

What Model for Entry in First-Price Auctions? A Nonparametric Approach*

Vadim Marmor
University of British Columbia

Artyom Shneyerov
Concordia University

Pai Xu
University of British Columbia

September 2, 2007

Abstract

We develop a model of selective entry in first-price auctions. The model nests the environments of Levin and Smith (1994) and Samuelson (1985) as limiting cases. We derive testable restrictions of the models and develop nonparametric tests of selective entry. We implement the tests on a dataset of highway procurement auctions in Oklahoma and find evidence of selective entry according to our model.

1 Introduction

A robust and well-documented feature of many real-world auctions is that not all bidders who are eligible to submit a bid choose to do so, suggesting that entry into the auction may be costly. In this paper, we develop nonparametric approaches that allow the empirical researcher to discriminate among different models of entry.

Most of the empirical auctions literature to date is based on the theoretical work of Levin and Smith (1994) (LS hereafter). In their model, potential bidders are initially uninformed about their valuations of the good, but may become informed and submit a bid at a cost. LS characterize a unique symmetric equilibrium in mixed strategies. The entrants randomize their entry decisions and earn zero expected profit.

The LS model is a natural starting point for empirical investigation of entry in auctions. It has a number of attractive features: both the distribution of valuations and entry cost are nonparametrically identified; the model is quite simple yet allows a new perspective on how competition affects expected revenues; typically produces realistic estimates of entry costs; can even be estimated in the presence of multiple equilibria.

Several empirical papers, most of them recent, estimated variants of this model. Bajari and Hortacsu (2003) study bidding behavior in eBay auctions, assuming common values. A Bayesian estimation method is implemented using a dataset of mint and proof sets of US

*We thank Ken Hendricks, Mike Peters and Larry Samuelson for useful comments, as well as participants of the CEA meeting in Montreal (2006) and the Summer Theory Workshop at UBC (2006). The second author thanks SSHRC for financial support made available through grants 12R27261 and 12R27788.

coins. They estimate the entry cost, simulate the expected seller revenues under different reserve prices and measure the extent of the winner’s curse. Athey, Levin, and Seira (2004) estimate a model of bidding in timber auctions with costly entry. An important novelty is the focus on bidder asymmetries, and the theoretical model is different from LS in that a random entry cost is assumed.¹ The model is estimated using a sample of sealed-bid auctions, and its predictions are compared to a control sample of ascending-bid auction. Athey, Levin and Seira find that the estimated model is able to explain several features of bidding behavior in ascending-bid auctions.

Li and Zheng (2005) study entry and bidding for lawn mowing contracts using the LS model. To our knowledge, it is the first paper in the literature that utilizes the number of planholders as a measure of potential competition. Li and Zheng (2005) propose and implement a Bayesian parametric estimation method and use their structural estimates to investigate the effect of restricting potential competition on the expected revenue. In addition, Li (2005) develops a general parametric approach to the estimation of auctions with entry. Krasnokutskaya and Seim study bid preference programs and bidder participation using California data. Their paper also uses the LS model, and as in Athey, Levin, and Seira (2004), the focus is on asymmetric equilibria. Bajari, Hong, and Ryan (2004) propose a parametric likelihood-based estimation strategy in the presence of multiple equilibria, and apply it to highway procurement auctions, using the LS model.

An alternative model of entry was developed in Samuelson (1985) (S hereafter). In this model, bidders make their entry decisions *after* they have learned their valuations. The entry cost is interpreted solely as the cost of preparing a bid, and bidders choose to enter if their valuations exceed a certain cutoff. The set of entrants is therefore a selected sample, biased towards bidders with higher valuations. We are not aware of any published work applying this model to data.²

Both LS and S models are stylized to capture the amount of information available to bidders at the entry stage: no information is available in LS, while the information is perfect in S. These polar assumptions lead to drastically differing policy implications. One of the most important and well-studied policy instruments in auctions is the reserve price. In a seminal paper, Riley and Samuelson (1981) show that, when the entry costs are null, the optimal policy for the seller is to set the reserve price above the level that he would be willing to accept. Moreover, the optimal reserve price does not depend on the number of potential bidders N . In the S model, while the optimal reserve price is also above the seller’s willingness to accept, it *increases* with N . LS, on the other hand, reach a striking conclusion that it is optimal to set the reserve price *at* the maximal willingness to accept level.

Given that policy implications are so different, it is important to be able to discriminate between these models empirically. We build on the insight in Haile, Hong, and Shum (2003) and propose to use exogenous variation in N as a basis for such test. We look at $F^*(v|N)$, the distribution of valuations conditional on entry. Following the approach of

¹Their theoretical model assumes that the entry cost is private information of potential bidders, but in the estimated model a fixed entry cost is assumed. They argue that in timber auctions, the heterogeneity in the entry cost is expected to be small.

²In a recent working paper, Xu (2007) adopts Samuelson’s model to estimate the entry cost in Michigan highway procurement auctions.

Guerre, Perrigne, and Vuong (2000) (GPV hereafter) we show that this distribution can be nonparametrically identified in both models if the number of potential bidders and all bids in each auction are observed. We show that, while $F^*(v|N)$ does not depend on N in the LS model, it does in the S model. The intuition here is simply that, in the S model, the valuations of active bidders are truncated by the entry cutoffs v_N^* that depend on N , but all share the same parent distribution across N .

We show that this leads to a stark testable restriction that can be derived in the following manner. Given that the distribution of active bidders valuations $F^*(v|N)$ is now simply a truncation of the parent distribution $F(v)$ of the potential bidders, $F^*(v|N) = (F(v) - F(v_N^*)) / (1 - F(v_N^*))$. Given that the truncation probability $1 - F(v_N^*)$ equals the bidding probability p_N that is directly identified from the data on bidding decisions, we have the testable restriction that $p_N F^*(v|N) + 1 - p_N$ is constant across N .

By exploiting the fact that the equilibrium bidding probability p_N is non-increasing in N , we can show that this restriction implies a weak ordering of actual bidders' conditional distribution functions (CDFs) $F^*(v|N)$, with larger values of N resulting in the distribution being tilted towards bidders with higher valuations,

$$F^*(v|N) \geq F^*(v|N') \text{ for } N' > N. \quad (1)$$

This is of course an intuitive implication of selective entry. It is also trivially satisfied by the LS model, with equality signs for all N . But both models have stronger testable implications as discussed above.

The LS and S models result in testable restrictions that are sharp equality constraints. It is important to have an alternative model that can explain data if these restrictions are rejected empirically. We propose a generalized model that results only in the weak order restriction (1). The model allows for selective entry but dispenses with the stark assumption that potential bidders perfectly know their valuations at the entry stage as in S, thus sharing with the LS model a costly valuation discovery stage. This model, which we refer to as the *generalized S* (GS hereafter), is as follows. At the entry stage, the potential bidders each observe a private signal correlated with their yet unknown valuation of the good. Based on this private signal, a bidder may learn the valuation upon incurring an entry cost k . The bidder who entered will only bid if the valuation exceeds the reserve price. The signals are assumed to be informative about the valuations, but to differentiate from the S model, not perfectly informative. The model does not formally nest LS and S models, but they can be viewed as its limiting cases: the LS model corresponds to uninformative signals, while the S model corresponds to perfectly informative signals.

Models similar to GS have been suggested, but not explored in detail, in the literature. Hendricks, Pinkse, and Porter (2003) estimate a model of bidding for off-shore oil. They sketch a model of entry that is in some respects similar to ours, but with a common-value component. The focus of their paper is however not on entry but on testing an equilibrium model of bidding. The model is also outlined in the concluding section of Ye (2005).

To implement the tests, we follow the approach of GPV and show that the distribution of entrants' valuations can be nonparametrically identified from the data if N and all bids in each auction are observed. This enables us to develop a nonparametric quantile-based test of selective entry in the spirit of Haile, Hong, and Shum (2003).

Although our approach shares with Haile, Hong, and Shum (2003) the basic idea that exogenous variation in the number of bidders can be used for testing the information environment of the game, there is a number of important differences. Haile, Hong, and Shum (2003) consider a different model in which bidders' valuations may have a common component. They propose a test for common values based on the variation in the number of *actual* bidders, while we test for selective entry using the variation in the number of *potential* bidders. Our approach is also different in the implementation: we use the estimation framework recently proposed in Marmor and Shneyerov (2006) that allows for arbitrary form of dependence on covariates.³

In our empirical application, we use a dataset of auctions conducted by the Oklahoma Department of Transportation (ODOT). In addition to all winning and losing bids and certain project characteristics, we also observe the planholders, a variable that can serve as a reasonable proxy for the number of potential bidders. We argue that, because the qualification process essentially selects bidders based on working capital requirements, the number of planholders may be assumed to be exogenous. Our robust empirical finding is that both LS and S models are strongly rejected, while the GS model is not rejected.

2 Three models of entry and their testable restrictions

The LS and S models share a common structure. There is an entry stage in which N potential bidders contemplate entry into the auction. At the auction stage, a bidding game transpires among those bidders that have entered. The auction is first-price sealed bid, possibly with a reserve price r . Only the bidders with valuations above the reserve price actually submit bids. We call them *actual* bidders. We assume the Independent Private Values (IPV) environment. The bidders' valuations are distributed according to the CDF $F(\cdot)$ that has support $[\underline{v}, \bar{v}]$ a corresponding density $f(\cdot)$, positive on the support. Entry is costly; only the bidders that have incurred the entry cost k can bid in the auction.

The two models differ in the information available at the entry stage. The LS model assumes that no information is available. Upon incurring the entry cost, the bidders learn their valuations and proceed to the bidding stage. Only the entrants with $v \geq r$ submit a bid. Levin and Smith characterize a symmetric perfect-Bayesian equilibrium of this game in which bidders submit a bid with probability $p \in [0, 1]$. We also sometimes index the bidding probability by N , i.e. write $p(N)$. The equilibrium is characterized in the following proposition. We assume that the reserve price is binding, but the result carries over with minor changes to the case when it is not binding.

Proposition 1 (*Levin and Smith, 1994; Milgrom, 2004*) *The bidding stage has a unique symmetric equilibrium, in which the bidding strategy $B_N(v)$ is an increasing and continuous function. The profit at the bidding stage, gross of the entry cost k , is given by*

$$\Pi(p) = \int_r^{\bar{v}} \bar{F}(v) (1 - p + pF^*(v))^{N-1} dv \quad (2)$$

³Haile, Hong, and Shum (2003) show how to incorporate covariates in a number of parametric specifications. We have not been able to extend this approach to the models with entry considered here.

where $F^*(v)$ is the CDF of actual bidders' valuations and $\bar{F}(v) = 1 - F(v)$ is the upper CDF of valuations. If k is smaller than the equilibrium profit when all rivals enter, i.e. $k < \Pi(1)$, then $p = 1$. Otherwise $p \in (0, 1)$, determined from the equation

$$k = \Pi(p). \quad (3)$$

(There is a qualification to be added to the above proposition, as well as to similar results for other models. Throughout the paper, we will assume away the uninteresting case of the entry cost so large that there is no entry, $p = 0$.) We do not state the equilibrium bidding strategies - see LS. The LS model has the following implications (we will show that these implications are testable). First, since the profit function in (2) is decreasing in the rival bidding probability p as well as in the number of potential rivals N , we can see that the probability of submitting a bid is at least non-increasing,

$$p(N) \geq p(N') \quad \forall N < N', \quad (4)$$

with *strict* inequality if N' is sufficiently large. Second, the distribution of entrants valuations coincides with the distribution of potential bidders valuations. The CDF of valuations conditional on entry

$$F^*(v) = \frac{F(v) - F(r)}{1 - F(r)},$$

which will be generally denoted as $F^*(v|N)$, has full support $[\underline{v}, \bar{v}]$ and does not depend on N ,

$$F^*(v|N) = F^*(v|N') \quad \forall N, N'. \quad (5)$$

In the S model, the potential bidders know their valuations at the entry stage. A bidder whose valuation is at the lower end of the support, $v = \underline{v}$, will, in any symmetric equilibrium, be unable to win with a positive probability. Consequently, it will not enter. Samuelson shows that there is a cutoff v_N^* such that a bidder strictly prefers to enter if and only if $v > v_N^*$, so that the probability of entry is $p = 1 - F(v_N^*)$. The equilibrium is formally characterized in the following proposition.

Proposition 2 (Samuelson (1985)) *The bidding stage has a unique symmetric equilibrium, in which the bidding strategy $B_N(v)$ is an increasing and continuous function. The profit at the bidding stage of the marginal entrant with valuation v_N^* is given by $(v_N^* - r)(1 - p)^{N-1}$, where $p = 1 - F(v_N^*)$ is the probability of bidding (here also equal to the probability of entry). The cutoff v_N^* is determined by the requirement that bidder with valuation v_N (who bids r in equilibrium) makes zero expected profit:*

$$k = (v_N^* - r)(1 - p)^{N-1}. \quad (6)$$

There is always entry with probability less than 1, i.e. $v_N^ \in (r, \bar{v} - k)$.*

The S model shares with the LS model the implication (4) that the bidding probabilities are non-increasing (they must actually be strictly decreasing in the S model). But the way

the entrants' CDFs depend on N is drastically different. We will denote the distribution of active bidders' valuations as $F^*(v|N)$ in all models considered. For $v > v_N^*$,

$$\begin{aligned} F^*(v|N) &= \frac{F(v) - F(v_N^*)}{1 - F(v_N^*)} \\ &= \frac{F(v) - (1 - p(N))}{p(N)} \end{aligned} \tag{7}$$

where we used the fact that the entry probability $p(N)$ is equal to $1 - F(v_N^*)$. Since the distribution F does not depend on N , a manipulation of (7) leads to the following restriction of the S model:

$$p(N) F^*(v|N) + 1 - p(N) = p(N') F^*(v|N') + 1 - p(N') \quad \forall N, N'. \tag{8}$$

A model of selective entry proposed in this paper occupies a middle ground between S and LS. Specifically, it shares with S the assumption that information about the valuation is available at the bidding stage, but dispenses with the stark assumption that this information is perfect. The game begins with the entry stage. N potential risk-neutral bidders obtain preliminary estimates (signals) S_i of their true values V_i ; it is assumed that this information is available to them for free. Upon observing S_i , a bidder may incur an entry cost k , which results in observing the true value V_i and entering the auction. Only the bidders that have learned V_i are eligible to submit a bid in the auction. Moreover, only those with valuations at or above the reserve price r submit a bid.

We assume that the pairs (V_i, S_i) are identically and independently distributed across bidders and are drawn from distribution $F(v, s)$ with density $f(v, s)$, positive on $[v, \bar{v}] \times [0, 1]$. For convenience, we assume that the marginal distribution of the signals is uniform on $[0, 1]$. Since the informational content of signals is preserved under a monotone transformation, this is without loss of generality.

The entry stage is followed by the bidding stage. Active bidders draw their values V_i , and then simultaneously and independently submit sealed bids. Active bidders do not know the number of active bidders, only the number of potential bidders N . The good is awarded to the highest bidder who pays its bid. We assume that the signals are informative and that higher signals are indicative of higher values. Formally, we make the following stochastic dominance assumption.⁴

Assumption 1 (Informative Signals) $\forall (v, s) \in [r, \bar{v}] \times [0, 1]$, $s < s'$ if and only if $F(v|s, V_i \geq r) > F(v|s', V_i \geq r)$.

