

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

Vadim Marmer
University of British Columbia

Artyom Shneyerov
Concordia University, CIREQ, and CIRANO

Pai Xu
University of Hong Kong

April 3, 2012

Abstract

We develop a nonparametric approach that allows for discrimination among alternative models of entry in first-price auctions. Three models of entry are considered: those of Levin and Smith (1994), Samuelson (1985), and a new model in which the information received at the entry stage is imperfectly correlated with bidders' valuations. We derive testable restrictions of these models based on how the pro-competitive *selection effect* shifts the quantiles of valuations' distribution in response to an increase in the number of potential bidders.

JEL Classification: C12; C13; C14

Keywords: First-price auctions; models of entry; selective entry; selection effect; nonparametric estimation; quantiles

1 Introduction

Economists have long recognized the importance of selectivity. In several subfields of economics, selection effects are among the most important theoretical and practical issues. This paper introduces and studies selection effects in auctions with endogenous entry.

*We thank Herman Bierens, Ken Hendricks, Elena Krasnokutskaya, Tong Li, Isabelle Perrigne, Mike Peters, Joris Pinkse, Andrea Prat, Larry Samuelson, Quang Vuong, Ken Wolpin, the Associate Editor and three anonymous referees, as well as seminar participants at University of Calgary, Concordia University, University of Pennsylvania, Penn State University, Vanderbilt University, the CEA meeting in Montreal (2006) and the Summer Theory Workshop at UBC (2006) for useful comments. Vadim Marmer gratefully acknowledges the financial support of the SSHRC under grants 410-2007-1998 and 410-2010-1394. Artyom Shneyerov gratefully acknowledges the financial support of the SSHRC under grants 410-2003-1366 and 410-2006-2320.

Entry in auctions is an important topic in itself. It is clear that equilibrium allocations and efficiency are affected by who enters and why. At present, however, there is no consensus on what model of entry best describes what occurs in most markets. Recent empirical papers, recognizing that it is necessary to endogenize entry in modeling auctions, have offered structural models. Since one attractive feature of the auction literature is that nonparametric identification can be achieved under certain assumptions, it would be ideal if models of entry could similarly be distinguished nonparametrically. The main contribution of this paper is to offer a more general model of entry that nests those proposed in earlier theoretical work; and then to offer nonparametric tests of these models. To the best of our knowledge, this is the first paper to adopt a nonparametric approach first developed in Laffont and Vuong (1996) and Guerre, Perrigne, and Vuong (2000, GPV hereafter) to endogenous entry in first-price auctions.

We consider two models of entry proposed in the theoretical literature, Levin and Smith (1994, LS hereafter) and Samuelson (1985, S hereafter), as well as a new model, the *selective entry model* (SEM). Most of the empirical auction literature is based on the LS model. In that model, potential bidders are initially uninformed about their valuations of the good, but may become informed and then submit a bid at a cost. In equilibrium, potential entrants randomize their entry decisions and earn zero expected profit.

In the S 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 select sample biased towards bidders with higher valuations.

Both the 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, whereas the information is perfect in S. However, the S model also has what we call the *selection effect*, which has not been studied previously. We show that as the number of potential bidders increases, those who enter, tend to have larger valuations.

Our new SEM model also allows for selective entry, but dispenses with the stark assumption that potential bidders know their valuations *perfectly* at the entry stage, as in S, thus sharing with the LS model a costly valuation discovery stage. This model formally nests the S and LS models. At the entry stage, the potential bidders each observe a private signal correlated with their as yet unknown valuation of the good. Based on this private signal, a bidder may learn the valuation upon incurring entry cost k . A bidder who has entered will bid only if the valuation exceeds the reserve price. Although the signals may be informative about the valuations, unlike in the S model, they are not perfectly informative. Both the LS and S models can be viewed as limit cases of our model: the LS model corresponds to uninformative signals, whereas the S model corresponds to perfectly informative signals.

How can one test for the presence of the selection effect in first-price auctions with endogenous entry? We build on the insight in Haile, Hong, and Shum (2003) and propose the use of exogenous variation in the number of potential bidders, N , as the basis for such a

test.¹ Let $Q^*(\tau|N)$ be the τ -th quantile of *active* bidders' valuations conditional on N . We show that, in the SEM, the selection effect manifests itself as the effect of N on $Q^*(\tau|N)$: in the face of greater potential competition, some potential entrants, who may be less efficient in the auction, will choose not to enter, and accordingly, the quantiles of those who do enter increase: $Q^*(\tau|N') \geq Q^*(\tau|N)$ for $N' > N$. We show that the inequality is strict in the S model, while $Q^*(\tau|N)$ does not depend on N in the LS model, so that there is *no* selection effect. Following the approach of GPV, we show that the quantiles, $Q^*(\tau|N)$, can be nonparametrically identified in any SEM model if the number of potential bidders and all bids in each auction are observed. Our tests are based on an asymptotically normal estimator of $Q^*(\tau|N)$ that we develop by following an approach similar to Marmor and Shneyerov (2012).

Models that are in some respects similar to our SEM have been considered in the literature. For example, Hendricks, Pinkse, and Porter (2003) estimate a bidding model for off-shore oil. They sketch a model of entry that is in some respects similar to ours, however, with a common-value component. The focus of their paper, however, is not on entry but rather on testing an equilibrium model of bidding. This model is also outlined in the concluding section of Ye (2007).

A closely related paper is Li and Zheng (2009), who develop Bayesian semiparametric approach for the LS and S models.² They focus on estimation, and address neither nonparametric identification nor nonparametric testing of the entry models. In this regard, our results are complementary.

An important contribution of Li and Zheng (2009) is that they identify the *entry effect*. This effect exists in addition to the usual competition effect: as the number of potential bidders increases, the probability of entry falls, which may lead to a reduction in the auction price. Using data on lawn mowing contracts, they estimate a variant of the LS model, and find that, under certain conditions, it may be optimal to restrict the competition. In fact, their counterfactual experiment shows that, for a typical auction, it is not optimal to allow more than three bidders to participate.

The entry effect is counter-competitive and therefore operates in the direction opposite to the pro-competitive selection effect. Which of the two is dominant depends on the application. We illustrate our approach using the same data as in Li and Zheng (2009). We propose a minimum distance estimation method directly based on our testing framework, where the nonparametric quantile estimators are used as a checking device. For a certain subset of auctions, our nonparametric test favors the newly proposed SEM. We then find that, in this application, the selection effect can overturn the entry effect.

The selection effect identified in this paper has been subsequently studied in Einav and Esponda (2008), Roberts and Sweeting (2010a,b), and Coviello and Mariniello (2010). Einav and Esponda (2008) and Roberts and Sweeting (2010a) investigate the effects of

¹Haile, Hong, and Shum (2003) consider a different model in which bidders' valuations may have a common component.

²Li (2005) develops a general parametric approach for auctions with entry.

bidder asymmetries; the latter paper considers second-price and open outcry auctions. Roberts and Sweeting (2010b) study sequential sales mechanisms in an environment with selective entry. Coviello and Mariniello (2010) explore how information revelation interacts with selective entry in Italian public procurements.

Li and Gentry (2012) discuss partial identification of our selective entry model. In particular, they show that exogenous variation in N is insufficient for full identification. This stands in contrast with the findings in Guerre, Perrigne, and Vuong (2009), where such variation is shown to be sufficient for the identification of a first-price auction model with risk aversion. On the other hand, they show that SEM can be *fully* identifiable if there is a continuous instrument that affects the entry cost, but not the valuations of the bidders.

Endogenous entry in auctions has been a subject of a number of recent papers. Bajari and Hortacsu (2003), for example, have investigated entry and bidding in eBay auctions using a common value framework. They implemented a Bayesian estimation method employing a dataset of mint and proof sets of US coins. The magnitude of the entry cost is estimated, and expected seller revenues are simulated under different reserve prices. Athey, Levin, and Seira (2011) estimate a model of timber auctions with costly entry. The entry cost is assumed to be the private information of potential bidders, who are selected into an entrant pool based on their entry cost draws. Krasnokutskaya and Seim (2011) explore bid preference programs and bidder participation using California data. The latter paper also adopts the LS model, but in an asymmetric context in which small firms are treated preferentially. Bajari, Hong, and Ryan (2010) propose a parametric likelihood-based estimation strategy in the presence of multiple equilibria, and apply it to highway procurement auctions, using the LS model.

To summarize, our contribution is fivefold. First, we propose a more general model of entry and perform a detailed study of its identification. Second, we derive testable implications of the new model. Third, based on those implications, we propose a flexible nonparametric framework for testing different models of entry. Fourth, based on this framework, we propose a structural estimation method for the proposed selective entry model. Fifth, in an empirical application, we show that the selection effect is economically significant and should not be ignored.

The paper proceeds as follows. Section 2 presents the SEM and discusses the selection effect. Section 3 contains results on nonparametric identification. Our testing framework is presented in Section 4. Section 5 contains an empirical application, and Section 6 concludes. An online supplement to this paper provides additional technical details: the results regarding smoothness of the distributions of values and bids, the proofs of consistency and asymptotic normality of the estimators, and the proof of the validity of the bootstrap procedure used in this paper.

2 Model and selection effect

2.1 Model

In our model, $N \geq 2$ risk-neutral potential bidders in a first-price auction observe (at no cost) a signal that is correlated with the true value of the object, and make costly entry decisions based on this signals. The model nests as special cases the LS model and the S model, as noted in the introduction. The LS model arises when the signals and valuations are statistically independent. The S model arises when the signals are perfectly informative about valuations.

This model, which will be referred to as the selective entry model (SEM), is formally described as follows. The game begins with the entry stage at which nature draws signal and valuation tuples (V_i, S_i) independently from the same distribution with joint cumulative distribution function (CDF) $F(v, s)$. At the entry stage, potential bidders privately observe their S_i 's at no cost, but they do not observe V_i . The latter become observable to those potential bidders who have paid an entry cost $k \geq 0$. Thus, at the bidding stage, the model has *independent private values*. The bidders who have paid this cost are called *entrants*, and only they are eligible to bid. Moreover, only those with valuations at or above the reserve price r actually submit a bid. These bidders are called *active*.

The entry stage is followed by the bidding stage. Active bidders simultaneously and independently submit sealed bids. They do not know the number of active bidders, only the number of potential bidders, N . The good is awarded to the highest bidder.

We assume that the marginal distributions of V_i and S_i are the same, and the marginal CDF is denoted as $F(\cdot)$.³ We make a number of regularity assumptions on the joint distribution of V_i and S_i , $F(\cdot, \cdot)$:

Assumption 1. *We assume: (i) the joint distribution $F(\cdot, \cdot)$ has support $[\underline{v}, \bar{v}]^2$; (ii) the marginal distribution $F(\cdot)$ is absolutely continuous, with a continuous density $f(\cdot)$ bounded away from 0 on $[\underline{v}, \bar{v}]$; (iii) the conditional distribution $F(\cdot|s, V_i \geq r)$ exists and satisfies the non-degeneracy assumption $F(0|s, V_i \geq r) < 1$ for $s \in (\underline{v}, \bar{v}]$; (iv) the conditional distribution*

$$F(v|S_i \geq s) = \frac{\int_s^{\bar{v}} F(v|\tilde{s})dF(\tilde{s})}{1 - F(s)} \quad (1)$$

has a support $[\underline{v}(s), \bar{v}]$, where $\underline{v} \leq \underline{v}(s) < \bar{v}$ for $s \in (\underline{v}, \bar{v})$, and has a continuous density positive on the support.

We assume that signals are informative and that higher signals are “good news”. For our purposes, it will be sufficient to assume the following.

³This is a normalization and is without loss of generality. In all our models, any strictly increasing transformation of the signals results in the same distribution of entry decisions and bids, and therefore is observationally equivalent. We adopt this normalization because, in the S model, the signals are perfectly informative. To minimize on notation, we use F to denote the joint distribution of V_i and S_i , and its marginals.

Assumption 2 (“Good news”). *For any $r \leq \bar{v}$, the conditional distribution of valuations conditional on bidding, $F(\cdot|s, V_i \geq r)$, satisfies first-order stochastic dominance: for any $s_2 \geq s_1$, $F(\cdot|s_2, V_i \geq r) \leq F(\cdot|s_1, V_i \geq r)$.*

The lemma below shows that the stochastic dominance relation appearing in Assumption 2 can be strengthened to $F(\cdot|S_i \geq s, V_i \geq r)$, the distribution of values conditional on entry with a signal *above* s and bidding. As we focus on equilibria in which bidders with signals above a certain threshold enter, this result implies that increasing the threshold will stochastically increase active bidders’ values. This fact will be fundamental for our testable restrictions.

Lemma 1. *Under Assumption 2, (i) $F(\cdot|S_i \geq s, V_i \geq r) \leq F(\cdot|s, V_i \geq r)$, and (ii) $F(\cdot|S_i \geq s, V_i \geq r)$ is nonincreasing in s .*

Both the LS and S models are included as special cases of SEM. In the LS model, there is no selectivity, and $F(v|s) = F(v)$. In the S model, entry is selective, and

$$F(v|s) = \begin{cases} 1 & , \quad v \geq s \\ 0 & , \quad v < s \end{cases} , \quad F(v|S_i \geq s) = \begin{cases} \frac{F(v)-F(s)}{1-F(s)} & , \quad v \geq s \\ 0 & , \quad v < s \end{cases} . \quad (2)$$

In particular, both models satisfy the “good news” assumption, with strict inequality for the S model (in the relevant range of values and signals).

