b|A)$ as the estimate for the bid densities. Note that the valuation estimates become negative within a low range of bids, and then at the upper range of bids, the valuations are decreasing in the bids, which violates a necessary condition of equilibrium bidding. These may be due to unreliability in estimating the bid densities $g(b|A)$ and $g(b|N^*)$ close to the bounds of the observed support of bids. Furthermore, in the estimates of $g(b|N^*)$, we see that the valuations rise steeply for low bids. This arises from the truncation procedure, which leads to a kink in the bid density at the point when the density changes from zero to a positive value.

[Figure 6 about here.]

Comparing the estimates of valuations using $g(b|N^*)$, and those obtained using $g(b|A)$, we see that the valuations using $g(b|N^*)$ are smaller than those using $g(b|A)$, for $N^* = 3, 4$ (but virtually indistinguishable for $N^* = 5$). As in the Monte Carlo results, this implies that the markups $(b - c)/b$ are larger using our estimates of $g(b|N^*)$. The differences in implied markups between these two approaches is economically meaningful, as illustrated in the right-hand-side graphs in Figure (6). For example, for $N^* = 4$, at a bid of $5 million, the corresponding markup using $g(b|A = 4)$ is around 15%, or $750,000, but using $g(b|N^* = 4)$ is around 40%, or $2 million. This suggests that failing to account for unobservability of $N^*$ can lead the researcher to understate bidders' profit margins.

**Additional Results** In figures 7 and 8, we present the estimated bid densities and valuations, for the grading and paving contracts. Generally, the results are qualitatively the same as in the highway work contracts discussed in detail above, but the results are not as clean. The expected equilibrium bids, for $N^* = 3, 4, 5$ are, respectively 0.948, 1.768, and 1.361 (in millions of 1989 dollars).

As in the highway work results, we see that the naïve estimates of the bid densities, in Figure 7, overweight the small bids relative to the estimates of $g(b|N^*)$. The estimated matrix $G_{A|N^*}$ for these auctions is

| | $N^* = 3$ | $N^* = 4$ | $N^* = 5$ |
|---|---|---|---|
| $A = (1, 2, 3)$ | 1.0000 | 0.2695 | 0.1202 |
| $A = 4$ | 0 | 0.7305 | 0.1937 |
| $A \geq 5$ | 0 | 0 | 0.6861 |

[Figure 7 about here.]

[Figure 8 about here.]

Up to this point, the empirical results have imposed upper-triangularity of the matrix on the left-hand side of Eq. (19), which is an implication of the ordering assumption that $A \leq N^*$. We next consider estimation of $g(b|N^*)$ without imposing upper-triangularity of this matrix, and use the alternative ordering condition that the top row of $G_{A|N^*}$ is decreasing from left to right; i.e.,

$$G(A = (1, 2, 3)|N^* = 3) > G(A = 4|N^* = 3) > G(A = 5|N^* = 3).$$

The estimates of the bid densities $g(b|N^*)$ under this alternative specification, for both project categories, are shown in Figure 9. These results are not as attractive as the results which imposed upper-triangularity. For example, note that the bid densities can be erratically shaped, as in the $N^* = 3$ case for the highway work contracts. Furthermore, when $G_{A|N^*}$ is not upper-triangular, there is difficulty interpreting what it means when $A \geq N^*$ (that is, the observed number of bidders exceeds the unobserved number of potential bidders).

[Figure 9 about here.]

# 6 Extension: Only Winning Bids are Recorded

In some first-price auction settings, only the winning bid is observed by the researcher. This is particularly likely for the case of descending price, or *Dutch* auctions, which end once a bidder signals his willingness to pay a given price. For instance, Laffont, Ossard, and Vuong (1995) consider descending auctions for eggplants where only the winning bid is observed, and van den Berg and van der Klaauw (2007) estimate Dutch flower auctions where only a subset of bids close to the winning bid are observed. Within the symmetric IPV setting considered here, Guerre, Perrigne, and Vuong (2000) and Athey and Haile (2002) argue that observing the winning bid is sufficient to identify the distribution of bidder valuations, provided that $N^*$ is known. Our estimation methodology can be applied to this problem even when the researcher does not know $N^*$, under two scenarios.

**First Scenario: Non-Binding Reserve Price** In the first scenario, we assume that there is no binding reserve price, but the researcher does not observe $N^*$. (Many Dutch auctions take place too quickly for the researcher to collect data on the number of participants.) Because there is no binding reserve price, the winning bid is the largest out of the $N^*$ bids in an auction. In this case, bidders' valuations can be estimated in a two-step procedure.

In the first step, we estimate $g_{WB}(\cdot|N^*)$, the equilibrium density of winning bids, conditional on $N^*$, using the methodology above.

In the second step, we exploit the fact that in this scenario, the equilibrium CDF of winning bids is related to the equilibrium CDF of the bids by the relation:

$$G_{WB}(b|N^*) = G(b|N^*)^{N^*}.$$

This implies that the equilibrium bid CDF can be estimated as $\hat{G}(b|N^*) = \hat{G}_{WB}(b|N^*)^{1/N^*}$, where $\hat{G}_{WB}(b|N^*)$ denotes the CDF implied by our estimates of $\hat{g}_{WB}(b|N^*)$. Subsequently, upon obtaining an estimate of $\hat{G}(b|N^*)$ and the corresponding density $\hat{g}(b|N^*)$, we can evaluate Eq. (3) at each $b$ to obtain the corresponding bidder valuation.

**Second Scenario: Binding Reserve Price, but A Observed** In the second scenario, we assume that the reserve price binds, but that the number of bidders who are willing to submit a bid above the reserve price are observed. The reason we require $A$ to be observed is that when reserve prices bind, the winning bid is not equal to $b^{N^*:N^*}$, the highest order statistic out of $N^*$ i.i.d. draws from $g(b|N^*)$, which is the equilibrium bid distribution truncated to $[r, +\infty)$. Rather, for a given $N^*$, it is equal to $b^{A:A}$, the largest out of $A$ i.i.d. draws from $g(b|N^*)$. Hence, because the density of the winning bid depends on $A$, even after conditioning on $N^*$, we must use $A$ as a conditioning covariate in our estimation.

For this scenario, we estimate $g(b|N^*)$ in two steps. First, treating $A$ as a conditioning covariate (as in Appendix A), we estimate $g_{WB}(\cdot|A, N^*)$, the conditional density of the winning bids conditional on both the observed $A$ and the unobserved $N^*$. Second, for a fixed $N^*$, we can recover the conditional CDF $G(b|N^*)$ via

$$\hat{G}(b|N^*) = \hat{G}_{WB}(b|A, N^*)^{1/A}, \ \forall A.$$

(That is, for each $N^*$, we can recover an estimate of $G(b|N^*)$ for each distinct value of $A$.)

In both scenarios, we need to find good candidates for the special variables $N$ and $Z$. Since typically many Dutch auctions are held in a given session, one possibility for $N$ could be the total number of attendees at the auction hall for a given session, while $Z$ could be an instrument (such as the time of day) which affects bidders' participation for a specific auction during the course of the day.[14]

[14]This corresponds to the scenario considered in the flower auctions in van den Berg and van der Klaauw (2007).

# 7 Conclusions

In this paper, we have explored the application of methodologies developed in the econometric measurement error literature to the estimation of structural auction models, when the number of potential bidders is not observed. The analysis is incomplete in a number of ways. First, we have not yet provided a fuller derivation of the asymptotic theory for our estimation procedure. Second, for the empirical application, we have not conditioned the observed bids on covariates. Third, we will consider the usefulness of these procedures for affiliated values auction models.