Note that although the LS and S models are not nested within GS, they can be viewed as limit cases of the GS. The LS model corresponds to signals being independent of the valuations; this would effectively purify the mixed-strategy equilibrium. The S model corresponds to the other extreme, namely the signals and valuations being perfectly correlated.

⁴Assumption 1 essentially means that S_i is "good news" about V_i , in the sense of Milgrom (1981). For our testable restriction, we need a stronger version of the good news assumption, namely that S_i is "good news" about V_i even conditional on $V_i \geq r$. This distinction is immaterial when the reserve price is not binding.

A symmetric equilibrium can be characterized in a manner similar to the LS model. Once again, we assume that the reserve price is binding, but the result carries over with minor changes to the case when it is not binding.

Proposition 3 *A symmetric equilibrium is characterized by a signal cutoff \bar{s} such that only the bidders for whom the news are sufficiently good, $S_i \geq \bar{s}$, enter. The equilibrium profit conditional on s is equal to*

$$\Pi(p, s) = \int_r^{\bar{v}} \bar{F}(v|s) (1 - p + pF^*(v|N))^{N-1} dv, \quad (9)$$

where

$$p = \Pr\{S_i \geq \bar{s}, V_i \geq r\} \quad (10)$$

is the probability of submitting a bid and

$$F^*(v|N) = \Pr\{V_i \leq v | S_i \geq \bar{s}, V_i \geq r\}$$

is the distribution of active bidders' valuations. If the entry cost is so small that $\bar{s} = 0$, all potential bidders enter and therefore $\bar{s} = 0$. If k is moderate so that $\bar{s} \in (0, 1)$ then the bidder with signal \bar{s} is indifferent between entering or not and the equilibrium cutoff \bar{s} is determined from

$$k = \Pi(p, \bar{s}). \quad (11)$$

where p depends on \bar{s} through (10). The bidding strategy is given by

$$B_N(v) = v - \frac{\int_r^v (1 - p + pF^*(\tilde{v}|N))^{N-1} d\tilde{v}}{(1 - p + pF^*(v|N))^{N-1}} \quad (12)$$

The proof of Proposition 3 parallels the argument in Milgrom (2004) and is sketched in the Appendix.

Corollary 1 *If we denote the signal cutoff as a function of N as $\bar{s}(N)$, then (9) implies that $\bar{s}(N)$ is non-decreasing, while (9) and (11) together imply that $\bar{s}(N)$ is an increasing function whenever $\bar{s}(N) \in (0, 1)$.*

The testable implications of the GS model are in a sense in between LS and S. A common restriction of the three models is that the probability of submitting a bid is decreasing in N , i.e. the restriction (4). But the restriction on active bidders' CDF $F^*(v|N)$ is different from either LS or S. Recall that this cdf is equal to $\Pr\{V_i \leq v | S_i \geq \bar{s}(N), V_i \geq r\}$. Since the cutoff $\bar{s}(N)$ is non-decreasing in N (strictly increasing when N becomes sufficiently large), and since the signals are informative about the valuations (Assumption 1), the $F^*(v|N)$'s are tilted towards bidders with higher valuations. Formally, the GS model has the following first-order stochastic dominance restriction :

$$F^*(v|N) \geq F^*(v|N') \quad \forall N < N' \quad (13)$$

with *strict* inequalities for sufficiently large N' . Note that this restriction is also true for the S model (compare to (8)), but is clearly weaker. We summarize the restrictions of all the models considered in this paper in the following proposition.

Proposition 4 *All three models, LS, S and GS, share the restriction that the probabilities of bidding $p(N)$ are non-increasing in N , as in (4). As N increases, the inequalities eventually become strict. The models impose different restrictions on the distribution of actual bidders' valuations $F^*(v|N)$. In the LS model, $F^*(v|N)$ does not change with N , as in (5). In the S model, $F^*(v|N)$ satisfy the restriction (8), implying that they are ordered in the sense of first-order stochastic dominance, as in (13). In GS, the only implication for $F^*(v|N)$ is this stochastic dominance restriction.*

3 Nonparametric identification of $p(N)$ and $F^*(v|N)$

This section deals with nonparametric identification of the distribution $F^*(v|N)$. It presents a unified treatment for all models - LS, S and GS. It is assumed that the econometrician can observe all the bids and therefore also the number of active bidders n . An important additional information that is assumed to be also available is the number of potential bidders N . In other words, we assume that the data generating process identifies $p(N)$ and $G^*(\cdot|N)$ where $p(N) = E[n|N]/N$ is the probability of submitting a bid and $G^*(b|N)$ is the distribution of entrants' bids conditional on N .

Bidders' valuations are not directly observable, but can be recovered from the first-order conditions following the approach of GPV. Consider first-order equilibrium conditions of the bidding game. A bidder with value v who submits a bid b has a probability of winning over a given rival equal to $1 - p(N) \bar{G}^*(b|N)$, i.e. it is complementary to the probability that the rival submits a bid times the probability that this bid is above b . Since there are $N - 1$ identical rivals, it follows by independence that the probability of winning is $(1 - p(N) \bar{G}^*(b|N))^{N-1}$, and the expected profit is

$$\tilde{\Pi}(b, v) = (b - v) (1 - p(N) \bar{G}^*(b|N))^{N-1}.$$

Writing out the first-order condition, i.e. taking the derivative of $\tilde{\Pi}(b, v)$ with respect to b and setting it equal to 0, gives the inverse bidding strategy

$$\xi_N(b) = b + \frac{1 - p(N) \bar{G}^*(b|N)}{(N - 1) p(N) g^*(b|N)}. \quad (14)$$

We can see that ξ_N is identified from the observables, and its inverse, the bidding strategy $B_N(v)$, is also identified. Then the distribution of entrants' valuations $F^*(v|N)$ is identified according to

$$F^*(v|N) = G^*(B_N(v)|N).$$

While the identification of $p(N)$ and $F^*(v|N)$ is sufficient for the purposes of this paper (since the focus is on testing), this may be a good place to provide some remarks about identification of model primitives in all three models considered in this paper.

Consider first the LS model. If the reserve price is absent or non-binding, then the distribution of valuations is identified since it is equal to $F^*(v|N)$. The entry cost is also identified from the zero profit condition, provided that $p(N) \in (0, 1)$. The presence of a binding reserve price makes identification of both $F(\cdot)$ and the entry cost difficult. Unless $p(N) = p(N')$ for some distinct N, N' , so that the auction profit is greater than the entry

cost and $p(N) = 1 - F(r)$, neither $F(\cdot)$ nor k is identified. The reason is that we cannot distinguish between not bidding due to no entry and due to entering but drawing the valuation below the reserve price r .

In the S model, the entry cost is identified from equation (6), regardless of whether the reserve price is binding or not. However, the distribution of valuations is only identified for the values of v above the cutoff v_N^* (the cutoffs v_N^* are also identified).⁵

Finally, in the model with selective entry that we propose, the primitives for non-parametric identification are the entry cost k and $F(v|s)$, the distribution of valuations conditional on s . Neither is nonparametrically identified, even if the probability of entry is strictly between 0 and 1. The reason is that the data generating process only reveals the distribution of bidders' valuations, i.e. $F^*(v|N) = \Pr\{V_i \leq v | S_i \geq \bar{s}_N, V_i \geq r\}$, but not $F(v|s)$. The knowledge of $F(v|s)$ would be needed however to identify the entry cost according to (11).

4 Econometric implementation

In what follows, we allow for auctions heterogeneity by introducing the vector of auctions specific covariates x . We assume now that the distribution of valuations can change from auction to auction depending on the value of x and is denoted by $F(v|x)$. Similarly, the distribution of valuations conditional on entry will be denoted by $F^*(v|N, x)$.

One of the reasons for introducing the covariates is that different models can be true for different values of x . For example, if x is the size of the project, it is possible that the LS is the correct model for the auctions for small projects, while the S or GS are the right models when the projects are large. Consequently, testing has to be performed given a fixed value of x .

4.1 Hypotheses

The previous section shows that the distributions of valuations conditional on entry are identified for the three alternative models considered here, and therefore model selection tests can be formulated in terms of the characteristics of those distributions. In this case, it is convenient to work with the quantiles of the distributions $F^*(v|N, x)$ for a given value of x . Define

$$Q^*(\tau|N, x) \equiv F^{*-1}(\tau|N, x)$$

to be the τ -th quantile of the distribution of entrants' valuations. In terms of the quantiles, the testable restriction of the randomized entry, model for a given value of x , is

$$H_{LS} : Q^*(\tau|\underline{N}, x) = \dots = Q^*(\tau|\bar{N}, x), \forall \tau \in [0, 1],$$

while GS implies the restriction

$$H_{GS} : Q^*(\tau|\underline{N}, x) \leq \dots \leq Q^*(\tau|\bar{N}, x), \forall \tau \in [0, 1]. \quad (15)$$

⁵In a recent working paper, Xu (2007) develops a nonparametric estimator of the entry cost for Samuelson's model.

The testable restriction (8) of the Samuelson model can also be expressed using the quantiles of the distributions of observable bids as follows. First, by the definition and since F has a compact support, for any $\tau \in [0, 1]$, $F(Q(\tau|x)|x) = \tau$. Next, for $\tau > 1 - p(N, x)$, equation (7) implies that

$$F^*(Q(\tau|x)|N, x) = \frac{\tau - (1 - p(N, x))}{p(N, x)},$$

which in turn implies that

$$Q(\tau|x) = Q^*\left(\frac{\tau - (1 - p(N, x))}{p(N, x)}|N, x\right). \quad (16)$$

Define a function

$$\alpha(\tau, N, x) = \frac{\tau - (1 - p(N, x))}{p(N, x)}.$$

The left-hand side of (16) does not depend on N because they correspond to the distribution of *potential* bidders' valuations, and we then have that $Q^*(\alpha(\tau, N, x)|N, x)$ must be constant across N 's for all $\tau > 1 - p(\bar{N}, x)$:

$$H_S : Q^*(\alpha(\tau, \underline{N}, x)|\underline{N}, x) = \dots = Q^*(\alpha(\tau, \bar{N}, x)|\bar{N}, x), \quad \forall \tau > 1 - p(\bar{N}, x).$$

The restriction in H_S is limited to a particular range of τ 's. A similar restriction, however with $\tau \in [0, 1]$, can be obtained from (8) directly. Define a function

$$\beta(\tau, N, x) = 1 - \frac{p(\bar{N}, x)}{p(N, x)}(1 - \tau).$$

Note that, since $p(\bar{N}, x) \leq p(N, x)$, $0 \leq \beta(\tau, N, x) \leq 1$ for all $\tau \in [0, 1]$, and therefore can be interpreted as a legitimate transformation of the quantile order τ .⁶ The condition in (8) implies that for all N ,

$$F^*(v|N, x) = \beta(F^*(v|\bar{N}, x), N, x),$$

and, by the same argument as before, we obtain the following restriction in terms of the transformed quantiles:

$$H'_S : Q^*(\beta(\tau, \underline{N}, x)|\underline{N}, x) = \dots = Q^*(\beta(\tau, \bar{N}, x)|\bar{N}, x), \quad \forall \tau \in [0, 1].$$

From the practical point of view, testing H'_S is similar to testing H_S ; however, the first does not require truncation of τ 's. Therefore, we focus only on H'_S . Note also that because $\beta(\tau, N, x)$ is decreasing in N , the restrictions under H_S and H'_S are consistent with the restriction of GS (15) on the quantiles $Q^*(\tau|N, x)$ without the transformation β , but are stronger.

In this section, we consider independent testing of H_{LS} , H_{GS} , and H'_S against their corresponding alternatives H_{LS}^A , H_{GS}^A , and H_S^A . The alternative hypotheses are defined as follows. Let H be the maintained hypothesis which does not impose any particular cross

⁶While any other fixed value of N can be used in the place of \bar{N} in the definition of β , the choice $N = \bar{N}$ ensures that β takes on values in the zero-one interval.

N restrictions on the quantile functions $Q^*(\tau|N, x)$. The alternative hypothesis H_{LS}^A is defined as $H_{LS}^A = H \setminus H_{LS}$; H_{GS}^A and H_S^A are defined similarly.

In addition to the hypotheses discussed above, we also consider testing whether the entry probabilities $p(N, x)$ are non-increasing in N . The null hypothesis, for a given value of x , is

$$H_p : 1 > p(\underline{N}, x) \geq \dots \geq p(\bar{N}, x) > 0,$$

while the alternative is given by $H_p^A = H \setminus H_p$, where in this case the maintained hypothesis H restricts all $p(N, x)$ only to be in the $(0, 1)$ interval.

4.2 Formal assumptions

We assume that a sample of L auctions is available, and index each auction by $l = 1, \dots, L$. In each auction l , we observe the set of (anonymous) potential bidders indexed by $i = 1, \dots, N_l$, where $N_l \in \mathcal{N} = \{\underline{N}, \dots, \bar{N}\}$ is the number of potential bidders. Each auction is characterized by a covariates vector $x_l \in \mathcal{X}$. Corresponding to each potential bidder there is a latent valuation V_{il} . Each model provides a structural link between the potential bids b_{il} of all potential bidders and their valuations V_{il} : $b_{il} = B(V_{il}|N_l, x_l)$. The model also provides a link between signals and the entry decisions $y_{il} \in \{0, 1\}$, where $y_{il} = 1$ if the potential bidder decided to enter and zero otherwise. Of course, the potential bids of those who have not entered remain latent, and only the entrants' bids are observable. We make the following assumptions concerning the data generating process.

Assumption 2 (a) $\{(N_l, x_l) : l = 1, \dots, L\}$ are *i.i.d.*

(b) *The marginal PDF of x_l , φ , is strictly positive, continuous on its compact support $\mathcal{X} \subset \mathbb{R}^d$, and admits at least $R \geq 2$ continuous and bounded partial derivatives on Interior(\mathcal{X}).*

(c) *The distribution of N_l conditional on x_l , $\pi(N|x)$, has support $\mathcal{N} = \{\underline{N}, \dots, \bar{N}\}$ for all $x \in \mathcal{X}$, $\underline{N} \geq 2$.*

(d) *V_{il} and N_l are independent conditional on x_l .*

(e) *$\{V_{il} : i = 1, \dots, N_l; l = 1, \dots, L\}$ are *i.i.d.* conditional on (N_l, x_l)*

(f) *For all $x \in \mathcal{X}$, the density of valuations $f(\cdot|\cdot)$ is strictly positive and bounded away from zero on its support, a compact interval $[\underline{v}(x), \bar{v}(x)] \subset \mathbb{R}_+$, and admits at least R continuous and bounded partial derivatives its interior.*

(g) *$\pi(N|\cdot)$ admit at least R continuous bounded derivatives on Interior(\mathcal{X}) for all $N \in \mathcal{N}$.*

(h) *The entry probability conditional on (N, x) , $p(N, x)$, is strictly positive for all $N \in \mathcal{N}$ and $x \in \mathcal{X}$, and $p(N, \cdot)$ admits at least R continuous derivatives bounded away from 0 on an open subset $\mathcal{X}^\dagger \in$ Interior(\mathcal{X}) and all $N \in \mathcal{N}$.*

Assumption 2(a) is the usual iid assumption on the data generating process for the covariates. Assumptions 2(b), (f), and smoothness of functions in (g) and (h) are standard in the nonparametric auctions literature (see, for example, GPV). Assumption 2(c) defines

the support of the distribution of N_l conditional on the covariates. Assumption 2(d) is one of the most important assumptions; it asserts that in the number of potential bidders N is exogenous conditional on $x_l = x$, which allows us to use the variation in N for the purpose of testing. Assumption 2(e) is the IPV assumption.