We restrict our attention to *symmetric* perfect Bayesian-Nash equilibria in which an equilibrium entry strategy can be characterized by a *cutoff* $\bar{s}(N)$ such that (i) a bidder weakly prefers to enter if $s > \bar{s}(N)$, and (ii) a bidder weakly prefers not to enter if $s < \bar{s}(N)$. In what follows, we will refer to such an equilibrium simply as *equilibrium*.

Fix an *arbitrary* cutoff $\tilde{s} \in [\underline{v}, \bar{v})$ and consider a first-price auction where bidders enter independently with probability $1 - F(\tilde{s})$, and draw their values independently from the distribution $F(v|S_i \geq \tilde{s})$. Let $B(v; \tilde{s}|N) : [\underline{v}(s), \bar{v}] \rightarrow \mathbb{R}_+$ be the corresponding equilibrium bidding strategy. Such an auction has been studied previously by Harstad, Kagel, and Levin (1990).⁴ Their results imply that the equilibrium bidding strategy is given by the (unique) solution to the differential equation

$$B'(v; \tilde{s}|N) = (v - B(v; \tilde{s}|N)) \frac{(N-1)(1-F(\tilde{s}))f(v|S_i \geq \tilde{s})}{F(\tilde{s}) + (1-F(\tilde{s}))F(v|S_i \geq \tilde{s})}, \quad (3)$$

subject to the boundary condition $B(\underline{v}(\tilde{s})|N) = r$.

In order to characterize the *equilibrium* cutoff $\bar{s}(N)$, consider $\Pi(s, \tilde{s}, N)$, the expected profit of a bidder with a signal s , who enters and bids “optimally” given his beliefs. In the Appendix, we prove the following crucial lemma.⁵

⁴See their Theorem 1.

⁵In this lemma and below, we use notation $x \vee y \equiv \max\{x, y\}$.

Lemma 2 (Monotonicity of Profit Function). *The expected profit is equal to*

$$\begin{aligned} \Pi(s, \tilde{s}, N) &= (\underline{v}(s) - r)F(\tilde{s})^{N-1} \vee 0 \\ &+ \int_{\underline{v}(s) \vee r}^{\bar{v}} (1 - F(v|s)) [F(\tilde{s}) + (1 - F(\tilde{s})) F(v|S_i \geq \tilde{s})]^{N-1} dv. \end{aligned} \quad (4)$$

The function $\Pi(s, \tilde{s}, N)$ is (i) nondecreasing in signal s ; (ii) nondecreasing in cutoff $\tilde{s} \in [\underline{v}, \bar{v}]$, and (iii) decreasing in N .

We now provide intuition for this lemma. To begin, note that the profits at the bidding stage, when the valuation v is known to the bidder, are increasing in v . Assuming a binding reserve price, $r > \underline{v}(s)$, our good news Assumption 2 implies that $\Pi(s, \tilde{s}, N)$ must be nondecreasing in s . (The proof also shows that this is the case when the reserve price is nonbinding). To gain intuition why $\Pi(s, \tilde{s}, N)$ is nondecreasing in \tilde{s} , consider $\lambda(v, \tilde{s}) = F(\tilde{s}) + F(v|S_i \geq \tilde{s})(1 - F(\tilde{s}))$, the probability that a bidder with value v wins over a rival who adopts cutoff \tilde{s} . This probability must be *nondecreasing* in \tilde{s} . Indeed, marginally increasing the cutoff \tilde{s} by $d\tilde{s}$ may only lead bidder i winning over some of the rival types v who originally had high v draws, but now simply do not enter. This leads to a weakly higher profit for bidder i . To understand why $\Pi(s, \tilde{s}, N)$ is decreasing in N , note that a bidder with value v wins the auction with probability $\lambda(v, \tilde{s})^{N-1}$. For $v < \bar{v}$, $\lambda(v, \tilde{s}) < 1$, and $\lambda(v, \tilde{s})^{N-1}$ is *decreasing* in N . The proof in the Appendix shows that the Envelope Theorem then implies that $\Pi(s, \tilde{s}, N)$ is then also decreasing in N .

We now introduce continuity and strict monotonicity assumptions that are likely to be satisfied in cases of practical interest.

Assumption 3 (Continuity). *The function $\Pi(s, s, N)$ is continuous in s .*

Assumption 4 (Strict Monotonicity). *The function $\Pi(s, s, N)$ is increasing in s .*

Under Assumption 3, the equilibrium cutoff under partial entry, i.e. whenever $\bar{s}(N) \in (\underline{v}, \bar{v})$, is characterized by an indifference property: given that all rivals adopt this cutoff, bidder i with a signal $S_i = \bar{s}(N)$ must be indifferent between entering or not entering. Then, the equilibrium cutoff $\bar{s}(N)$ is given by a solution to the equation $k = \Pi(\bar{s}(N), \bar{s}(N), N)$, and is unique if Assumption 4 holds.

Both S and LS models satisfy Assumptions 3 and 4. *These assumptions will be maintained from this point on.* The proposition below formally describes properties of an equilibrium.

Proposition 1 (SEM equilibrium). *There exists a unique symmetric perfect Bayesian equilibrium. The bidding strategy $B(v|N) \equiv B(v; \bar{s}(N)|N)$ is strictly increasing and continuously differentiable. The entry strategy is characterized by a cutoff $\bar{s}(N)$ that satisfies the following properties:*

(a) Large entry cost: *If $\Pi(\bar{v}, \bar{v}, N) < k$, there is no entry: $\bar{s}(N) = \bar{v}$.*

- (b) Small entry cost: If $k \leq \Pi(\underline{v}, \underline{v}, N)$, there is entry with probability 1, or “full” entry: $\bar{s}(N) = \underline{v}$.
- (c) Intermediate entry cost: If $\Pi(\underline{v}, \underline{v}, N) < k < \Pi(\bar{v}, \bar{v}, N)$ and continuity Assumption 3 holds, entry is “partial”, $\bar{s}(N) \in (\underline{v}, \bar{v})$, and the cutoff $\bar{s}(N)$ must satisfy the indifference condition

$$k = \Pi(\bar{s}(N), \bar{s}(N), N). \quad (5)$$

The cutoff is increasing in N when entry is partial: $\bar{s}(N) \in (\underline{v}, \bar{v})$ and $N' > N$ implies $\bar{s}(N') > \bar{s}(N)$.

The main result in the above proposition is the cutoff monotonicity result. Simply put, lower entrant profits, arising ceteris paribus from more competition, eliminate entry incentives of bidders with sufficiently low signals. This results in a higher cutoff $\bar{s}(N)$.

We define N_* as the minimal N such that the cutoff $\bar{s}(N)$ implies partial entry,

$$N_* \equiv \min\{N \geq 2 : \bar{s}(N) > \underline{v}\}$$

(we let $N_* = 2$ if the entry cost is large enough so that $\bar{s}(N) > \underline{v}$ for all $N \geq 2$). Consequently, the cutoff $\bar{s}(N)$ is increasing in N for $N \geq N_*$.

Remark 1. Proposition 1(b) implies that entry in the S model cannot be “full”, as (2) implies $\Pi(\underline{v}, \bar{v}) = 0$. Furthermore, the signal cutoff $\bar{s}(N) \in (\underline{v}, \bar{v})$ in the S model is determined as the unique solution of

$$(\bar{s}(N) - r) F(\bar{s}(N))^{N-1} = k \quad (6)$$

(the entrant with the lowest valuation, $\bar{s}(N)$, wins the auction if and only if all rivals do not enter, the event that occurs with probability $F(\bar{s}(N))^{N-1}$).

2.2 Selection effect

We now turn to the *selection effect*, which is the main subject of our paper. Denote the distribution of valuations conditional on bidding as

$$F^*(v|N) = F(v|S_i \geq \bar{s}(N), V_i \geq r). \quad (7)$$

Because the cutoff $\bar{s}(N)$ is nondecreasing, the result in Lemma 1(ii) ensures that for $N' > N$,

$$F^*(v|N') \leq F^*(v|N),$$

so that quantiles of $F^*(v|N)$, defined for $\tau \in (0, 1)$ as

$$Q^*(\tau|N) = F^{*-1}(\tau|N) \equiv \inf\{v : F^*(v|N) \geq \tau\},$$

weakly increase:

$$Q^*(\tau|N') \geq Q^*(\tau|N).$$

We define a *selection effect* $\Delta(\tau, N, N')$ as

$$\Delta(\tau, N, N') \equiv Q(\tau|N') - Q(\tau|N) \geq 0. \quad (8)$$

The inequality above is weak in general; for example, it becomes equality in the LS model, so that there is no selection effect. On the other hand, the S model features a positive selection effect for all $\tau \in (0, 1)$. A further question arises: what other SEM models also have a positive selection effect? A sufficient condition for a positive selection effect is given below.

Assumption 5. For all $s \in (\underline{v}, \bar{v})$, and v in the interior of the support of $F(v|s, V_i \geq r)$, $v \in (r \vee \underline{v}(s), \bar{v})$,

$$F(v|S_i \geq s, V_i \geq r) < F(v|s, V_i \geq r).$$

This assumption mirrors the condition in Lemma 1, but with a strict inequality, and is satisfied if, for example, the vector (V_i, S_i) has a strictly affiliated density $f(v, s) > 0$ on the support. However, it is weaker since it is also satisfied in the S model, where the joint density of values and signals does not exist.⁶

We then have the following proposition.

Proposition 2 (Selection effect). *Under Assumption 5, the SEM features a positive selection effect: for any $N' > N \geq N_*$ and $\tau \in (0, 1)$, $\Delta(\tau, N, N') > 0$.*

3 Nonparametric identification

3.1 Main identification result

We now show that the valuation quantiles $Q^*(\tau|N)$, which are necessary for the identification of the selection effect, are nonparametrically identifiable if the econometrician can observe the reserve price r and all bids in each auction. The number of potential bidders N is also assumed to be observable. In the SEM, the equilibrium probability of bidding is

$$p(N) = 1 - F(\bar{s}(N), V_i \geq r), \quad (9)$$

and the distribution of the number of active bidders is binomial with parameters N and $p(N)$. Both the probability of entry $p(N)$ and the distribution of active bidders' bids,

$$\begin{aligned} G^*(b|N) &= F(\xi(b|N) | S_i \geq \bar{s}(N), V_i \geq r) \\ &= F^*(\xi(b|N) | N), \end{aligned} \quad (10)$$

⁶We would have liked instead to impose a strict version of the stochastic dominance condition appearing in our “good news” Assumption 2. However, such a strict version would be violated in the S model, where the selection effect is obviously present.

are identifiable, where $\xi(\cdot|N) : [r, \bar{b}(N)] \rightarrow [v^*(N), \bar{v}]$ is the inverse bidding strategy, which is defined as the unique solution to the equation $B(\xi|N) = b$. (We use the notation $v^*(N)$ for the lower bound of the support of $F^*(v|N)$, i.e. $v^*(N) = r \vee \underline{v}(\bar{s}(N))$.) The probability density of $G^*(\cdot|N)$ is denoted as $g^*(\cdot|N)$. Our identification strategy follows GPV.

Proposition 3. *In both models, the inverse bidding strategy is identifiable for all $b \in [r, \bar{b}(N)]$ as*

$$\xi(b|N) = b + \frac{1}{N-1} \left(\frac{G^*(b|N)}{g^*(b|N)} + \frac{1-p(N)}{p(N)} \frac{1}{g^*(b|N)} \right), \quad (11)$$

and $B(v|N)$ is identifiable on $[v^*(N), \bar{v}]$. The quantiles of active bidders' valuations (and therefore the selection effect $\Delta(\tau, N', N)$) are also identifiable according to

$$Q^*(\tau|N) = \xi(q^*(\tau|N)|N),$$

where $q^*(\tau|N) \equiv G^{*-1}(b|N)$ are the quantiles of bids.

Remark 2. Proposition 3 implies that the distribution of active bidders' valuations is identifiable in both models as $F^*(v|N) = G^*(\xi^{-1}(v|N)|N)$, with support

$$[\xi(r|N), \xi(\bar{b}(N)|N)]. \quad (12)$$

The intuition for this result is as follows. Consider the 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 $1 - p(N) + p(N) G^*(b|N)$. As there are $N - 1$ identical rivals, it follows by independence that the probability of winning is $(1 - p(N) + p(N) G^*(b|N))^{N-1}$, and the expected profit is

$$(v - b) (1 - p(N) + p(N) G^*(b|N))^{N-1}.$$

Writing out the first-order condition, i.e. taking the derivative of the expected profit with respect to b and setting it equal to 0, gives the inverse bidding strategy (11).

Remark 3 (Identification using only winning bids). The assumption that all bids are observable is not necessary. If only the winning bid is observable, then the distribution $G^*(b|N)$ can be recovered from the distribution of the winning bid $G_Y^*(b|N)$ according to the formula

$$G_Y^*(b|N) = \frac{(1 - p(N) + p(N) G^*(b|N))^N - (1 - p(N))^N}{1 - (1 - p(N))^N}.$$

Therefore, the inverse bidding strategy $\xi(b|N)$ as well as the quantiles of active bidders valuations $Q^*(\tau|N)$ are identifiable as in Proposition 3.

