Journal of Econometrics 157 (2010) 328–341

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimating first-price auctions with an unknown number of bidders: A misclassification approach☆

Yonghong An [a], Yingyao Hu [a,*], Matthew Shum [b]

[a] Johns Hopkins University, Department of Economics, 440 Mergenthaler Hall, 21218 Baltimore, MD, United States
[b] Caltech, United States

ARTICLE INFO

Article history:
Received 16 April 2008
Received in revised form
5 February 2010
Accepted 17 February 2010
Available online 25 February 2010

Keywords:
Auction models
Nonparametric identification
Misclassification

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, but observed by bidders. 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 the unknown N^*. Monte Carlo results illustrate that the procedure works well in practice. We present illustrative evidence from a dataset of procurement auctions, which shows that accounting for the unobservability of N^* can lead to economically 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 N^* to be unknown to the researcher 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 observed by bidders, but not by the researcher. 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 et al., 2000; Li et al., 2002) to obtain estimates of bidders' valuations.

Accommodating the possibility that the researcher does not know N^* is important for drawing valid policy implications from auction model estimates. Because N^* is the level of competition in an auction, not knowing N^*, or 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. Indeed, a naïve approach where the number of observed bids is used as a proxy for N^* will tend to overstate competition, because the unknown N^* is always (weakly) larger than the number of observed bids. This bias will be shown in the empirical illustration below.

Not knowing the potential number of bidders N^* has been an issue since the earliest papers in the structural empirical auction literature. 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 knowing N^* need not be a problem. For instance, Laffont et al. (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 integrated 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 knowing N^* (or observing N^* with error) raises difficulties. Within the independent private-values (IPV) framework, and under the additional assumption that the unknown N^* is fixed across all auctions (or fixed across a known subset of the auctions),

☆ We thank an associate editor, two anonymous referees, Ken Hendricks, Harry Paarsch, Isabelle Perrigne, Jean-Marc Robin, Quang Vuong, and seminar participants at Brown, Caltech, FTC, Harvard-MIT, UC-Irvine, Iowa, NC State, Toronto, Yale, and SITE (Stanford) for helpful comments. Guofang Huang provided exceptional research assistance.

* Corresponding author. Tel.: +1 04105167610.
E-mail address: yhu@jhu.edu (Y. Hu).

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.02.002

Guerre et al. (2000) showed how to identify N^* and the equilibrium bid distribution in the range of bids exceeding the reserve price. Hendricks et al. (2003) allowed N^* to vary across auctions, and assumed that $N^* = L$, where L is a measure of the number of potential bidders which they construct.

The main contribution of this paper is to present a solution for the 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 develop a nonparametric procedure for recovering the distribution of bids conditional on unknown N^* which requires neither N^* to be fixed across auctions, nor for an (assumed) perfect measure of N^* to be available. Our procedure applies results from the recent econometric literature on models with misclassification error, such as e.g. Mahajan (2006) and Hu (2008).

As a specific case, our method is, as far as we aware, the first to solve the identification problem for IPV first-price auctions with reserve prices when the unobserved number of potential bidders N^* is a random variable. Previously, Guerre et al. (2000) also considered identification for first-price IPV auctions with reserve prices. However, they assumed that the observed number of potential bidders N^* is fixed across auctions, so that it could be estimated as a parameter.

For first-price auctions, allowing the unknown N^* to vary randomly across auctions is not innocuous. Because N^* is observed by the bidders, it affects their equilibrium bidding strategies. Hence, when N^* is not known by the researcher, and varies across auctions, 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 econometric methods developed for models with a misclassified regressor, where (likewise) the observed outcomes are drawn from a mixture distribution.

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 (and varies in the sample). She showed that the distribution of valuations can be recovered from observation of any two valuations of which rankings from the top are known.[1] However, her approach cannot be applied to first-price auctions, which are the focus of this paper. The reason for this is that, in IPV first-price auctions (but not in ascending- or second-price auctions), even if the distribution of bidders' valuations do not vary across the unknown N^*, the equilibrium distribution of bids still vary across N^*. Hence, because the researcher does not know N^*, the observed bids are drawn from a mixture distribution, and estimating the model requires deconvolution methods which have been developed in the econometric literature on measurement error.[2]

In a different context, Li et al. (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 (forthcoming) also used deconvolution results to estimate auction models with unobserved heterogeneity. To our knowledge, however, our paper is the first application of (discrete) measurement error results to estimate an auction model where the number of potential bidders is unknown.

The issues considered in this paper are close to those considered in the literature on entry in auctions: e.g. Li (2005); Li and Zheng (2009), Athey et al. (2005), Krasnokutskaya and Seim (2005) and Haile et al. (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 et al. (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^*.[3]

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, using data from procurement auctions in New Jersey. In Section 6, we consider extensions of the approach to both scenarios where only the winning bid is observed, and models of endogenous entry. Section 7 concludes. Proofs of the asymptotic properties of our estimator are presented in the Appendix.

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, Ch. 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 $F_{N^*}(x)$ which has support $[\underline{x}, \bar{x}]$. Furthermore, we assume the density of the private valuation $f_{N^*}(x)$ is bounded away from zero on $[\underline{x}, \bar{x}]$.[4] N^* can vary freely across the auctions, and while it is observed by the bidders, it is not known by the researcher. We allow the distribution of valuations $F_{N^*}(x)$ to vary across N^*.[5] There is a reserve price r, assumed to be fixed across all auctions, where $r > \underline{x}$.[6] The equilibrium bidding function for bidder i with valuation x_i is

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

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

[1] Adams (2007) also considers estimation of ascending auctions when the distribution of potential bidders is unknown.

[2] Song (2006) showed that the top two bids are also enough to identify first-price auctions where the number of active bidders is *not observed* by bidders. Under her assumptions, however, the observed bids are i.i.d. samples from a homogeneous distribution, so that her estimation methodology would not work for the model considered in this paper.

[3] 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.

[4] This assumption guarantees that the density of bids $g(b|N^*, b > r)$ is also bounded away from zero. See Guerre et al. (2000, Section 3.1) for detailed discussions.

[5] This is consistent with some models of endogenous entry. See Section 6.2.

[6] 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.

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

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

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 (2008).

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

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

Notice that F_{N^*}, the valuation distributions, can also be recovered after we identify $g(b|N^*)$ for different N^*.

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 (2008) to show the identification of the first-price auction model with unknown N^*. The procedure requires two auxiliary variables:

1. a proxy N, which is a mismeasured version of N^*; and
2. an instrument Z, which could be a discretized second bid.

We observe a random sample of $\{\vec{b}_t, N_t\}$, 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. In what follows, we use b to denote a randomly chosen bid from each auction.

We assume that the variables N, and N^* are both discrete, and that they have the same support $\mathcal{N} = \{2, \ldots, K\}$ as the discretized second bid Z. Here K can be interpreted as the maximum number of bidders, which is fixed across all auctions.[7]

For convenience, we first define the following matrices which we shall use repeatedly. We use the notation $g(\cdots)$ to denote, generically, a probability mass or density function.

$$G_{b,N,Z} \equiv [g(b, N = i, Z = j)]_{i,j},$$
$$G_{N|N^*} \equiv \left[g\left(N = i|N^* = k\right)\right]_{i,k},$$
$$G_{N^*,Z} \equiv \left[g\left(N^* = k, Z = j\right)\right]_{k,j},$$
$$G_{N,Z} \equiv [g\left(N = i, Z = j\right)]_{i,j},$$

and

$$G_{b|N^*} \equiv \begin{pmatrix} g(b|N^* = 2) & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & g(b|N^* = K) \end{pmatrix}. \quad (4)$$

All of these are $(K - 1)$-dimensional square matrices.

The five conditions required for our identification argument are given here:

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

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

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

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 $i \neq j$.*

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

In this section, we will show how Conditions 1–5 lead to the identification of the unknown elements $G_{b|N^*}$, $G_{N|N^*}$ and $G_{N^*,Z}$ (the former pointwise in b). The conditions will be discussed as they arise in the identification argument.

Condition 1 implies that N or Z affects the equilibrium density of bids only through the unknown number of potential bidders N^*. In the econometric literature, this is known as the "non-differential" 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 to estimate $g(b|N^*)$ which circumvents this problem.

Condition 2 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

Before proceeding with the identification argument, we consider several examples of auxiliary variables (N, Z) which satisfy Conditions 1 and 2.

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. In particular, 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 Condition 1 is satisfied.[8]

A good candidate for the instrument Z is a discretized second bid, and it depends on N^* through Eq. (1):

$$Z = b(N^*, x_z)$$

where x_z denotes the valuation of the bidder who submits the second bid Z. In order to satisfy Conditions 1 and 2, we would require $b \perp Z|N^*$, and also $A \perp Z|N^*$, which are both satisfied in the IPV setting. The use of a second bid in the role of the instrument Z echoes the use of two bids per auction in the earlier identification results of Li et al. (2002) and Krasnokutskaya (forthcoming). Hence, just as in those papers, our identification and estimation approach is applicable to any IPV auction with two or more bidders.

Because we are focused on the symmetric IPV model in this paper, we will consider this example in the remainder of this section, and also in our Monte Carlo experiments and in the empirical illustration.