4.3 Estimation of quantiles

We now turn to the identification and estimation of the quantiles $Q^*(\tau|N, x)$. Since the bidding strategies in both models are increasing, equation (14) for the inverse bidding strategy implies that the (unobservable) value distribution quantiles are related to the (observable) bids distribution quantiles $q^*(\tau|N, x)$,

$$\begin{aligned} q^*(\tau|N, x) &= G^{*-1}(\tau|N, x) \\ &\equiv \inf \{b : G^*(b|N, x) \geq \tau\}. \end{aligned}$$

according to the formula

$$Q^*(\tau|N, x) = q^*(\tau|N, x) + \frac{1 - p(N, x)(1 - \tau)}{(N - 1)p(N, x)g^*(q^*(\tau|N, x)|N, x)}. \quad (17)$$

This relationship will form a basis for nonparametric estimators of the quantiles, $\hat{Q}^*(\tau|N, x)$, which in turn will form a basis for our nonparametric test of selective entry.

In order to implement our test, we need a consistent and asymptotically normal estimators of value quantiles $Q^*(\tau|N, x)$. Our approach to the estimation of $Q^*(\tau|N, x)$ follows Marmor and Shneyerov (2006).⁷ We first estimate $g^*(b|N, x)$, $q^*(\tau|N, x)$ and $p(N, x)$ by kernel methods, and then insert them into equation (17). Let K be a kernel satisfying the following usual assumption (see, for example, Newey (1994)).

Assumption 3 *The kernel K has at least R continuous and bounded derivatives on \mathbb{R} , compactly supported on $[-1, 1]$ and is of order R : $\int K(u) du = 1$, $\int u^j K(u) du = 0$ for $j = 1, \dots, R - 1$.*

The corresponding kernel estimators of $\varphi(x)$, $\pi(N|x)$, and $p(N, x)$ are

$$\begin{aligned} \hat{\varphi}(x) &= \frac{1}{h^{dL}} \sum_{l=1}^L \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right), \\ \hat{\pi}(N|x) &= \frac{1}{\hat{\varphi}(x) h^{dL}} \sum_{l=1}^L 1\{N_l = N\} \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right), \text{ and} \\ \hat{p}(N, x) &= \frac{1}{\hat{\varphi}(x) N \hat{\pi}(N|x) L h^d} \sum_{l=1}^L n_l 1\{N_l = N\} \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right), \end{aligned}$$

where h is the bandwidth parameter.

⁷Marmor and Shneyerov (2006) consider a setting that corresponds to our "uninformative signal" case, i.e. the Levin and Smith model. Our model shares the same crucial i.i.d. assumption on the valuations $v_{i\ell}$, implying that all their asymptotic results in Section 5 are also valid in our more general setting.

Denote the number of actual bidders in auction l as $n_l = \sum_{i=1}^{N_l} y_{il}$. We propose to estimate the conditional bids densities and distributions by standard kernel methods, with an adjustment needed to account for a random number of observations within each auction. We propose to estimate first the *expected* number of bid observations that correspond to N -bidder auctions in the sample with covariates x as

$$\hat{e}(N, x) = \hat{p}(N, x) \hat{\pi}(N|x) NL.$$

The proposed estimators of g^* and G^* are

$$\begin{aligned} \hat{g}^*(b|N, x) &= \frac{1}{h^{d+1} \hat{e}(N, x) \hat{\varphi}(x)} \sum_{l=1}^L \sum_{i=1}^{N_l} y_{il} \mathbf{1}\{N_l = N\} K\left(\frac{b_{il} - b}{h}\right) \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right), \\ \hat{G}^*(b|N, x) &= \frac{1}{h^{d+1} \hat{e}(N, x) \hat{\varphi}(x)} \sum_{l=1}^L \sum_{i=1}^{N_l} y_{il} \mathbf{1}\{N_l = N\} \mathbf{1}(b_{il} \leq b) \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right), \end{aligned}$$

The estimators \hat{g}^* and \hat{G}^* are essentially standard nonparametric conditional density and CDF estimators with the number of bids observations $\sum_{l=1}^L \sum_{i=1}^{N_l} y_{il} \mathbf{1}\{N_l = N\}$ replaced by its estimated expected value $\hat{e}(N, x)$.

Our estimator of bid quantiles $q^*(\tau|N, x)$ is given by

$$\hat{q}^*(\tau|N, x) = \inf \left\{ b : \hat{G}(b|N, x) \geq \tau \right\}.$$

The quantiles $Q^*(\tau|N, x)$ in (17) can be consistently estimated by plugging in the estimators \hat{g}^* , \hat{q}^* and \hat{p} :

$$\hat{Q}^*(\tau|N, x) = \hat{q}^*(\tau) + \frac{1 - \hat{p}(N, x)(1 - \tau)}{(N - 1)\hat{p}(N, x)\hat{g}^*(\hat{q}^*(\tau|N, x)|x)}. \quad (18)$$

The transformation $\beta(\tau, N, x)$ can be similarly estimated by using the estimators $\hat{p}(N, x)$,

$$\hat{\beta}(\tau, N, x) = 1 - \frac{\hat{p}(\bar{N}, x)}{\hat{p}(N, x)} (1 - \tau),$$

and the transformed quantiles estimated as $\hat{Q}^*(\hat{\beta}(\tau, N, x)|N, x)$.

Lemma 2 in the Appendix establishes, for an appropriately chosen bandwidth sequence h , consistency of $\hat{\pi}$, $\hat{\varphi}$, and uniform consistency of $\hat{g}^*(\cdot|N, x)$, $\hat{q}^*(\cdot|N, x)$, $\hat{Q}^*(\cdot|N, x)$, $\hat{\beta}(\cdot, N, x)$, as well as $\hat{Q}^*(\hat{\beta}(\cdot, N, x)|N, x)$. It also shows that \hat{g}^* has the slowest rate of convergence among all the components of \hat{Q}^* in (18). Consequently, the rate of convergence and asymptotic distribution of \hat{g}^* determine those of \hat{Q}^* . Using the results of Lemma 2, one obtains the following expansion of \hat{Q}^* :

$$\begin{aligned} \hat{Q}^*(\tau|N, x) &= Q^*(\tau|N, x) - \frac{1 - p(N, x)(1 - \tau)}{(N - 1)p(N, x)\tilde{g}^{*2}(q^*(\tau|N, x)|N, x)} \\ &\quad \times (\hat{g}^*(q^*(\tau|N, x)) - g^*(q^*(\tau|N, x))) + o_p\left(\frac{1}{\sqrt{Lh^{d+1}}}\right), \end{aligned} \quad (19)$$

where \tilde{g}^* is a mean-value between g^* and \hat{g}^* for $b = q^*(\tau|N, x)$. An analogous expansion can be obtained in the case of $\hat{Q}^*(\hat{\beta}(\tau, N, x)|N, x)$. Lemma 3 in the Appendix shows that, provided that the bandwidth h satisfies $Lh^{d+1} \rightarrow \infty$ and $\sqrt{Lh^{d+1}}h^R \rightarrow 0$, and Assumptions 2 and 3 hold, the estimator $\hat{g}^*(b|N, x)$ is asymptotically normal with variance

$$V_g(N, b, x) = \frac{g^*(b|N, x)}{Np(N, x)\pi(N|x)\varphi(x)} \left(\int K(u)^2 du \right)^{d+1}. \quad (20)$$

It follows then from the decomposition in (19) that our quantile estimator $\hat{Q}^*(\tau|N, x)$ is asymptotically normal.

Proposition 5 *Suppose that $\tau \in (0, 1)$ and $x \in \mathcal{X}^\dagger$. Assume that the bandwidth h satisfies as $L \rightarrow \infty$: $Lh^{d+1} \rightarrow \infty$ and $\sqrt{Lh^{d+1}}h^R \rightarrow 0$, where R is the order of the kernel. Then, under Assumptions 2 and 3, $\sqrt{Lh^{d+1}}(\hat{Q}^*(\tau|N, x) - Q^*(\tau|N, x))$ converges in distribution to a normal random variable with mean zero and variance $V_Q(N, \tau, x)$ and $\sqrt{Lh^{d+1}}(\hat{Q}^*(\hat{\beta}(\tau, N, x)|N, x) - Q^*(\beta(\tau, N, x)|N, x))$ converges in distribution to a normal random variable with mean zero and variance $V_Q(N, \beta(\tau, N, x), x)$, where*

$$V_Q(N, \tau, x) = \left(\frac{1 - p(N, x)(1 - \tau)}{(N - 1)p(N, x)g^{*2}(q^*(\tau|N, x)|N, x)} \right)^2 V_g(N, q^*(\tau|N, x), x).$$

Moreover, for any distinct $N, N' \in \{\underline{N}, \dots, \bar{N}\}$, $\tau, \tau' \in \Upsilon$, and $x, x' \in \mathcal{X}^\dagger$, the estimators $\hat{Q}^*(\tau|N, x)$ are asymptotically independent, as well as $\hat{Q}^*(\hat{\beta}(\tau, N, x)|N, x)$.

The standard nonparametric regression arguments imply that the estimator of entry probabilities $\hat{p}(N, x)$ is asymptotically normal as well (see, for example, Pagan and Ullah (1999), Theorem 3.5, page 110):

Proposition 6 *Suppose that $x \in \mathcal{X}^\dagger$. Assume that the bandwidth h satisfies as $L \rightarrow \infty$: $Lh^d \rightarrow \infty$ and $\sqrt{Lh^d}h^R \rightarrow 0$, where R is the order of the kernel. Then, under Assumptions 2 and 3, $\sqrt{Lh^d}(\hat{p}(N, x) - p(N, x))$ is asymptotically normal with mean zero and variance*

$$V_p(N, x) = \frac{p(N, x)(1 - p(N, x))}{N\pi(N|x)\varphi(x)} \left(\int K(u)^2 du \right)^d.$$

Moreover, the estimators $\hat{p}(N, x)$ are asymptotically independent for any distinct $N, N' \in \{\underline{N}, \dots, \bar{N}\}$ and $x, x' \in \mathcal{X}^\dagger$.

Consistent estimators of the asymptotic variances V_g , V_Q , and V_p can be constructed by the plug-in method, as it follows from the results of Lemma 2 in the Appendix.

4.4 Tests

In view of the results of Section 4.3, quantile restrictions derived from the LS, S, and GS models can be tested in a standard manner as equality or inequality constraints. We

implement the tests using a finite set of τ 's from $(0, 1)$ interval, $\Upsilon = \{\tau_1, \tau_2, \dots, \tau_k\}$. The tests of H_{LS} against H_{LS}^A and H_S against H_S^A proceed as follows. Let

$$\begin{aligned}\bar{Q}^*(\tau|x) &= \frac{\sum_{N \in \mathcal{N}} \hat{Q}^*(\tau|N, x) / \hat{V}_Q(N, \tau, x)}{\sum_{N \in \mathcal{N}} 1 / \hat{V}_Q(N, \tau, x)}, \text{ and} \\ \bar{Q}_\beta^*(\tau|x) &= \frac{\sum_{N \in \mathcal{N}} \hat{Q}^*(\beta(\tau, N, x) | N, x) / \hat{V}_Q(N, \beta(\tau, N, x), x)}{\sum_{N \in \mathcal{N}} 1 / \hat{V}_Q(N, \beta(\tau, N, x), x)}\end{aligned}$$

where \hat{V}_Q is the plug-in estimator of V_Q . Define the following distance statistics:

$$\begin{aligned}T^{LS}(x) &= Lh^{d+1} \sum_{N \in \mathcal{N}, \tau \in \Upsilon} \frac{1}{\hat{V}_Q(N, \tau, x)} \left(\hat{Q}^*(\tau|N, x) - \bar{Q}^*(\tau|x) \right)^2, \text{ and} \\ T^S(x) &= Lh^{d+1} \sum_{N \in \mathcal{N}, \tau \in \Upsilon} \frac{1}{\hat{V}_Q(N, \beta(\tau, N, x), x)} \left(\hat{Q}^*(\beta(\tau, N, x) | N, x) - \bar{Q}_\beta^*(\tau|x) \right)^2.\end{aligned}$$

The LS model should be rejected for the auctions with covariates' values x whenever $T^{LS}(x) > \chi_{(\#\mathcal{N}-1)k, 1-\alpha}^2$, where $\chi_{(\#\mathcal{N}-1)k, 1-\alpha}^2$ denotes the $1 - \alpha$ quantile of the chi-square distribution with degrees of freedom $(\#\mathcal{N} - 1)k$, where $\#\mathcal{N}$ denotes the number of elements in \mathcal{N} . Similarly, one rejects H_S in favor of H_S^A if $T^S(x) > \chi_{(\#\mathcal{N}-1)k, 1-\alpha}^2$. From the results of Proposition 5, the $T^{LS}(x)$ and $T^S(x)$ tests have asymptotic size α and are consistent against the corresponding alternatives. Note that due to asymptotic normality and independence of the quantile estimators, $T^{LS}(x)$ and $T^S(x)$ can be also viewed as likelihood ratio (LR) statistics.

We now turn to testing H_{GS} against H_{GS}^A . In this case, the distance or LR statistic is

$$T^{GS}(x) = \min_{y_N, \tau \leq \dots \leq y_{N, \tau}, \tau \in \Upsilon} Lh^{d+1} \sum_{N \in \mathcal{N}, \tau \in \Upsilon} \frac{1}{\hat{V}_Q(N, \tau, x)} \left(\hat{Q}^*(\tau|N, x) - y_{N, \tau} \right)^2, \quad (21)$$

and one should reject the null of GS in favor of the general alternative when $T^{GS}(x)$ takes on large values. Again, by Proposition 5 such a test is consistent. The following proposition describes how to construct an asymptotic size α test.

