INFORMED ENTRY IN FIRST-PRICE AUCTIONS

JUN MA, VADIM MARMER, AND PAI XU

ABSTRACT. We study the role of information in first-price auctions with selective entry. In this model, bidders do not know their actual values before deciding to enter and only observe imperfect signals. We use a semiparametric copula-based model to identify and estimate the information effects. More informative signals are favorable to the seller because the value distribution for entering bidders is more stochastically dominant. However, due to the entry cost, the equilibrium effect of more informative signals may in fact reduce entry and reduce the seller's expected revenue. We estimate the model using data on Texas Department of Transportation (TDoT) procurement auctions and find that there is an optimal level of signals' informativeness and TDoT should prefer well (although not perfectly) informed bidders.

KEYWORDS: First-Price Auctions, Selective Entry, Value of Information, Semi-parametric Estimation.
JEL CLASSIFICATION: C12; C13; C14

1. Introduction

Entry is a core and fundamental issue in industrial organization and has been extensively studied in various markets. In the context of auction design, it entails addressing an interesting observation from markets - some potential bidders chose not to bid even though they have previously indicated their interests. Much research effort attributed the insufficient participation to costly entry. Yet little is known whether at all and how the quality and accuracy of the information a bidder possesses before entering an auction may play a role in the auction outcomes. This work fills this gap in the literature.

SCHOOL OF ECONOMICS, RENMIN UNIVERSITY OF CHINA
VANCOUVER SCHOOL OF ECONOMICS, UNIVERSITY OF BRITISH COLUMBIA
BUSINESS SCHOOL, UNIVERSITY OF HONG KONG
E-mail addresses: jun.ma@ruc.edu.cn, vadim.marmer@ubc.ca, paixu@hku.hk.
Date: February 27, 2023. First version: August 2022.
We thank Michael Peters, Xiaohong Chen, and Li Hao for helpful comments and discussions. We acknowledge the financial support from the Social Sciences and Humanities Research Council of Canada under grant 435-2021-0189 (Marmer), HK Grant Council under General Research Funding (Xu), and HKU Business School under its visiting scholar program.

For our purpose, we build our analysis on the selective model of entry for first-price auctions studied in Marmer, Shneyerov, and Xu (2013) (MSX hereafter) and Gentry and Li (2014); Chen, Gentry, Li, and Lu (2020). In this framework, a risk-averse bidder first receives a free private signal on the auctioned object's value, based on which she decides on whether to enter. Entry is costly, signals are imperfectly informative, and only after incurring the entry cost, the bidder learns her true private value of the object and submits a sealed bid. In the equilibrium, only bidders with sufficiently favorable signals above a certain cutoff will enter. At the time of bidding, the bidder only knows the number of potential bidders and her valuation; she does not know the number of entering or active bidders or their bids.

The signals in this model are informative in the sense that they are correlated with the valuations. We extend MSX by focusing on signals' informativeness and its effects on entry and the seller's expected revenue.

The econometrician observes many independent auctions with all submitted bids and the number of potential bidders in each auction. Data may also contain additional variables capturing observed auction-specific heterogeneity. The entry cost is unknown, and signals are unobserved. As a result, the joint distribution of signals and values is unidentified nonparametrically.

Our analytical approach is based on the parametrization of the affiliation between bidders' signals and their valuations to capture how well bidders are informed before making their entry decisions. Specifically, we assume a parametric copula with a single parameter that connects the marginal distributions of signals and values with their joint distribution; the marginal distributions are treated nonparametrically. The copula's parameter naturally captures the informativeness of signals, which allows us to study the effect of bidders' informedness on entry and the seller's expected revenue.

A priori, it is unclear whether the seller would always prefer more informed bidders or less informed bidders. On the one hand, under more informative signals, entering bidders have a more stochastically dominant value distribution, which benefits the seller by increasing her expected revenue. We refer to this as the "information effect". On the other hand, more informative signals change entry by raising or lowering the equilibrium entry cutoff for signals. We refer to this as the "cutoff effect", and its direction is ambiguous. As a result, more informative signals may result in a lower expected revenue for the seller if the cutoff effect works in the opposite direction from the information effect and dominates the latter.

For example, when signals are sufficiently uninformative, and the entry cost is low, one can expect close to full entry with a high expected revenue for the seller. However, even with a low entry cost, when signals are informative enough many bidders may choose not to enter if they draw low signals. This can substantially reduce the seller's expected revenue. On the other hand, when the entry cost is high and, as a result, the probability of entry is low, more informative signals may improve the seller's revenue by affecting which bidders participate in the bidding stage. Such considerations become policy-relevant when the seller can affect signals' informativeness, for example, by reducing or increasing the signals' variance through information releases.

Our empirical approach is semiparametric. First, following Guerre, Perrigne, and Vuong (2000) and Marmer, Shneyerov, and Xu (2013), we nonparametrically identify and estimate the CDF of values conditional on entry and the number of potential bidders. In the next step, we estimate the copula's parameter by matching the estimated CDF of values conditional on entry with that implied by the model, the number of potential bidders, and observed entry probabilities. Given the copula parameter, we can identify and estimate the marginal distribution of values and the entry cost. Once the model's fundamentals are estimated, we can perform counterfactuals such as changing the signals' informativeness by varying the copula's parameter and estimating the corresponding changes in the seller's revenue.