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

$$N = l(N^*, \upsilon)$$
$$N^* = k(Z, \nu). \quad (5)$$

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^*.

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

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

Furthermore, in this example, 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 ν is a source of unobserved auction-specific heterogeneity.[9] Note that ν will generally be (unconditionally) correlated with the bids b, which our assumptions allow for. □

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

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

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

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

Eq. (7) can be written as

$$G_{b,N,Z} = G_{N|N^*}G_{b|N^*}G_{N^*,Z}. \tag{8}$$

Condition 2 implies that

$$g(N, Z) = \sum_{N^*=2}^{K} g(N|N^*)g(N^*, Z), \tag{9}$$

which, using the matrix notation above, is equivalent to

$$G_{N,Z} = G_{N|N^*}G_{N^*,Z}. \tag{10}$$

Eqs. (8) and (10) 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 sides 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 unknown matrices $G_{N|N^*}$, $G_{N^*,Z}$ and $G_{b|N^*}$ 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.

Eq. (10) implies that

$$\text{Rank}\left(G_{N,Z}\right) \leq \min\left\{\text{Rank}\left(G_{N|N^*}\right), \text{Rank}\left(G_{N^*,Z}\right)\right\}. \tag{11}$$

Hence, it follows from Condition 3 that $\text{Rank}\left(G_{N|N^*}\right) = K-1$ and $\text{Rank}\left(G_{N^*,Z}\right) = K-1$. In other words, the matrices $G_{N,Z}$, $G_{N|N^*}$, and $G_{N^*,Z}$ are all invertible.[10] Therefore, postmultiplying both sides of Eq. (8) by $G_{N,Z}^{-1} = G_{N^*,Z}^{-1}G_{N|N^*}^{-1}$, we obtain the key equation

$$G_{b,N,Z}G_{N,Z}^{-1} = G_{N|N^*}G_{b|N^*}G_{N|N^*}^{-1}. \tag{12}$$

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 (see Eq. (4)), the right-hand side matrix represents an eigenvalue–eigenvector decomposition of the left-hand side matrix, with $G_{b|N^*}$ being the diagonal matrix of eigenvalues, and $G_{N|N^*}$ being the corresponding matrix of eigenvectors. This is the key representation which will identify and facilitate estimation of the unknown matrices $G_{N|N^*}$ and $_{b|N^*}$.

In order to make the eigenvalue–eigenvector decomposition in Eq. (12) unique, Condition 4 is required. This condition, which is actually implied by equilibrium bidding, guarantees that the eigenvalues in $G_{b|N^*}$ are distinctive for some bid b, which ensures that the eigenvalue decomposition in Eq. (12) exists and is unique, for some bid b. Moreover, it guarantees that all the linearly independent eigenvectors are identified from the decomposition in Eq. (12).[11]

Given Condition 4, Eq. (12) shows that an eigenvalue decomposition of the observed $G_{b,N,Z}G_{N,Z}^{-1}$ matrix 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 constant 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 nonnegative.

The appropriate ordering of the columns of $G_{N|N^*}$ is less clear, and in order to complete the identification, we need an additional condition which pins down the ordering of these columns. Condition 5, which posits that $N \leq N^*$, is one example of such an ordering condition. It is natural, and automatically satisfied, when $N = A$, the observed number of bidders. This condition implies that for any $i, j \in \mathcal{N}$

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

In other words, $G_{N|N^*}$ is an upper-triangular matrix. Since the triangular matrix $G_{N|N^*}$ must be invertible (by Eq. (11)), its diagonal entries are all nonzero, i.e.,

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

In other words, Condition 5 implies that, once we have the columns of $G_{N|N^*}$ obtained as the eigenvectors from the matrix decomposition (12), the right ordering can be obtained by rearranging these columns so that they form an upper-triangular matrix.

Hence, the arguments in this section have shown the following result:

Theorem 1. *Under Conditions 1–5, $G_{b|N^*}$, $G_{N|N^*}$ and $G_{N^*,Z}$ are identified (the former pointwise in b).*

3. Nonparametric estimation: 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. In the key Eq. (12), the matrix $G_{N|N^*}$ is identical for all b.[12] This suggests a convenient two-step procedure for estimating the unknown matrices $G_{N|N^*}$ and $G(b|N^*)$.

Step one

In Step 1, we estimate the eigenvector matrix $G_{N|N^*}$. To maximize the convergence rate in estimating $G_{N|N^*}$, we average

[9] In the case when N^* is observed, correlation between bids and the participation shock ν can be accommodated, given additional restriction on the $k(\cdots)$ function. See Guerre et al. (2009) and Haile et al. (2003) for details. However, when N^* is unobserved, as is the case here, it is not clear how to generalize these results.

[10] Note that Condition 3 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).

[11] Specifically, 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. In this case, the two corresponding eigenvectors cannot be uniquely identified, because any linear combination of the two eigenvectors is still an eigenvector. Condition 4 guarantees that there exists another value $\bar{b}$ such that $g(\bar{b}|N^* = i) \neq g(\bar{b}|N^* = j)$. Because Eq. (12) holds for every b, implying that $g(b|N^* = i)$ and $g(\bar{b}|N^* = i)$ correspond to the same eigenvector, as do $g(\bar{b}|N^* = j)$ and $g(\widetilde{b}|N^* = j)$, we can use the value $\bar{b}$ to identify the two eigenvectors corresponding to $N^* = i$ and $N^* = j$.

[12] This also implies that there is a large degree of overidentification in this model, and 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.

across values of the bid b. Specifically, from Eq. (7), we have

$$E(b|N,Z)g(N,Z) = \sum_{N^*=2}^{K} E(b|N^*)g(N|N^*)g(N^*,Z) \tag{15}$$

where $E[\cdot|\cdot]$ denote conditional expectation. Define the matrices

$$G_{Eb,N,Z} \equiv [E\,(b|N=i,Z=j)\,g(N=i,Z=j)]_{i,j}\,, \tag{16}$$

and

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

Then

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

and, as before, postmultiplying both sides of this equation by $G_{N,Z}^{-1} = G_{N^*,Z}^{-1}G_{N|N^*}^{-1}$, we obtain an integrated version of the key equation:

$$G_{Eb,N,Z}G_{N,Z}^{-1} = G_{N|N^*}G_{Eb|N^*}G_{N|N^*}^{-1}. \tag{17}$$

This implies

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

where $\psi(\cdot)$ denotes the mapping from a square matrix to its eigenvector matrix following the identification procedure in the previous section.[13] As mentioned in Hu (2008), the function $\psi(\cdot)$ is a non-stochastic analytic function. Therefore, we may estimate $G_{N|N^*}$ as follows:

$$\widehat{G}_{N|N^*} := \psi\left(\widehat{G}_{Eb,N,Z}\widehat{G}_{N,Z}^{-1}\right), \tag{18}$$

where $\widehat{G}_{Eb,N,Z}$ and $\widehat{G}_{N,Z}$ may be constructed directly from the sample. In our empirical example, we estimate $\widehat{G}_{Eb,N,Z}$ using a sample average:

$$\widehat{G}_{Eb,N,Z} = \left[\frac{1}{T}\sum_t \frac{1}{N_j}\sum_{i=1}^{N_j} b_{it}\mathbf{1}(N_t=N_j, Z_t=Z_k)\right]_{j,k}. \tag{19}$$

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 proceed to estimate $g(b|N^*)$, pointwise in b. First, consider

$$g(b,N) = \sum_{N^*} g(N|N^*)g(b,N^*)$$

which, in matrix form, is

$$\overrightarrow{g}\,(b,N) = G_{N|N^*}\,\overrightarrow{g}\,(b,N^*),$$

where the vector of densities $\overrightarrow{g}\,(b,N) \equiv [g(b,N=2), g(b,N=3), \ldots, g(b,N=K)]^T$.