The reserve price r is identifiable as the infimum of the bids in all our models, regardless of whether it is observable or not. (If there is no reserve price, the aforementioned infimum is 0). Formally, in our notation, $r = \lim_{b \downarrow 0} q^*(b|N)$.

Note that (11) is parallel to equation (25) in GPV, who consider the identification in a standard first-price auction model ($k = 0$) with a binding reserve price. In the standard model, a failure to bid is attributed solely to the truncating effect of the reserve price, and $p(N) = 1 - F(r)$. If $k > 0$, then there is an additional channel of costly entry, and $p(N)$ is given by (9). As $\bar{s}(N)$ is nondecreasing in N ,

$$p(N) \leq p(N'), \quad \text{for } N > N'.$$

This is another restriction that is shared by all models in this paper and will be tested empirically.

3.2 Additional identification results

The focus of our paper is on the identification of the selection effect using variation in N . However, in view of the identification results in GPV for the standard model ($k = 0$) obtained for N fixed, it is also worth investigating the question what is identified under this assumption in models with endogenous entry. The following proposition provides an answer to this question.

Proposition 4. *Given a number of potential bidders N , consider any absolutely continuous CDF $G(\cdot)$ with the support $[r, \bar{b}]$, a probability mass function $\rho(\cdot)$ with the support $\{0, 1, \dots, N\}$, and $p \in (0, 1)$.*

- (a) *There exists a SEM that rationalizes the data, i.e. such that the probability of bidding is p , the induced equilibrium distribution of bids is $G(\cdot)$ and the distribution of the number of active bidders n is $\rho(\cdot)$, if and only if: (i) $\rho(\cdot)$ is Binomial with parameters (N, p) , (ii) the observed bids are i.i.d. according to $G(\cdot)$, and (iii) the function $\xi(\cdot|N)$ defined in (11) with $G^*(\cdot|N) = G(\cdot)$ and $p(N) = p$ is strictly increasing on $[r, \bar{b}]$ and its inverse is differentiable on $[\xi(r|N), \xi(\bar{b}|N)]$.*

Furthermore, when conditions (i)-(iii) in (a) hold,

- (b) *Any SEM is observationally equivalent to the standard model ($k = 0$); therefore $F(v) = F^*(v|N)$ is identifiable for $v \geq r$, while the entry cost is not identifiable.*
- (c) *An S model rationalizes the data if and only if $\xi(r|N) > r$. The distribution $F(v)$ is identifiable according to*

$$F(v) = p(N) F^*(v|N) + 1 - p(N)$$

on $[\xi(r|N), \xi(\bar{b}(N)|N)]$, and k is identifiable according to (6):

$$k = (\xi(r|N) - r) (1 - p(N))^{N-1}.$$

Remark 4 (Entry Cost Bias). It is clear that model misspecification will lead to biased estimation of the entry cost. The direction of the bias can also be determined. Suppose that k is identifiable and is estimated according to LS. However, the data are generated according to the SEM with selective entry (Assumption 5 holds). Then k is estimated using the wrong (LS) factor $1 - F^*(v)$ instead of the correct (SEM) factor $1 - F(v|\bar{s}(N))$ in (4) and (5). Since selective entry implies $1 - F^*(v) > 1 - F(v|\bar{s}(N))$, the true k is smaller than the estimated one. The intuition for the presence of this bias is straightforward. The entry cost in the LS model is equal to the equilibrium expected profit of the average potential bidder, while it is equal to the expected profit of the marginal potential bidder (with signal $\bar{s}(N)$) in the SEM. If there is a positive selection effect, as implied by the selective entry condition in Assumption 5, the marginal bidder has a smaller expected profit than the average bidder, which leads to over estimation of the entry cost.

The following proposition shows that the variation in N enhances the identification of the S and LS models. Let $\mathcal{N} = [\underline{N}, \bar{N}]$, where $2 \leq \underline{N} < \bar{N}$, denote the support of the distribution of the number of potential bidders. In the S model, the set of v for which $F(v)$ is identified is $[\bar{s}(\underline{N}), \bar{v}] \supset [\bar{s}(N), \bar{v}]$ for all $N > \underline{N}$. In the LS model, $F(v)$ is still identified if and only if $v \in [r, \bar{v}]$. However, now the entry cost is identifiable if \mathcal{N} has at least three elements and $N_* < \bar{N}$. This is because $N_* < \bar{N}$ and the cutoff monotonicity property in Proposition 1 implies

$$p(\underline{N}) = \dots = p(N_*) > \dots > p(\bar{N}). \quad (13)$$

When N belongs to the flat segment, $N \leq N_*$, we are certain that bidders enter with probability one and non-participation is due to the truncating effect of the reserve price only, and therefore are able to identify $F(r) = 1 - p(\underline{N})$. When N belongs to the decreasing segment, $N > N_*$, we are certain that bidders are indifferent between entering or not, and are able to identify the entry cost from the indifference condition given the knowledge of $F(r)$.⁷

Proposition 5. *In the S model,*

(a) $F(v)$ is identifiable for $v \geq \xi(r|\underline{N})$.

(b) k is identifiable according to (6), $k = (\xi(r|N) - r)(1 - p(N))^{N-1}$.

In the LS model,

(c) $F(v)$ is identifiable for $v \geq r$.

(d) N_* is identifiable as

$$N_* = \min \{N \in \mathcal{N} : p(N) > p(N + 1)\}. \quad (14)$$

⁷Xu (2012) develops a nonparametric estimator of the entry cost in S model.

(e) If \mathcal{N} has at least three elements and $N_* \geq \underline{N} + 1$, k is identifiable according to

$$k = (1 - F(r)) \int_r^{\bar{v}} (1 - F^*(v))(1 - p(N) + p(N)F^*(v))^{N-1} dv. \quad (15)$$

Remark 5. Even if \mathcal{N} has at least three elements, the variation in N is insufficient for identification of the entry cost in the LS model when $N_* = \underline{N}$. In this case, $p(N) < p(N')$ for all $N, N' \in \mathcal{N}$, and (15) holds for all $N \in \mathcal{N}$. Then only the ratio $k/(1 - F(r))$ is identifiable according to (15), but k and $F(r)$ are not separately identifiable. Also, the result in (d) above applies to the SEM.

4 Econometric implementation

In what follows, we allow for observable auction heterogeneity by introducing a vector of auction-specific covariates $x \in \mathcal{X} \subset \mathbb{R}^d$. We now assume that the distribution of (v, s) can change from auction to auction depending on the covariate value x , and is denoted by $F(v, s|x)$. Other objects are also indexed by x in a similar notation. The reserve price $r(x)$ is assumed to be binding. The SEM's differential equation (3) for the bidding strategy takes the following form:

$$B'(v|N, x) = (v - B(v|N, x)) \times \frac{(N-1)(1 - F(\bar{s}(N, x)|x))f(v|S_i \geq \bar{s}(N, x), x)}{F(\bar{s}(N, x)|x) + (1 - F(\bar{s}(N, x)|x))F(v|S_i \geq \bar{s}(N, x), x)}, \quad (16)$$

with the boundary condition $B(r(x)|N, x) = r(x)$, whereas the counterpart of equation (5) for the equilibrium cutoff $\bar{s}(N, x)$ is

$$\int_{r(x)}^{\bar{v}(x)} (1 - F(v|\bar{s}(N, x), x)) \times (F(\bar{s}(N, x)|x) + (1 - F(\bar{s}(N, x)|x))F(v|S_i \geq \bar{s}(N, x), x))^{N-1} dv = k(x), \quad (17)$$

where $k(x)$ denotes the entry cost function.

4.1 Data generating process (DGP)

We assume that a sample of L auctions is available, and index the auctions by $l = 1, \dots, L$. Each auction is characterized by the vector of covariates $x_l \in \mathcal{X}$. We make the following formal assumptions concerning the DGP.

Assumption 6. (a) $\{(N_l, x_l) : l = 1, \dots, L\}$ are *i.i.d.*

- (b) Let $\varphi(x)$ denote the marginal PDF of x_l . We assume that $\varphi(\cdot) > 0$ on its compact support $\mathcal{X} \subset \mathbb{R}^d$ and admits up to $R \geq 2$ continuous partial derivatives on the interior of \mathcal{X} .
- (c) Let $\pi(N|x)$ denote the distribution of N_l conditional on x_l . We assume that $\pi(\cdot|x)$ has support $\mathcal{N} = \{\underline{N}, \dots, \bar{N}\}$ for all $x \in \mathcal{X}$, $\underline{N} \geq 2$, and $\pi(N|\cdot)$ admits R continuous bounded derivatives on the interior of \mathcal{X} for all $N \in \mathcal{N}$.
- (d) (V_{il}, S_{il}) and N_l are independent conditional on x_l .
- (e) $\{(V_{il}, S_{il}) : i = 1, \dots, N_l; l = 1, \dots, L\}$ are i.i.d. conditional on (N_l, x_l) .
- (f) In an SEM other than the S model, the density $f(v, s|x)$ is strictly positive and bounded away from zero on its support, a compact interval $[\underline{v}(x), \bar{v}(x)]^2 \subset \mathbb{R}_+^2$ for all $x \in \mathcal{X}$; $f(\cdot, \cdot|x)$ admits up to $R - 1$ continuous partial derivatives on $[\underline{v}(x), \bar{v}(x)]^2$ for all $x \in \mathcal{X}$, and $f(v, s|\cdot)$ admits up to R continuous partial derivatives on the interior of \mathcal{X} for all $v, s \in [\underline{v}(x), \bar{v}(x)]^2$. In the S model, the density $f(v|x)$ is strictly positive and bounded away from zero on its support, a compact interval $[\underline{v}(x), \bar{v}(x)] \subset \mathbb{R}_+$ for all $x \in \mathcal{X}$; $f(\cdot|x)$ admits up to $R - 1$ continuous derivatives on $[\underline{v}(x), \bar{v}(x)]$ for all $x \in \mathcal{X}$, $f(v|\cdot)$ admits up to R continuous partial derivatives on the interior of \mathcal{X} for all $v \in [\underline{v}(x), \bar{v}(x)]$.
- (g) The reserve price $r(x)$ is binding for all $x \in \mathcal{X}$, $N \in \mathcal{N}$, i.e. $\underline{v}(x) < r(x)$, and $r(\cdot)$ admits up to R continuous partial derivatives on the interior of \mathcal{X} .
- (h) The entry cost $k(x)$ admits up to R continuous partial derivatives on the interior of \mathcal{X} .

Central to our approach is the assumption (d): conditional on x_l , the number of potential bidders N_l is *exogenous*. This assumption allows us to use the variation in the number of potential bidders for testing purposes. In Section 5, we explain why this assumption is plausible in the context of our empirical application.

Assumption 6(a) is the usual i.i.d. assumption on the DGP for the covariates. The smoothness conditions in Assumptions 6(b), (c), (f), (g) and (h) are standard in the non-parametric literature and used for deriving the asymptotic properties of the kernel estimators. Assumption 6(c) also defines the support of the distribution of N_l conditional on the covariates. Assumption 6(e) is the IPV assumption.

The bid b_{il} corresponding to the valuation V_{il} is generated according to the bidding strategy

$$b_{il} = B(V_{il}|N_l, x_l),$$

where $B(\cdot|N, x)$ is a solution to (16). Decisions to submit a bid, $y_{il} \in \{0, 1\}$, are generated according to the cutoff strategy

$$y_{il} = 1 \text{ if } S_{il} \geq \bar{s}(N_l, x_l) \text{ and } V_{il} \geq r_l.$$

The bidding strategy B and the cutoff function \bar{s} depend on the model's primitives f and k through the equilibrium conditions of each model; neither B nor \bar{s} is available in closed form.

4.2 Hypotheses

Our main hypotheses are based on the identification of the selection effect in the active bidders' valuation quantiles. Recall that, for $\tau \in (0, 1)$, $Q^*(\tau|N, x)$ denotes the conditional quantile function of the distribution of active bidders' valuations. Conditional on $x_l = x$, the restrictions of the SEM and LS model can be stated respectively as:

$$\begin{aligned} H_{SEM} : Q^*(\tau|\underline{N}, x) &\leq \dots \leq Q^*(\tau|\bar{N}, x), \\ H_{LS} : Q^*(\tau|\underline{N}, x) &= \dots = Q^*(\tau|\bar{N}, x). \end{aligned} \quad (18)$$

In the S model, $F^*(v|N, x)$ is derived by truncation from the common "parent" distribution $F(v|x)$:

$$\begin{aligned} F^*(v|N, x) &= \frac{F(v|x) - F(\bar{s}(N, x)|x)}{1 - F(\bar{s}(N, x))} \\ &= \frac{F(v|x) - (1 - p(N, x))}{p(N, x)}. \end{aligned} \quad (19)$$

Because the distribution F does not depend on N , this leads to the restriction

$$p(N, x) F^*(v|N, x) + 1 - p(N, x) = p(N', x) F^*(v|N', x) + 1 - p(N', x). \quad (20)$$

To express restriction (20) in terms of quantiles, we define

$$\beta(\tau, N, x) = 1 - \frac{p(\bar{N}, x)}{p(N, x)} (1 - \tau). \quad (21)$$

Since $p(\bar{N}, x) \leq p(N, x)$, it follows that $0 \leq \beta(\tau, N, x) \leq 1$ for all $\tau \in [0, 1]$, and $\beta(\tau, N, x)$ can therefore be interpreted as a legitimate transformation of the quantile order τ .⁸ One can easily show that in the S model, condition (20) implies the following restriction in terms of the transformed quantiles:

$$H_S : Q^*(\beta(\tau, \underline{N}, x) | \underline{N}, x) = \dots = Q^*(\beta(\tau, \bar{N}, x) | \bar{N}, x). \quad (22)$$

In this paper, we consider the independent testing of H_{SEM} , H_{LS} , and H_S against their corresponding unrestricted alternatives.⁹ In addition, we also test whether the entry

⁸Although any other fixed value of N can be used in 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.