More broadly, one maintained assumption in this paper that $N^*$ is observed and deterministic from bidders' point of view, but not known by the researcher. The empirical literature has also considered models where the number of bidders $N^*$ is stochastic from the bidders' perspective: e.g., Athey and Haile (2002); Hendricks, Pinkse, and Porter (2003); Bajari and Hortacsu (2003); Li and Zheng (2006); and Song (2006). It will be interesting to explore whether the methods used here can be useful in these models. Finally, these methodologies developed in this paper may also be applicable to other structural models in industrial organization, where the number of participants is not observed by the researcher. We are considering these applications in future work.

# References


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

Athey, S., and P. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70, 2107–2140.

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.

Efromovich, S. (1999): *Nonparametric Curve Estimation: Methods, Theory, and Applications.* Springer-Verlag.

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

——— (2005): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," Manuscript, Pennsylvania State University.

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.

Hong, H., and M. Shum (2002): "Increasing Competition and the Winner's Curse: Evidence from Procurement," *Review of Economic Studies*, 69, 871–898.

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

Krasnokutskaya, E. (2005): "Identification and Estimation in Highway Procurement Auctions under Unobserved Auction Heterogeneity," Manuscript, University of Pennsylvania.

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

Laffont, J. J., H. Ossard, and Q. Vuong (1995): "Econometrics of First-Price Auctions," *Econometrica*, 63, 953–980.

Li, T. (2005): "Econometrics of First Price Auctions with Entry and Binding Reservation Prices," *Journal of Econometrics*, 126, 173–200.

Li, T., I. Perrigne, and Q. Vuong (2000): "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 98, 129–161.

——— (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.

Mahajan, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631–665.

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

Paarsch, H., and H. Hong (2006): *An Introduction to the Structural Econometrics of Auction Data*. MIT Press, with M. Haley.

Song, U. (2004): "Nonparametric Estimation of an E-Bay Auction Model with an Unknown Number of Bidders," working paper, University of British Columbia.

——— (2006): "Nonparametric Identification and Estimation of a First-Price Auction Model with an Uncertain Number of Bidders," working paper, University of British Columbia.

van den Berg, G., and B. van der Klaauw (2007): "If Winning Isn't Everything, Why do they Keep Score? A Structural Empirical Analysis of Dutch Flower Auctions," mimeo, Free University Amsterdam.

# A Appendix: asymptotic properties of the two step estimator

## A.1 Uniform consistency

In the first step, we estimate $\widehat{G}_{N|N^*}$ from

$$\widehat{G}_{N|N^*} := \psi\left(\widehat{G}_{Eb,N,Z}\widehat{G}_{N,Z}^{-1}\right), \tag{27}$$

where $\psi(\cdot)$ is an analytic function as shown in Hu (2007) and

$$\begin{aligned}\widehat{G}_{Eb,N,Z} &= \left[\frac{1}{T}\sum_t \frac{1}{N_j}\sum_{i=1}^{N_j} b_{it}\mathbf{1}(N_t = N_j, Z_t = Z_k)\right]_{j,k}, \\ \widehat{G}_{N,Z} &= \left[\frac{1}{T}\sum_t \mathbf{1}(N_t = N_j, Z_t = Z_k)\right]_{j,k}.\end{aligned}$$

We summarize the uniform convergence of $\widehat{G}_{N|N^*}$ as follows:

**Lemma 6** *Suppose that $Var(b|N,Z) < \infty$. Then,*

$$\widehat{G}_{N|N^*} - G_{N|N^*} = O_p\left(T^{-1/2}\right).$$

**Proof.** It is straightforward to show that $\widehat{G}_{Eb,N,Z} - G_{Eb,N,Z} = O_p\left(T^{-1/2}\right)$ and $\widehat{G}_{N,Z} - G_{N,Z} = O_p\left(T^{-1/2}\right)$. As shown in Hu (2007), and the function $\psi(\cdot)$ is an analytic function. Therefore, the result holds. ■

In the second step, we have

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

where

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

Let $\omega := (b, N, Z)$. Define the norm $\|\cdot\|_\infty$ as

$$\|\widehat{g}(\cdot|N^*) - g(\cdot|N^*)\|_\infty = \sup_b \left|\widehat{g}_{b|N^*}(b|N^*) - g_{b|N^*}(b|N^*)\right|.$$

The uniform convergence of $\widehat{g}(\cdot|N^*)$ is established as follows:

**Lemma 7** *Suppose:*

*3.1)* $\omega \in \mathcal{W}$ *and* $\mathcal{W}$ *is a compact set.*

*3.2)* $g_{b,N,Z}(\cdot, N_j, Z_k)$ *is continuously differentiable to order* $R$ *with bounded derivatives on an open set containing* $\mathcal{W}$.

*3.3)* $K(u)$ *is differentiable of order* $R$*, and the derivatives of order* $R$ *are bounded.* $K(u)$ *is zero outside a bounded set.* $\int_{-\infty}^{\infty} K(u)du = 1$ *, and there is a positive integer* $m$ *such that for all* $j < m$, $\int_{-\infty}^{\infty} K(u)u^j du = 0$. *And the characteristic function of* $K$ *is absolutely integrable.*

*3.4)* $h \to 0$ *and* $nh \to \infty$, *as* $n \to \infty$.

*Then, for all* $j$,

$$\|\widehat{g}(\cdot|N^*) - g(\cdot|N^*)\|_\infty = O_p\left[\left(\frac{T}{\ln T}h^{1+2R}\right)^{-1/2} + h^m\right]. \tag{28}$$

The convergence properties of the kernel estimate $g(b|N, Z)$ relies on its smoothness. This conditional distribution is a mixture of distributions $g(b|N^*)$ with a different support for a different $N^*$. Therefore, the smoothness of $g(b, N, Z)$ requires that of $g(b|N^*)$ on not just its support $[r, u_{N^*}]$ but the largest support $[r, u_K]$. When the supports of $g(b|N^*)$ are known, we only require the smoothness of $g(b|N^*)$ on its own support $[r, u_{N^*}]$ because the distribution $g(b|N, Z)$ can be estimated piecewise on $[r, u_2], [u_2, u_3], ..., [u_{K-1}, u_K]$. When the supports of $g(b|N^*)$ are unknown, we require the density $g(b|N^*)$ for each value of $N^*$ to be smooth at the upper boundary.