Define $e_{N^*} = (0,\ldots,0,1,0,\ldots,0)^T$, where 1 is at the N^*-th position in the vector. This relation suggests that we may estimate the joint density $g(b,N^*)$ as follows:

$$\widehat{g}(b,N^*) = e_{N^*}^T\widehat{G}_{N|N^*}^{-1}\,\overrightarrow{\widehat{g}}\,(b,N),$$

where $\widehat{G}_{N|N^*}$ is estimated in Step 1, and we use a kernel estimate for each element of the vector $\overrightarrow{\widehat{g}}\,(b,N) = [\widehat{g}(b,N=2), \widehat{g}(b,N=3), \ldots, \widehat{g}(b,N=K)]^T$:

$$\widehat{g}\left(b,N_j\right) = \left[\frac{1}{Th}\sum_t \frac{1}{N_t}\sum_{i=1}^{N_t} K\left(\frac{b-b_{it}}{h}\right)\mathbf{1}(N_t=N_j)\right]. \tag{20}$$

Given this estimate of $\widehat{g}(b,N^*)$, it is straightforward to estimate $g(b|N^*)$. Define $\overrightarrow{g}_N$, and $\overrightarrow{g}_{N^*}$ as the vectors of distributions for N and N^*, respectively.[14] Then,

$$\overrightarrow{g}_N = G_{N|N^*}\,\overrightarrow{g}_{N^*}.$$

We may then estimate

$$\widehat{\Pr(N^*)} = e_{N^*}^T\widehat{G}_{N|N^*}^{-1}\,\overrightarrow{\widehat{g}}\,(N),$$

where $\overrightarrow{\widehat{g}}\,(N) \equiv \left[\frac{1}{T}\sum_t \mathbf{1}_{N_t=2}, \ldots, \frac{1}{T}\sum_t \mathbf{1}_{N_t=K}\right]$ can be recovered directly from the sample. Therefore, the conditional bid densities $g(b|N^*)$ may be estimated as

$$\widehat{g}(b|N^*) = \frac{e_{N^*}^T\widehat{G}_{N|N^*}^{-1}\,\overrightarrow{\widehat{g}}\,(b,N)}{e_{N^*}^T\widehat{G}_{N|N^*}^{-1}\,\overrightarrow{\widehat{g}}\,(N)}. \tag{21}$$

Analogously, we can also recover $F(b|N^*)$, the empirical conditional CDFs for the bids, using the conditional empirical CDF:

$$\widehat{F}(b|N^*) = \frac{e_{N^*}^T\widehat{G}_{N|N^*}^{-1}\,\overrightarrow{\widehat{F}}\,(b,N)}{e_{N^*}^T\widehat{G}_{N|N^*}^{-1}\,\overrightarrow{\widehat{g}}\,(N)}, \tag{22}$$

where $\widehat{F}(b,N)$ denotes the vector of empirical CDFs with elements

$$\widehat{F}\left(b,N_j\right) = \frac{1}{N_t}\sum_{i=1}^{N_t}\mathbf{1}\left(b_{it}<b, N_t=N_j\right), \quad N_j=2,\ldots,K \tag{23}$$

which can be recovered from the sample.

In the Monte Carlo experiments and empirical application, we estimated both bid CDFs (using Eq. (22)) and bid densities (using Eq. (21)) to assess the performance of our estimation procedure. An advantage of empirical CDFs over kernel density estimates is that we do not need to worry about the effects of bandwidth choice on the performance of our estimator.

Because $\Pr(N^*=K|A=K)=1$, and $G_{N|N^*}$ is an upper-triangular matrix, our estimates of $F(b|N^*=K)$ and $g(b|N^*=K)$ are identical to, respectively, $F(b|A=K)$ and $g(b|A=K)$. Our estimation requires a value for K, the upper bound for the number of potential bidders. In practice, K is unknown, but we set it to be the maximum number of observed bidders, which is a super-consistent estimate.[15]

The bid b may have a different unknown support for different N^*. That is,

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

where u_{N^*}, the upper bound of the support of $g(b|N^*)$, may not be known by the researcher. In practice, we estimate the upper bound u_{N^*} as follows:

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

In general, using the supremum to estimate the upper bound of an observed random sample is somewhat naïve. Estimation of the support of an observed random sample has been extensively studied in the statistics literature (see Cuevas and Rodríguez-Casal (2004) for i.i.d. data, and Delaigle and Gijbels (2006a,b) for data measured with error), and our estimate of u_{N^*} can be improved by employing these methods. However, because an unbiased and

[13] In order for $G_{N|N^*}$ to be recovered from this eigenvector decomposition, Condition 4 from the previous section must be strengthened so that the conditional means $E[b|N^*]$, which are the eigenvalues from this decomposition, are distinct for every N^*.

[14] For example, if $N^*=\{2,3,4\}$, then $\overrightarrow{g}_{N^*} = \{\Pr(N^*=2), \Pr(N^*=3), \Pr(N^*=4)\}^T$.

[15] This is obvious if the reserve price is zero. However, this is also valid when the reserve price is greater than zero because, even when $r>0$, the probability that the observed number of bidders is equal to K is still strictly positive.

consistent estimator of u_{N^*} is all we need, the naïve estimator $\widehat{u}_{N^*}$ is sufficient for our purposes, and we do not consider more sophisticated estimators in this paper.[16]

The asymptotic properties of our estimator are analyzed in detail in the Appendix. Here we provide a brief summary. Given the discreteness of N, Z, and the use of a sample average to construct $\widehat{G}_{Eb,N,Z}$ (via Eq. (19)), the estimates of $\widehat{G}_{N|N^*}$ (obtained using Eq. (18)) 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. (21), will be determined by the convergence properties of the kernel estimate of $g(b, N)$ in Eq. (20), which converges at a rate slower than $\sqrt{T}$. In the Appendix, we show that, pointwise in b, $(Th)^{1/2}\,[\widehat{g}(b|N^*) - g(b|N^*)]$ converges to a normal distribution. We also present a uniform convergence rate for $\widehat{g}(b|N^*)$. As for the empirical distribution $\widehat{F}(b|N^*)$, it is well known that $T^{1/2}[\widehat{F}(b, N) - F(b, N)]$ converges to a normal distribution with mean zero. Because $\widehat{G}_{N|N^*}$ converges at a $\sqrt{T}$-rate, $\widehat{F}(b|N^*)$ also converges at $\sqrt{T}$-rate. We omit the proof of this as the argument is similar to the proof for $\widehat{g}(b|N^*)$.

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. In the scenario where 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 (see Paarsch (1997)) that

$$A|N^* \sim \text{Binomial}(N^*, 1 - F_{N^*}(r)) \tag{24}$$

where $F_{N^*}(r)$ denotes the CDF of bidders' valuations in auctions with N^* potential bidders, evaluated at the reserve price. This suggests that the recovered matrix $G_{A|N^*}$ can be useful in two respects. First, using Eq. (24), the truncation probability $F_{N^*}(r)$ could be estimated, for each value of N^*. 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 $\Pr(A|N^*)$ for a fixed N^*, are consistent with the binomial distribution in Eq. (24).

4. Monte Carlo evidence

In this section, we present some Monte Carlo evidence for our estimation procedure. 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

$$b^*(x; N^*) = \mathbf{1}_{x \geq r} \left\{ \left(\frac{N^* - 1}{N^*} \right) x + \frac{1}{N^*} \left(\frac{r}{x} \right)^{N^* - 1} r \right\}. \tag{25}$$

For each auction t, we generate the equilibrium bids b_{jt}, for $j = 1, \ldots, N_t^*$, as well as (N_t^*, N_t, Z_t). The proxy N_t is taken to be the number of observed bidders A_t, and Z_t is a discretized second bid. The number of potential bidders N_t^* for each auction t is generated uniformly on $\{2, 3, \ldots, K\}$, where K, the maximum number of bidders, is set at 4. For each auction t, and each bidder $j = 1, \ldots, N_t^*$, we draw valuations $x_j \sim U[0, 1]$, and construct the corresponding equilibrium bids using Eq. (25). Subsequently, the number of observed bidders is determined as the number of bidders whose valuations exceed the reserve price: $A_t = \sum_{j \in \mathcal{N}_t^*} \mathbf{1}(x_j \geq r)$.

The estimation procedure in Section 3 requires $A_t \geq 2$ for each t, so that the supports of A_t and N_t^* coincide. For this reason, we discard all the auctions with $A_t = 1$[17]; for each of the remaining auctions, we randomly pick a pair of bids (b_{1t}, b_{2t}), and use a discretized version of the second bid b_{2t} in the role of Z_t.[18]

4.1. Results

We present results from $S = 400$ replications of a simulation experiment. The performance of our estimation procedure is illustrated in Figs. 1 and 3. The estimator performs well for all values of $N^* = 2, 3, 4$, and for modest-sized datasets of $T = 1000$ and $T = 400$ auctions, especially for the empirical bid distribution functions. Across the Monte Carlo replications, the estimated distribution and density functions track the actual densities quite closely. In these graphs, we also plot the bid CDFs (labelled "$G(b|A)$") and densities ($g(b|A)$) conditional on A, which are "naïve" estimators for $F(b|N^*)$, and $g(b|N^*)$, respectively. For $N^* = 2, 3$, our estimator outperforms the naïve estimator, especially for the case of $N^* = 2$. As we mentioned earlier, for $N^* = 4$, our estimates coincide with the naïve estimates.

In Figs. 2 and 4, we present estimates of bidders' valuations. In each graph on the left-hand side of the figure, 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 "True values"; (ii) the estimated valuations using our estimates of $g(b|N^*)$, labeled "Estimated value"[19]; and (iii) naïve estimates of the values, computed using $g(b|A)$, the observed bid densities conditional on the observed number of bidders.[20]

The graphs show that there are sizable differences between the value estimates, across all values of the bids. For all values of N^*, we see that our estimator tracks the true values quite closely. In contrast, the naïve approach underestimates the valuations. This is to be expected—because $N^* \geq A$, 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, which leads to underestimation of bidders' markdowns $(x - b)/x$. The markdowns implied by our valuation estimates are shown in the right-hand-side graphs in Figs. 2 and 4.

5. Empirical illustration

In this section, we illustrate our methodology using 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. Moreover, Hong and Shum's (2002) analysis allows for common values, whereas we just have a simpler IPV model in the application here.[21]

[16] This naïve estimator for the upper bound of the support of bids is commonly used in the auction literature; e.g., see Donald and Paarsch (1993) and Guerre et al. (2000), among others.