⁹Our testing framework is designed to capture the selection effect, and the resulting tests may not be consistent against all conceivable alternative models.

probabilities $p(N, x)$ are nonincreasing in N , as implied by all the models. In this case, for a given value of x , the null hypothesis is

$$H_p : 1 > p(\underline{N}, x) \geq \dots \geq p(\bar{N}, x) > 0. \quad (23)$$

Hypothesis H_p is also tested against its corresponding unrestricted alternative.

The fact that the equilibrium probabilities of submitting a bid decline with the number of potential bidders is probably a common feature of many other models of entry. Whenever a model with costly entry is used to explain why some potential bidders do not bid, an alternative explanation must be confronted. Following Paarsch (1997), even with no entry cost, non-participation can still be explained by the fact that some bidders draw valuations below the reserve price. In that case, however, the probability of bidding is equal to probability $P(V_i \leq r|x)$ and therefore does not depend on the number of potential bidders under Assumption 6(e). Thus, the null hypothesis of costless entry ($k = 0$) can be formulated as

$$H_{k=0} : p(\underline{N}, x) = \dots = p(\bar{N}, x). \quad (24)$$

Note that because $k = 0$ is formally included as a special case of LS, H_{LS} is also a testable restriction of the standard model.¹⁰

Our model implies that the distribution of the number of active bidders conditional on (N, x) is binomial:

$$\rho(n|N, x) = \binom{N}{n} p(N, x)^n (1 - p(N, x))^{N-n}. \quad (25)$$

Unobserved heterogeneity or the presence of multiple equilibria (see below) may result in a more complicated DGP that violates this restriction.¹¹ It is therefore desirable to have a nonparametric test of the binomial restriction. The restriction is testable because the conditional probability $\rho(n|N, x)$ is directly identifiable, and its test is proposed in Section 4.4.

Remark 6 (Multiple Equilibria). As can be seen from the proof of Proposition 1, in the SEM, there are also asymmetric equilibria, where a subset of $\min\{N, N_*\}$ bidders enters with probability one.^{12,13} These equilibria also induce a binomial distribution of the number

¹⁰This result also follows from equation (6) in Guerre, Perrigne, and Vuong (2009), who consider identification in a more general model with risk-averse bidders.

¹¹Krasnokutskaya (2011) explores identification and estimation in first-price auctions under unobserved heterogeneity. In a recent working paper, Roberts (2009) proposes an alternative approach. He shows that, to the extent that the reserve price is optimally chosen by the sellers, conditioning on the reserve controls for unobserved heterogeneity.

¹²Since a subset of firms enters with probability 1, there is no selective entry in these equilibria.

¹³The equilibria of this form do not exist in the S model because the bidders with $v = \underline{v}$ make zero expected profit from bidding, and are thus unable to recover the entry cost. It is unknown under which conditions multiple equilibria exist in the S model (Tan and Yilankaya (2006) obtain some results for second-price auctions).

of active bidders, conditional on (N, x) : for $n = 0, 1, \dots, \min\{N, N_*\}$,

$$\tilde{\rho}(n|N, x) = \binom{\min\{N, N_*\}}{n} (1 - F(r|x))^n F(r|x)^{\min\{N, N_*\} - n}.$$

However, this binomial distribution is different from that corresponding to the symmetric equilibrium in (25), whenever there is partial entry for some $N \in \mathcal{N}$. This is because partial entry means that $N_* < \bar{N}$, and therefore $\min\{N, N_*\} < N$ and $1 - F(r|x) > p(N, x)$, which implies $\tilde{\rho}(n|N, x) \neq \rho(n|N, x)$. If such asymmetric equilibria are present in the data with probability $\alpha(N, x) > 0$ for some $N > N_*$, then the conditional distribution of the number of active bidders is a *mixture* of two different binomial distributions $\rho(n|N, x)$ and $\tilde{\rho}(n|N, x)$:

$$\alpha(N, x) \tilde{\rho}(n|N, x) + [1 - \alpha(N, x)] \rho(n|N, x),$$

and therefore itself is *not* binomial. Our binomial test can thus be used to uncover the presence of such asymmetric equilibria in the data.

4.3 Nonparametric estimation of quantiles

In this section, we present our nonparametric estimation method for $Q^*(\tau|N, x)$. This method is based on the fact that, because the bidding strategies are increasing, the quantiles of valuations $Q^*(\tau|N, x)$ and bids $q^*(\tau|N, x)$ for active bidders, where

$$q^*(\tau|N, x) = G^{*-1}(\tau|N, x),$$

are linked through the (inverse) bidding strategy $Q^*(\tau|N, x) = \xi(q^*(\tau|N, x)|N, x)$. Because both $\xi(b|N, x)$ and $q^*(\tau|N, x)$ can be estimated nonparametrically, we consider a natural plug-in estimator:

$$\hat{Q}^*(\tau|N, x) = \hat{\xi}(\hat{q}^*(\tau|N, x)|N, x), \quad (26)$$

where $\hat{\xi}(b|N, x)$ and $\hat{q}^*(\tau|N, x)$ denote the nonparametric estimators of $\xi(b|N, x)$ and $q^*(\tau|N, x)$ respectively, which we describe below.

Recalling that the inverse bidding strategy $\xi(b|N, x)$ is given by

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

(see Proposition 3), our estimator $\hat{\xi}(b|N, x)$ is obtained by replacing $p(N, x)$, $G^*(b|N, x)$, and $g^*(b|N, x)$ with their corresponding nonparametric estimators, $\hat{p}(N, x)$, $\hat{G}^*(b|N, x)$, and $\hat{g}^*(b|N, x)$. The conditional quantile $q^*(\tau|N, x)$ is estimated by inverting the nonparametric estimator for the bids' CDF, $\hat{G}^*(b|N, x)$:

$$\hat{q}^*(\tau|N, x) = \hat{G}^{*-1}(\tau|N, x) = \inf \left\{ b : \hat{G}^*(b|N, x) \geq \tau \right\}.$$

Our nonparametric estimators for $g^*(b|N, x)$, $G^*(b|N, x)$, and $p(N, x)$ are based on the kernel method. Let K denote a kernel functions satisfying the following assumption.

Assumption 7. The kernel K has at least $R \geq 2$ 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 conditional distribution of the number of potential bidders and the conditional probability of bidding (given N and x) are estimated as

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

where h is the bandwidth parameter, and $n_l = \sum_{i=1}^{N_l} y_{il}$ is the number of active bidders in auction l . As the probability of observing N conditional on $x_l = x$ and the probability of submitting a bid conditional on $N_l = N$ and $x = x_l$ can be written as $\pi(N|x) = E[1 \{N_l = N\} | x]$ and $p(N, x) = E[n_l | N, x] / N$ respectively, their estimators are standard nonparametric regression estimators.

In the supplement, we show that the estimator $\hat{p}(N, x)$ is asymptotically normal and derive its asymptotic variance:

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

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 the interior of \mathcal{X} .

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

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

where $\hat{\varphi}(x)$ is the usual multivariate kernel density estimator of the PDF φ :¹⁴

$$\hat{\varphi}(x) = (Lh^d)^{-1} \sum_{l=1}^L \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right).$$

¹⁴The estimators \hat{g}^* and \hat{G}^* are standard kernel estimators of the conditional density and conditional CDF, see Pagan and Ullah (1999), equation (2.127) on page 58.

The transformed quantiles $Q^*(\beta(\tau, N, x) | N, x)$, which are needed for the tests of the S model, can be estimated by $\hat{Q}^*(\hat{\beta}(\tau, N, x) | N, x)$, where $\hat{\beta}(\tau, N, x)$ is a plug-in estimator of $\beta(\tau, N, x)$ in (21):

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

In the supplement, we prove that the estimators $\hat{Q}^*(\tau | N, x)$ and $\hat{Q}^*(\hat{\beta}(\tau, N, x) | N, x)$ are consistent and asymptotically normal. More specifically, we prove that, under certain technical but standard assumptions, $\sqrt{Lh^{d+1}}(\hat{Q}^*(\tau | N, x) - Q^*(\tau | N, x))$ is asymptotically normal with mean zero and the variance given by

$$V_Q(N, \tau, x) = \left(\int K(u)^2 du \right)^{d+1} \frac{(1 - p(N, x)(1 - \tau))^2}{(N - 1)^2 N p^3(N, x) g^{*3}(q^*(\tau | N, x) | N, x) \pi(N | x) \varphi(x)}.$$

Further, the estimation of $\beta(\tau, N, x)$ has no effect on the asymptotic distribution of the transformed quantiles estimator $\hat{Q}^*(\hat{\beta}(\tau, N, x) | N, x)$, since, as we also show in the supplement, $\hat{\beta}(\tau, N, x)$ converges at a faster rate than $\hat{Q}^*(\cdot | N, x)$. A consistent estimator $\hat{V}_Q(N, \tau, x)$ can be obtained by replacing $p(N, x)$, $q^*(\tau | N, x)$, and the other unknown functions by their estimators. Moreover, quantiles' estimators are independent across the values of $\tau \in (0, 1)$, $N \in \mathcal{N}$, and x in the interior of \mathcal{X} .

Remark 7 (Circumventing the curse of dimensionality). Although nonparametric methods are attractive in principle, they may be unsuitable for applications with small samples and a high-dimensional set of covariates. Indeed, most auction papers having made nonparametric identification arguments, almost always follow the approach of Haile, Hong, and Shum (2003) and apply them to “homogenized” bids, thus finessing the covariate problem. Unfortunately, a standard trick of assuming linearity in the covariates, estimating a first-stage regression model to get rid of the covariates, and then using “homogenized” bids in the second stage does not work in our model with endogenous entry. However, following the suggestion in Paarsch and Hong (2006, Chapter 3.3), a variant of the quantile single index model can be applied to reduce the curse of dimensionality in our setting. This approach is explained further in the supplement.

4.4 Tests

In this section, we propose nonparametric tests based on the models' restrictions. (Refer to Table 1 for a summary of these tests.) Recall the definition of the selection effect (8) from Section 2.2. Conditional on $x_l = x$, the estimated selection effect is

$$\hat{\Delta}(\tau, N, N', x) = \hat{Q}^*(\tau | N', x) - \hat{Q}^*(\tau | N, x).$$

We propose to test the affiliated model using the following statistic

$$T^{SEM}(x) = \sup_{\tau \in \Upsilon} T^{SEM}(\tau, x),$$

Table 1: Test statistics

Hypothesis	$\hat{\Delta}$	Test statistic
SEM	$\hat{Q}^*(\tau N', x) - \hat{Q}^*(\tau N, x)$	$\sqrt{Lh^{d+1}} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{[\hat{\Delta}(\tau, N, N', x)]_-}{\hat{\sigma}(\tau, N, N', x)}$
LS	$\hat{Q}^*(\tau N', x) - \hat{Q}^*(\tau N, x)$	$\sqrt{Lh^{d+1}} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{ \hat{\Delta}(\tau, N, N', x) }{\hat{\sigma}(\tau, N, N', x)}$
S	$\hat{Q}^*(\hat{\beta}(\tau, N', x) N', x) - \hat{Q}^*(\hat{\beta}(\tau, N, x) N, x)$	$\sqrt{Lh^{d+1}} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{ \hat{\Delta}(\tau, N, N', x) }{\hat{\sigma}(\tau, N, N', x)}$
Monotone Prob. of Entry	$\hat{p}(N', x) - \hat{p}(N, x)$	$\sqrt{Lh^d} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{[\hat{\Delta}_p(N, N', x)]_+}{\hat{\sigma}_p(N, N', x)}$
Costless Entry	$\hat{p}(N', x) - \hat{p}(N, x)$	$\sqrt{Lh^d} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{ \hat{\Delta}_p(N, N', x) }{\hat{\sigma}_p(N, N', x)}$
Binomial	$\hat{\rho}(n N, x) - \binom{N}{n} \hat{p}(N, x)^n (1 - \hat{p}(N, x))^{N-n}$	$\sqrt{Lh^d} \sum_{N=\underline{N}}^{\bar{N}} \sum_{n=0}^N \frac{ \hat{\Delta}_{Binomial}(n, N, x) }{\hat{\sigma}_{Binomial}(n, N, x)}$

where $\Upsilon = \{\tau_1, \dots, \tau_k\} \subset (0, 1)$ is a finite set of quantile values, and $T^{SEM}(\tau, x)$ is defined as

$$T^{SEM}(\tau, x) = \sqrt{Lh^{d+1}} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{[\hat{\Delta}(\tau, N, N', x)]_-}{\hat{\sigma}(\tau, N, N', x)}.$$

In the above expression,

$$\hat{\sigma}^2(\tau, N, N', x) = \hat{V}_Q(b, N, x) + \hat{V}_Q(b, N', x) \quad (27)$$

is the estimator of the asymptotic variance of $\hat{\Delta}(\tau, N, N', x)$, and the function $[\cdot]_-$ is defined as

$$[u]_- = \begin{cases} -u, & u < 0, \\ 0, & u \geq 0. \end{cases}$$

Note that under the hypothesis H_{SEM} , $Q^*(\tau|N', x) - Q^*(\tau|N, x) \geq 0$ for $N' \geq N$, and through the function $[\cdot]_-$, the statistic $T^{SEM}(\tau, x)$ captures only pairwise differences between the sample quantiles corresponding to different N 's that violate the model's restriction. We rescale each term by its standard error to give more weight to differences that are estimated more precisely.