**Proof.** By lemma 6, it is straightforward to show that

$$\widehat{g}(b|N^*) = e_{N^*}^T \left[ G_{N|N^*}^{-1} \left( \widehat{G}_{b,N,Z}(b)\, G_{N,Z}^{-1} \right) G_{N|N^*} \right] e_{N^*} + O_p\left(T^{-1/2}\right).$$

In order to show the consistency of our estimator $\widehat{g}(b|N^*)$, we need the uniform convergence of $\widehat{g}_{b,N,Z}(\cdot, N_j, Z_k)$. The kernel density estimator has been studied extensively. Following results from lemma 8.10 in Newey and McFadden (1994, handbook), we have for all $j$ and $k$

$$\sup_b \left| \widehat{g}_{b,N,Z}(\cdot, N_j, Z_k) - g_{b,N,Z}(\cdot, N_j, Z_k) \right| = O_p\left[ \left( \frac{T}{\ln T} h^{1+2R} \right)^{-1/2} + h^m \right]. \tag{29}$$

The uniform convergence of $\widehat{g}_{b|N^*}$ then follows. ■

## A.2 Asymptotic Normality

Next, we show the asymptotic normality of $\widehat{g}(b|N^*)$ for a given value of $b$. Moreover, we define $\gamma_0(b) = vec\{G_{b,N,Z}(b)\}$, which is a column vector containing all the elements in the matrix $G_{b,N,Z}(b)$. Similarly, we define $\widehat{\gamma}(b) = vec\left\{\widehat{G}_{b,N,Z}(b)\right\}$. Therefore, the proof of lemma 7 suggests that

$$\widehat{g}(b|N^*) = \varphi(\widehat{\gamma}(b)) + O_p\left(T^{-1/2}\right)$$

where

$$\varphi(\widehat{\gamma}(b)) \equiv e_{N^*}^T \left[ G_{N|N^*}^{-1} \left( \widehat{G}_{b,N,Z}(b)\, G_{N,Z}^{-1} \right) G_{N|N^*} \right] e_{N^*}.$$

Notice that the function $\varphi(\cdot)$ is linear in each entry of the vector $\widehat{\gamma}(b)$. Therefore, we may have

$$\widehat{g}(b|N^*) - g(b|N^*) = \left( \frac{d\varphi}{d\gamma} \right)^T (\widehat{\gamma}(b) - \gamma_0(b)) + o_p\left(1/\sqrt{Th}\right),$$

where $\frac{d\varphi}{d\gamma}$ is nonstochastic because it is a function of $G_{N|N^*}$ and $G_{N,Z}$ only. The asymptotic distribution of $\widehat{g}(b|N^*)$ then follows that of $\widehat{\gamma}(b)$. We summarize the results as follows:

**Lemma 8** *Suppose that assumptions in lemma 7 hold with $R = 2$ and that*

1. *there exists some* $\delta$ *such that* $\int |K(u)|^{2+\delta} du < \infty$,

2. $(Th)^{1/2} h^2 \to 0$, *as* $T \to \infty$.

*Then, for a given* $b$ *and* $N^*$,

$$(Th)^{1/2} [\widehat{g}(b|N^*) - g(b|N^*)] \xrightarrow{d} N(0, \Omega),$$

*where*

$$\begin{aligned}\Omega &= \left(\frac{d\varphi}{d\gamma}\right)^T V(\widehat{\gamma}) \left(\frac{d\varphi}{d\gamma}\right), \\ V(\widehat{\gamma}) &= \lim_{T\to\infty} (Th) E\left[(\widehat{\gamma} - E(\widehat{\gamma}))(\widehat{\gamma} - E(\widehat{\gamma}))^T\right].\end{aligned}$$

**Proof.** As discussed before Lemma, the asymptotic distribution of $\widehat{g}(b|N^*)$ is derived from that of $\widehat{\gamma}(b)$. In order to proof that the asymptotic distribution of the vector $\widehat{\gamma}(b)$ is multivariate normal $N(0, V(\widehat{\gamma}))$, we show that the scalar $\lambda^T \widehat{\gamma}(b)$ for any vector $\lambda$ has a normal distribution $N(0, \lambda^T V(\widehat{\gamma}) \lambda)$. For a given value of $b$, it is easy to follow the proof of theorems 2.9 and 2.10 in Pagan and Ullah (1999) to show that

$$(Th)^{1/2} \left[\lambda^T \widehat{\gamma}(b) - \lambda^T \gamma_0(b)\right] \xrightarrow{d} N\left(0, Var\left(\lambda^T \widehat{\gamma}(b)\right)\right),$$

where $Var\left(\lambda^T \widehat{\gamma}(b)\right) = \lambda^T V(\widehat{\gamma}(b)) \lambda$ is the variance of the scalar $\lambda^T \widehat{\gamma}(b)$. The asymptotic distribution of $\widehat{g}(b|N^*)$ then follows. ■

## A.3 Extension to the estimated private values

Let $v$ denote the private value. As shown in GPV (2000), we have

$$\xi(b, N^*) = b + \frac{1}{N^* - 1}\left[\frac{\int_{-\infty}^{b} g_{b|N^*}\left(\widetilde{b}|N^*\right) d\widetilde{b}}{g_{b|N^*}(b|N^*)} + \frac{F_v(r)}{1 + F_v(r)} \frac{1}{g_{b|N^*}(b|N^*)}\right].$$

The truncation probability $F_v(r)$ may be recovered from $g(N|N^*)$ as follows:

$$F_v(r) = 1 - [g(N|N^*)]^{1/N^*}.$$

A "plug-in" estimator of $\xi$ is

$$\widehat{\xi}(b, N^*) = b + \frac{1}{N^* - 1}\left[\frac{\int_{-\infty}^{b} \widehat{g}_{b|N^*}\left(\widetilde{b}|N^*\right) d\widetilde{b}}{\widehat{g}_{b|N^*}(b|N^*)} + \frac{\widehat{F}_v(r)}{1 + \widehat{F}_v(r)} \frac{1}{\widehat{g}_{b|N^*}(b|N^*)}\right].$$

and

$$\widehat{F}_v(r) = 1 - \left[\widehat{g}(N|N^*)\right]^{1/N^*}.$$

Notice that $\widehat{F}_v$ may converge at a $\sqrt{T}$-rate. The uniform consistency of $\widehat{\xi}(b, N^*)$ follows from that of $\widehat{g}_{b|N^*}$. The asymptotic distribution of $\widehat{\xi}(b, N^*)$ may be derived in a similar way as that of $\widehat{g}(b|N^*)$. Define

$$\begin{aligned}
\xi\left(b, N^*; g_{b|N^*;t}\right) &= b + \frac{1}{N^*-1}\left[\frac{\int_{-\infty}^{b} g_{b|N^*;t}\left(\widetilde{b}|N^*\right) d\widetilde{b}}{g_{b|N^*;t}(b|N^*)} + \frac{F_v(r)}{1+F_v(r)}\frac{1}{g_{b|N^*;t}(b|N^*)}\right]. \\
g_{b|N^*;t}(b|N^*) &= g_{b|N^*}(b|N^*) + t\left(\widehat{g}_{b|N^*}(b|N^*) - g_{b|N^*}(b|N^*)\right)
\end{aligned}$$

We will then consider the Taylor expansion at $t = 0$

$$\begin{aligned}
\xi\left(b, N^*; \widehat{g}_{b|N^*}\right) &= \xi\left(b, N^*; g_{b|N^*}\right) + \left.\frac{\partial \xi\left(b, N^*; g_{b|N^*,t}\right)}{\partial t}\left[\widehat{g}_{b|N^*} - g_{b|N^*}\right]\right|_{t=0} \\
&\quad + \left.\frac{\partial^2 \xi\left(b, N^*; g_{b|N^*;t}\right)}{\partial t^2}\left[\widehat{g}_{b|N^*} - g_{b|N^*}, \widehat{g}_{b|N^*} - g_{b|N^*}\right]\right|_{t=\tau},
\end{aligned}$$

for $\tau \in [0, 1]$ where the first-order pathwise derivative is

$$\begin{aligned}
&\left.\frac{\partial \xi\left(b, N^*; g_{b|N^*,t}\right)}{\partial t}\left[\widehat{g}_{b|N^*} - g_{b|N^*}\right]\right|_{t=0} \\
= &\ \frac{1}{N^*-1}\frac{\int_{-\infty}^{b}\left[\widehat{g}_{b|N^*}\left(\widetilde{b}|N^*\right) - g_{b|N^*}\left(\widetilde{b}|N^*\right)\right] d\widetilde{b}}{g_{b|N^*}(b|N^*)} \\
&+ \frac{1}{N^*-1}\left[\frac{\int_{-\infty}^{b} g_{b|N^*}\left(\widetilde{b}|N^*\right) d\widetilde{b}}{g_{b|N^*}(b|N^*)} + \frac{F_v(r)}{1+F_v(r)}\frac{1}{g_{b|N^*}(b|N^*)}\right]\frac{-\left[\widehat{g}_{b|N^*}(b|N^*) - g_{b|N^*}(b|N^*)\right]}{g_{b|N^*}(b|N^*)}.
\end{aligned}$$