[17] Because of Condition 1, ignoring the auctions with $A_t = 1$ does not affect the consistency of the estimates of the bid distributions $g(b|N^*)$. There is only an efficiency impact from using fewer observations.

[18] Specifically, in this experiment, bids are distributed on [0.3, 0.75], and both $N^*, A \in \{2, 3, 4\}$. Hence, the discretized second bid Z_t also takes values $\{2, 3, 4\}$ as follows: if $b_{2t} \in [0.3, 0.55], Z = 2$; $b_{2t} \in [0.55, 0.675], Z = 3$; $b_{2t} \in [0.675, 0.55], Z = 4$.

[19] 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 $\hat{F}(r) = 1 - \left[\hat{G}(N^*|N^*)\right]^{1/N^*}$.

[20] In computing the values for the naïve approach, 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.

[21] We are uncertain how to extend our estimation approach to common (or affiliated) value settings, and are exploring this in ongoing work.

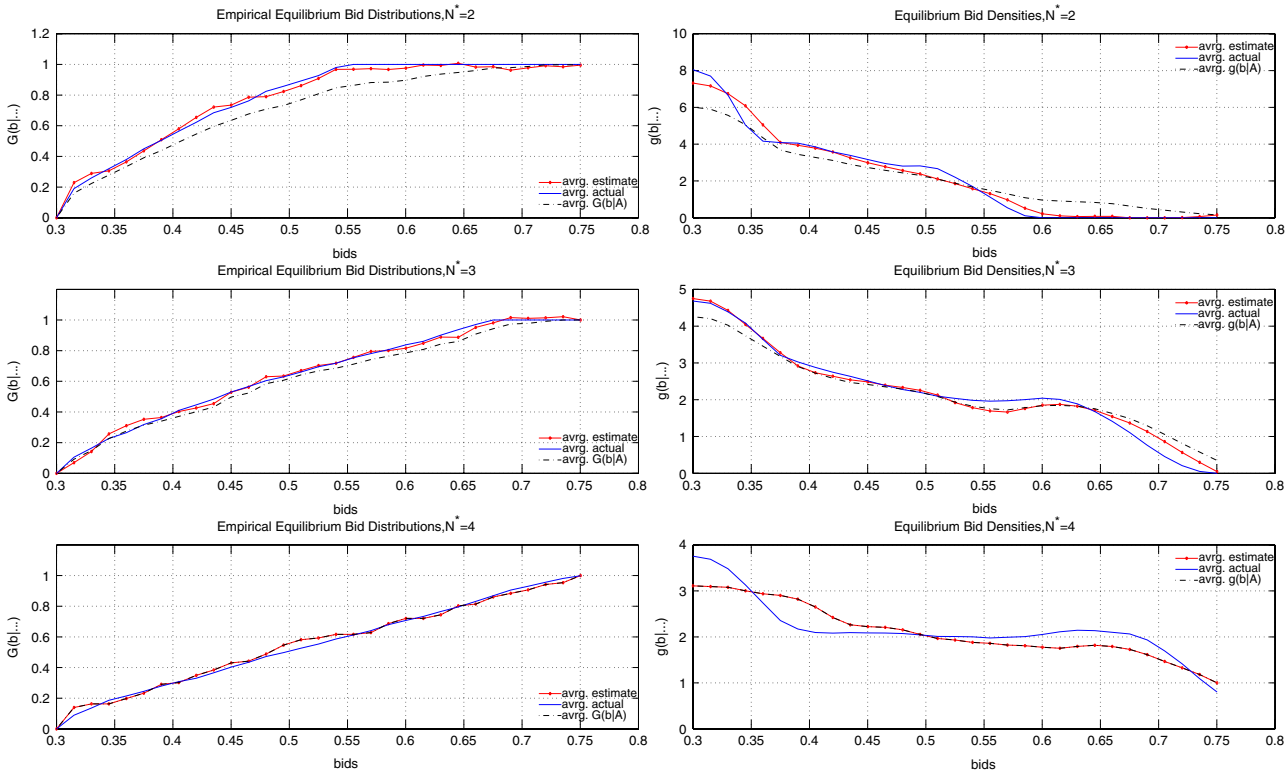

Fig. 1. Estimates of bid distribution functions and densities: $K = 4, T = 1000$.

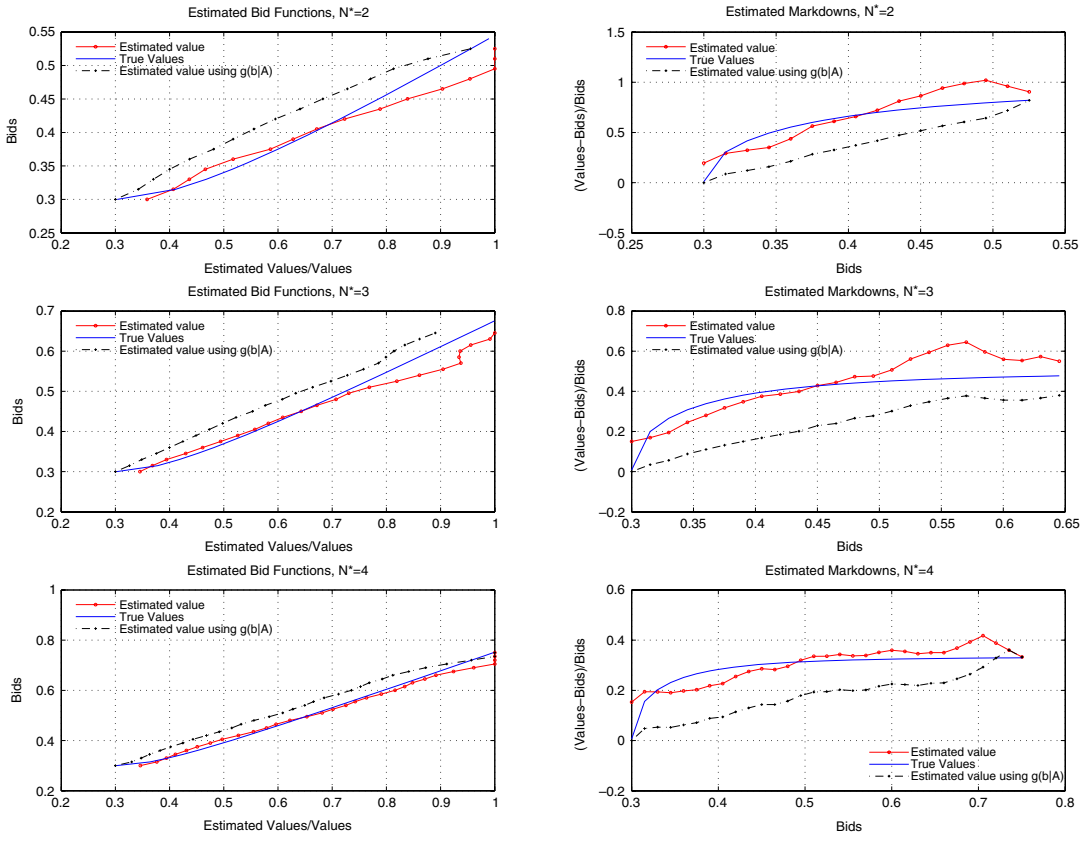

Fig. 2. Estimates of bid functions and implied markdowns, $K = 4, T = 1000$.

Among all the auctions in our dataset, we focus on highway work construction projects, for which the number of auctions is the largest. In Table 1, we present some summary statistics on the auctions used in the analysis. Note that there were six auctions with just one bidder, in which non-infinite bids were submitted. If the observed number of bidders A is equal to N^*, the