The hypothesis H_{SEM} is rejected for large values of $T^{SEM}(x)$. Thus, the asymptotic size α test is

$$\text{Reject } H_{SEM} \text{ when } T^{SEM}(x) > c_{L,1-\alpha}^{SEM}(x), \quad (28)$$

where $c_{L,1-\alpha}^{SEM}(x)$ is an appropriately chosen critical value. We suggest computing the critical values $c_{L,1-\alpha}^{SEM}(x)$ using the bootstrap procedure described below.

To generate bootstrap samples, we first draw randomly with replacement, L auctions from the original sample of auctions $\{(N_l, x_l) : l = 1, \dots, L\}$. In the second step, we draw bootstrap bids randomly with replacement from the bid data corresponding to each selected auction. Thus, if auction \bar{l} is selected in the first step, then in the second step, we draw $N_{\bar{l}}$ values from the set $\{b_{i\bar{l}} : i = 1, \dots, N_{\bar{l}}\}$. Here we redefine the $b_{i\bar{l}}$'s so that they take values in the augmented set $\{\mathbb{R}_{++}, \text{"no bid"}\}$. Let $\{(b_{1l}^\dagger, \dots, b_{N_{\bar{l}}^\dagger l}^\dagger, N_l^\dagger, x_l^\dagger) : l = 1, \dots, L\}$ be a bootstrap sample, and M be the number of bootstrap samples. Then, in each bootstrap sample $m = 1, \dots, M$, compute $T_m^{\dagger, SEM}(x) = \sup_{\tau \in \Upsilon} T_m^{\dagger, SEM}(\tau, x)$, where

$$T_m^{\dagger, SEM}(\tau, x) = \sqrt{Lh^{d+1}} \sum_{N=\underline{N}}^{\bar{N}-1} \sum_{N'=N+1}^{\bar{N}} \frac{[\hat{\Delta}_m^\dagger(\tau, N, N', x) - \hat{\Delta}(\tau, N, N', x)]_-}{\hat{\sigma}_m^\dagger(\tau, N, N', x)}.$$

Here, $\hat{\Delta}_m^\dagger(\tau, N, N', x)$ is the bootstrap analogue of the selection effect $\hat{\Delta}(\tau, N, N', x)$:

$$\hat{\Delta}_m^\dagger(\tau, N, N', x) = \hat{Q}_m^{*,\dagger}(\tau|N', x) - \hat{Q}_m^{*,\dagger}(\tau|N, x),$$

$\hat{Q}_m^{*\dagger}(\tau|N, x)$ is the bootstrap analogue of estimator $\hat{Q}^*(\tau|N, x)$ for the bootstrap sample m , and $\hat{\sigma}_m^\dagger(\tau, N, N', x)$ is the bootstrap analogue of $\hat{\sigma}(\tau, N, N', x)$ computed using (27) in bootstrap sample m . To compute the required statistics, we also define the bootstrap variable y_{il}^\dagger , where $y_{il}^\dagger = 0$ if b_{il}^\dagger takes the value “no bid”, and $y_{il}^\dagger = 1$ otherwise. The critical value $c_{L,1-\alpha}^{SEM}(x)$ is the $(1 - \alpha)$ -th sample quantile of $\{T_m^{\dagger, SEM} : m = 1, \dots, M\}$.

A similar approach is used to test the restrictions of the LS and S models and those of the standard model (no costly entry). In the case of the LS model, the statistic is given by $T^{LS}(x) = \sup_{\tau \in \Upsilon} T^{LS}(\tau, x)$, where $T^{LS}(\tau, x)$ is constructed similarly to $T^{SEM}(\tau, x)$, however, with the function $[\cdot]_-$ replaced by the absolute value $|\cdot|$. This is because the LS model’s restrictions are equalities, and we need to capture the deviations from H_{LS} in either direction. For the S model, we define the quantile difference by using the transformed quantiles,

$$\hat{\Delta}(\tau, N, N', x) = \hat{Q}^*(\hat{\beta}(\tau, N', x)|N', x) - \hat{Q}^*(\hat{\beta}(\tau, N, x)|N, x),$$

and then proceed exactly as before.

The hypothesis H_p in (23) can be tested using the statistic

$$\begin{aligned} T_p(x) &= \sqrt{Lh^d} \sum_{N=\underline{N}}^{\bar{N}} \sum_{N'=\underline{N}}^{\bar{N}} \frac{[\hat{\Delta}_p(N, N', x)]_+}{\hat{\sigma}_p(N, N', x)}, \text{ where} \\ \hat{\Delta}_p(N, N', x) &= \hat{p}(N', x) - \hat{p}(N, x), \\ [u]_+ &= \begin{cases} 0, & u < 0, \\ u, & u \geq 0. \end{cases} \end{aligned}$$

Here, $\hat{\sigma}_p^2$ is the estimator of the asymptotic variance of $\hat{\Delta}_p$. The null hypothesis of costless entry in (24) can be tested using a statistic constructed similarly to $T_p(x)$, however, with the function $[\cdot]_+$ replaced by the absolute value $|\cdot|$.

To test the binomial restriction in (25), let

$$\hat{\Delta}_{Binomial}(n, N, x) = \hat{\rho}(n|N, x) - \binom{N}{n} \hat{p}(N, x)^n (1 - \hat{p}(N, x))^{N-n}$$

be the deviation of the directly estimated probability of entry from the corresponding binomial probability, where

$$\hat{\rho}(n|N, x) = \frac{\sum_{l=1}^L \mathbf{1}\{n_l = n\} \mathbf{1}\{N_l = N\} \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right)}{\sum_{l=1}^L \mathbf{1}\{N_l = N\} \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right)} \quad (29)$$

is a consistent nonparametric estimator of $\hat{\rho}(n|N, x)$. To test the binomial restriction, we employ the following statistic:

$$T_{Binomial}(x) = \sqrt{Lh^d} \sum_{N=\underline{N}}^{\bar{N}} \sum_{n=0}^N \frac{|\hat{\Delta}_{Binomial}(n, N, x)|}{\hat{\sigma}_{Binomial}(n, N, x)},$$

where $\hat{\sigma}_{Binomial}^2(n, N, x)$ is the plug-in estimator of the asymptotic variance of the difference $\hat{\Delta}_{Binomial}(n, N, x)$:

$$\sigma_{Binomial}^2(n, N, x) = V_\rho(n, N, x) + D^2(n, N, x) V_p(N, x) - 2D(n, N, x) \sigma_{\rho, p}(n, N, x),$$

where V_ρ and V_p are the asymptotic variances of $\hat{\rho}$ and \hat{p} respectively:

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

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

$D(n, N, x)$ is the derivative of $\binom{N}{n} p(N, x)^n (1 - p(N, x))^{N-n}$ with respect to $p(N, x)$, and $\sigma_{\rho, p}(n, N, x)$ is the asymptotic covariance between $\hat{\rho}$ and \hat{p} :

$$\sigma_{\rho, p}(n, N, x) = \frac{E \{ (1(n_l = n) - \rho(n|N, x)) (n_l/N - p(N, x)) | N, x \} \left(\int K^2(u) du \right)^d}{N\pi(N|x) \varphi(x)}.$$

In the above expression, the term $E \{ (1(n_l = n) - \rho(n|N, x)) (n_l/N - p(N, x)) | N, x \}$ can be estimated as

$$\frac{\sum_{l=1}^L (1(n_l = n) - \hat{\rho}(n|N, x)) (n_l/N - \hat{p}(N, x)) 1\{N_l = N\} \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right)}{\sum_{l=1}^L 1\{N_l = N\} \prod_{k=1}^d K\left(\frac{x_{kl} - x_k}{h}\right)}.$$

As described above, we test each hypothesis for a given value x of the covariates. The reason for this is that a null hypothesis can be true for one value of x and false for another. If, however, one wishes to test a hypothesis over a range of values of x , then our approach can be modified as follows. In the case of H_{SEM} , for example, one can select a grid of values of x from the interval of interest and compute $T^{SEM}(x)$ for each value in the grid. $T^{SEM}(x)$ can then be averaged over the chosen grid of values. The critical values for the test can be simulated using the averaged over the grid bootstrap statistic $T_m^{\dagger, SEM}(x)$. Alternatively, one can use the maximum of $T^{SEM}(x)$ over the chosen grid for x . In this case, the critical values should be simulated using the maximum of $T_m^{\dagger, SEM}(x)$ over the same grid.

We establish the asymptotic validity of the bootstrap tests in the supplement. Note that to ensure a valid bootstrap approximation, the bootstrap statistics must use re-centered differences. Thus, the statistic $T^{SEM}(\tau, x)$ is based on $\hat{\Delta}(\tau, N, N', x)$, and its bootstrap counterpart $T_m^{\dagger, SEM}(\tau, x)$ is based on the re-centered differences $\hat{\Delta}_m^{\dagger}(\tau, N, N', x) - \hat{\Delta}(\tau, N, N', x)$. This is to ensure that the bootstrap critical values are simulated from the null distribution.

A Monte Carlo study of our test is contained in Appendix B. We find that our proposed tests are close in size to the nominal levels, and also have good power properties.

5 Empirical application

5.1 Auctions for highway maintenance jobs

We use the same data as those in Li and Zheng (2009, LZ hereafter) in our empirical application. This dataset consists of 540 auctions for “mowing highway right-of-way” maintenance jobs held by the Texas Department of Transportation (TDoT) from January 2001 to December 2003. We observe all of the bids and the identities of the bidders. Moreover, the available data contain information on the project characteristics, include the engineer’s estimates, the length of the contract (number of working days), the number of items in the proposal, the acreage of full-width mowing, the acreage of other mowing, whether it is a state job, and whether the job is on an interstate highway.¹⁵ LZ provide further details of this dataset. We focus only on those aspects relevant for this study.

Importantly, we observe the list of eligible bidders (planholders) for each auction. In the nomenclature of this paper, these eligible bidders constitute the potential bidders. A firm becomes a planholder through the following process. All projects to be auctioned are advertised by the TDoT three to six weeks prior to the auction date. These advertisements include the engineer’s estimate, a brief summary of the project, and the location and type of work involved. However, they lack detailed schedules of the work items that are revealed only in the construction plans. Interested firms then submit a request for plans and bidding proposals, and these documents contain the project specifics (such as the item schedule). The list of planholders is made available to all bidders prior to the bid submission deadline. All bids must be submitted by the deadline in a sealed envelope. Submitted bids and active bidders’ identities are released only after bid opening. In this application, we do not observe the reserve price.

Our testing framework shares a number of assumptions with that in LZ, the first of which is the exogeneity of the number of potential bidders. They note that “only (and usually all of) those contractors in the mowing sub-industry who are located in the same county as where the job is or nearby counties would request the official bidding proposals”.¹⁶ Thus, the exogeneity of N is supported by the local nature of participating firms and the variation in the number of local firms across counties.¹⁷

Next, we also assume that the firms are ex-ante symmetric. LZ argue that the local participating firms are likely to be small and not much differentiated from one another. However, dynamic considerations such as those in Jofre-Bonet and Pesendorfer (2003) may lead to asymmetries even within such firms. To address this issue, LZ construct the backlog variable as the amount of incomplete work, measured in dollars, that remains from the previously won projects. This backlog variable contains substantial variation, but has sta-

¹⁵Refer to Table 1 in LZ for the basic summary statistics of the data.

¹⁶See the first paragraph on page 6 of LZ.

¹⁷The assumption of an exogenous number of potential bidders has also been employed in other empirical work on auctions. See, for example, Athey, Levin, and Seira (2011).

tistically insignificant effects on both the bidding and bid-submission decisions (see Tables 3 and 4 in LZ).

Before turning to our nonparametric tests, we first discuss the importance of the effects of various auction characteristics on bid levels and the decisions to submit a bid found in the random-effect and probit regressions in LZ.¹⁸ In the former regression, the dependent variable is $\log(\text{bid})$, where bid is the amount of the bid in dollars. The size of the project has a strongly positive effect on the bids, and clearly, is the most important variable. The impact of the other variables is much smaller. The effect of project size on the probability of submitting a bid is also positive, although not statistically significant. This positive effect may reflect the fact that bidders have some information when deciding whether to submit a bid, which further implies the possibility of selective entry.

The next potentially important variable is the number of items in the construction plan, which captures the complexity of the project. This variable is statistically significant in both regressions, although its degree of variation is very small. The standard deviation is 0.82, and the mean is 2.01. This is not surprising, as lawn mowing jobs are relatively simple. The estimates for “acreage of other mowing” are also statistically significant in both regressions. However, the effects are small: a 1% increase in the “acreage of other mowing” is associated with only a 0.0075% decrease in the bids.