Under regularity conditions, the uniform convergence of $\widehat{g}_{b|N^*}$ implies that

$$\xi\left(b, N^*; \widehat{g}_{b|N^*}\right) - \xi\left(b, N^*; g_{b|N^*}\right) = \left.\frac{\partial \xi\left(b, N^*; g_{b|N^*,t}\right)}{\partial t}\left[\widehat{g}_{b|N^*} - g_{b|N^*}\right]\right|_{t=0} + O_p\left(\left\|\widehat{g}_{b|N^*} - g_{b|N^*}\right\|_\infty^2\right).$$

We may then show that

$$(Th)^{1/2}\left[\widehat{\xi}(b, N^*) - \xi(b, N^*)\right] \xrightarrow{d} N(0, \Omega_\xi),$$

where

$$\Omega_\xi = \lim_{T\to\infty}(Th)\,Var\left(\left.\frac{\partial \xi\left(b, N^*; g_{b|N^*,t}\right)}{\partial t}\left[\widehat{g}_{b|N^*} - g_{b|N^*}\right]\right|_{t=0}\right).$$

# B An Alternative Approach to Estimate $G_{N|N^*}$

As an alternative to step one in Section 3, 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 $k \geq 2$,

$$H_k(x) = \frac{x^{k-1} - \sum_{j=1}^{k-1} \left\langle x^{k-1}, H_j \right\rangle_\omega H_j(x)}{\sqrt{\int_{\mathbb{R}} [x^{k-1} - \sum_{j=1}^{k-1} \langle x^{k-1}, H_j \rangle_\omega H_j(x)]^2 \exp\{-x^2\}dx}}. \tag{30}$$

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

$$g_1(N^*, w) = \sum_{i=1}^{\infty} \sum_{j=2}^{K} \beta_{i,j} \times I\left(N^* = j\right) \times p_i(w). \tag{31}$$

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

$$g_{1n}(N^*, w) = \sum_{i=1}^{I_{1n}} \sum_{j=2}^{K} \beta_{i,j} \times I\left(N^* = j\right) \times p_i(w). \tag{32}$$

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

$$\sum_{N=1}^{K} g_{2n}(N|N^*, w) = 1 \text{ for any } N^* \text{ and } w \tag{33}$$

and

$$\sum_{N^*=2}^{K} g_{3n}(N^*|Z, w) = 1 \text{ for any } Z \text{ and } w \tag{34}$$

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

We define an alternative values of $\alpha_0$ as follows:

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

We also define the data as $D_t = (b_{kt}, k = 1, \dots, A_t; A_t, L_t)$. 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 $\alpha_0$ 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

$$\left(\widehat{\beta}_{i,j}, \widehat{\gamma}_{i,j,k}, \widehat{\delta}_{i,j,k}\right) = \underset{\left(\beta_{i,j}, \gamma_{i,j,k}, \delta_{i,j,k}\right) \text{ in } \alpha=(g_{1n}, g_{2n}, g_{3n})}{\arg\max} \sum_{t=1}^{T} \left[b_t - m\left(D_t, \alpha\right)\right]^2,$$
$$\text{such that } \left(\beta_{i,j}, \gamma_{i,j,k}, \delta_{i,j,k}\right) \text{ satisfies conditions 33,34.}$$

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

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

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.