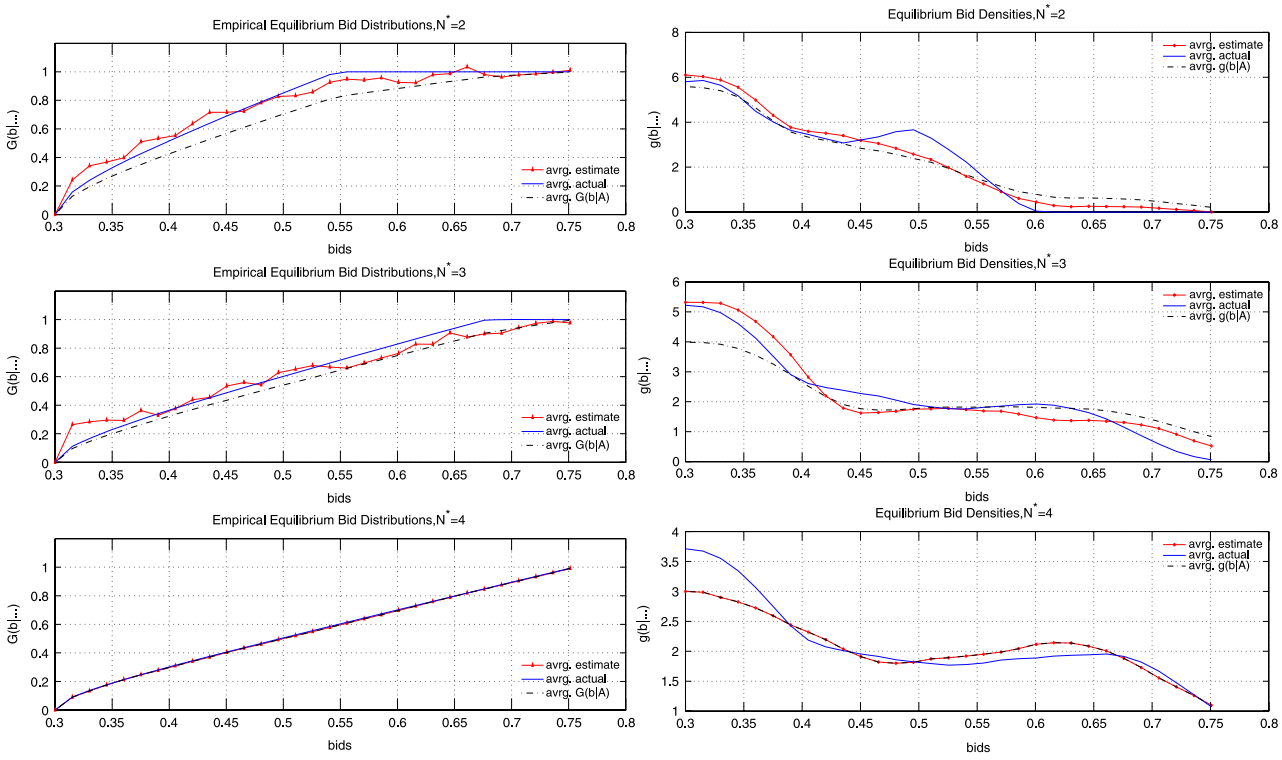

Fig. 3. Estimates of bid distribution functions and densities: $K = 4, T = 400$.

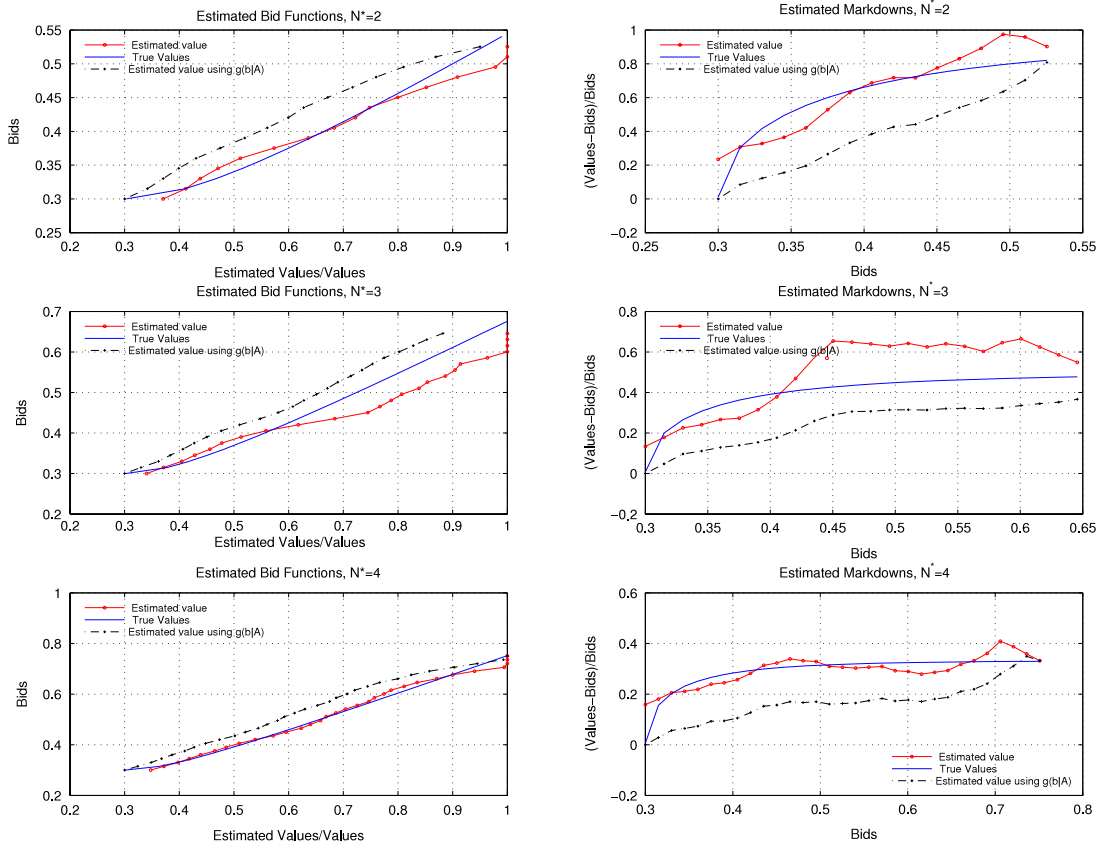

Fig. 4. Estimates of bid functions and implied markdowns, $K = 4, T = 400$.

Table 1
Summary statistics of procurement auction data. Highway work auctions, worktype = 4. Only auctions with $A = 2, 3, 4$ were used in empirical analysis.

Observed # bidders (A)	# auctions	Frequency	Average bid[a]
1	6	2.96	0.575
2	11	5.42	1.495
3	31	15.27	1.692
4	46	22.67	1.843
5+	109	53.69	4.034

[a] In millions of 1989$.

number of potential bidders observed 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.[22] However, occurrences of one-bidder auctions are a sign that the observed number of bidders is less than the potential number of bidders, perhaps due to an implicit reserve price. The methodology developed in this paper allows for this possibility.

For the two auxiliary variables, we used A, the number of observed bidders, in the role of the noisy measure N. We only analyze auctions with $A = 2, 3, 4$. Correspondingly, N^* also takes three distinct values from $\{2, 3, 4\}$. Because we focus on this range of small A, we assume that all the auctions are homogeneous.[23] In the role of the instrument Z, we use a second bid, discretized to take three values, so that the support of Z is the same as that of A.[24]

Furthermore, we use the ordering Condition 5, which implies that $A \leq N^*$, which is consistent with the story that bidders decide not to submit a bid due to an implicit reserve price. By an implicit reserve price, we mean a reserve price that bidders observe at the time of bidding, while not the econometrician. While there was no explicit reserve price in these auctions, there may have been an implicit reserve price, which can be understood as bidders' common beliefs regarding the upper bound of bids that the auctioneer is willing to consider.[25]

Because we model these auctions in a simplified setting, we do not attempt a full analysis of these auctions. Rather, this exercise highlights some practical issues in implementing the estimation methodology. There are three important issues. First, the assumption that $A \leq N^*$ implies that the matrix on the right-hand side of the key equation (17) 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. In practice, this matrix may not be upper-triangular. However, we do not impose upper-triangularity on $G_{Eb,N,Z}G_{N,Z}^{-1}$ in the first step of estimation. Instead, we constrain the estimated matrix $\widehat{G}_{A|N^*}$ to be upper-triangular in the second step of estimation. Doing so has no effect on the asymptotic consistency and convergence properties of $\widehat{G}_{A|N^*}$ since $G_{Eb,N,Z}G_{N,Z}^{-1}$ is upper-triangular asymptotically, i.e., with probability 1, the lower-triangular elements of $G_{Eb,N,Z}G_{N,Z}^{-1}$ vanish.[26]

Second, even after imposing upper-triangularity on estimated $G_{A|N^*}$, 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.[27] 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.[28]

Third, for low-bid procurement auctions, the optimal bidding strategy, analogous to Eq. (1) above, is

$$b(x_i; N^*) = \begin{cases} x_i + \dfrac{\int_{x_i}^{r}(1 - F_{N^*}(s))^{N^*-1}\mathrm{d}s}{(1 - F_{N^*}(x_i))^{N^*-1}} & \text{for } x_i \leq r; \\ 0 & \text{for } x_i > r. \end{cases} \tag{26}$$

Correspondingly, the valuation x is obtained by

$$\xi\left(b, N^*\right) = b - \frac{1}{N^* - 1} \times \frac{1 - F_{N^*}(r)G(b|N^*)}{F_{N^*}(r)g(b|N^*)}. \tag{27}$$

Results: highway work auctions

Fig. 5 contains the graphs of the estimated densities $g(b|N^*)$ for $N^* = 2, 3, 4$, for the highway work auctions. In each column of this table, we present three estimates of each $g(b|N^*)$: (i) the normalized estimate with the negative portions removed, just following the remedy we mentioned above, labeled "trunc est"; (ii) the unnormalized estimate, which includes the negative values for the density, labeled "Orig est"; and (iii) the naïve estimate, given by $g(b|A)$. In each plot, we also include the 5% and 95% pointwise confidence intervals, calculated using bootstrap resampling.[29]

Fig. 5 shows that the naïve bid density estimates, using A in place of N^*, overweight 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, these economically sensible differences between the naïve estimates (using $g(b|A)$) and our estimates (using $g(b|N^*)$) serve as a confirmatory reality check on the assumptions underlying our estimator. In order to observe the performances of these estimators closely by comparisons, we also include estimated empirical CDFs and densities for $N^* = 2, 3, 4$ in Fig. 6.