In the implementation of our testing approach, we are confronted with the usual bias and variance trade-off of our nonparametric estimators. Including a greater number of variables reduces the bias, but at the same time increases the sample variability of the estimators. The reduced-form regressions in LZ indicate that, apart from project size, the auction characteristics are unlikely to matter much. In fact, their structural estimates suggest that the project size is the only significant factor.

5.2 Tests

We perform our tests conditional on the median project size, $x = \$138,000$.¹⁹ We consider a grid Υ of 200 uniformly spaced quantile values with $\tau_{\min} = 0.05$ and $\tau_{\max} = 0.95$, and use the tri-weight kernels. The bandwidth chosen is $h = 1.978 \times 1.06 \times (\text{std.err.}) \times (\text{sample size})^{-(1/5)}$. A well-understood practical issue is that nonparametric estimators suffer from a substantial loss of precision when the sample size is very small. When we tried to include all of the auctions, however, the estimates of the quantiles $Q^*(\tau|N, x)$ were highly erratic. Because the data are sparse, for some N , the estimated probabilities $\hat{\pi}(N|x)$ are very close to zero.²⁰ To ensure the stability of our estimators, we decided to exclude those N where the number of observations with $X_i \in [x - h, x + h]$ is less than 100. The working sample

¹⁸Refer to Tables 3 and 4 in LZ for details.

¹⁹We believe our choice of median project size is natural and reasonable. The models are not expected to vary with x in this application.

²⁰The estimated probability of N , conditional on x , is presented in Figure 3. There are no auctions with 19 or 21 potential bidders conditional on $x = 138,000$. Note also that there remains considerable variation in N .

Table 2: Test results: Hypotheses, test statistics, and critical values corresponding to asymptotic significance levels 0.10, 0.05, and 0.01

Hypothesis	Test statistic	Critical values		
		0.10	0.05	0.01
Costless Entry	14.50	13.56	14.92	17.05
Monotone Prob. of Entry	2.86	9.00	10.40	13.35
Binomial	327.27	391.53	470.76	759.94
SEM model	8.40	37.09	41.45	49.99
LS model	64.87	51.72	55.09	63.01
S model	60.22	67.82	71.23	78.90

ultimately consisted of auctions with 8-14 potential bidders, and all of the results discussed in the following were obtained using this smaller sample.

The results of our tests are reported in Table 2. We first test the standard model ($k = 0$), i.e. the equality of bid submission probabilities across N . The null hypothesis is rejected at the 10% significance level. We then test the prediction shared by all remaining models considered in this paper. As can be seen from Figure 1, the bid submission probabilities $p(N, x)$ are apparently nonincreasing in N . The formal test of the monotonicity restriction on the bidding probabilities fails to reject the null at any reasonable significance level.²¹ The DGP passes the binomial test, thereby offering no evidence of multiple equilibria or unobserved auction heterogeneity in the data.

The quantile test fails to reject the SEM restriction $Q^*(\tau|N, x) \leq Q^*(\tau|N', x)$ for $N \geq N'$ uniformly over $\tau \in \Upsilon$.²² We now turn to tests of the two extreme models: LS and S. We first note that the LS model is rejected even at a 1% significance level. We therefore reach a different conclusion from that in LZ, who find a better fit for the LS model. The S model passes all of our tests. However, based on the p -values of the S and SEM tests, the SEM has more empirical support in this application.

²¹It is worth noting that the effect of the number of potential bidders is statistically significant in the probit regression in LZ. Having more potential bidders reduces the bid submission rate. For example, increasing N by 1 reduces the odds of submitting a bid by about 4%.

²²Note that here, the order of N and N' is reversed because, in our application, we are dealing with low-bid auctions.

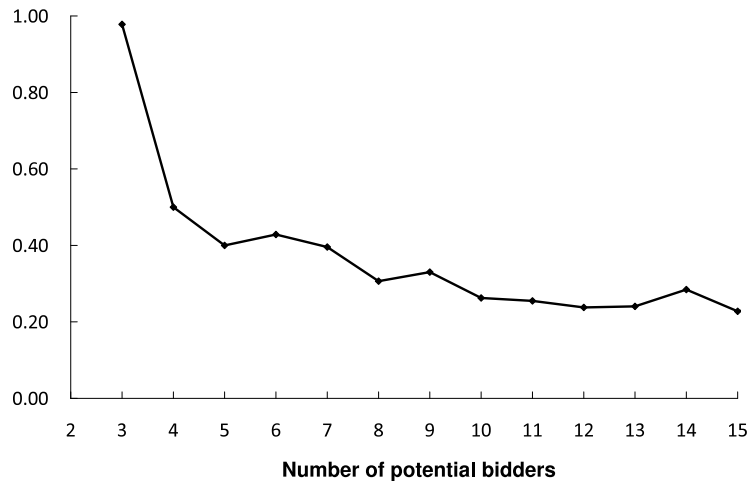


Figure 1: Estimated probabilities of bidding

5.3 Should the TDoT restrict competition?

An interesting empirical exercise is to fit an affiliated model to our data and then to compare its predictions to those in LZ, who investigate, in a counterfactual experiment, the effect of the number of potential bidders on the procurement price. In the LS model they estimated, the entry effect overwhelms the direct effect of potential competition. Indeed, for a typical project in the sample, they found it optimal to limit the competition to only $N = 3$ potential bidders. An interesting question is whether the selection effect is strong enough to overturn this conclusion.

To facilitate the counterfactual calculations, we rely on a parametric model specification in the spirit of Bajari (2001) and Hong and Shum (2002). We, however, use nonparametric estimates of quantiles for matching the model to the data.²³ Suppose that the conditional distribution of valuations can be described by a parametric family of PDFs $\{f(v, s|x; \theta) : \theta \in \Theta \subset \mathbb{R}^d\}$, and let $Q^*(\tau|N, x; \theta)$ be the implied quantile function of valuations conditional on entry. We follow Bajari, Hong, and Ryan (2010) and normalize the costs and signals by the engineer's estimate. The joint distribution of costs and signals is specified to (truncated) normal:

$$\frac{c_{il}}{x_l} = \mu_0 + \varepsilon_{il}, \quad \frac{s_{il}}{x_l} = \mu_0 + \eta_{il},$$

²³We thank the Associate Editor for suggesting such an approach.

where

$$(\varepsilon_{il}, \eta_{il}) \sim N \left(\mathbf{0}, \sigma_0^2 \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \right).$$

Since the reserve price r is unobservable, we treat it as an unknown parameter.²⁴ The vector of parameters to be estimated is thus given by

$$\theta_0 = (\mu_0, \sigma_0^2, \rho_0, k_0, r_0),$$

where k_0 and r_0 are the normalized entry cost and reserve price respectively.

To estimate the parameters, we adopt a minimum distance (MD) estimation approach (Newey and McFadden, 1994), where we match our nonparametric estimator $\hat{Q}^*(\tau|N, x)$ developed in Section 4.3 with the parametric conditional quantile function $Q^*(\tau|N, x; \theta)$ for a grid of pre-chosen quantile values $\tau_j, j = 1, \dots, k$:

$$\hat{\theta}(x) = \arg \min_{\theta \in \Theta} J_L(\theta, x), \quad \text{where}$$

$$J_L(\theta, x) = \sum_{j=1}^k \sum_{N=\underline{N}}^{\bar{N}} \frac{(\hat{Q}^*(\tau_j|N, x) - Q^*(\tau_j|N, x; \theta))^2}{N^{-1} \hat{V}_Q(N, \tau_j, x)}.$$

Note that $\hat{Q}^*(\tau_j|N, x)$ are asymptotically independent across j 's and N 's, and the reciprocals of their asymptotic variances $N^{-1} \hat{V}_Q(N, \tau_j, x)$ are the efficient weights in this case. By a usual formula, the variance-covariance matrix of $\hat{\theta}(x)$ can be computed as:

$$\frac{1}{Lh^{d+1}} \left[\sum_{j=1}^k \sum_{N=\underline{N}}^{\bar{N}} \frac{1}{N^{-1} \hat{V}_Q(N, \tau_j, x)} \frac{\partial Q^*(\tau_j|N, x; \hat{\theta}(x))}{\partial \theta} \frac{\partial Q^*(\tau_j|N, x; \hat{\theta}(x))}{\partial \theta'} \right]^{-1}.$$

The parametric model is overidentified if the number of N 's multiplied by k exceeds the number of parameters (five). In this case, the specification can be tested with an overidentifying restrictions test. As usual, the hypothesis of correct specification should be rejected when $Lh^{d+1} J_L(\hat{\theta}(x), x)$ exceeds a χ^2 critical value for the number of degrees of freedom equal to the degree of overidentification (the number of N 's multiplied by k minus five).

The estimation results are given in Table 3. We match 19 equally spaced quantile values $\{0.05, 0.1, \dots, 0.95\}$ for $N = 8, \dots, 14$ (i.e. we chose the same values for N as those in the previous section). As the estimate $\hat{\rho} = 0.80$ indicates, the signals contain a substantial amount of information concerning the construction cost. This is realistic because, as we have remarked previously, lawn-mowing jobs are relatively small and simple in nature, meaning that the plans are likely to be quite informative about the true project cost. The

²⁴Alternatively, this feature can be taken into account by considering a model with a secret reserve price as in Li and Perrigne (2003). However, this complicates significantly the derivation of the equilibrium bidding strategy.

Table 3: Estimates of SEM structural parameters

Parameter	Estimate	Standard error
Mean of normalized cost, μ	0.991*	0.162
Standard deviation, σ	0.233*	0.838
Correlation, ρ	0.805*	0.302
Normalized reserve price, r	1.220*	0.091
Normalized entry cost, k	0.045*	0.012
MD objective	124.48	

* Significant at 5%

estimate of the entry cost $\hat{k} = 4.5\%$ is given as a percentage of the project size for a median-sized project, $x = \$138,000$.

The affiliated model is expected to yield a smaller estimate of the entry cost than the LS model. Indeed, the entry cost in the latter is equal to the equilibrium expected profit of the average potential bidder, whereas it is equal to the expected profit of the marginal potential bidder (with signal $\bar{s}(N)$) in the former. Because the signals in the SEM are positively correlated with the costs, the marginal bidder has a smaller entry cost than the average bidder.

Formally, the fit of the model is assessed by the J test. Since we have fitted 19 quantiles, the J statistic has a limiting χ^2 distribution with $19 \times 7 - 5$ degrees of freedom when the specification is correct. The model fits reasonably well. The J test p -value is 0.57, and therefore our parametric model is not rejected.

In the counterfactual experiment, we find that the selection effect leads to a *reduction* in the procurement price. Figures 2 and 3 depict the counterfactual effect of the number of potential bidders on the median entrant's cost and price for a median-sized project.²⁵ The graphs demonstrate the selection effect at work. Increasing the number of potential bidders by 1 results, on average, in a 1.74% reduction in the cost and 4.35% reduction in the price.²⁶ Therefore, our results suggest that, in order to reduce the cost of procurement,

²⁵LZ perform their counterfactual experiment for the 123rd auction in the sample, which has a project size close to the mean value of \$165,349. As in our nonparametric test, we have decided here to use the median-sized project. Our results take the same form for other project sizes, including that considered by LZ.

²⁶The confidence bands in the graphs are computed using the parametric bootstrap.

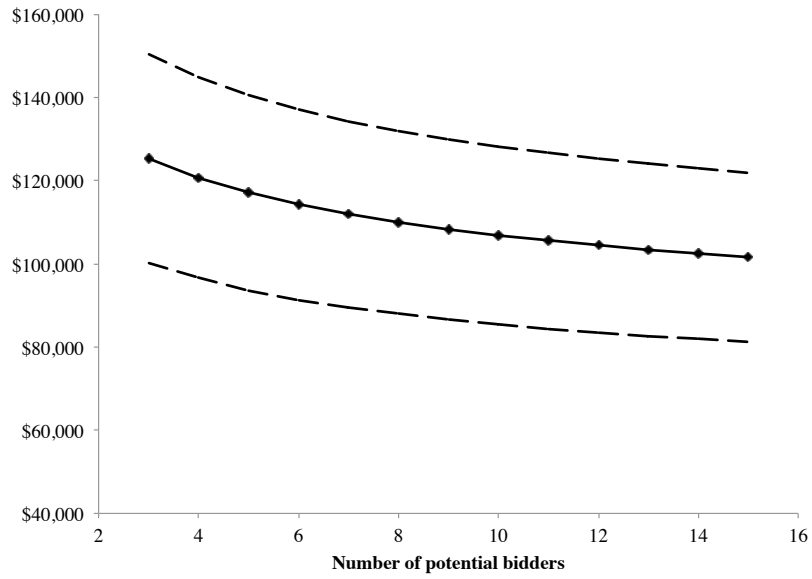


Figure 2: Estimated median cost conditional on entry (solid line) and its 95% confidence band (dashed lines)

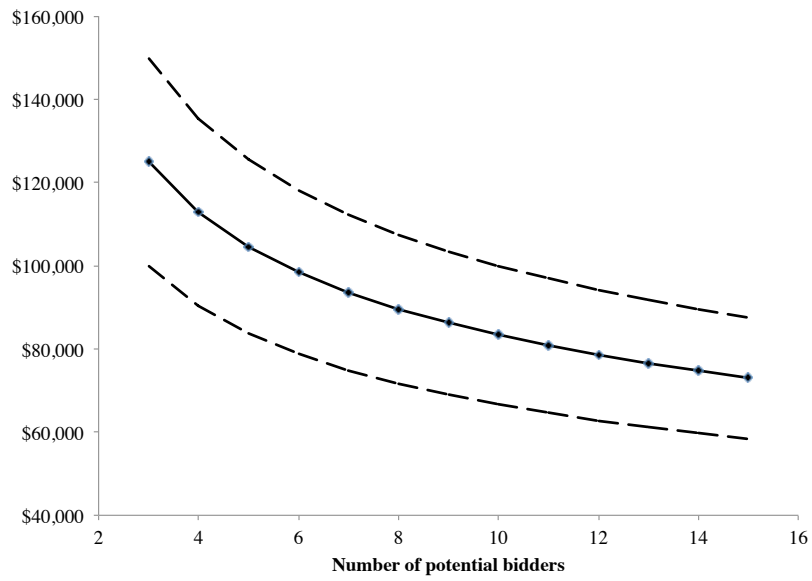


Figure 3: Counterfactual price (solid line) and its 95% confidence band (dashed lines)

the TDoT should *not* limit the competition in lawn-mowing auctions.