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

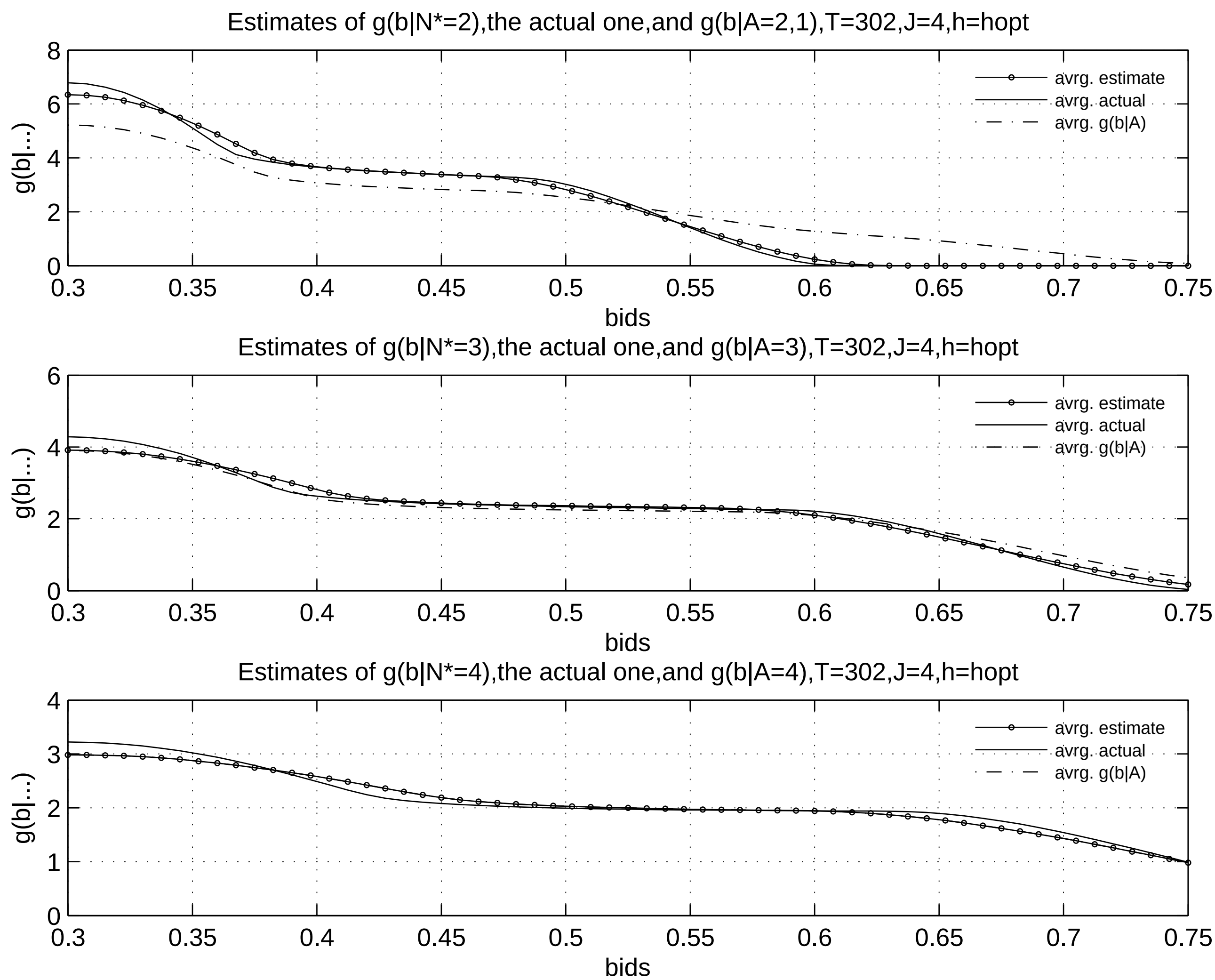


Figure 2: Estimates of bid functions and implied markdowns, $K = 4$ experiments

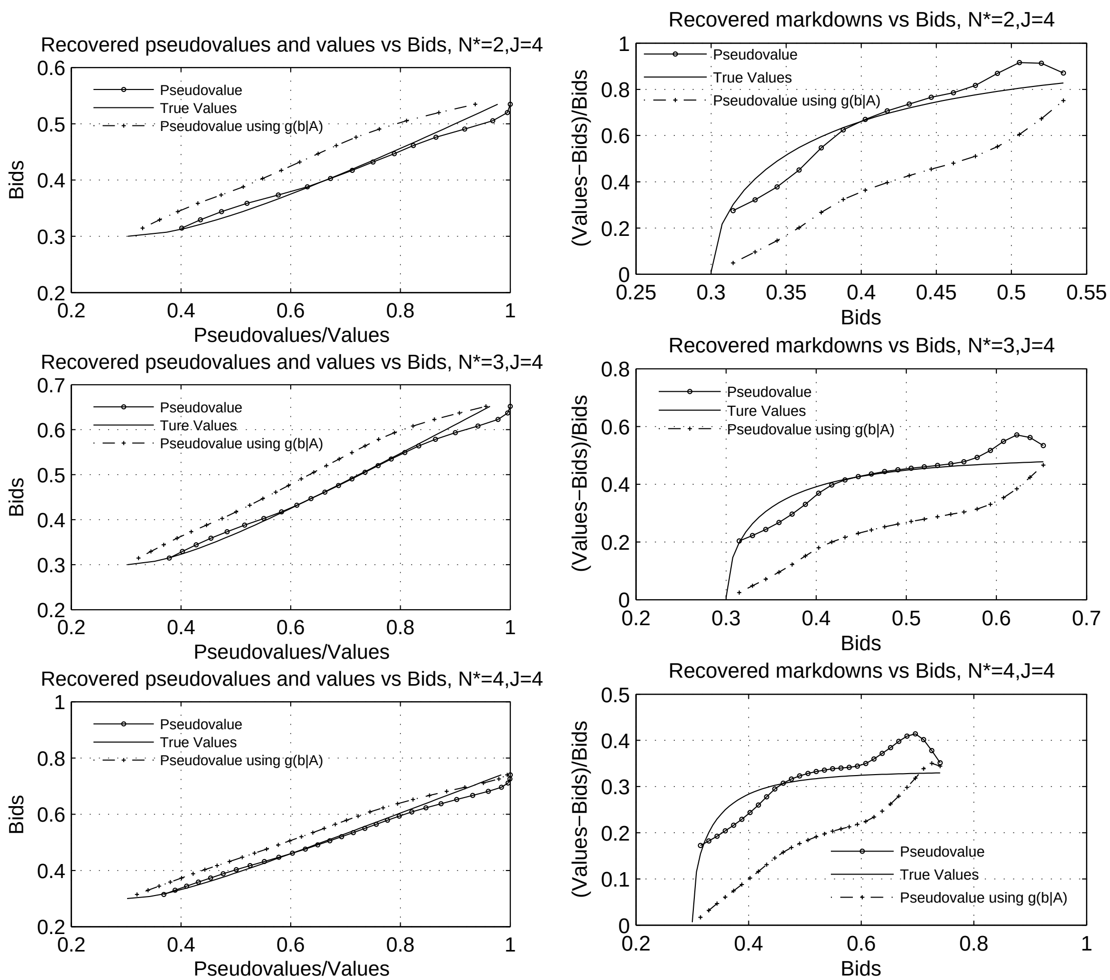


Figure 3: Monte Carlo Evidence: $K = 6$

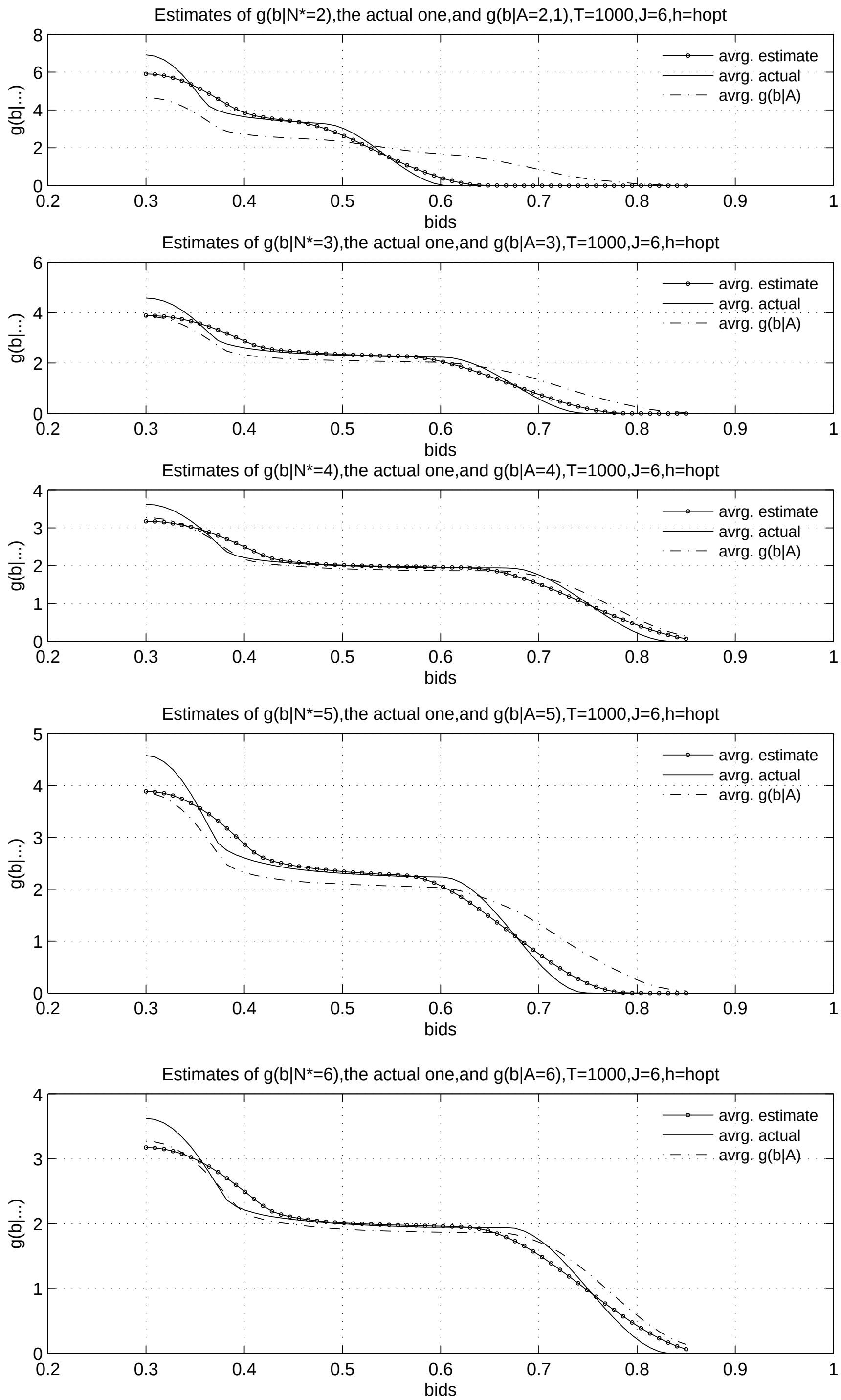


Figure 4: Estimates of bid functions, $K = 6$ experiments