Proposition 7 *Let $x \in \mathcal{X}^\dagger$. Assume that $Lh^d \rightarrow \infty$ and $\sqrt{Lh^d}h^R \rightarrow 0$ as $L \rightarrow \infty$, and Assumptions 2 and 3 hold. Then*

$$\sup P_{H_{GS}}(T^{GS}(x) > c) = P_{H_{LS}}(T^{GS}(x) > c) \quad (22)$$

$$\rightarrow P(\mathcal{T}^{GS}(x) > c), \quad (23)$$

where $P_{H_{GS}}$ and $P_{H_{LS}}$ denotes probabilities under the inequality restrictions of GS and equality restrictions of LS respectively,

$$\mathcal{T}^{GS}(x) = \sum_{\tau \in \Upsilon} \min_{\Omega^{1/2}(\tau, x)\mu \leq 0} \|Z_\tau - \mu\|^2,$$

$Z_\tau \sim N(0, I_{\#\mathcal{N}-1})$ and independent across τ 's, $\Omega(\tau, x) = RV_Q(\tau, x)R'$ with $V_Q(\tau, x) = \text{diag}(V_Q(\underline{N}, \tau, x), \dots, V_Q(\bar{N}, \tau, x))$, and R is the $(\#\mathcal{N}-1) \times (\#\mathcal{N})$ matrix

$$R = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}.$$

The GS model does not determine uniquely the null distribution of the $T^{GS}(x)$ statistic; however, as the above result says, the probability of type I error is maximized when all inequalities are replaced with the equalities, i.e. the same restrictions as in the LS model. Consequently, a test that rejects H_{GS} when $T^{GS}(x) > c(x)_{GS,1-\alpha}$, where $c(x)_{GS,1-\alpha}$ is the $1-\alpha$ quantile of the distribution of $T^{GS}(x)$, has asymptotic size α . The distribution of $T^{GS}(x)$ depends on asymptotic variances of the quantile estimators; however, the critical values can be simulated as follows. First, from $\hat{V}_Q(\underline{N}, \tau, x), \dots, \hat{V}_Q(\bar{N}, \tau, x)$ construct the matrices $\hat{V}_Q(\tau, x)$ and $\hat{\Omega}(\tau, x)$ for τ_1, \dots, τ_k as defined in Proposition 7. Second, for $m = 1, \dots, M$, generate independent $N(0, I_{\#\mathcal{N}-1})$ vectors $Z_{\tau_1, m}, \dots, Z_{\tau_k, m}$, and compute

$$\hat{T}_m^{GS}(x) = \sum_{\tau \in \Upsilon} \min_{\hat{\Omega}^{1/2}(\tau, x)\mu \leq 0} \|Z_{\tau, m} - \mu\|^2.$$

The simulated critical value for a test with asymptotic size α , say $\hat{c}(x)_{GS,1-\alpha}$, is then computed as the $1-\alpha$ sample quantile of $\{\hat{T}_m^{GS}(x) : m = 1, \dots, M\}$. One should reject the null of GS when $T^{GS}(x) > \hat{c}(x)_{GS,1-\alpha}$.

Testing whether the entry probabilities are non-increasing in N , the H_p hypothesis, can be performed similarly to testing H_{GS} . Define

$$T^p(x) = \min_{y_N \geq \dots \geq y_{\bar{N}}} Lh^d \sum_{N \in \mathcal{N}} \frac{1}{\hat{V}_p(N, x)} (\hat{p}(N, x) - y_N)^2.$$

One should reject H_p in favour of H_p^A when $T^p(x) > c(x)_{p,1-\alpha}$, where $c(x)_{p,1-\alpha}$ is the $1-\alpha$ quantile of $T^p(x) = \min_{\Omega_p^{1/2}(x)\mu \geq 0} \|Z - \mu\|^2$, $\Omega_p(x) = RV_p(x)R'$, $V_p(x)$ is a diagonal matrix with the main diagonal elements $V_p(\underline{N}, x), \dots, V_p(\bar{N}, x)$, and $Z \sim N(0, I_{\#\mathcal{N}-1})$. Such a test has asymptotic size α and is consistent. As in the case of $T^{GS}(x)$, the critical values for the $T^p(x)$ test can be simulated following the steps described above.

5 Monte-Carlo experiment

In this section using Monte-Carlo experiments, we study the small sample properties of the tests discussed in the previous section. In particular, we are interested in the way the choice of quantiles τ affects size and power of the tests. In our simulations, we focus on testing the GS model without covariates x . We simulate the random signals S and valuations V using the Gaussian copula. Let (Z_1, Z_2) be bivariate normal with zero means, variances equal to one, and the correlation coefficient ρ . Let Φ denote the standard normal CDF. A pair (S, V) is generated as $S = \Phi(Z_1)$, and $V = \Phi(Z_2)$. Nonzero values of ρ correspond to

the case of informative signals and selective entry; while $\rho = 0$ corresponds to the case of the LS model.

Given the values of S and V the bids can be computed as follows. First, recall that $Z_2|Z_1 \sim N(\rho z_1, 1 - \rho^2)$, and, consequently, the conditional distribution of V given S is given by

$$\begin{aligned} F(v|S) &= P(V \leq v|S) \\ &= P(Z_2 \leq \Phi^{-1}(v) | \Phi^{-1}(S)) \\ &= \Phi\left(\frac{\Phi^{-1}(v) - \rho\Phi^{-1}(S)}{\sqrt{1 - \rho^2}}\right). \end{aligned}$$

Next, note that the marginal distribution of S is uniform on the $[0, 1]$ interval, and

$$\begin{aligned} F^*(v|N) &= F(v|S \geq \bar{s}(N)) \\ &= \frac{1}{1 - \bar{s}} \int_{\bar{s}(N)}^1 \Phi\left(\frac{\Phi^{-1}(v) - \rho\Phi^{-1}(s)}{\sqrt{1 - \rho^2}}\right) ds, \end{aligned}$$

where the cutoff signal $\bar{s}(N)$ can be found, given the value of N , as a solution to equation (11). Lastly, for $S \geq \bar{s}(N)$, the bids are computed according to the bidding strategy (12).

In our simulations, we set $L = 250$, $\mathcal{N} = \{2, 3, 4, 5\}$, $\pi(N) = 1/4$ for all $N \in \mathcal{N}$, and $k = 0.17$. The number of Monte Carlo replications is 1,000; in each replication, the critical values for the T^{GS} test are obtained using 999 replications. We use the triweight kernel function $K(u) = (35/32)(1 - u^2)^3 1\{|u| \leq 1\}$ for nonparametric kernel estimation. To reflect the fact that the number of active bidders varies from auction to auction depending on the number of potential bidders in the auction N and $p(N)$, we decided to use a bandwidth that depends on N . After some simulations, we determined that the following bandwidth performs reasonably well $h = (LN\hat{p}(N))^{-2/5}$. Thus, effectively different choices of bandwidth are used in estimation of $Q^*(\cdot|N)$ depending on N ; also, when computing T^{GS} , different values of h are used to multiply the elements of the sum in (15).

Table 1 reports the results of size simulations for $\rho = 0, 0.5$, and the following sets of quantiles: $\{0.5\}$, $\{0.3, 0.5, 0.7\}$, and $\{0.3, 0.4, 0.5, 0.6, 0.7\}$. While the asymptotic approximation works reasonable well for a small number quantiles, it appears that the finite sample size properties of the test deteriorate when the number of quantiles used to construct the test increases. For example, it appears that the T^{GS} test over rejects the null when $\rho = 0$ and $\tau \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$: for the nominal size of 10%, 5%, and 1% the simulated rejection rates are approximately 16%, 10% and 4% respectively. Note also that the rejection rates for $\rho = 0.5$ are smaller than for $\rho = 0$ and below the nominal rejection rates. This reflects the fact that the probability of type I error is maximized at $\rho = 0$.

Table 2 reports the *size corrected* power results (the critical values are computed from the simulated distribution of the test statistic under the null). To address the power issue, it is necessary to come up with an alternative. Ideally, this would be achieved by considering a structural model. Absent a structural model, however, we are allowed to consider any configuration of bidding quantiles, in particular we may reverse their order, making *decreasing* as opposed to increasing in N . To do this in the simplest fashion possible, we

multiply each quantile by minus one and then add a constant to all quantiles to assure that they are positive.

Table 2 shows that the power increases with the distance from the null. The power also increases when we use quantiles 0.3 and 0.7 in addition to the median. However, we also observe that in some cases the size corrected power may decrease with the number of quantiles, when the number of quantiles used to construct the test is large. For example, in the case of $\{0.3, 0.5, 0.7\}$ quantiles, $\rho = 0.9$, and the nominal size of 5%, the simulated rejection rate is about 20%; however, when we use in addition the quantiles 0.4 and 0.6, the rejection rate is only about 18%. In practice, given samples of moderate size, we recommend using 3 fixed quantiles in order to maintain good size and reasonable power.

6 Empirical application

Our dataset consists of 547 auctions for surface paving and grading contracts let by Oklahoma Department of Transportation (ODOT) for the period of January, 2002 to December, 2005.⁸ The available data items include all bids, the engineer’s estimate, the time length of the contract (in days), the number of items in the proposal and the length of the road. The ODOT implements a policy under which all bids over 7% of the engineer’s estimate are typically rejected, so there is a binding reserve price. In reality, we do observe bids above the reserve price (although extremely few winning bids were above the reserve price). We treat these bids as non-serious and eliminate them from the sample.

Importantly, we observe the list of eligible bidders (planholders) for each auction. The list of planholders is published on the ODOT website prior to bidding. A sample list of planholders is exhibited in the Appendix. A firm becomes a planholder through the following process. All projects to be auctioned are advertised by the ODOT 4 to 10 weeks prior to the letting date. These advertisements include a brief summary of the project, including the general location of the work and the type of the work involved. A sample advertisement page is exhibited in the Appendix; one can see that the information in the advertisements is quite imprecise. It only contains the location of the project, the number of calendar days and sometimes the quality of asphalt to be used, but lacks the most important information: the engineer’s estimate and the detailed schedule of items.

Interested companies can then submit a request for plans and bidding proposals, the documents that contain the specifics of the project (in particular, the engineer’s estimate and items schedule). An important feature of the qualification process is that only eligible firms are allowed access to these documents. A firm is deemed eligible if it satisfies certain qualification requirements. We take the set of potential bidders to be equal the set of eligible firms.

The goal of the qualification process is to ensure that a potential bidder has sufficient expertise and capacity to undertake the project. While the expertise part is typically determined at the pre-qualification stage, the capacity part is typically project-specific. An important requirement is that the prospective bidder is not qualified for more than $2\frac{1}{2}$ times

⁸Our choice of surface paving and grading contracts is motivated by the fact that Hong and Shum (2002), in their study of highway procurement auctions in New Jersey, find little support for common values for this type of contracts. This is important because in this paper, we assume independent private values (costs).

its current working capital. Moreover, the prospective bidders are not permitted to bid on individual projects that in total exceed this working capital requirement. It is therefore plausible that the variation in the set of planholders occurs for exogenous reason, most prominently due to exogenous variation in the capacity relative to the project size.

Another variable deserving special mention is *Nitems*, the number of pay items in the project advertisement. (The pay items are the various inputs needed for the construction process.) On average, the projects have about 70 pay items, although even twice this number is not uncommon- the standard deviation is also about 70 items. The reason why we included this variable is because the number of pay items can affect the cost of information acquisition, and therefore the entry cost k . One would expect that projects that have more pay items would be more difficult to evaluate, because a potential bidder would need to search for more prices. The theoretical prediction therefore is that the probability of entry is decreasing in *Nitems*.

A number of explanatory variables are included; their description is given in Table 3.⁹ The results of the entry logit regression and OLS bidding regression are presented in Table 4. In the logit regression, *EngEst* is not significant, only N and *Nitems* are significant. They have signs that are consistent with the predictions of our model. The effect of adding one more potential bidder is to reduce the odds of submitting a bid by about 4%. Increasing *Nitems* by one standard deviation (adding about 70 pay items) reduces the odds by about 1%. Although statistically significant, the effect of *Nitems* is empirically quite small.

In the OLS regression, the dependent variable is $\log(\text{bid})$, where bid is the dollar bid divided by the engineer's estimate. The most important explanatory variable is the engineer's estimate. Using it alone produced R^2 of about 0.79, so the impact of the other variables is much smaller. In the order of importance, the next variable is the number of potential bidders N ; if it is included in the regression, R^2 increases to about 0.94.

Theory also predicts that increasing the cost of entry should, ceteris paribus, result in a lower entry probability and therefore less aggressive bidding. This is confirmed by the results of the OLS regression. Increasing *Nitems* by one standard deviation increases the bids by about 3%, a statistically significant but numerically small effect. A similar significant but small effect of *Nitems* is present in the logit regression. Because of this smallness, we do not condition on *Nitems*, or any other exogenous variables, in our implementation of the tests.

Included in both regressions are the dummy variables for top 20 firms. We define them as firms that appear on the planholders list most frequently. Observe that, although not all firms enter at the same rate and bid similarly, the empirical evidence of asymmetries is strong only for out-of-state firms (the firms with headquarters outside the state of Oklahoma) that enter less frequently and also bid less, and for the following three firms: Broce Construction, Glover Construction and Becco Contractors. In our effort to make potential bidders symmetric, we eliminate the auctions in which either out-of-state firms or these three firms were on the planholders list. We also excluded auctions that have more than $N = 13$ potential bidders (as the histogram in Figure ... shows, there are very few auctions for each $N > 13$). The working sample was thereby reduced significantly, to 258 auctions,

⁹The covariates are basically the same as in other papers on procurement auctions (e.g. Bajari and Ye (2003); Pesendorfer and Jofre-Bonet (2003); Krasnokutskaya (2003); Krasnokutskaya and Seim (2006); Li and Zheng (2005)).

and all the results discussed below this point were obtained using this smaller sample.

We begin by testing whether bid submission probabilities are declining in N , the prediction shared by all models considered in this paper. The average rate of bid submission among the planholders is about 62%. The plot of the frequency of bidding as a function of N is exhibited in Figure 2. Table 5 shows the estimated bidding probabilities as well as their standard errors. A strongly declining pattern is evident. The probability of submitting a bid is the highest (83%) when $N = 2$, and is reduced to about 30% when $N = 13$. The p-value of the formal statistical test of monotonicity is 0.998, so that the declining pattern is not rejected.

We now turn to the tests of the models - LS, S and GS. First note that, because the procurement auctions are low-bid, the null hypothesis that corresponds to the GS model must be changed accordingly, i.e. the quantiles must decreasing rather than increasing.

Following the standard approach in the literature, we work with bids divided by the engineer's estimate, i.e. with so-called normalized bids.¹⁰ The estimated quantiles and their standard errors, as well as transformed quantiles and their standard errors are shown in Tables 6,7 and 8,9 respectively. The median quantiles, together with 95% confidence bands, are exhibited graphically in Figures 3 and 4 respectively. In the figures, one can observe discernible declining patterns for the medians. The results of statistical tests are contained in Table 10. We have tried various combinations of quantiles, and were unable to reject the GS model even when we considered a relatively dense set of quantiles - $\{0.3, 0.4, 0.5, 0.6, 0.7\}$. (Recall that our simulation results suggest that there might even be over rejection with many quantiles, so that the size-corrected p-values may even be larger than the asymptotic ones reported in Table 10 for $\{0.3, 0.4, 0.5, 0.6, 0.7\}$.) On the other hand, there is robust evidence against both LS and S. These models are strongly rejected for all considered combinations of quantiles.

7 Concluding remarks

In this paper, we have proposed nonparametric tests to discriminate among alternative models of entry in first-price auctions. Among the models considered are: (a) the Levin and Smith (1994) model that assumes randomized entry strategies, (b) the Samuelson (1985) model that assumes that bidders are perfectly informed about their valuations at the entry stage, and select into the pool of entrants based on this information, and (c) a new model that allows for selective entry but in a less stark form than Samuelson (1985). Specifically, our model assumes that bidders receive signals that are informative about their valuations and make their entry decisions based on these signals.

Applying these tests to data from the Oklahoma Department of Transportation (ODOT) highway procurement auctions, we have found strong evidence for selective entry according to our model. This research could be extended in a number of directions. One obvious extension would be developing more powerful tests. This could be achieved in at least two ways.