22 Indeed, (Li and Zheng, 2009, p. 9) point out that even when bidders are uncertain about the number of competitors they are facing, finite bids cannot be explained when bidders face a non-zero probability that they could be the only bidder.

23 We also considered an alternative specification where we control for observed auction-specific heterogeneity via preliminary regressions of bids on auction characteristics, and then perform the analysis using the residuals from these regressions. The resulting estimates of the bid distributions (available from the authors upon request) were qualitatively similar to, but noisier than, the results presented here. This may be due to the weak correlation between the residuals and N^*. Our identification scheme relies critically on the correlation between bids and N^*, and if the auction characteristics were strongly related with, and affect the bids through N^*, using the residuals from the regressions in place of the bids may eliminate much of the correlation, leading to noisier estimates.

24 Namely, we set $Z_t = 2$ if the second bid b_t is less than the 25th percentile of all the second bids; between the 25th and the 75th percentile, $Z_t = 3$; greater that the 75th, $Z_t = 4$. We tried several other alternatives, to ensure that the results are robust. In general, even if the support of Z exceeds that of A, the rank of $G_{A,Z}$ remains the same, but the model is overidentified in the sense that there are more instruments than needed. Our estimation approach can be extended to this case by using the generalized inverse of $G_{A,Z}$, but we did not pursue this possibility here.

25 In conversations with an NJDOT authority, we were told that bids which were deemed excessive could be rejected outright at the discretion of the auction officials, which is consistent with an implicit reserve price.

26 Indeed, in the Monte Carlo simulations, we sometimes also had to impose this on the simulated data, as the $G_{Eb,N,Z}G_{N,Z}^{-1}$ matrix could be non-upper-triangular due to small sample noise. In a previous version of the paper, we imposed upper-triangularity directly on $G_{Eb,N,Z}G_{N,Z}^{-1}$. Both methods have no effect on the asymptotic consistency and convergence properties on our estimator, but clearly the method in current version is more plausible since we did not impose any restriction on data-driven matrix $G_{Eb,N,Z}G_{N,Z}^{-1}$.

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

28 Here we follow the recommendation of Efromovich (1999, p. 63). This remedy does not affect the asymptotic properties of our estimator in that asymptotically $g(b|N^*)$ is bounded away from zero on its support, as we mentioned in footnote 4.

29 The asymptotic variance is derived analytically in the Appendix. However, it is tedious to compute in practice, which is why we use the bootstrap to approximate the pointwise variance of the density estimates.

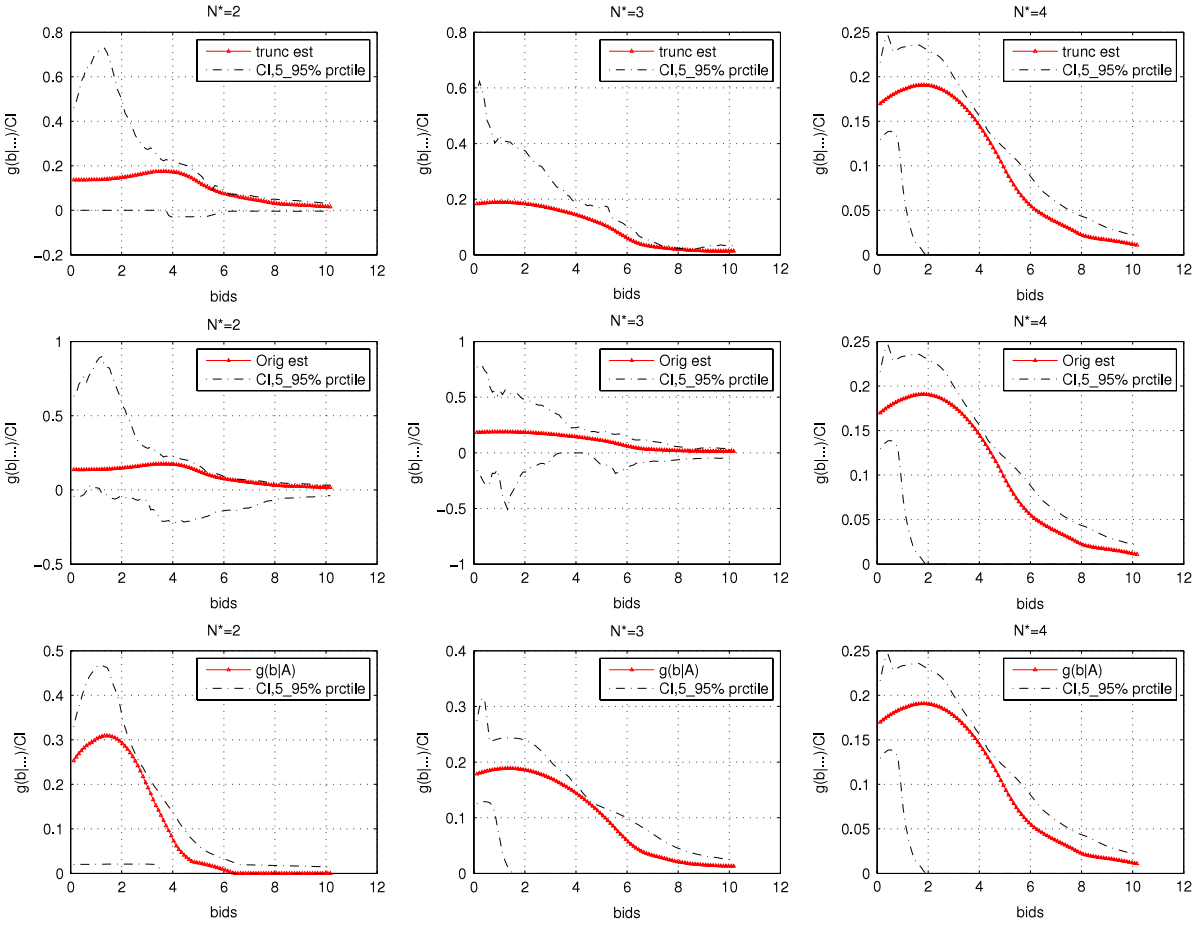

Fig. 5. Highway work projects, estimated densities: bootstrap 90% CI of the adjusted estimator.

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

	$N^*=2$	$N^*=3$	$N^*=4$
$A=2$	1.0000	0.1300	0.4091
$A=3$	0	0.8700	0.1041
$A=4$	0	0	0.4868

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 3.6726, 3.1567, 3.1776 for, respectively, $N^*=2, 3, 4$ (in millions of dollars).

The corresponding valuation estimates, obtained by solving Eq. (27) pointwise in b using our bid distribution and density estimates, are graphed in Fig. 7. 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.

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 $N^*=2, 3, 4$. 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 Fig. 7. For example, for $N^*=4$, at a bid of \$ 2 million, the corresponding markup using $g(b|A=4)$ is around 30%, or \$ 600,000, but using $g(b|N^*=4)$ is around 55%, or \$ 1.1 million. This suggests that failing to account for unobservability of N^* can lead the researcher to understate bidders' profit margins.

6. Extensions

6.1. 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 et al. (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 et al. (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 know 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.

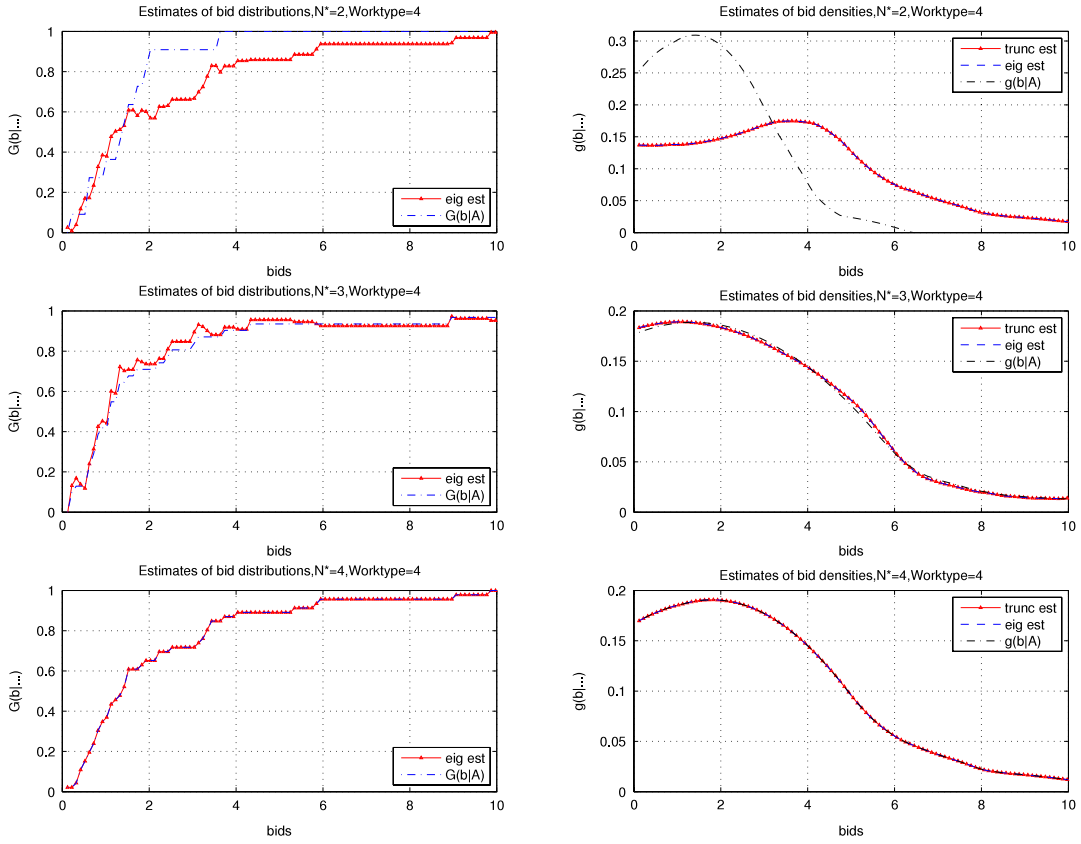

Fig. 6. Highway work projects, estimated distribution functions and densities.