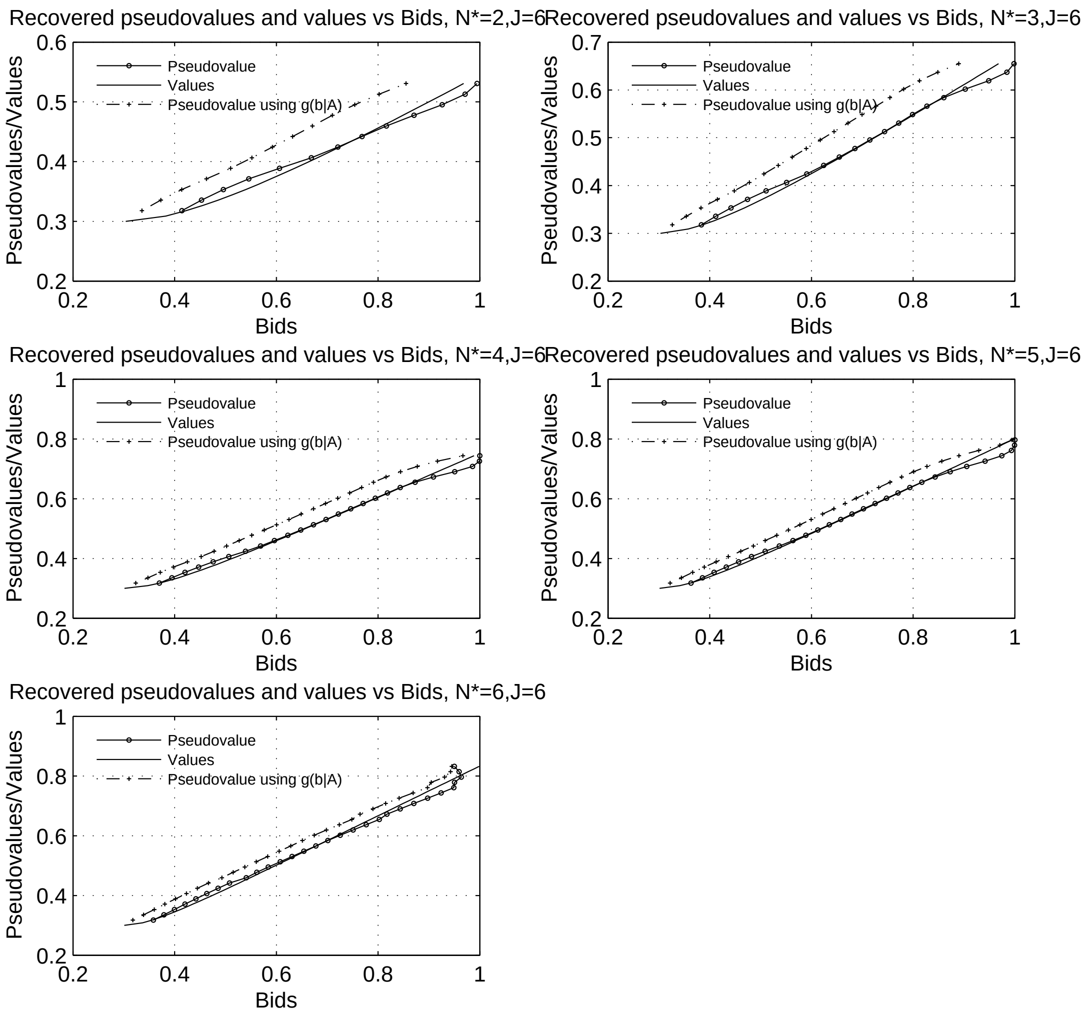


Figure 5: Highway work projects, impose upper-triangularity

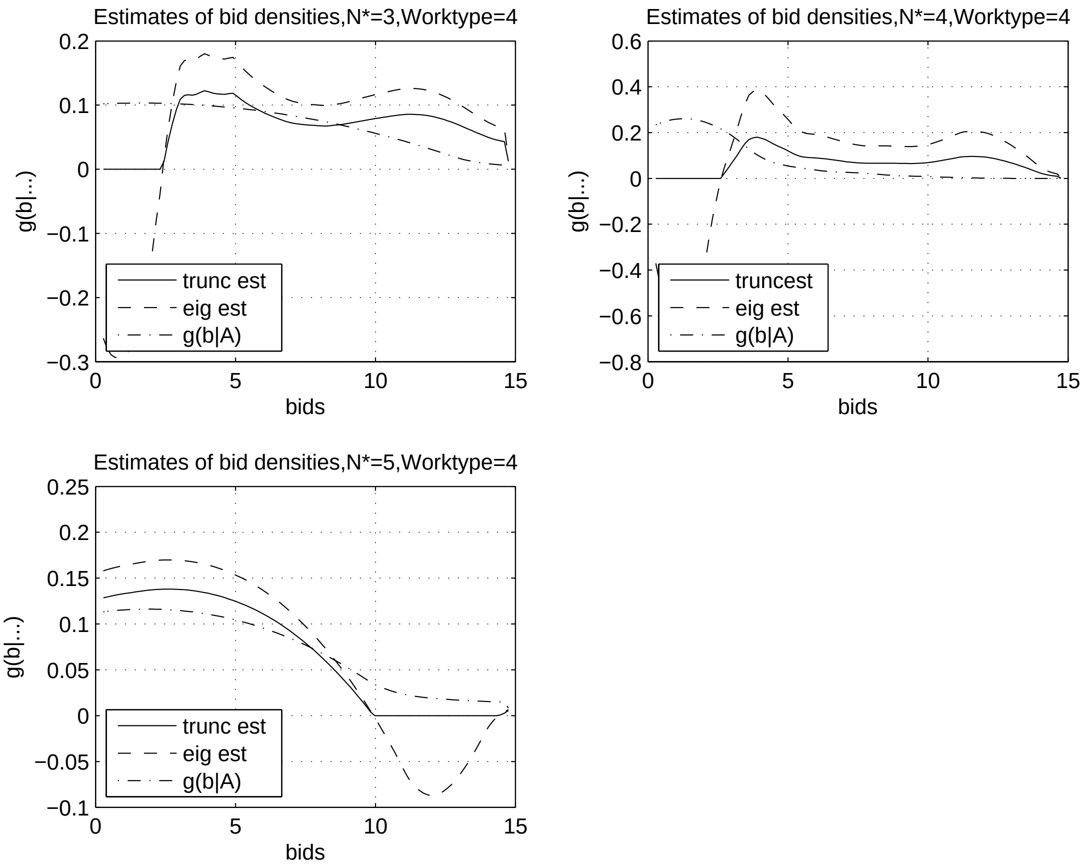


Figure 6: Highway work projects, pseudovalues

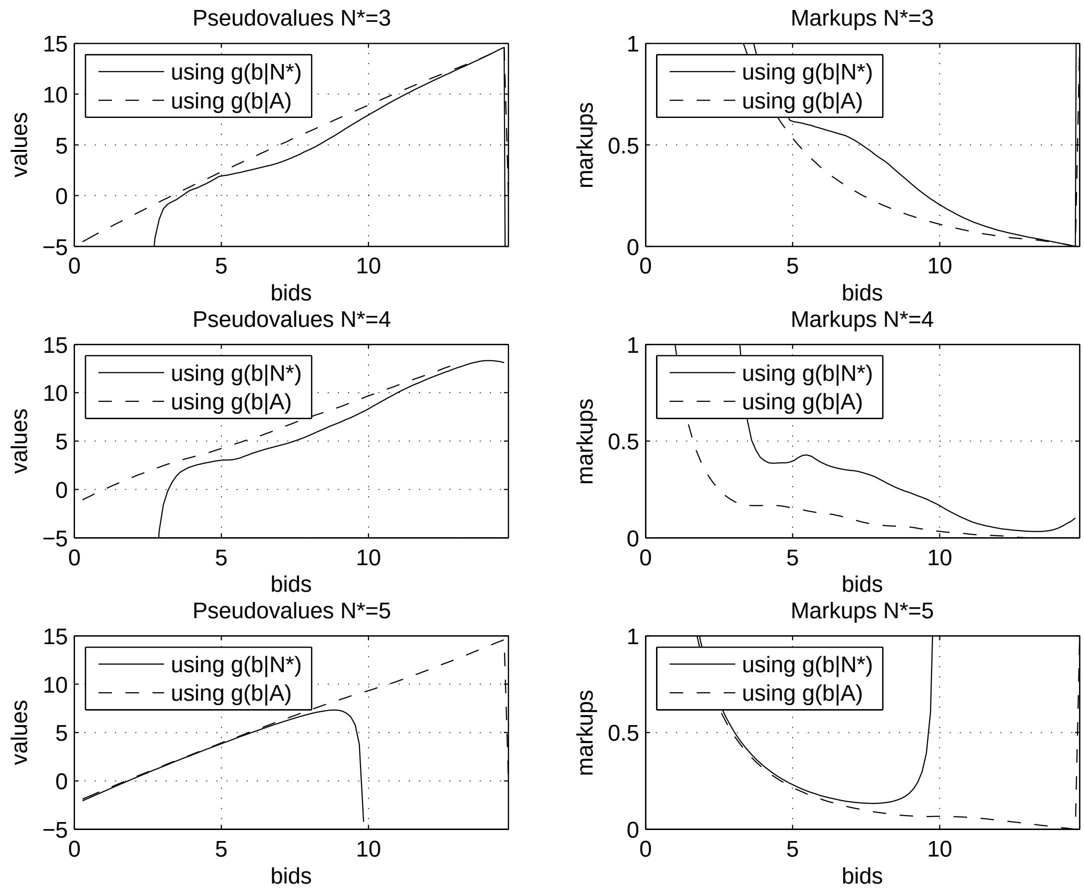


Figure 7: Paving/grading projects, impose upper-triangularity

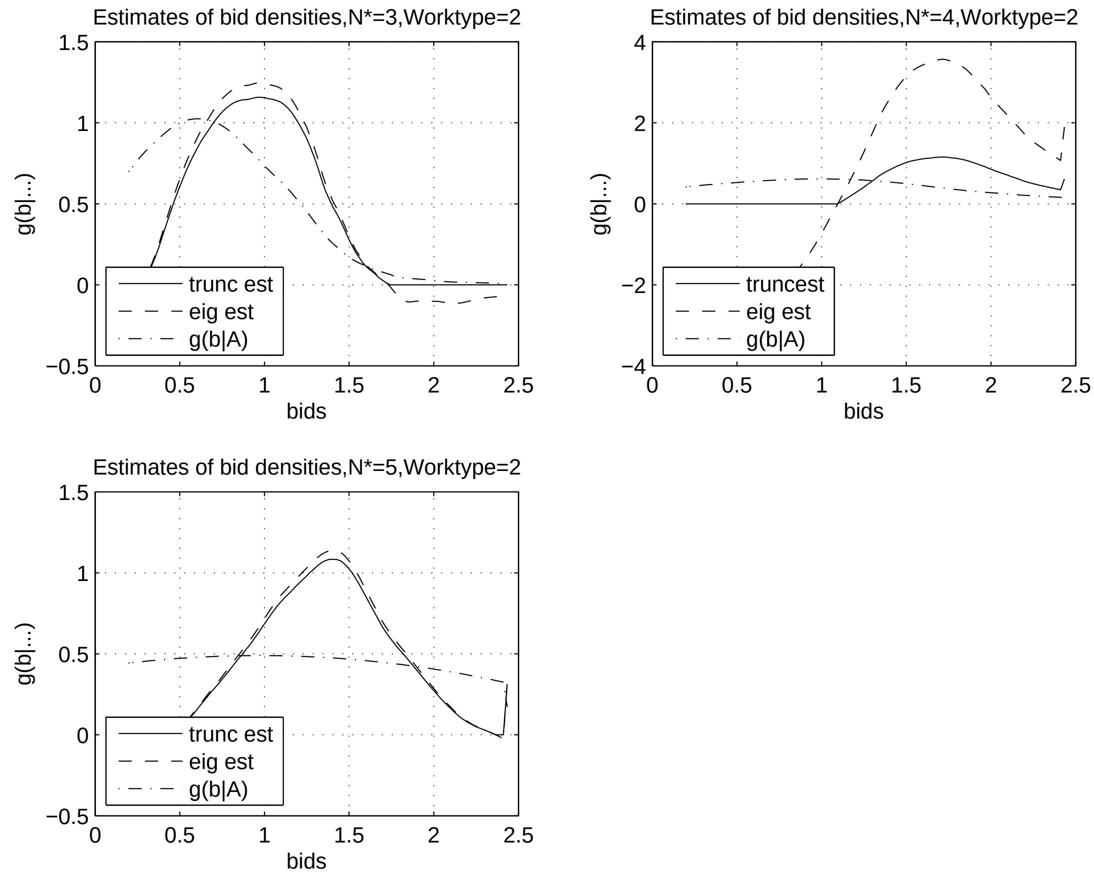


Figure 8: Grading/paving projects, pseudovalues