First, one can improve power by considering restricted alternatives for the LS and S models. This can be implemented using a sequential testing strategy. In the first step,

¹⁰This approach is pursued, for example, in Krasnokutskaya (2003).

one would test the (composite) null hypothesis that corresponds to our generalized model. If the weak ordering of the CDFs is not rejected, then one proceeds to the next step (if it is rejected, then all models are wrong). In the next step, one would test the equality restrictions of the remaining models against the inequality restrictions that correspond to the alternative of the weak ordering of the CDFs. This would roughly correspond to the idea of one-sided testing. We have not pursued this approach for several reasons. In our empirical application, there is sufficient evidence against the restrictions imposed by the LS or S model even within the two-sided testing framework that we have developed. Therefore, introducing a sharper alternative would not lead to new empirical findings. Moreover, in considering a sharper alternative for the S model, we cannot rely on any available econometric results. The difficulty here is that the null hypothesis of the S model consists of equality constraints of the transformed quantiles, while the alternative consists of inequality constraints on the original quantiles.

Second, one can improve power by considering tests based on a continuum of quantiles rather than a finite set. Haile, Hong, and Shum (2003) have pursued this approach, developing a Kolmogorov-Smirnov type test. This would be an important extension of our approach left for future work. Similar hypotheses are considered in the recent literature on tests of stochastic dominance and monotonicity (see, for example, Lee, Linton, and Whang (2006)); however, their approach cannot be applied directly in our case, since private valuations are unobservable, and our statistics are based on kernel density estimators.

Another extension would be to allow bidder asymmetries (e.g., a recent working paper by Krasnokutskaya and Seim (2006)). The obvious difficulty here would be the necessity to deal with multiple equilibria. Bajari, Hong, and Ryan (2004) obtain a number of identification results in this direction and estimate a parametric model with multiple equilibria for highway procurement auctions.

Yet another important extension would be incorporating unobserved heterogeneity in the estimates and the tests. Krasnokutskaya (2003) has found some evidence of the importance of unobserved heterogeneity in highway procurement auctions conducted by the Michigan Department of Transportation. It is presently unclear how her method would generalize to a model with endogenous entry such as here, but more work would clearly be welcome on this topic.

The auction model itself could also be generalized in a number of ways, such as allowing affiliated private values, as in Li, Perrigne, and Vuong (2002) or dynamic features as in Pesendorfer and Jofre-Bonet (2003). We leave these potential extensions for future work.

8 Appendix of proofs

Proof of Proposition 3. One can show that a symmetric equilibrium is characterized by a signal cutoff \bar{s} such that only the bidders for whom the news are sufficiently good, $S_i \geq \bar{s}$, enter. Since the bidders who have entered know their true values V_i , the bidding game is essentially a first-price auction with a random number of bidders.

It is easy to see that, in a symmetric equilibrium, the probability of winning the auction for a bidder with valuation v is

$$(1 - p + pF^*(v|N))^{N-1}. \quad (24)$$

To see why, note that because of independence, a bidder wins against a given potential rival either if the rival does not bid, or bids but his valuation conditional on bidding is less than v . This probability is equal to $1 - p + pF^*(v|N)$. By independence, we arrive at (24) as the total probability of winning the auction.

A standard envelope formula then applies: the slope of the profit function equals the probability of winning the auction. Since the expected profit obtained by the lowest bidder type is 0 the profit at the bidding stage as a function of v is equal to

$$\Pi^*(v) = \int_r^v (1 - p + pF^*(\tilde{v}|N))^{N-1} d\tilde{v}.$$

(Notice that this profit is gross of entry cost.) Since this profit is also equal to

$$(v - B_N(v))(1 - p + pF^*(v|N))^{N-1},$$

we obtain the bidding strategy equation (12). Standard single-crossing arguments (e.g., Athey (2002)) imply that that $B_N(v)$ is strictly increasing in v .

Multiplying $\Pi^*(v)$ by $f(v|s)$, the density of valuations conditional on the signal that is obtained at the entry stage, we get the expected equilibrium profit at the bidding stage as a function of the signal:

$$\Pi(p, s) = \int_r^{\bar{v}} f(v|s) \int_r^v (1 - p + pF^*(\tilde{v}|N))^{N-1} d\tilde{v}.$$

Changing the order of integration on the right-hand side of the above equation gives equation (9) in the statement of the proposition. The rest follows simply by observing that the bidder will make his entry decision by comparing $\Pi(p, s)$ with the entry cost k . Since $\Pi(p, s)$ is increasing in s , there are three possibilities depending on the magnitude of the entry cost k . First, the entry cost can be so small that $\bar{s} = 0$ and all potential bidders enter (but not necessarily submit a bid). Second, k may be moderate so that $\bar{s} \in (0, 1)$. In this case, the bidder with signal \bar{s} is indifferent between entering or not, which gives equation (11) in the statement of the proposition. ■

Lemma 1 *Under Assumption 2(f), for all $N \in \mathcal{N}$ and $x \in \mathcal{X}$, the distribution of bids has the compact support $[\underline{b}(N, x), \bar{b}(N, x)]$, and $g^*(\cdot|N, \cdot)$ has at least $R + 1$ continuous partial derivatives on its interior. Furthermore, $g^*(b|N, x)$ is bounded away from zero.*

Proof of Lemma 1. The proof is similar to that of Proposition 1 of GPV. ■

Lemma 2 Let $[\underline{v}(N, x), \bar{v}(N, x)]$ denote the support of $F^*(v|N, x)$. Define $\Lambda(N, x) = [v_1(N, x), v_2(N, x)] \subset [\underline{v}(N, x), \bar{v}(N, x)]$, and $\Upsilon(N, x) = [\tau_1(N, x), \tau_2(N, x)]$, such that $\tau_i(N, x) = F^*(v_i(N, x)|N, x)$ for $i = 1, 2$. Define further $\Theta(N, x) = [b_1(N, x), b_2(N, x)]$, where $b_i(N, x) = q^*(\tau_i(N, x)|N, x)$, $i = 1, 2$. Lastly, let $\Upsilon^\beta(N, x) = [\tau_1^\beta(N, x), \tau_2^\beta(N, x)]$ such that $\tau_1^\beta(N, x) = \inf\{\tau|\beta(\tau, N, x) \geq \tau_1(N, x), \tau \in [0, 1]\}$ and similarly $\tau_2^\beta(N, x) = \sup\{\tau|\beta(\tau, N, x) \leq \tau_2(N, x), \tau \in [0, 1]\}$. Then, under Assumptions 2 and 3, for all $x \in \text{Interior}(\mathcal{X})$ and $N \in \mathcal{N}$,

- (a) $\hat{\varphi}(x) - \varphi(x) = O_p\left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R\right)$.
- (b) $\hat{\pi}(N|x) - \pi(N|x) = O_p\left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R\right)$.
- (c) $\hat{p}(N, x) - p(N, x) = O_p\left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R\right)$.
- (d) $\sup_{b \in [b(N, x), \bar{b}(N, x)]} |\hat{G}^*(b|N, x) - G^*(b|N, x)| = O_p\left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R\right)$.
- (e) $\sup_{\tau \in \Upsilon(N, x)} |\hat{q}^*(\tau|N, x) - q^*(\tau|N, x)| = O_p\left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R\right)$.
- (f) $\sup_{b \in \Theta(N, x)} |\hat{g}^*(b|N, x) - g^*(b|N, x)| = O_p\left(\left(\frac{Lh^{d+1}}{\log L}\right)^{-1/2} + h^R\right)$.
- (g) $\sup_{\tau \in \Upsilon(N, x)} |\hat{Q}^*(\tau|N, x) - Q^*(\tau|N, x)| = O_p\left(\left(\frac{Lh^{d+1}}{\log L}\right)^{-1/2} + h^R\right)$.
- (h) $\sup_{\tau \in [0, 1]} |\hat{\beta}(\tau, N, x) - \beta(\tau, N, x)| = O_p\left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R\right)$.
- (i) $\sup_{\tau \in \Upsilon^\beta(N, x)} |\hat{Q}^*(\hat{\beta}(\tau, N, x)|N, x) - Q^*(\beta(\tau, N, x)|N, x)| = O_p\left(\left(\frac{Lh^{d+1}}{\log L}\right)^{-1/2} + h^R\right)$.

Proof of Lemma 2. Parts (a)-(c) of the lemma follow from Lemma B.3 of Newey (1994).

For part (d), define a function

$$G_0^*(b, N, x) = Np(N, x)\pi(N|x)G^*(b|N, x)\varphi(x),$$

and its estimator as

$$\hat{G}_0^*(b, N, x) = \frac{1}{h^d L} \sum_{l=1}^L \sum_{i=1}^{N_l} y_{il} 1\{N_l = N\} 1(b_{il} \leq b) K_{*h}(x_l - x),$$

where

$$K_{*h}(x_l - x) = \frac{1}{h^d} K_d\left(\frac{x_l - x}{h}\right), \text{ and}$$

$$K_d \left(\frac{x_l - x}{h} \right) = \prod_{k=1}^d K \left(\frac{x_{kl} - x_k}{h} \right). \quad (25)$$

Next,

$$\begin{aligned} E\hat{G}_0^*(b, N, x) &= E \left(1 \{N_l = N\} K_{*h}(x_l - x) \sum_{i=1}^{N_l} y_{il} 1(b_{il} \leq b) \right) \\ &= NE(1 \{N_l = N\} K_{*h}(x_l - x) y_{il} 1(b_{il} \leq b)) \\ &= NE(E(1(b_{il} \leq b) | N, x_l, y_{il} = 1) y_{il} 1 \{N_l = N\} K_{*h}(x_l - x)) \\ &= NE(G^*(b|N, x_l) p(N, x_l) \pi(N|x_l) K_{*h}(x_l - x)) \\ &= N \int G^*(b|N, u) p(N, u) \pi(N|u) K_{*h}(x - u) \varphi(u) du \\ &= \int G_0^*(b, N, x + hu) K_d \left(\frac{u}{h} \right) du. \end{aligned}$$

By Lemma 1, $G^*(b|N, \cdot)$ admits at least $R+1$ continuous derivatives. Then, as in the proof of Lemma B.2 of Newey (1994), Assumptions 2(b), (g) and (h) imply that there exists a constant $c > 0$ such that

$$\left| G_0^*(b, N, x) - E\hat{G}_0^*(b, N, x) \right| \leq ch^R \left(\int \left| K_d \left(\frac{u}{h} \right) \right| \|u\|^R du \right) \|vec(D_x^R G_0^*(b, N, x))\|,$$

where $\|\cdot\|$ denotes the Euclidean norm, and $D_x^R G_0^*$ denotes the R -th partial derivative of G_0^* with respect to x . It follows then that

$$\sup_{b \in [\underline{b}(N, x), \bar{b}(N, x)]} \left| G_0^*(b, N, x) - E\hat{G}_0^*(b, N, x) \right| = O(h^R). \quad (26)$$

Now, we show that

$$\sup_{b \in [\underline{b}(N, x), \bar{b}(N, x)]} |\hat{G}_0^*(b, N, x) - E\hat{G}_0^*(b, N, x)| = O_p \left(\left(\frac{Lh^d}{\log L} \right)^{-1/2} \right). \quad (27)$$

We follow the approach of Pollard (1984). Consider, for given $N \in \mathcal{N}$ and $x \in Interior(\mathcal{X})$, a class of functions \mathcal{Z} indexed by h and b , with a representative function

$$z_l(b, N, x) = \sum_{i=1}^{N_l} y_{il} 1 \{N_l = N\} 1(b_{il} \leq b) h^d K_{*h}(x_l - x).$$

By the result in Pollard (1984) (Problem 28), the class \mathcal{Z} has polynomial discrimination. Theorem 37 in Pollard (1984) (see also Example 38) implies that for any sequences δ_L, α_L such that $L\delta_L^2 \alpha_L^2 / \log L \rightarrow \infty$, $Ez_l^2(b) \leq \delta_L^2$,

$$\alpha_L^{-1} \delta_L^{-2} \sup_{b \in [\underline{b}(N, x), \bar{b}(N, x)]} \left| \frac{1}{L} \sum_{l=1}^L z_l(b, N, x) - Ez_l(b, N, x) \right| \rightarrow 0 \quad (28)$$

almost surely. We claim that this implies

$$\left(\frac{Lh^d}{\log L}\right)^{1/2} \sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} |\hat{G}_0^*(b, N, x) - E\hat{G}_0^*(b, N, x)|.$$

is bounded as $L \rightarrow \infty$ almost surely. This implies that

$$\sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} |\hat{G}_0^*(b, N, x) - E\hat{G}_0^*(b, N, x)| = O_p \left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} \right).$$

The proof is by contradiction. Suppose not. Then there exist a sequence $\gamma_L \rightarrow \infty$ and a subsequence of L such that along this subsequence

$$\sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} |\hat{G}_0^*(b, N, x) - E\hat{G}_0^*(b, N, x)| \geq \gamma_L \left(\frac{Lh^d}{\log L}\right)^{-1/2}. \quad (29)$$

on a set of events $\Omega' \subset \Omega$ with a positive probability measure. Now if we let $\delta_L^2 = h^d$ and $\alpha_L = \left(\frac{Lh^d}{\log L}\right)^{-1/2} \gamma_L^{1/2}$, then the definition of z implies that, along the subsequence, on a set of events Ω' ,

$$\begin{aligned} & \alpha_L^{-1} \delta_L^{-2} \sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} \left| \frac{1}{L} \sum_{l=1}^L z_l(b, N, x) - E z_l(b, N, x) \right| \\ &= \left(\frac{Lh^d}{\log L}\right)^{1/2} \gamma_L^{-1/2} h^{-d} \sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} \left| \frac{1}{L} \sum_{l=1}^L z_l(b, N, x) - E z_l(b, N, x) \right| \\ &= \left(\frac{Lh^d}{\log L}\right)^{1/2} \gamma_L^{-1/2} \sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} |\hat{G}_0^*(b, N, x) - E\hat{G}_0^*(b, N, x)| \\ &\geq \left(\frac{Lh^d}{\log L}\right)^{1/2} \gamma_L^{-1/2} \gamma_L \left(\frac{Lh^d}{\log L}\right)^{-1/2} \\ &= \gamma_L^{1/2} \rightarrow \infty, \end{aligned}$$

where the inequality follows by (29), a contradiction to (28). This establishes (27), so that (26), (27) and the triangle inequality together imply that

$$\sup_{b \in [\underline{b}(N,x), \bar{b}(N,x)]} |\hat{G}_0^*(b, N, x) - G_0^*(b, N, x)| = O_p \left(\left(\frac{Lh^d}{\log L}\right)^{-1/2} + h^R \right). \quad (30)$$

To complete the proof, recall that, from the definitions of $G_0^*(b, N, x)$ and $\hat{G}_0^*(b, N, x)$,

$$G^*(b|N, x) = \frac{G_0^*(b, N, x)}{p(N, x) \pi(N|x) \varphi(x)}, \text{ and } \hat{G}^*(b|n, x) = \frac{\hat{G}_0^*(b, N, x)}{\hat{p}(N, x) \hat{\pi}(N|x) \hat{\varphi}(x)},$$

so that by the mean-value theorem,

$$\left| \hat{G}^*(b|N, x) - G^*(b|N, x) \right| \leq \tilde{C}(b, N, x) \left\| \begin{pmatrix} \hat{G}_0^*(b, N, x) - G_0^*(b, N, x) \\ \hat{p}(N, x) - p(N, x) \\ \hat{\pi}(N|x) - \pi(N|x) \\ \hat{\varphi}(x) - \varphi(x) \end{pmatrix} \right\|, \quad (31)$$

where $\tilde{C}(b, N, x)$ is given by

$$\left\| \frac{1}{\tilde{p}(N, x) \tilde{\pi}(N, x) \tilde{\varphi}(x)} \right\| \left\| \begin{pmatrix} \tilde{G}_0^*(b, N, x) \\ \tilde{p}(N, x) \\ \tilde{\pi}(N, x) \\ \tilde{\varphi}(x) \end{pmatrix} \right\|,$$

and $\left\| \left(\tilde{G}^0 - G^0, \tilde{p} - p, \tilde{\pi} - \pi, \tilde{\varphi} - \varphi \right) \right\| \leq \left\| \left(\hat{G}^0 - G^0, \hat{p} - p, \hat{\pi} - \pi, \hat{\varphi} - \varphi \right) \right\|$ for all (b, N, x) . Further, by Assumption 2(b), (c) and (h), and the results in parts (a)-(c) of the lemma, with the probability approaching one \tilde{p} , $\tilde{\pi}$ and $\tilde{\varphi}$ are bounded away from zero. The desired result follows from (30), (31) and parts (a)-(c) of the lemma.