6 Concluding remarks

In this paper, we propose nonparametric tests to discriminate among alternative models of entry in first-price auctions. Those considered are: (a) the Levin and Smith (1994) (LS) model with randomized entry strategies; (b) the Samuelson (1985) (S) model that assumes that bidders are perfectly informed about their valuations at the entry stage, and selected into the pool of entrants based on this information; and (c) a new model (the selective entry model, or SEM), which allows for selective entry but in a less stark form than that in Samuelson (1985). More specifically, our model captures the selection effect by assuming that bidders receive signals that are informative about their valuations and make entry decisions based on these signals.

We estimate an SEM and find a strong degree of correlation between costs and signals. We then repeat the counterfactual experiment of Li and Zheng (2009), and, using our new model, find that increasing the number of potential bidders would be desirable: on average, the price falls by 1% with each additional potential bidder in the auction.

Our work can be extended in a number of directions. First, it would be interesting to extend our approach to other auction mechanisms and data structures as in Athey and Haile (2002).²⁷ Second, our models of entry do not fully endogenize the information acquisition process, the nature of which may have important consequences for auction design (see, e.g., Persico (2000) and Bergemann and Valimaki (2002)). Also, incorporating risk aversion, as Guerre, Perrigne, and Vuong (2009) and Campo, Guerre, Perrigne, and Vuong (2011) do in the standard first-price auction model, would clearly be very interesting.

Another extension would be to address an important limitation of our model, that is to allow for bidder asymmetries, as in a recent working paper by Krasnokutskaya and Seim (2011). The obvious difficulty here would be the necessity to deal with multiple equilibria head-on. Bajari, Hong, and Ryan (2010) obtain a number of identification results in this direction, and estimate a parametric model with multiple equilibria for highway procurement auctions. Another extension would be to incorporate dynamic features, as in Jofre-Bonet and Pesendorfer (2003). These extensions are left for future research.

²⁷See Athey and Haile (2007) for a survey of the literature.

Appendix

A Proofs of the results in Sections 2 and 3

Proof of Lemma 1. To show part (i),

$$\begin{aligned} F(v|S_i \geq s, V_i \geq r) &= \frac{1}{1 - F(s|V_i \geq r)} \int_s^{\bar{v}} F(v|s', V_i \geq r) f(s'|V_i \geq r) ds' \\ &\leq F(v|s, V_i \geq r) \end{aligned}$$

where the last inequality holds because, by Assumption 2, $F(v|s', V_i \geq r) \leq F(v|s, V_i \geq r)$ for all $s' \geq s$. To show part (ii),

$$\frac{\partial F(v|S_i \geq s, V_i \geq r)}{\partial s} = \frac{f(s|V_i \geq r)}{1 - F(s|V_i \geq r)} (F(v|S_i \geq s, V_i \geq r) - F(v|s, V_i \geq r)),$$

and the result follows by part (i) of the lemma.

Proof of Lemma 2. The expected profit (4) follows from the Envelope Theorem: in a symmetric equilibrium of the bidding game, by independence, the bidder with valuation v has the probability of winning the auction

$$(F(\tilde{s}) + (1 - F(\tilde{s}))F(v|S_i \geq \tilde{s}))^{N-1} \equiv \lambda(v, \tilde{s})^{N-1}.$$

Therefore, his expected profit in the auction is $(\underline{v}(s) - r)F(\tilde{s})^{N-1} \vee 0 + \int_{\underline{v}(s) \vee r}^v \lambda(x, \tilde{s})^{N-1} dx$. (The first term reflects a positive profit of the bidder with the lowest valuation \underline{v} .) At the entry stage, the expected profit is

$$\begin{aligned} \Pi(s, \tilde{s}, N) &= (\underline{v}(s) - r)F(\tilde{s})^{N-1} \vee 0 + \int_{r \vee \underline{v}(s) \vee r}^{\bar{v}} \int_{\underline{v}(s) \vee r}^v \lambda(x, \tilde{s})^{N-1} dx dF(v|s) \\ &= (\underline{v}(s) - r)F(\tilde{s})^{N-1} \vee 0 + \int_{\underline{v}(s) \vee r}^{\bar{v}} (1 - F(v|s)) \lambda(v, \tilde{s})^{N-1} dv, \end{aligned} \quad (30)$$

where the last line follows from integration by parts. This proves (4).

Next, $\Pi(s, \tilde{s}, N)$ is nondecreasing in \tilde{s} because $\lambda(v, \cdot)$ is nondecreasing: for all $v \in [\underline{v}(s), \bar{v}]$,

$$\begin{aligned} \frac{\partial \lambda(v, \tilde{s})}{\partial \tilde{s}} &= f(\tilde{s}) + \frac{\partial}{\partial \tilde{s}} P\{V_i \leq v, S_i \geq \tilde{s}\} \\ &= f(\tilde{s}) - f(\tilde{s})F(v|\tilde{s}) \geq 0. \end{aligned}$$

To show that $\Pi(s, \tilde{s}, N)$ is also nondecreasing in s , observe that the integral that appears in (30) depends on s in two ways: first, through the integrand term $1 - F(v|s)$, and second

– through the lower limit $\underline{v}(s)$ if $\underline{v}(s) > r$. The term $1 - F(v|s)$ is nondecreasing in s by Assumption 2. Also, for $\underline{v}(s) > r$,

$$\frac{\partial}{\partial y} \bigg|_{y=\underline{v}(s)} \left[(y-r)F(\tilde{s})^{N-1} + \int_y^{\bar{v}} (1-F(v|s)) \lambda(v, \tilde{s})^{N-1} dv \right] = F(\underline{v}(s)|s) \lambda(\underline{v}(s), \tilde{s})^{N-1} \geq 0.$$

Therefore the r.h.s. of (30) is nondecreasing in $\underline{v}(s)$. Since Lemma 1 implies that $\underline{v}(s)$ is nondecreasing in s , it follows that $\Pi(s, \tilde{s}, N)$ is nondecreasing in s .

Lastly, to show that $\Pi(s, \tilde{s}, N)$ is decreasing in N , observe that for $v < \bar{v}$, $\lambda(v, \tilde{s}) < 1$, which implies that $\lambda(v, \tilde{s})^{N-1}$ is decreasing in N . Assumption 1 implies that the interval $[\underline{v}(s) \vee r, \bar{v}]$ has positive length, and therefore the integral in (30) must be decreasing in N . Since $F(\tilde{s})^{N-1}$ is decreasing in N , $\Pi(s, \tilde{s}, N)$ is decreasing in N . ■

Proof of Proposition 1. We have already argued in text that existence of a symmetric *bidding* equilibrium follows from known results in the literature. To show existence of an equilibrium with endogenous entry, we need to verify that bidders' entry decisions are optimal. We show this separately depending on whether the entry cost is small, large, or intermediate, corresponding to cases (a) – (c) in the proposition. In case (a), $\Pi(\bar{v}, \bar{v}, N) < k$, $\bar{s}(N) = \bar{v}$, because strict monotonicity of $\Pi(s, s, N)$ in s implies that a bidder with $v = \bar{v}$, the highest valuation possible, is unable to recover the entry cost even when he is the single bidder in the auction. In case (b), $\Pi(\underline{v}, \underline{v}, N) \geq k$, we must have $\bar{s}(N) = \underline{v}$ because a bidder with $v = \underline{v}$, the lowest valuation possible, is able to recover the entry cost even when all his rivals enter with probability 1. On the other hand, strict monotonicity of $\Pi(s, s, N)$ in s implies that bidders with signals strictly above \underline{v} strictly prefer to enter. In case (c), $\Pi(\underline{v}, \underline{v}, N) \leq k < \Pi(\bar{v}, \bar{v}, N)$, strict monotonicity and continuity of $\Pi(s, s, N)$ in s implies that there exists a unique cutoff $\bar{s}(N)$ that solves the indifference condition $\Pi(\bar{s}(N), \bar{s}(N), N) = k$.

The cutoff $\bar{s}(N)$ must be strictly increasing in N whenever entry is partial. Otherwise, if for $N' > N$, $\bar{s}(N') \leq \bar{s}(N)$, Lemma 2 would imply $\Pi(\bar{s}(N'), \bar{s}(N'), N') < k$, at least whenever $\bar{s}(N) \in (\underline{v}, \bar{v})$. This would be in violation of indifference condition (5). ■

Proof of Proposition 2. Using the definition of $F(v|S_i \geq s, V_i \geq r)$ in (1), one can calculate its partial derivative with respect to s ,

$$\begin{aligned} \frac{\partial F(v|S_i \geq s, V_i \geq r)}{\partial s} &= -\frac{f(s)}{1-F(s)} F(v|s, V_i \geq r) + \frac{f(s)}{(1-F(s))^2} \int_s^{\bar{v}} F(v|\tilde{s}, V_i \geq r) dF(\tilde{s}) \\ &= \frac{f(s)}{1-F(s)} (F(v|S_i \geq s, V_i \geq r) - F(v|s, V_i \geq r)). \end{aligned} \quad (31)$$

Consider a v in the interior of the support of $F(v|s, V_i \geq r)$. Then under Assumption 5, the above partial derivative is negative for all $s \in (\underline{v}, \bar{v})$. This implies that $F(v|S_i \geq s', V_i \geq r) < F(v|S_i \geq s, V_i \geq r)$ for $s' > s$. Because $N' > N \geq N_*$, $\bar{s}(N') > \bar{s}(N)$ and

$\bar{s}(N) \in (\underline{v}, \bar{v})$, it follows that $F(v|\bar{s}(N'), V_i \geq r) < F(v|\bar{s}(N), V_i \geq r)$ for v in the interior of $\text{supp}F(\cdot|\bar{s}(N), V_i \geq r)$. This implies $Q^*(\tau|N') > Q^*(\tau|N)$ for $\tau \in (0, 1)$, which in turn implies the statement in the proposition. ■

Proof of Proposition 3. For every $b \in [r, \bar{b}(N)]$, we have

$$\begin{aligned} G^*(b|N) &= F(\xi(b|N)|N, S_i \geq \bar{s}(N), V_i \geq r) \\ &= \frac{F(\xi(b|N)|S_i \geq \bar{s}(N)) - F(r|S_i \geq \bar{s}(N))}{1 - F(r|S_i \geq \bar{s}(N))}. \end{aligned} \quad (32)$$

where $\xi(\cdot|N) : [r, \bar{b}(N)] \rightarrow \mathbb{R}_+$ denotes the inverse bidding strategy, defined as the unique solution to the equation $B(\xi|N) = b$. Since the probability of submitting a bid is

$$\begin{aligned} p(N) &= P(S_i \geq \bar{s}(N), V_i \geq r) \\ &= P(S_i \geq \bar{s}(N)) P(V_i \geq r|S_i \geq \bar{s}(N)) \\ &= (1 - F(\bar{s}(N))) (1 - F(r|S_i \geq \bar{s}(N))), \end{aligned}$$

we see that

$$p(N) G^*(b|N) = (1 - F(\bar{s}(N))) [F(\xi(b|N)|S_i \geq \bar{s}(N)) - F(r|S_i \geq \bar{s}(N))]$$

After simplification,

$$1 - p(N) + p(N) G^*(b|N) = F(\bar{s}(N)) + (1 - F(\bar{s}(N))) F(\xi(b|N)|S_i \geq \bar{s}(N)).$$

Differentiating (32) with respect to b gives the density of bids

$$g^*(b|N) = \frac{f(\xi(b|N)|S_i \geq \bar{s}(N))}{1 - F(r|S_i \geq \bar{s}(N))} \xi'(b|N),$$

and we see that

$$p(N) g^*(b|N) = (1 - F(\bar{s}(N))) f(\xi(b|N)|S_i \geq \bar{s}(N)) \xi'(b|N).$$

Therefore,

$$\frac{1 - p(N) + p(N) G^*(b|N)}{p(N) g(b|N)} = \frac{F(\bar{s}(N)) + (1 - F(\bar{s}(N))) F(\xi(b|N)|S_i \geq \bar{s}(N))}{(1 - F(\bar{s}(N))) f(\xi(b|N)|S_i \geq \bar{s}(N))} \frac{1}{\xi'(b|N)}.$$

Since $\xi'(b|N) = 1/B'(v|N)$ for $v = \xi(b|N)$, (11) follows from the differential equation (3) for the bidding strategy. ■

Proof of Proposition 4. The proof of the results in (a) is omitted because it parallels the proofs of Theorems 1 and 4 in GPV.