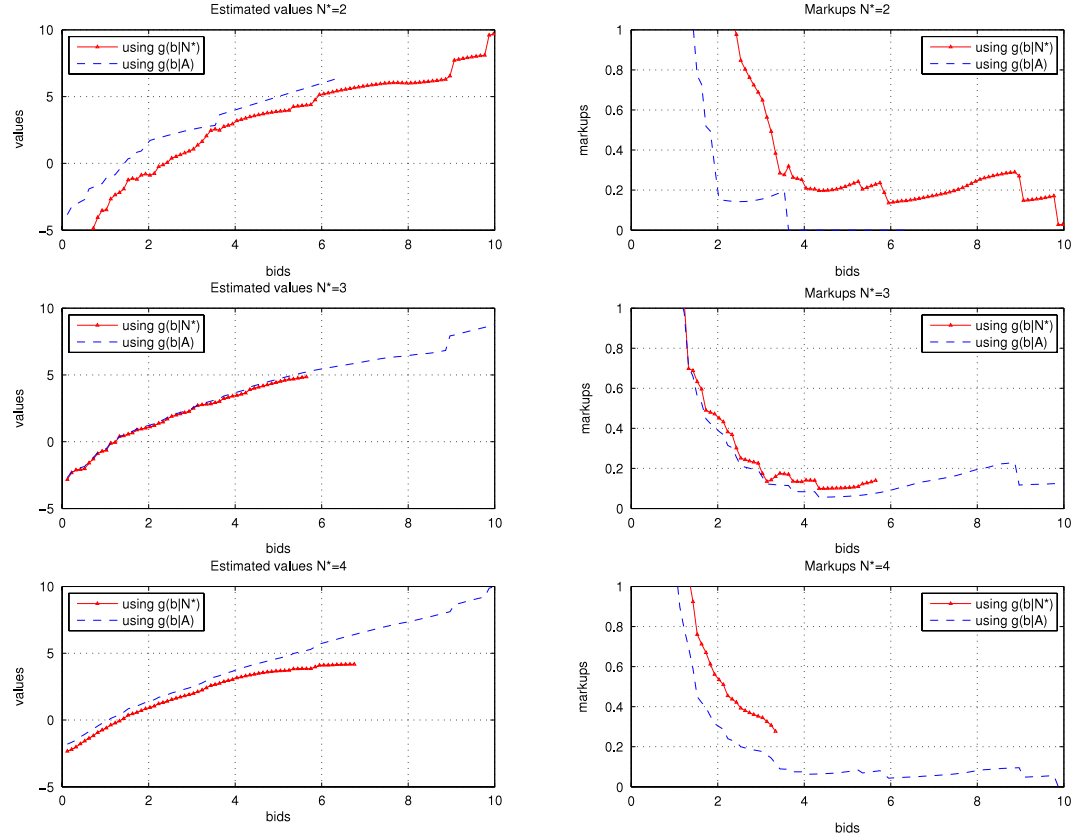

Fig. 7. Highway work projects, estimated values and markups.

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 value.

Second scenario: binding reserve price, but A observed

In the second scenario, we assume that the reserve price binds, but that A, the number of bidders who are willing to submit a bid above the reserve price, is 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 $g(b|N^*, b > r)$, 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 $g(b|N^*, b > r)$. 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 $g(b|N^*, b > r)$ in two steps. First, treating A as a conditioning covariate, 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 $G(b|N^*, b > r)$ via

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

(That is, for each N^*, we can recover an estimate of $G(b|N^*, b > r)$ for each distinct value of A. Since the model implies that these distributions should be identical for all A, we can, in principle, use this as a specification check of the model.)

In both scenarios, we need to find good candidates for the auxiliary 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.[30]

6.2. Endogenous entry

A second possible extension of our approach is to models of endogenous entry. In Samuelson's (1985) model, N^* potential entrants observe their valuations, and must decide whether or not to pay an entry cost $k > 0$ to bid in the auction. In this model (see Li and Zheng (2009) and Marmer et al. (2009)), the distribution of the valuations of the bidders who enter the auction, $F_{N^*}(v)$, varies depending on N^*. As Marmer et al. (2009) show, the inverse bidding strategy for this model, analogous to Eq. (3), is

$$\xi\left(b, N^*\right) = b + \frac{1 - p(N^*) + p(N^*)G(b|N^*)}{(N^* - 1)p(N^*)g(b|N^*)}, \tag{28}$$

where $p(N^*)$ denotes the equilibrium entry probability with N^* potential entrants.

We can apply our methodology to identify and estimate the valuation distributions $F_{N^*}(v)$ in this model, even when the number of potential entrants N^* is not observed. Let A denote the number of bidders who enter, which we assume to be observed.[31] First, using our procedure, the equilibrium bid distributions $G(b|N^*)$ and misclassification probabilities $G_{A|N^*}$ can be estimated using A as the proxy for N^* and a second bid in each auction in the role of Z. For recovering the valuations, note that, corresponding to Eq. (24), in equilibrium we have

$$A|N^* \sim \text{Binomial}(N^*, p(N^*)), \tag{29}$$

implying that $p(N^*)$ can be recovered for each value of N^* from the misclassification probability matrix $G_{A|N^*}$. Once $p(N^*)$ is known, the valuations can be identified for each b in the support of $G(b|N^*)$ using Eq. (28).

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. We have developed a nonparametric approach for estimating first-price auctions when N^*, the number of potential bidders, is unknown to the researcher, and varies in an unknown way among the auctions in the dataset. To our knowledge, our approach is the first solution to estimating such a model. Accommodating unknown N^* is also important for the policy implications of auction estimates, and the Monte Carlo and empirical results illustrate that ignoring the problem can lead to economically meaningful in differences the estimates of bidders' markups.

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 and unobserved from the bidders' perspective: e.g., Athey and Haile (2002), Hendricks et al. (2003), Bajari and Hortacsu (2003), Li and Zheng (2009) and Song (2006). It will be interesting to explore whether the methods used here can be useful for estimating these models.

More broadly, 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. These could include search models, or entry models. We are considering these possibilities in future work.

Appendix. Asymptotic properties of the two step estimator

Proof of uniform consistency of $\widehat{g}(b|N^*)$. In the first step, we estimate $\widehat{G}_{N|N^*}$ from

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

where $\psi(\cdot)$ is an analytic function as mentioned in Hu (2008) and

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

$$\widehat{G}_{N,Z} = \left[\frac{1}{T}\sum_t \mathbf{1}(N_t = N_j, Z_t = Z_k)\right]_{j,k}.$$

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

Lemma 1. *Suppose that* $\text{Var}(b|N, Z) < \infty$. *Then,*
$\widehat{G}_{N|N^*} - G_{N|N^*} = O_p\left(T^{-1/2}\right)$.

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

[31] In this model, a reserve price is irrelevant, because all bidders with valuations below the reserve price will never enter the auction. Hence, we do not need to distinguish between the number of bidders who enter and those who enter and submit a nonzero bid.

Proof. It is straightforward to show that $\widehat{G}_{Eb,N,Z} - G_{Eb,N,Z} = O_p(T^{-1/2})$ and $\widehat{G}_{N,Z} - G_{N,Z} = O_p(T^{-1/2})$. As mentioned in Hu (2008), the function $\psi(\cdot)$ is an analytic function. Therefore, the result holds. □

In the second step, we have

$$\widehat{g}(b|N^*) = \frac{e_{N^*}^T \widehat{G}_{N|N^*}^{-1} \overrightarrow{\widehat{g}}(b,N)}{e_{N^*}^T \widehat{G}_{N|N^*}^{-1} \overrightarrow{\widehat{g}}(N)},$$

where

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

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

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

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

Lemma 2. *Suppose:*

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

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

(2.3) $K(u)$ *is differentiable of order R, and the derivatives of order R are bounded.* $K(u)$ *is zero outside a bounded set.* $\int_{-\infty}^{\infty} K(u)\mathrm{d}u = 1$, *and there is a positive integer m such that for all* $j < m$, $K^{(j)}(u)$ *is absolutely continuous,* $\int_{-\infty}^{\infty} K(u)u^j\mathrm{d}u = 0$, *and* $\int_{-\infty}^{\infty} |u|^m |K(u)|\mathrm{d}u < \infty$.