For part (e) of the lemma, since $\hat{G}^*(\cdot|N, x)$ is monotone by construction,

$$\begin{aligned} P(\hat{q}^*(\tau_1(N, x)|N, x) < \underline{b}(N, x)) &= P\left(\inf_b \left\{ b : \hat{G}^*(b|N, x) \geq \tau_1(N, x) \right\} < \underline{b}(N, x)\right) \\ &= P\left(\hat{G}^*(\underline{b}(N, x)|N, x) > \tau_1(N, x)\right) \\ &= o(1), \end{aligned}$$

where the last equality is by the result in part (d). Similarly,

$$\begin{aligned} P(\hat{q}^*(\tau_2(N, x)|N, x) > \bar{b}(N, x)) &= P\left(\hat{G}(\bar{b}(N, x)|N, x) < \tau_2(N, x)\right) \\ &= o(1). \end{aligned}$$

Hence, for all $x \in \text{Interior}(\mathcal{X})$ and $N \in \mathcal{N}$, with the probability approaching one, $\underline{b}(N, x) \leq \hat{q}^*(\tau_1(N, x)|N, x) < \hat{q}^*(\tau_2(N, x)|N, x) \leq \bar{b}(N, x)$. Since the distribution $G^*(b|N, x)$ is continuous in b , $G^*(q^*(\tau|N, x)|N, x) = \tau$, and, for $\tau \in \Upsilon(N, x)$, we can write the identity

$$G^*(\hat{q}^*(\tau|N, x)|N, x) - G^*(q^*(\tau|N, x)|N, x) = G^*(\hat{q}^*(\tau|N, x)|N, x) - \tau. \quad (32)$$

Using Lemma 21.1(ii) of van der Vaart (1998), and by the definition of \hat{G}^* ,

$$0 \leq \hat{G}^*(\hat{q}^*(\tau|N, x)|N, x) - \tau \leq \frac{1}{\hat{p}(N, x) \hat{\pi}(N|x) \hat{\varphi}(x) NLh^d},$$

and by the results in (a)-(c),

$$\hat{G}^*(\hat{q}^*(\tau|N, x)|N, x) = \tau + O_p\left(\left(Lh^d\right)^{-1}\right) \quad (33)$$

uniformly over τ . Combining (32) and (33), and applying the mean-value theorem to the left-hand side of (32), we obtain

$$\hat{q}^*(\tau|N, x) - q^*(\tau|N, x)$$

$$= \frac{G^*(\hat{q}^*(\tau|N, x)|N, x) - \hat{G}^*(\hat{q}^*(\tau|N, x)|N, x)}{g^*(\tilde{q}^*(\tau|N, x)|N, x)} + O_p\left(\left(Lh^d\right)^{-1}\right), \quad (34)$$

where \tilde{q}^* lies between \hat{q}^* and q^* for all (τ, N, x) . Now, by Lemma 1, $g^*(b|N, x)$ is bounded away from zero, and the result in part (e) follows from (34) and part (d) of the lemma.

To prove part (f), by Lemma 1, $g^*(\cdot|N, \cdot)$ admits at least $R + 1$ continuous bounded partial derivatives. Let

$$g_0^*(b, N, x) = p(N, x) \pi(N|x) \varphi(x) g^*(b|N, x), \text{ and} \quad (35)$$

$$\hat{g}_0^*(b, N, x) = \hat{p}(N, x) \hat{\pi}(N|x) \hat{\varphi}(x) \hat{g}^*(b|N, x). \quad (36)$$

By Lemma B.3 of Newey (1994), $\hat{g}_0^*(b, N, x)$ is uniformly consistent over $b \in \Theta(N, x)$:

$$\sup_{b \in \Theta(N, x)} |\hat{g}_0^*(b, N, x) - g_0^*(b, N, x)| = O_p\left(\left(\frac{Lh^{d+1}}{\log L}\right)^{-1/2} + h^R\right). \quad (37)$$

By the results in parts (a)-(c), the estimators $\hat{p}(N, x)$, $\hat{\pi}(N|x)$ and $\hat{\varphi}(x)$ converge at the rate faster than that in (37). The desired result follows by the same argument as in the proof of part (d), equation (31).

Next, we prove part (g). By Lemma 1, $g^*(b|N, x) > c_g > 0$. Then

$$\begin{aligned} & \left| \hat{Q}^*(\tau|N, x) - Q^*(\tau|N, x) \right| \\ \leq & \left| \hat{q}^*(\tau|N, x) - q^*(\tau|N, x) \right| + 2 \frac{|\hat{g}^*(\hat{q}^*(\tau|N, x)|N, x) - g^*(q^*(\tau|N, x)|N, x)|}{p(N, x) \hat{g}^*(\hat{q}^*(\tau|N, x)|N, x) c_g} \\ & + \frac{|\hat{p}(N, x) - p(N, x)|}{\hat{p}(N, x) p(N, x) \hat{g}^*(\hat{q}^*(\tau|N, x)|N, x)} \\ \leq & \left(1 + \frac{2 \sup_{b \in \Theta(N, x)} |\partial g^*(b|N, x) / \partial b|}{p(N, x) \hat{g}^*(\hat{q}^*(\tau|N, x)|N, x) c_g} \right) |\hat{q}(\tau|n, x) - q(\tau|n, x)| \\ & + 2 \frac{|\hat{g}^*(\hat{q}^*(\tau|N, x)|N, x) - g^*(\hat{q}^*(\tau|N, x)|N, x)|}{p(N, x) \hat{g}^*(\hat{q}^*(\tau|N, x)|N, x) c_g} \\ & + \frac{|\hat{p}(N, x) - p(N, x)|}{\hat{p}(N, x) p(N, x) \hat{g}^*(\hat{q}^*(\tau|N, x)|N, x)}. \end{aligned} \quad (38)$$

Define an event

$$E_L(N, x) = \{\hat{q}^*(\tau_1(N, x)|N, x) \geq b_1(N, x), \hat{q}^*(\tau_2(N, x)|N, x) \leq b_2(N, x)\},$$

and let $\beta_L = \left(\frac{Lh^{d+1+2k}}{\log L}\right)^{1/2} + h^{-R}$. By the result in part (e), $P(E_L^c(N, x)) = o(1)$. Hence, it follows from part (e) of the lemma the estimator $\hat{g}^*(\hat{q}^*(\tau|N, x)|N, x)$ is bounded away from zero with the probability approaching one. Consequently, it follows by Lemma 1 and part (e) of this lemma that the first summand on the right-hand side of (38) is $O_p(\beta_L^{-1})$ uniformly over $\Upsilon(N, x)$. Next,

$$P\left(\sup_{\tau \in \Upsilon(N, x)} \beta_L |\hat{g}^*(\hat{q}^*(\tau|N, x)|N, x) - g^*(\hat{q}^*(\tau|N, x)|N, x)| > M\right)$$

$$\begin{aligned}
&\leq P \left(\sup_{\tau \in \Upsilon(N,x)} \beta_L |\hat{g}^*(\hat{q}^*(\tau|N,x)|N,x) - g^*(\hat{q}^*(\tau|N,x)|N,x)| > M, E_L(x) \right) \\
&\quad + P(E_L^c(x)) \\
&\leq P \left(\sup_{b \in \Theta(N,x)} \beta_L |\hat{g}^*(b|N,x) - g^*(b|N,x)| > M \right) + o(1). \tag{39}
\end{aligned}$$

The result of part (g) follows from parts (c) and (f) of the lemma and (39).

For part (h), by Assumption 2(h) and part (c) of the lemma, $\hat{\beta}(\tau, N, x) \rightarrow_p \beta(\tau, N, x)$ for all τ, N , and x . The result of part (h) follows since β is linear in τ (see Andrews (1992); also Theorems 21.9 and 21.10 on pages 337-339 of Davidson (1994)).

Lastly, we prove part (i). We have

$$\begin{aligned}
&\sup_{\tau \in \Upsilon^\beta(N,x)} \left| \hat{q}^*(\hat{\beta}(\tau, N, x)|N, x) - q^*(\beta(\tau, N, x)|N, x) \right| \\
&= \hat{q}^*(\hat{\beta}(\tau, N, x)|N, x) - q^*(\hat{\beta}(\tau, N, x)|N, x) \\
&\quad + q^*(\hat{\beta}(\tau, N, x)|N, x) - q^*(\beta(\tau, N, x)|N, x) \\
&\leq \sup_{\tau \in \Upsilon(N,x)} |\hat{q}^*(\tau|N, x) - q^*(\tau|N, x)| + O_p \left(\left(\frac{Lh^d}{\log L} \right)^{-1/2} + h^R \right) \\
&= O_p \left(\left(\frac{Lh^d}{\log L} \right)^{-1/2} + h^R \right), \tag{40}
\end{aligned}$$

where the inequality follows from part (h) of the lemma and Lemma 1. The result of part (i) follows from the definition of \hat{Q}^* in (18) and (40). ■

Lemma 3 *Let $\Theta(N, x)$ be as in Lemma 2. Suppose that Assumptions 2 and 3 hold, and that the bandwidth h is such that $Lh^{d+1} \rightarrow \infty$, $\sqrt{Lh^{d+1}}h^R \rightarrow 0$. Then*

$$\sqrt{Lh^{d+1}} (\hat{g}^*(b|N, x) - g^*(b|N, x)) \rightarrow_d N(0, V_g(b, N, x))$$

for $b \in \Theta(N, x)$, $x \in \text{Interior}(\mathcal{X})$, and $N \in \mathcal{N}$, where $V_g(b, N, x)$ is defined in (20). Furthermore, $\hat{g}^*(b|N_1, x)$ and $\hat{g}^*(b|N_2, x)$ are asymptotically independent for all $N_1 \neq N_2$, $N_1, N_2 \in \mathcal{N}$.

Proof of Lemma 3. Consider $g_0^*(b, n, x)$ and $\hat{g}_0^*(b, n, x)$ defined in (35) and (36) respectively. It follows from parts (a)-(c) of Lemma 2,

$$\begin{aligned}
&\sqrt{Lh^{d+1}} (\hat{g}^*(b|N, x) - g^*(b|N, x)) \\
&= \frac{1}{p(N, x) \pi(N|x) \varphi(x)} \sqrt{Lh^{d+1}} (\hat{g}_0^*(b, N, x) - g_0^*(b, N, x)) + o_p(1). \tag{41}
\end{aligned}$$

Furthermore, as in Lemma B2 of Newey (1994), $E\hat{g}_0^*(b, N, x) - g_0^*(b, N, x) = O(h^R)$ uniformly in $b \in \Theta(N, x)$ for all $x \in \text{Interior}(\mathcal{X})$ and $N \in \mathcal{N}$. Thus, it remains to establish asymptotic normality of $\sqrt{Lh^{d+1}} (\hat{g}_0^*(b, N, x) - E\hat{g}_0^*(b, N, x))$.

Define

$$w_{il,N} = \sqrt{\frac{1}{h^{d+1}}} y_{il} 1\{N_l = N\} K\left(\frac{b_{il} - b}{h}\right) K_d\left(\frac{x_l - x}{h}\right),$$

$$\bar{w}_{L,N} = \frac{1}{NL} \sum_{l=1}^L \sum_{i=l}^{N_l} w_{il,N},$$

where K_d is defined in (25). With above definitions we have that

$$\sqrt{NLh^{d+1}} (\hat{g}_0^*(b, N, x) - E\hat{g}_0^*(b, N, x)) = \sqrt{NL} (\bar{w}_{L,N} - E\bar{w}_{L,N}). \quad (42)$$

Then, by the Liapunov CLT (see, for example, Corollary 11.2.1 on page 427 of Lehman and Romano (2005)),

$$\sqrt{NL} (\bar{w}_{L,N} - E\bar{w}_{L,N}) / \sqrt{NL \text{Var}(\bar{w}_{L,N})} \rightarrow_d N(0, 1), \quad (43)$$

provided that $Ew_{il,N}^2 < \infty$, and for some $\delta > 0$,

$$\lim_{L \rightarrow \infty} \frac{1}{L^{\delta/2}} E |w_{il,N} - Ew_{il,N}|^{2+\delta} = 0.$$

The last condition follows from the Liapunov's condition (equation (11.12) on page 427 of Lehman and Romano (2005)) and because $w_{il,N}$ are iid Next, $Ew_{il,N}$ is given by

$$\begin{aligned} & \sqrt{\frac{1}{h^{d+1}}} E \left(p(N, x_l) \pi(N|x_l) \int K\left(\frac{u-b}{h}\right) g^*(u|N, x_l) du K_d\left(\frac{x_l-x}{h}\right) \right) \\ &= \sqrt{\frac{1}{h^{d+1}}} \int \int p(N, y) \pi(N|y) K\left(\frac{u-b}{h}\right) g^*(u|N, y) K_d\left(\frac{y-x}{h}\right) \varphi(y) dudy \\ &= \sqrt{h^{d+1}} \\ & \quad \times \int \int p(N, x+hy) \pi(N|x+hy) K(u) g^*(b+hu|N, x+hy) K_d(y) \varphi(x+hy) dudy \\ & \rightarrow 0. \end{aligned}$$

Further, $Ew_{il,N}^2$ is given by

$$\begin{aligned} & \frac{1}{h^{d+1}} \int \int p(N, y) \pi(N|y) K^2\left(\frac{u-b}{h}\right) g^*(u|N, y) K_d^2\left(\frac{y-x}{h}\right) \varphi(y) dudy \\ &= \int \int p(N, x+hy) \pi(N|x+hy) K^2(u) g^*(b+hu|N, x+hy) K_d^2(y) \varphi(x+hy) dudy \\ & < \infty. \end{aligned}$$

Hence,

$$NL \text{Var}(\bar{w}_{L,N}) \rightarrow p(N, x) \pi(N|x) g^*(b|N, x) \varphi(x) \left(\int K^2(u) du \right)^{d+1} du. \quad (44)$$

Next, $E |w_{il,N}|^{2+\delta}$ is bounded by

$$\begin{aligned}
& \frac{1}{h^{(d+1)(1+\delta/2)}} \int \int \left| K \left(\frac{u-b}{h} \right) \right|^{2+\delta} g^*(u|N, y) \left| K_d \left(\frac{y-x}{h} \right) \right|^{2+\delta} \varphi(y) \, dud y \\
&= \frac{1}{h^{(d+1)\delta/2}} \int \int |K(u)|^{2+\delta} g^*(b+hu|N, x+hy) |K_d(y)|^{2+\delta} \varphi(x+hy) \, dud y \\
&\leq \frac{1}{h^{(d+1)\delta/2}} \sup_{u \in [-1,1]} |K(u)|^{(d+1)(2+\delta)} \sup_{x \in \mathcal{X}} \varphi(x) \sup_{b \in \Theta(N,x)} g^*(b|N, x) \\
&= \frac{C}{h^{(d+1)\delta/2}}.
\end{aligned}$$

Lastly,

$$\begin{aligned}
\frac{1}{L^{\delta/2}} E |w_{il,N} - E w_{il,N}|^{2+\delta} &\leq \frac{2^{1+\delta}}{L^{\delta/2}} E |w_{il,N}|^{2+\delta} \\
&\leq \frac{2^{1+\delta} C}{(Lh^{d+1})^{\delta/2}} \\
&\rightarrow 0,
\end{aligned} \tag{45}$$

since $Lh^{d+1} \rightarrow \infty$ by the assumption. The first result of the lemma follows now from (41)-(45).