For (b), suppose the data are generated by an SEM with some $k > 0$ and $F(\cdot|\cdot)$ such that $F(r) \in (0, 1)$. Consider any standard model ($k = 0$) with the distribution of valuations given by

$$F_0(v) = F_0(v|v < r) F_0(r) + F^*(v|N) [1 - F_0(r)]$$

where $F_0(r) = 1 - p(N)$ and $F_0(\cdot|v < r)$ is an arbitrary absolutely continuous CDF with support $[\underline{v}_0, r]$ for some $\underline{v}_0 \in (0, r)$. One can check that the differential equation (3) for the bidding strategy remains the same as for the original SEM, so the inverse bidding strategy is also $\xi(\cdot|N)$. Since the distribution of active bidders' valuations $F^*(v|N)$ is the same by construction, it follows that this standard model induces, in the Bayesian-Nash equilibrium, the same distribution of observable bids $G^*(\cdot|N)$. Furthermore, since each bidder becomes active with probability $1 - F(r) = p(N)$, the distribution of the number of active bidders is also the same.

Going in the other direction, suppose the data $(G^*(\cdot|N), p(N))$ are generated by some standard model with the distribution of valuations F_0 above. Since LS model is included in the SEM class, it is sufficient to construct an observationally equivalent LS model. Consider an LS model with the distribution of valuations

$$F_{LS}(v) = F_{LS}(v|v < r) F_{LS}(r) + F^*(v|N) [1 - F_{LS}(r)]$$

where $F_{LS}(\cdot|v < r)$ is an arbitrary absolutely continuous CDF with support $[\underline{v}_{LS}, r]$ for some $\underline{v}_{LS} \in (0, r)$, and $F_{LS}(r)$ is chosen to assure that the probability of bidding is still $p(N)$: $\alpha [1 - F_{LS}(r)] = p(N)$ for some *entry probability* $\alpha \in (0, 1)$. To complete the construction, let

$$k = \int_r^{\bar{v}} (1 - F_{LS}(v) (\alpha + (1 - \alpha) F_{LS}(v)))^{N-1} dv.$$

Again, it is straightforward to show that this LS model's differential equation for the bidding strategy is the same as in the original standard model. Since $F_{LS}(v|v \geq r) = F_0(v \geq r)$, the induced distribution of bids is the same. Also, according to (5), bidders are indifferent between entering and not entering provided the rivals enter with probability $\alpha \in (0, 1)$, while the probability of submitting a bid is equal to $p(N)$ as in the original standard model.

For part (c), if $\xi(r|N) > r$, similar arguments imply that any S model with $k = (\xi(r|N) - r) (1 - p(N))^{N-1}$ and

$$F_S(v) = F_S(v|v < \xi(r|N)) [1 - p(N)] + F^*(v|N) p(N)$$

where $F_S(\cdot|v < \xi(r|N))$ is any absolutely continuous CDF with support $[\underline{v}_{LS}, r]$ for some $\underline{v}_{LS} \in (0, r)$, rationalizes the data. ■

Proof of Proposition 5. In the S model, $F(v)$ is identified for $v \geq \bar{s}(N) > r$, using, for example, (19):

$$F(v) = p(N) F^*(v|N) + 1 - p(N) \tag{33}$$

Consequently, $F(v)$ is identified for $v \geq \min\{\bar{s}(N) : N \in \mathcal{N}\} = \bar{s}(\underline{N})$. The entry cost is also identified:

$$k = (\xi(r|N) - r)(1 - p(N))^{N-1}. \quad (34)$$

In the LS model, k is smaller than the expected profit from bidding even under full entry for $N = \underline{N}, \dots, N_*$, and is strictly greater than the expected profit for $N = N_* + 1, \dots, \bar{N}$. Equivalently,

$$\begin{aligned} \int_r^{\bar{v}} (1 - F(v))(F(r) + [1 - F(r)]F(v|V_i \geq r))^{N_*-1} dv &> k \\ &> \int_r^{\bar{v}} (1 - F(v))(F(r) + [1 - F(r)]F(v|V_i \geq r))^{N_*} dv. \end{aligned} \quad (35)$$

For $N = \underline{N}, \dots, N_*$, non-participation in the auction is caused by the truncating effect of the reserve price. In other words, $p(N) = 1 - F(r)$ for $N = \underline{N}, \dots, N_*$, and $F(r) = 1 - p(N)$ is identifiable. This implies the identification of both $F(v) = [1 - F(r)]F^*(v|N) + F(r)$ and k according to (15). ■

Remark 8. Note that (35) is a condition on the primitives of the model sufficient for the pattern (13) to arise.

B Monte-Carlo experiment

In this section we present a Monte-Carlo study of the small sample properties of the tests. In particular, we are interested in seeing how a dense grid of quantiles τ affects size and power of the tests.

We simulate S and V using a Gaussian copula. Let (Z_1, Z_2) be bivariate normal with zero means, variances equal to one, and correlation coefficient $\rho \in [0, 1)$. Informative signals corresponds to $\rho > 0$, while the LS model corresponds to $\rho = 0$. Let Φ denote the standard normal CDF. A pair (S, V) is generated as $S = \Phi(Z_1)$ and $V = \Phi(Z_2)$.

Details of the computation of the distributions $F(v|S)$ and $F^*(v|N)$ that are needed in order to solve for the equilibrium of the auction are as follows. First, $Z_2|Z_1 \sim N(\rho z_1, 1 - \rho^2)$, and therefore the conditional distribution of V given S is given by

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

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

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

Table 4: Simulated size and size-corrected power of models' tests

Number of auctions						
$L = 250$						
$L = 50$						
Nominal size						
Nominal size						
Model	0.1	0.05	0.01	0.1	0.05	0.01
Size						
Size						
LS	0.0570	0.0250	0.0050	0.0720	0.0220	0.0070
SEM	0.0880	0.0470	0.0140	0.0810	0.0430	0.0090
S	0.0850	0.0510	0.0120	0.0790	0.0480	0.0150
Size-corrected power						
Size-corrected power						
LS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SEM	0.9930	0.9700	0.8550	0.3012	0.2745	0.2201
S	0.9950	0.9810	0.9300	0.7643	0.7133	0.7002

where the cutoff signal $\bar{s}(N)$ can be found, given the value of N , as a solution to equation (5). Lastly, for $S \geq \bar{s}(N)$, the bids are computed according as a solution to (3).

In our simulations, we set $L = 250$ or $L = 50$, $\mathcal{N} = \{2, 3, 4, 5\}$, $\pi(N) = 1/4$ 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 tests are obtained using 999 replications. We use the tri-weight kernel function $K(u) = (35/32)(1 - u^2)^3 1_{\{|u| \leq 1\}}$ and $1.978 \times 1.06 \times (\text{std.err.}) \times (\text{sample size})^{-(1/5)}$ for the bandwidth. Our grid consists of 200 quantiles equally spaced between 0.05 and 0.95.

We generate three random samples. The first one is $\rho = 0$, which corresponds to LS model. The second is an SEM sample with $\rho = 0.5$. The last sample is the S sample. Table 4 reports the results of size simulations for the tests. The test based on T^{SEM} performs reasonably well in the finite samples. Some under rejection is observed for the LS sample ($\rho = 0$). Consider $L = 250$. For nominal sizes of 0.1, 0.05, and 0.01, the simulated rejection rates are approximately 0.057, 0.025 and 0.005 respectively. For the SEM ($\rho = 0.5$), the simulated rejection probabilities are very close to their nominal values: for the same nominal

sizes, they are approximately 0.088, 0.047, and 0.014. Finally, the rejection probabilities in the S experiment are also close to their nominal values. Roughly similar conclusions can be drawn for $L = 50$.

Table 1 also 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 for the SEM test, it is necessary to come up with an alternative. We reverse the order of quantiles, making them *decreasing* 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.

The results in Table 4 show that the power properties are good. Consider $L = 250$. For $\rho = 0.5$ and for nominal sizes of 10%, 5%, and 1%, the simulated rejection rates are 99.3%, 97% and 85.5% respectively. Similar results hold for LS ($\rho = 0$) and S models. For a smaller sample $L = 50$, the power for LS is still excellent. It is good for S, but there is a substantial power loss for SEM. Taken together, the simulation results indicate that our bootstrap method has both good size and power properties, even in samples of moderate size.

References

- ATHEY, S., AND P. A. HAILE (2002): "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107–2140.
- (2007): "Nonparametric Approaches to Auctions," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6, part 1, chap. 60, pp. 3847–3965. North-Holland, Amsterdam.
- ATHEY, S., J. LEVIN, AND E. SEIRA (2011): "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions," *Quarterly Journal of Economics*, 126(1).
- BAJARI, P. (2001): "Comparing Competition and Collusion: A Numerical Approach," *Economic Theory*, 18(1), 187–205.
- BAJARI, P., H. HONG, AND S. RYAN (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78(5), 1529–1568.
- 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.
- BERGEMANN, D., AND J. VALIMAKI (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70(3), 1007–1033.

- CAMPO, S., E. GUERRE, I. PERRIGNE, AND Q. VUONG (2011): “Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders,” *Review of Economic Studies*, 78(1), 112–147.
- COVIELLO, D., AND M. MARINIELLO (2010): “The Role of Publicity Requirements on Entry and Auctions Outcomes,” Working Paper, Harvard University.
- EINAV, L., AND I. ESPONDA (2008): “Endogenous Participation and Local Market Power in Highway Procurement,” Working Paper, Stanford University.
- GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): “Optimal Nonparametric Estimation of First-Price Auctions,” *Econometrica*, 68(3), 525–74.
- (2009): “Nonparametric Identification of Risk Aversion in First-Price Auctions under Exclusion Restrictions,” *Econometrica*, 77(4), 1193–1227.
- HAILE, P. A., H. HONG, AND M. SHUM (2003): “Nonparametric Tests for Common Values in First-Price Sealed Bid Auctions,” NBER Working Paper 10105.
- HARSTAD, R. M., J. KAGEL, AND D. LEVIN (1990): “Equilibrium Bid Functions for Auctions with an Uncertain Number of Bidders,” *Economic Letters*, 33(1), 35–40.
- 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.
- JOFRE-BONET, M., AND M. PESENDORFER (2003): “Estimation of a Dynamic Auction Game,” *Econometrica*, 71(5), 1443–1489.
- KRASNOKUTSKAYA, E. (2011): “Identification and Estimation in Procurement Auctions under Unobserved Auction Heterogeneity,” *Review of Economic Studies*, 78(1), 293–327.
- KRASNOKUTSKAYA, E., AND K. SEIM (2011): “Bid Preference Programs and Participation in Highway Procurement Auctions,” *American Economic Review*, 101(6), 2653–2686.
- LAFFONT, J. J., AND Q. VUONG (1996): “Structural Analysis of Auction Data,” *American Economic Review*, 86(2), 414–420.
- LEVIN, D., AND J. L. SMITH (1994): “Equilibrium in Auctions with Entry,” *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., AND M. GENTRY (2012): “Partial Identification in Auctions with Selective Entry,” Working Paper.
- LI, T., AND I. PERRIGNE (2003): “Timber Sale Auctions With Random Reserve Prices,” *Review of Economics and Statistics*, 85(1), 189–200.
- LI, T., AND X. ZHENG (2009): “Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions,” *Review of Economic Studies*, 76(4), 1397–1429.
- MARMER, V., AND A. SHNEYEROV (2012): “Quantile-Based Nonparametric Inference for First-Price Auctions,” *Journal of Econometrics*, 167(2), 345–357.
- NEWKEY, W. K., AND D. L. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier, Amsterdam.
- PAARSCH, H. J. (1997): “Deriving an Estimate of the Optimal Reserve Price: An application to British Columbian Timber Sales,” *Journal of Econometrics*, 78(2), 333–357.
- PAARSCH, H. J., AND H. HONG (2006): *An Introduction to the Structural Econometrics of Auction Data*. The MIT Press, Cambridge, Massachusetts.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press, New York.
- PERSICO, N. (2000): “Information Acquisition in Auctions,” *Econometrica*, 68(1), 135–148.
- ROBERTS, J. W. (2009): “Unobserved Heterogeneity and Reserve Prices in Auctions,” Working Paper, Duke University.
- ROBERTS, J. W., AND A. SWEETING (2010a): “Entry and Selection in Auctions,” Working Paper, Duke University.
- (2010b): “When Should Sellers Use Auctions?,” Working Paper, Duke University.
- SAMUELSON, W. F. (1985): “Competitive Bidding with Entry Costs,” *Economics Letters*, 17(1-2), 53–57.
- TAN, G., AND O. YILANKAYA (2006): “Equilibria in Second Price Auctions with Participation Costs,” *Journal of Economic Theory*, 130(1), 205–219.
- XU, P. (2012): “Nonparametric Estimation of the Entry Cost in First-Price Procurement Auctions,” *Journal of Applied Econometrics*, forthcoming.
- YE, L. (2007): “Indicative Bidding and a Theory of Two-Stage Auctions,” *Games and Economic Behavior*, 58(1), 181–207.