(2.4) $h = cT^{-\delta}$ *for* $0 < \delta < 1/2$, *and* $c > 0$.

Then, for all j,

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

The most important assumption for Lemma 2 is (2.2), which places smoothness restrictions on the joint density $g(b,N)$. Via Eq. (7), this distribution is a mixture of conditional distributions $g(b|N^*)$, which possibly have a different support for different N^*. When the supports of $g(b|N^*)$ are known, condition (2.2) only requires the smoothness of $g(b|N^*)$ on its own support $[r,u_{N^*}]$ because the distribution $g(b|N)$ can be estimated piecewise on $[r,u_2],[u_2,u_3],\ldots,[u_{K-1},u_K]$. When the supports of $g(b|N^*)$ are unknown, condition (2.2) would require the density $g(b|N^*)$ for each value of N^* to be smooth at the upper boundary.[32]

Proof. By Lemma 1, it is straightforward to show that

$$\begin{aligned}\widehat{\Pr(N^*)} &= e_{N^*}^T \widehat{G}_{N|N^*}^{-1} \overrightarrow{\widehat{g}}(N) \\ &= e_{N^*}^T G_{N|N^*}^{-1} \overrightarrow{g}(N) + O_p\left(T^{-1/2}\right).\end{aligned}$$

Taking into account the fact that $\overrightarrow{\widehat{g}}(b,N)$ is bounded above, and $\widehat{\Pr(N^*)}$ is of order 1, we conclude that

$$\widehat{g}(b|N^*) = \frac{e_{N^*}^T G_{N|N^*}^{-1} \overrightarrow{\widehat{g}}(b,N)}{e_{N^*}^T G_{N|N^*}^{-1} \overrightarrow{g}(N)} + O_p\left(T^{-1/2}\right).$$

In order to show the consistency of our estimator $\widehat{g}(b|N^*)$, we need the uniform convergence of $\widehat{g}(\cdot,N_j)$. The kernel density estimator has been studied extensively. Following results from Lemma 5.4 and the discussion followed in Fan and Yao (2005) (which is based on the results in Bickel and Rosenblatt (1973)), under assumptions of Lemma 2, we have for all j[33]

$$\sup_b \left|\widehat{g}_{b,N}(\cdot,N_j) - \mathbb{E}\widehat{g}_{b,N}(\cdot,N_j)\right| = O_p\left(\frac{Th}{\ln T}\right)^{-1/2}. \tag{A.3}$$

According to the discussion on page 205 in Fan and Yao (2005), assumption (2.3) implies that the bias

$$\mathbb{E}\widehat{g}_{b,N}(\cdot,N_j) - g_{b,N}(\cdot,N_j) = O_p\left(h^m\right). \tag{A.4}$$

Consider that

$$\begin{aligned}\left|\widehat{g}_{b,N}(\cdot,N_j) - g_{b,N}(\cdot,N_j)\right| \le{}& \left|\widehat{g}_{b,N}(\cdot,N_j) - \mathbb{E}\widehat{g}_{b,N}(\cdot,N_j)\right| \\ &+ \left|\mathbb{E}\widehat{g}_{b,N}(\cdot,N_j) - g_{b,N}(\cdot,N_j)\right|.\end{aligned}$$

From (A.3) and (A.4), we immediately conclude that

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

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

Remark. Another technical issue pointed out in Guerre et al. (2000) is that the density $g(b|N^*)$ may not be bounded at the lower bound of its support, which is the reserve price r. They suggest using the transformed bids $b_\dagger \equiv \sqrt{b-r}$. Our identification and estimation procedures remain the same if b replaced by $b_\dagger$, where an estimate of the reserve price r could be the lowest observed bid in the dataset (given our assumption that the reserve price is fixed in the dataset). □

Proof of asymptotic normality of $\widehat{g}(b|N^*)$. In this proof, we show the asymptotic normality of $\widehat{g}(b|N^*)$ for a given value of b. Define $\gamma_0(b) = \{g_{b,N}(b)\}$, a column vector containing all the elements in the matrix $g(b,N)$. Similarly, we define $\widehat{\gamma}(b) = \{\widehat{g}_{b,N}(b)\}$. The proof of Lemma 2 suggests that

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

where

$$\varphi(\widehat{\gamma}(b)) \equiv \frac{e_{N^*}^T G_{N|N^*}^{-1} \overrightarrow{\widehat{g}}(b,N)}{e_{N^*}^T G_{N|N^*}^{-1} \overrightarrow{g}(N)}.$$

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

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

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

Lemma 3. *Suppose that the assumptions in Lemma 2 hold with* $R = 2$ *and that*

1. *there exists some* δ *such that* $\int |K(u)|^{2+\delta}\,\mathrm{d}u < \infty$,
2. $(Th)^{1/2}h^2 \to 0$, *as* $T \to \infty$.

Then, for a given b and N^*,

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

where

[32] In ongoing work, we are exploring alternative methods, based on wavelet methods (e.g. Hall et al. (1996)), to estimate the joint density $g(b,N)$ when there are unknown points of discontinuity, which can be due to the non-smoothness of the individual densities $g(b|N^*)$ at the upper boundary of their supports.

[33] The results in Fan and Yao (2005) are for $m = 2$ but they also hold for $m > 2$.

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

Proof. As discussed above, the asymptotic distribution of $\widehat{g}(b|N^*)$ is derived from that of $\widehat{\gamma}(b)$. In order to prove 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 λ 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

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

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

References

Adams, C., 2007. Estimating demand from eBay prices. International Journal of Industrial Organization 25, 1213–1232.

Athey, S., Haile, P., 2002. Identification of standard auction models. Econometrica 70, 2107–2140.

Athey, S., Levin, J., Seira, E., 2005. Comparing open and sealed bid auctions: theory and evidence from timber auctions. Working paper, Harvard University.

Bajari, P., Hortacsu, A., 2003. Winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions. RAND Journal of Economics 34, 329–355.

Bickel, P., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. The Annals of Statistics 1, 1071–1095.

Cuevas, A., Rodríguez-Casal, A., 2004. On boundary estimation. Advances in Applied Probability 36, 340–354.

Delaigle, A., Gijbels, I., 2006a. Data-driven boundary estimation in deconvolution problems. Computational Statistics and Data Analysis 50, 1965–1994.

Delaigle, A., Gijbels, I., 2006b. Estimation of boundary and discontinuity points in deconvolution problems. Statistica Sinica 16, 773–788.

Donald, S., Paarsch, H., 1993. Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. International Economic Review 34, 121–148.

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

Fan, J., Yao, Q., 2005. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.

Guerre, E., Perrigne, I., Vuong, Q., 2000. Optimal nonparametric estimation of first-price auctions. Econometrica 68, 525–574.

Guerre, E., Perrigne, I., Vuong, Q., 2009. Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. Econometrica 77 (4), 1193–1227.

Haile, P., Hong, H., Shum, M., 2003. Nonparametric tests for common values in first-price auctions. NBER working paper #10105.

Hall, P., McKay, I., Turlach, B., 1996. Performance of wavelet methods for functions with many discontinuities. Annals of Statistics 24, 2462–2476.

Hendricks, K., Pinkse, J., Porter, R., 2003. Empirical implications of equilibrium bidding in first-price, symmetric, common-value auctions. Review of Economic Studies 70, 115–145.

Hong, H., Shum, M., 2002. Increasing competition and the winner's curse: evidence from procurement. Review of Economic Studies 69, 871–898.

Hu, Y., 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution. Journal of Econometrics 144, 27–61.

Krasnokutskaya, E., Identification and estimation in highway procurement auctions under unobserved auction heterogeneity. Review of Economic Studies (forthcoming).

Krasnokutskaya, E., Seim, K., 2005. Bid preference programs and participation in highway procurement auctions. Working paper, University of Pennsylvania.

Laffont, J.J., Ossard, H., Vuong, Q., 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., Perrigne, I., Vuong, Q., 2000. Conditionally independent private information in OCS wildcat auctions. Journal of Econometrics 98, 129–161.

Li, T., Perrigne, I., Vuong, Q., 2002. Structural estimation of the affiliated private value auction model. RAND Journal of Economics 33, 171–193.

Li, T., Zheng, X., 2009. Entry and competition effects in first-price auctions: theory and evidence from procurement auctions. Review of Economic Studies 76, 1397–1429.

Mahajan, A., 2006. Identification and estimation of regression models with misclassification. Econometrica 74, 631–665.

Marmer, V., Shneyerov, A., Xu, P., 2009. What model for entry in first-price auctions? A nonparametric approach. Mimeo, University of British Columbia.

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

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press.

Samuelson, W.F., 1985. Competitive bidding with entry costs. Economics Letters 17, 53–57.

Song, U., 2004. Nonparametric estimation of an e-Bay auction model with an unknown number of bidders. Working paper, University of British Columbia.

Song, U., 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., van der Klaauw, B., 2007. If winning isn't everything, why do they keep score? A structural empirical analysis of dutch flower auctions. Mimeo, Free University Amsterdam.