Next, note that the asymptotic covariance of \bar{w}_{L,N_1} and \bar{w}_{L,N_2} involves a product of the two indicator functions, $1\{N_l = N_1\} 1\{N_l = N_2\}$, which is zero for all $N_1 \neq N_2$. The joint asymptotic normality and asymptotic independence of $\hat{g}^*(b|N_1, x)$ and $\hat{g}(b|N_2, x)$ follows then by the Cramér-Wold device. ■

Proof of Proposition 5. The result will follow by (19) and Lemma 3. To show (19), note that, by Lemma 2 (c), (e) and (f),

$$\hat{Q}^*(\tau|N, x) = q^*(\tau|N, x) + \frac{1-p(N, x)(1-\tau)}{(N-1)p(N, x)\hat{g}^*(q^*(\tau|N, x)|x)} + o_p\left(\frac{1}{\sqrt{Lh^{d+1}}}\right),$$

and (19) follows by the mean-value theorem. ■

Proof of Proposition 7. The result in (22) follows by Lemma 8.2 of Perlman (1969). In order to show (23), consider first the case of $k = 1$. By the results in Chapter 21.3.3 of Gourieroux and Monfort (1995), $T^{GS}(x)$ is asymptotically equivalent to

$$\tilde{T}^{GS}(x) = \min_{\Omega^{1/2}(\tau, x)\mu \leq 0} Lh^{d+1} \|\hat{\gamma} - \mu\|^2,$$

where $\sqrt{Lh^{d+1}}\hat{\gamma} \rightarrow_d N(0, I_{\#N-1})$; however,

$$\begin{aligned}
\tilde{T}^{GS}(x) &= \min_{\Omega^{1/2}(\tau, x)\mu/\sqrt{Lh^{d+1}} \leq 0} \left\| \sqrt{Lh^{d+1}}\hat{\gamma} - \mu \right\|^2 \\
&= \min_{\Omega^{1/2}(\tau, x)\mu \leq 0} \left\| \sqrt{Lh^{d+1}}\hat{\gamma} - \mu \right\|^2,
\end{aligned}$$

and the result follows by the Continuous Mapping Theorem. Extension to the case of $k > 1$ is straightforward since there are no cross τ restrictions in (21), and the quantile estimators are asymptotically independent across τ . ■

References

- ANDREWS, D. W. K. (1992): “Generic Uniform Convergence,” *Econometric Theory*, 8(2), 241–257.
- ATHEY, S., J. LEVIN, AND E. SEIRA (2004): “Comparing Sealed Bid and Open Auctions: Theory and Evidence from Timber Auctions,” Department of Economics, Stanford University.
- BAJARI, P., H. HONG, AND S. RYAN (2004): “Identification and Estimation of Discrete Games of Complete Information,” NBER Working paper.
- BAJARI, P., AND A. HORTACSU (2003): “Winner’s Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions,” *RAND Journal of Economics*, 34(2), 329–355.
- BAJARI, P., AND L. YE (2003): “Deciding Between Competition and Collusion,” *Review of Economics and Statistics*, 85(4), 971–989.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, New York.
- GOURIEROUX, C., AND A. MONFORT (1995): *Statistics and Econometric Models*. Cambridge University Press, Cambridge.
- GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): “Optimal Nonparametric Estimation of First-Price Auctions,” *Econometrica*, 68(3), 525–74.
- HAILE, P., H. HONG, AND M. SHUM (2003): “Nonparametric Tests for Common Values at First-Price Sealed-Bid Auctions,” NBER Working Paper 10105.
- HENDRICKS, K., J. PINKSE, AND R. H. PORTER (2003): “Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions,” *Review of Economic Studies*, 70(1), 115–145.
- HONG, H., AND M. SHUM (2002): “Increasing Competition and the Winner’s Curse: Evidence from Procurement,” *Review of Economic Studies*, 69(4), 871–898.
- KRASNOKUTSKAYA, E. (2003): “Identification and Estimation in Highway Procurement Auctions under Unobserved Auction Heterogeneity,” Working Paper, University of Pennsylvania.
- KRASNOKUTSKAYA, E., AND K. SEIM (2006): “Bid Preference Programs and Participation in Highway Procurement Auctions,” Working paper, University of Pennsylvania.
- LEE, S., O. LINTON, AND Y. WHANG (2006): “Testing for Stochastic Monotonicity,” Working Paper, LSE.
- LEHMAN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Springer, New York.

- LEVIN, D., AND J. L. SMITH (1994): “Equilibrium in Auctions with Entry,” *The American Economic Review*, 84(3), 585–599.
- LI, T. (2005): “Econometrics of first-price auctions with entry and binding reservation prices,” *Journal of Econometrics*, 126(1), 173–200.
- LI, T., I. PERRIGNE, AND Q. VUONG (2002): “Structural Estimation of the Affiliated Private Value Auction Model,” *The RAND Journal of Economics*, 33(2), 171–193.
- LI, T., AND X. ZHENG (2005): “Procurement Auctions with Entry and an Uncertain Number of Actual Bidders: Theory, Structural Inference, and an Application,” Working Paper, Indiana University.
- MARMER, V., AND A. SHNEYEROV (2006): “Quantile-Based Nonparametric Inference for First-Price Auctions,” Working paper, UBC.
- MILGROM, P. (1981): “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics*, 12(2), 380–391.
- MILGROM, P. (2004): *Putting Auction Theory to Work*. Cambridge University Press.
- NEWHEY, W. K. (1994): “Kernel Estimation of Partial Means and a General Variance Estimator,” *Econometric Theory*, 10, 233–253.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*, Themes in Modern Econometrics. Cambridge University Press, New York.
- PERLMAN, M. D. (1969): “One-Sided Testing Problems in Multivariate Analysis,” *The Annals of Mathematical Statistics*, 40(2), 549–567.
- PESENDORFER, M., AND M. JOFRE-BONET (2003): “Estimation of a Dynamic Auction Game,” *Econometrica*, 71(5), 1443–1489.
- POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag, New York.
- RILEY, J., AND W. SAMUELSON (1981): “Optimal auctions,” *The American Economic Review*, 71, 58–73.
- SAMUELSON, W. (1985): “Competitive Bidding with Entry Costs,” *Economics Letters*, 17(1), 2.
- VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- YE, L. (2005): “Indicative bidding and a theory of two-stage auctions,” *Games and Economic Behavior*.

Table 1: Size of the GS test

nominal size	Quantiles		
	0.5	0.3, 0.5, 0.7	0.3, 0.4, 0.5, 0.6, 0.7
	rho=0		
0.10	0.0760	0.1520	0.1580
0.05	0.0310	0.0730	0.1030
0.01	0.0070	0.0200	0.0360
	rho=0.5		
0.10	0.0420	0.0660	0.0640
0.05	0.0240	0.0350	0.0310
0.01	0.0010	0.0130	0.0100

Table 2: Size-corrected power of the GS test

nominal size	Quantiles		
	0.5	0.3, 0.5, 0.7	0.3, 0.4, 0.5, 0.6, 0.7
rho=0.5			
0.10	0.1622	0.2800	0.2683
0.05	0.0941	0.2030	0.1812
0.01	0.0210	0.0630	0.0651
rho=0.9			
0.10	0.2693	0.3624	0.4124
0.05	0.1842	0.2853	0.2983
0.01	0.0661	0.1231	0.1552

Table 3: Description of Variables

Variable	Description	Mean	Std. Dev	Min	Max
EngEst	The engineer's estimate for the project, in mil. Dollars	3.647	4.488	0.066	24.800
Bid	Bid divided by the engineer's estimate	1.067	0.173	0.385	2.106
Nitems	Number of pay items in the project ad	71.736	70.704	1.000	363.000
Ndays	Number of business days to complete the project	195.995	142.370	10.000	681.000
Length	Length of the road (in miles)	4.699	4.794	0.000	36.63
Distance	Distance in miles from the headquarters of the firm of the bidding firm to the project site	344.237	382.469	0.000	1702.016
Backlog	The total amount of unfinished work on a given day and normalized by the bidder-specific maximum, the value is between 0 and 1	0.219	0.297	0.000	1.000
Npotential	number of planholders	8.299	4.244	2.000	26.000
Nactual	number of actual bidders	3.451	1.428	0.000	7.000
out-of-state	dummy =1 if the firm has headquarters outside the state of Oklahoma	0.136	0.343	0	1

Table 4: Logit and OLS regressions

Variable	Logit		OLS	
	Coefficient	s.e.	Coefficient	s.e.
Intercept	3.565	1.148	0.497	0.123
Log(EngEst)	-0.010	0.034	0.970	0.010
Npotential	-0.043	0.019	-0.008	0.002
Length	0.009	0.012	0.001	0.001
Ndays	0.000	0.001	0.000	0.000
Nitems	-2.700E-04	0.000	4.580E-04	0.000
Distance	0.000	0.000	0.000	0.000
Backlog	0.137	0.170	0.019	0.019
Out-of-state	-0.359	0.166	-0.034	0.020
Fringe firm	0.250	0.175	0.027	0.034
<u>Firm</u>				
APAC-OKLAHOMA, INC.	0.303	0.497	0.013	0.025
THE CUMMINS CONST. CO., INC.	0.266	0.385	0.005	0.023
HASKELL LEMON CONST. CO.	-0.206	0.351	0.028	0.023
BROCE CONSTRUCTION CO., INC.	-1.803	0.313	0.052	0.029
WESTERN PLAINS CONSTRUCTION COMPANY	-1.338	0.756	-0.011	0.031
BELLCO MATERIALS, INC.	-0.672	0.411	-0.086	0.277
OVERLAND CORPORATION	-0.726	0.399	0.057	0.029
GLOVER CONST. CO., INC.	-0.919	0.395	0.061	0.003
T & G CONSTRUCTION, INC.	-0.974	0.992	-0.036	0.028
TIGER INDUSTRIAL TRANS. SYS., INC.	-0.061	0.479	-0.004	0.029
HORIZON CONST. CO., INC.	-0.181	0.473	0.009	0.033
CORNELL CONST. CO., INC.	-0.481	0.450	0.007	0.030
SEWELL BROTHERS, INC	-0.075	0.472	-0.018	0.034
BECCO CONTRACTORS, INC.	-1.818	0.876	-0.068	0.033
EVANS & ASSOC. CONST. CO., INC.	-0.228	0.582	-0.044	0.036
SHERWOOD CONST. CO., INC.	-0.561	0.416	0.006	0.035
VANTAGE PAVING, INC.	-1.619	0.962	0.030	0.047
ALLEN CONTRACTING, INC.	0.245	0.493	0.014	0.035
DUIT CONSTRUCTION CO., INC.	1.275	0.707	0.027	0.037
MUSKOGEE BRIDGE CO., INC.	-1.122	0.809	0.006	0.037
Observations	4485		1860	
Log-Likelihood	-1543.750			
R2			0.983	

Notes: Significant coefficients (at 5% level) are marked in bold. For the OLS regression, the bids were normalized by the engineer's estimate.

Table 5: Estimated probability of bidding

Number of potential bidders	Estimate of $p(N)$	s.e.
2	0.8333	0.0761
3	0.7174	0.0383
4	0.6795	0.0374
5	0.5935	0.0395
6	0.5952	0.0339
7	0.4436	0.0431
8	0.3807	0.0366
9	0.4012	0.0385
10	0.4333	0.0522
11	0.3077	0.0386
12	0.3690	0.0527
13	0.3077	0.0484

Testing that probabilities are decreasing in N:

Test statistic: 0.867
P-value: 0.998

Table 6: Estimated quantiles

Number of potential bidders	Quantiles						
	0.25	0.3	0.4	0.5	0.6	0.7	0.75
2	1.1859	1.2372	1.2369	1.2394	1.2691	1.2975	1.3666
3	1.1450	1.1574	1.1934	1.2326	1.2915	1.3714	1.5044
4	1.0717	1.0803	1.1151	1.1911	1.2869	1.3680	1.4221
5	1.0946	1.0945	1.1232	1.1689	1.2266	1.3727	1.4332
6	1.0133	1.0355	1.0772	1.1404	1.1599	1.2076	1.2652
7	1.1212	1.1545	1.1608	1.1873	1.2706	1.2891	1.3010
8	1.0545	1.0869	1.1050	1.1281	1.3068	1.3614	1.3739
9	1.0432	1.0673	1.1492	1.1573	1.1979	1.4898	1.3549
10	1.0547	1.0609	1.0930	1.1529	1.1745	1.1951	1.2320
11	1.0222	1.0468	1.0865	1.0852	1.1119	1.4101	1.3014
12	1.0023	1.0240	1.0543	1.1214	1.1323	1.2508	1.3851
13	1.0219	1.0174	1.0397	1.0534	1.0699	1.0843	1.0885

Table 7: Standard errors of estimated quantiles

Number of potential bidders	Quantiles						
	0.25	0.3	0.4	0.5	0.6	0.7	0.75
2	0.1000	0.1140	0.0711	0.0502	0.0561	0.0639	0.0856
3	0.0381	0.0313	0.0284	0.0304	0.0330	0.0467	0.0806
4	0.0137	0.0135	0.0148	0.0253	0.0402	0.0495	0.0596
5	0.0288	0.0201	0.0187	0.0208	0.0273	0.0635	0.0696
6	0.0093	0.0101	0.0145	0.0179	0.0137	0.0184	0.0281
7	0.0344	0.0417	0.0230	0.0210	0.0389	0.0290	0.0305
8	0.0301	0.0332	0.0228	0.0252	0.0754	0.0582	0.0552
9	0.0389	0.0417	0.0507	0.0237	0.0259	0.1675	0.0440
10	0.0167	0.0164	0.0187	0.0313	0.0196	0.0185	0.0282
11	0.0383	0.0399	0.0345	0.0187	0.0215	0.1832	0.0693
12	0.0164	0.0173	0.0220	0.0248	0.0210	0.0752	0.1391
13	0.0401	0.0291	0.0206	0.0191	0.0128	0.0123	0.0128