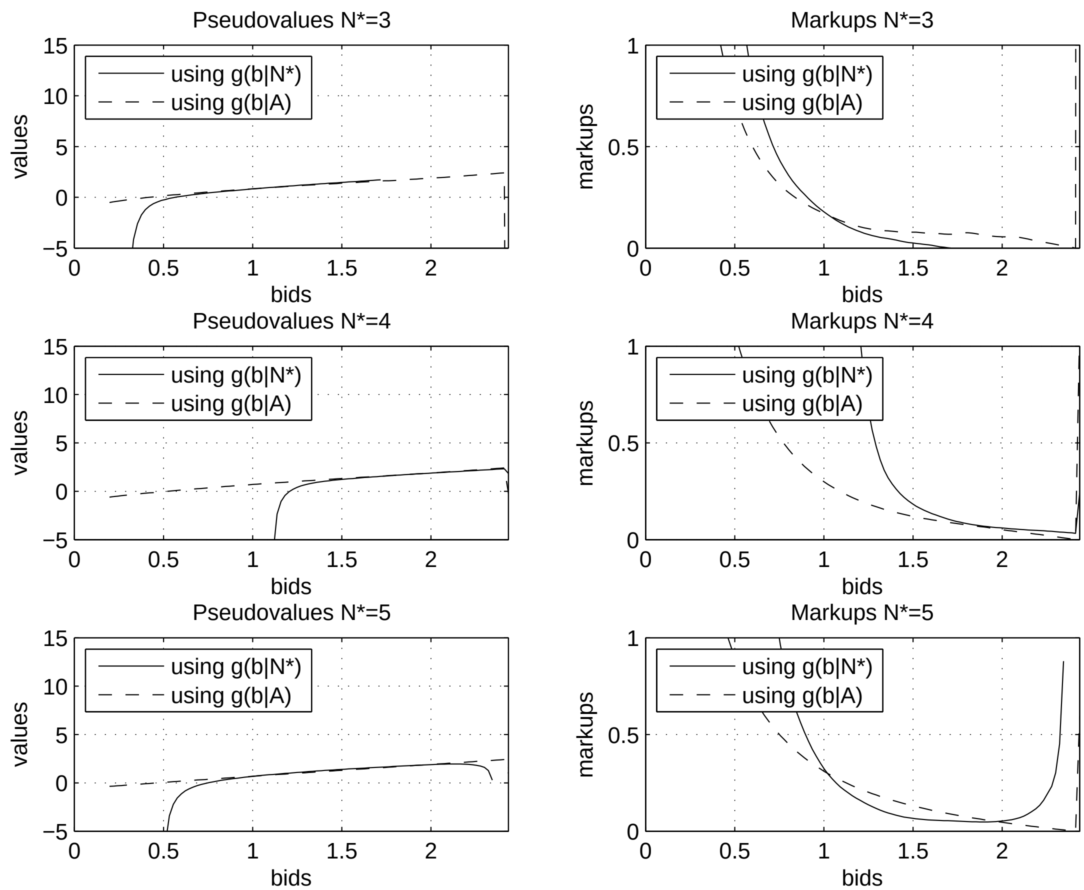
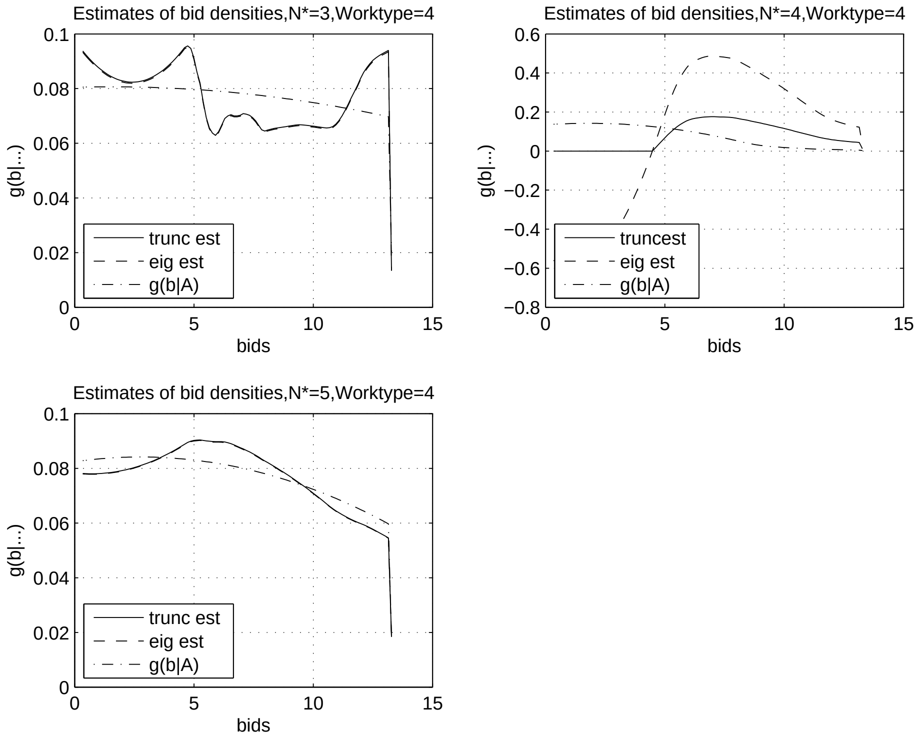
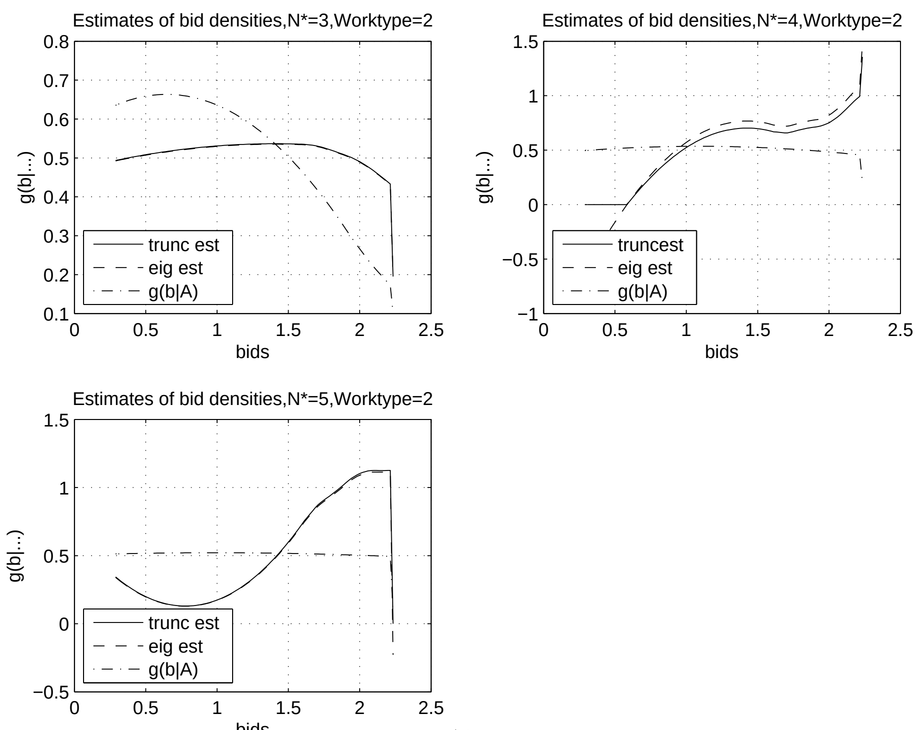


Figure 9: Estimated bid densities, without imposing upper-triangularity




Highway work contracts






Grading/paving contracts

Table 1: Summary statistics of procurement auction data

| | **Highway work** | | | **Grading/paving** | | |
|---|---|---|---|---|---|---|
| # bidders | # aucs. | Freq. | avg bid[a] | # aucs. | Freq. | avg bid[a] |
| 1 | 6 | 1.42 | 0.575 | 5 | 3.23 | 0.464 |
| 2 | 12 | 2.84 | 5.894 | 9 | 5.81 | 0.737 |
| 3 | 31 | 7.33 | 1.692 | 14 | 9.03 | 0.629 |
| 4 | 46 | 10.87 | 1.843 | 23 | 14.84 | 1.086 |
| 5+ | 338 | 77.54 | 7.920 | 104 | 77.10 | 1.248 |

[a]: in millions of 1989$