Table 8: Estimated transformed quantiles

Number of potential bidders	Quantiles						
	0.25	0.3	0.4	0.5	0.6	0.7	0.75
2	1.1859	1.2372	1.2369	1.2394	1.2691	1.2975	1.3666
3	1.1450	1.1574	1.1934	1.2326	1.2915	1.3714	1.5044
4	1.0717	1.0803	1.1151	1.1911	1.2869	1.3680	1.4221
5	1.0946	1.0945	1.1232	1.1689	1.2266	1.3727	1.4332
6	1.0133	1.0355	1.0772	1.1404	1.1599	1.2076	1.2652
7	1.1212	1.1545	1.1608	1.1873	1.2706	1.2891	1.3010
8	1.0545	1.0869	1.1050	1.1281	1.3068	1.3614	1.3739
9	1.0432	1.0673	1.1492	1.1573	1.1979	1.4898	1.3549
10	1.0547	1.0609	1.0930	1.1529	1.1745	1.1951	1.2320
11	1.0222	1.0468	1.0865	1.0852	1.1119	1.4101	1.3014
12	1.0023	1.0240	1.0543	1.1214	1.1323	1.2508	1.3851
13	1.0219	1.0174	1.0397	1.0534	1.0699	1.0843	1.0885

Table 9: Standard errors of estimated quantiles

Number of potential bidders	Quantiles						
	0.25	0.3	0.4	0.5	0.6	0.7	0.75
2	0.1000	0.1140	0.0711	0.0502	0.0561	0.0639	0.0856
3	0.0381	0.0313	0.0284	0.0304	0.0330	0.0467	0.0806
4	0.0137	0.0135	0.0148	0.0253	0.0402	0.0495	0.0596
5	0.0288	0.0201	0.0187	0.0208	0.0273	0.0635	0.0696
6	0.0093	0.0101	0.0145	0.0179	0.0137	0.0184	0.0281
7	0.0344	0.0417	0.0230	0.0210	0.0389	0.0290	0.0305
8	0.0301	0.0332	0.0228	0.0252	0.0754	0.0582	0.0552
9	0.0389	0.0417	0.0507	0.0237	0.0259	0.1675	0.0440
10	0.0167	0.0164	0.0187	0.0313	0.0196	0.0185	0.0282
11	0.0383	0.0399	0.0345	0.0187	0.0215	0.1832	0.0693
12	0.0164	0.0173	0.0220	0.0248	0.0210	0.0752	0.1391
13	0.0401	0.0291	0.0206	0.0191	0.0128	0.0123	0.0128

Table 10: Test results

Quantiles	GS statistic	GS p-value	LS statistic	LS p-value	S statistic	S p-value
0.5	5.02	0.7798	54.84	0.0000	146.40	0.0000
0.3, 0.5, 0.7	30.97	0.2352	226.88	0.0000	583.93	0.0000
0.3, 0.4, 0.5, 0.6, 0.7	52.85	0.1652	372.21	0.0000	909.91	0.0000
0.25	13.91	0.1031	39.14	0.0001	320.38	0.0000
0.75	8.13	0.4805	162.34	0.0000	176.63	0.0000
0.25, 0.50, 0.75	27.06	0.3904	256.32	0.0000	643.40	0.0000

Figure 1: Histogram of number of potential bidders

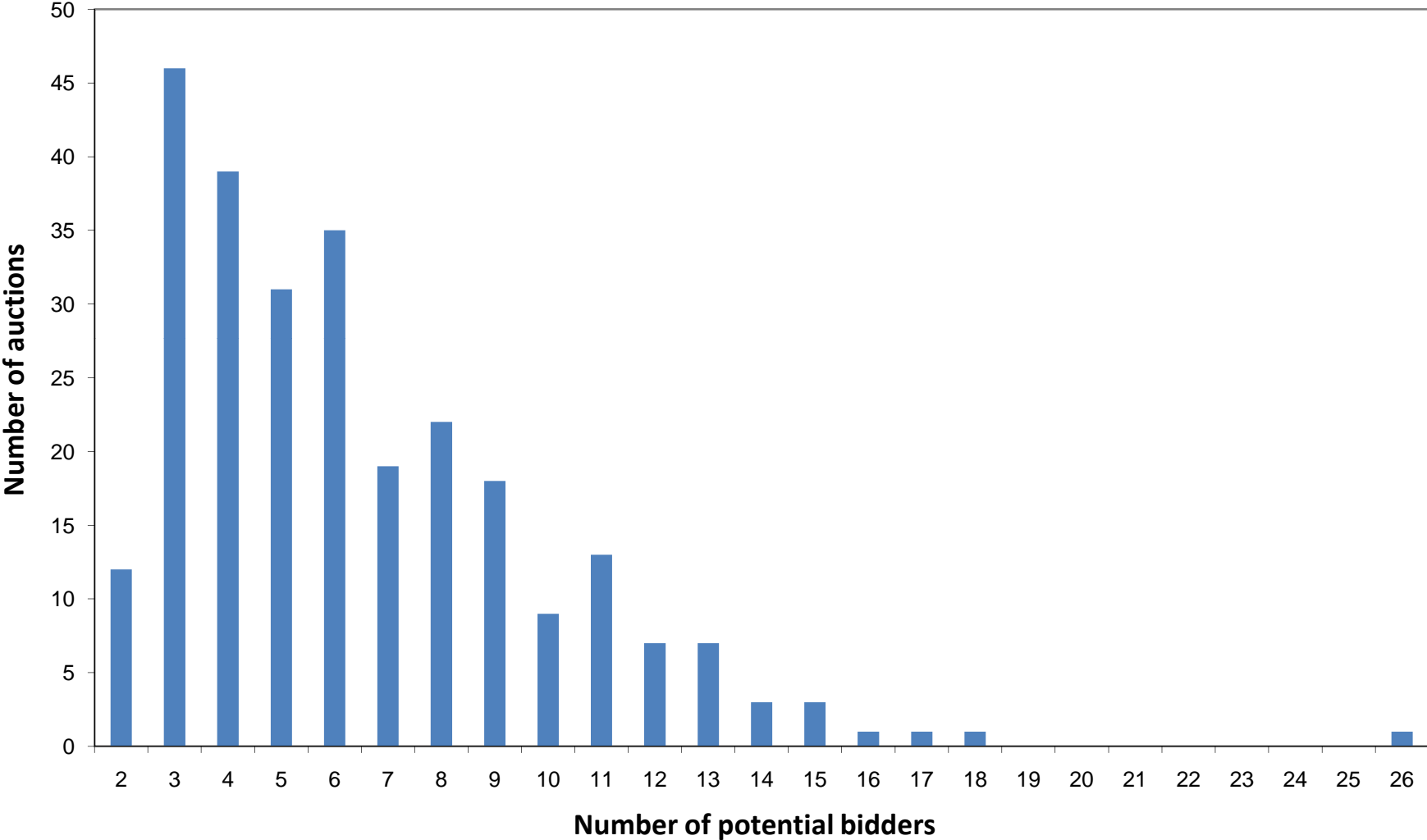
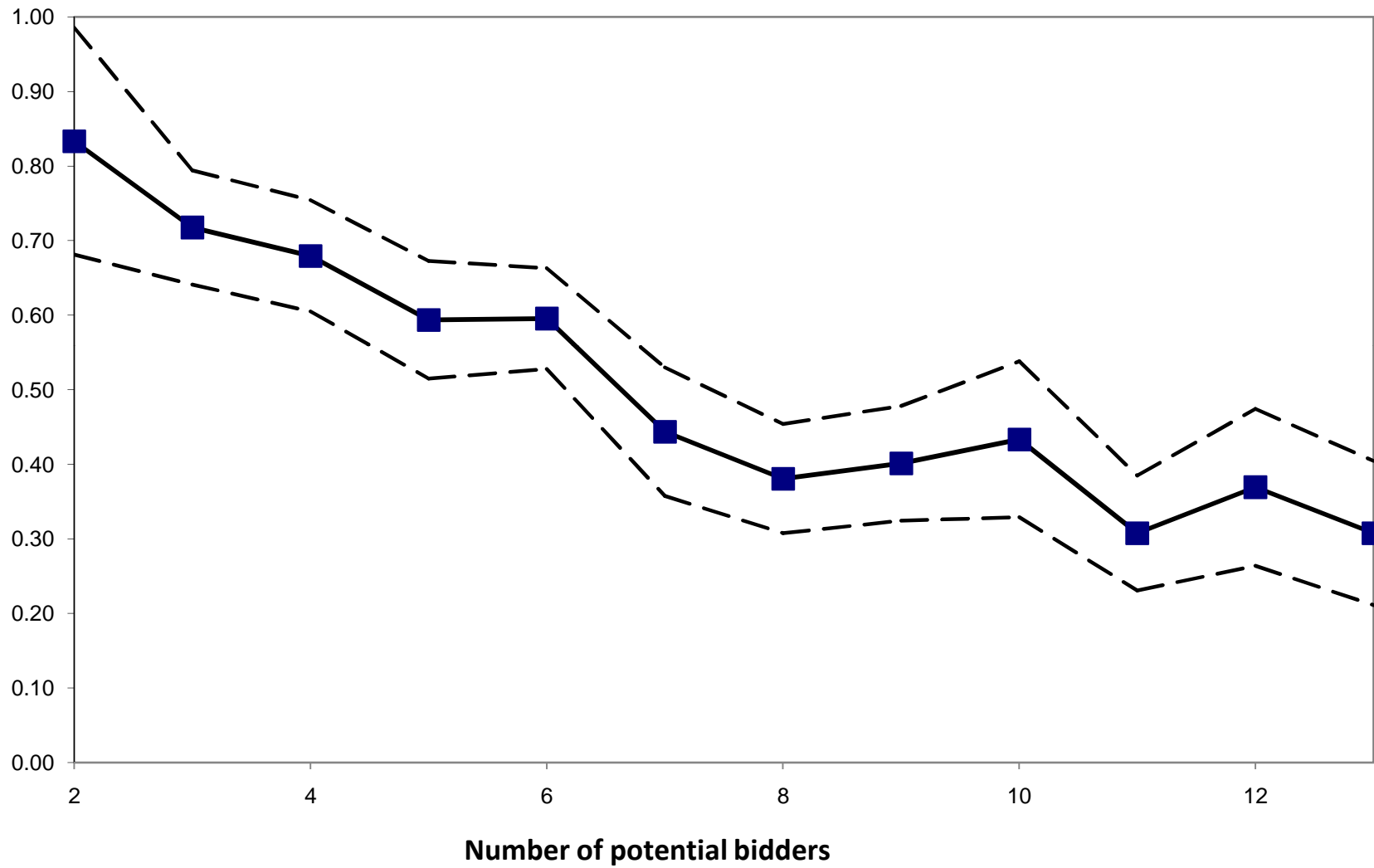
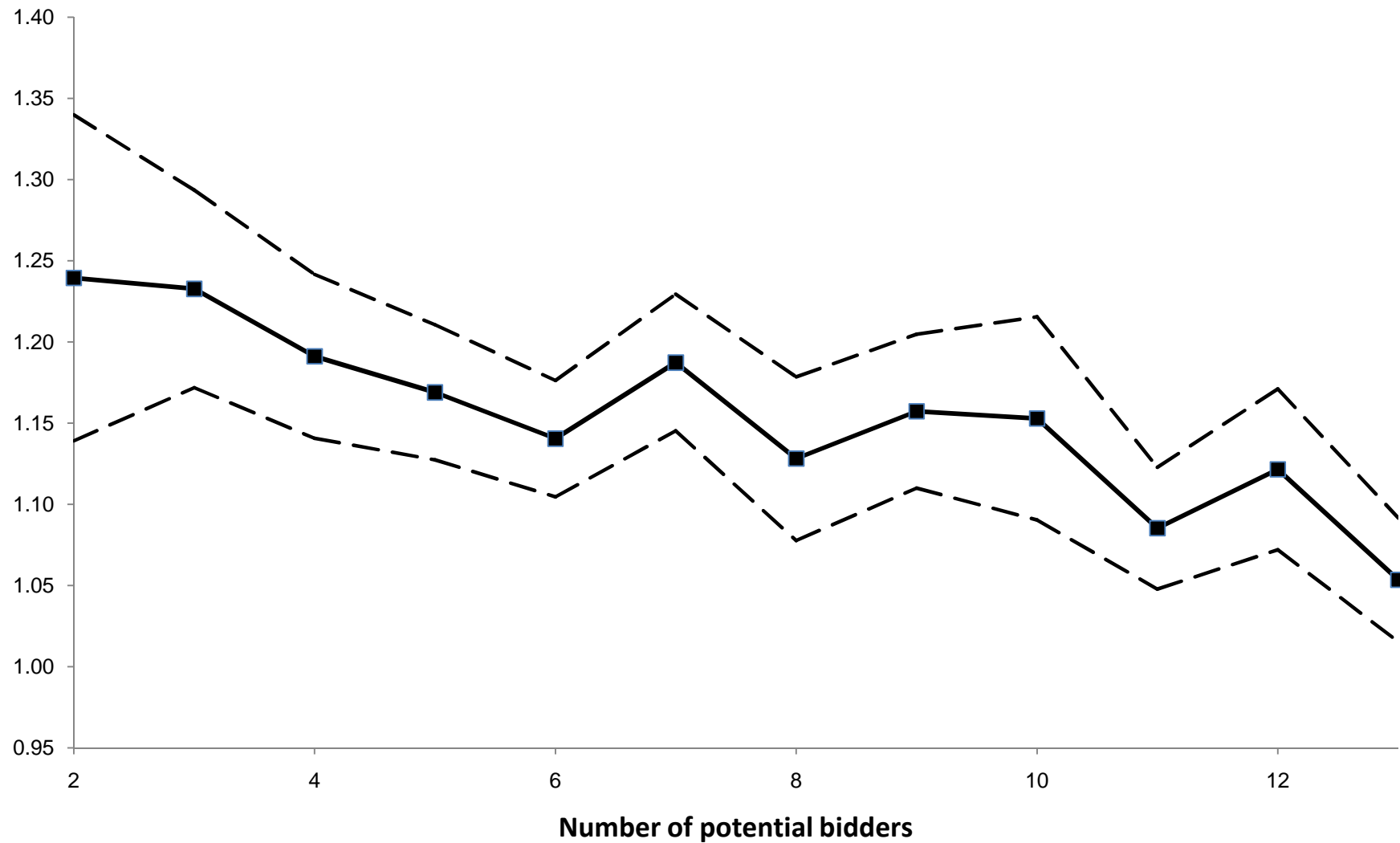


Figure 2: Estimated probability of bidding (with 95% confidence intervals)



**Figure 3: Median costs
(with 95% confidence intervals)**



**Figure 4: Transformed median costs
(with 95% confidence intervals)**