We applied our methodology to the lawn-mowing jobs in Texas Department of Transportation (TDoT) highway procurement auctions studied in Li and Zheng (2009). It is natural to focus on the rank correlation between values and signals in this context. Given a parametric copula family with a single parameter, there is a one-to-one correspondence between the Spearman rank correlation coefficient and the copula's parameter. We found the rank correlation between signals and valuations varies with the number of potential bidders. Moreover, this relationship is not monotonic. The correlation is the weakest with about 0.371 for the auctions with 11-14 potential bidders. The signals are most informative with the correlation being 0.85 for the auctions with 15-17 potential bidders. The entry cost is about 5% of engineer's estimate for auctions with 6-14 potential bidders. However, this cost comes down to 3% for the auctions with 15-17 potential bidders.

We find that in all cases, the seller's expected revenue is maximized when the rank correlation between values and signals is around 0.9. Hence, in the case of highway maintenance procurement auctions, TDoT should prefer well but not perfectly informed bidders.

We also study if TDoT could achieve the same outcome by offering entry cost subsidies to encourage entry instead of trying to change the level of bidders' informedness. In our counterfactual experiments, we find that the required subsidy is substantial and may erase any benefits from increased participation. It appears that reducing bidders' uncertainty about the cost and complexity of TDoT projects is a more effective channel. Bidders' uncertainty can be plausibly reduced by releasing more information on past contracts and their actual costs.

Our paper contributes to the growing literature on entry in auctions. Li and Zheng (2009) found that a larger number of potential bidders does not necessarily benefits the seller as it may reduce entry in the equilibrium. Marmer, Shneyerov, and Xu (2013) developed nonparametric tests for distinguishing between different models of entry that are nested under the selective entry model. Xu (2013) proposed a procedure for estimation of the entry cost when signals are perfectly informative.[1] Gentry and Li (2014) studied nonparametric set identification of the selective entry model. Chen, Gentry, Li, and Lu (2020) consider the identification of the selective entry model with risk-averse bidders. They also develop a copula-based approach in the model's context. Our paper adds to the literature by investigating the role of the bidder's informedness.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 presents our empirical approach, and Section 4 covers our empirical results.

2. Selective entry model

2.1. The basic model

There are $N \geq 2$ risk-neutral potential bidders in an auction. Bidders draw values V and signals S from the joint distribution $F(v, s)$. The draws (V, S) are independent across the bidders and remain private. At the entry stage, the potential bidders observe their signals, but not their values. A potential bidder can learn her value and become an active bidder after paying the entry cost κ. Only active bidders participate in the bidding stage.

The following assumption describes the model's fundamentals and combines Assumptions 1–2 in Marmer, Shneyerov, and Xu (2013) (MSX hereafter).

Assumption 2.1.

(i) The joint CDF $F(\cdot, \cdot)$ is absolutely continuous with a density $f(\cdot, \cdot)$ and has support $[v_l, v_u] \times [s_l, s_u]$.

(ii) The marginal CDF of signals, $F_S(\cdot)$ is absolutely continuous with a density $f_S(\cdot)$. The density $f_S(\cdot)$ is bounded from above and away from zero on its support $[s_l, s_u]$.

(iii) The following conditional distribution exists:

$$F_{V|S}(v \mid s) = \Pr(V \leq v \mid S = s), \tag{2.1}$$

and has a density $f_{V|S}(\cdot \mid s)$ on the support $[v_l, v_u]$ for $s \in [s_l, s_u]$.

[1] When signals are perfectly informative, the selective model of entry reduces to that of Samuelson (1985).

(iv) The family of conditional distributions $\{F_{V|S}(\cdot \mid s),\ s \in [s_l, s_u]\}$ satisfies a stochastic dominance relationship:

$$F_{V|S}(v \mid s_1) \geq F_{V|S}(v \mid s_2) \text{ for } s_1 \leq s_2 \text{ and all } v. \tag{2.2}$$

(v) There is a binding reserve price $v_l \leq r \leq v_u$.

Assumption 2.1(iv) is the key of the model - signals are informative about valuations. MSX called it "good news": a more favorable signal S is associated with a higher value V. Therefore, in equilibrium, the bidders all adopt a cut-off strategy at the entry stage - they will participate in bidding and become active bidders when

$$S \geq \bar{s}, \tag{2.3}$$

where $S = \bar{s}$ defines the marginal entrant expecting zero profit from entry. We use $p(\bar{s}, r)$ to denote the probability of entry and bidding given the threshold $\bar{s}$ and the reserve price r.

Let $F^*(v \mid \bar{s}, r)$ be the distribution of values conditional on entry and bidding. For an active bidder with $V = v$, the probability of winning is denoted by $\Lambda(v \mid \bar{s})^{N-1}$, where $\Lambda(\cdot \mid \bar{s})$ is

$$\begin{aligned} \Lambda(v \mid \bar{s}) &= 1 - p(\bar{s}, r) + p(\bar{s}, r) F^*(v \mid \bar{s}, r) && (2.4) \\ &= F_S(\bar{s}) + \int_{\bar{s}}^{s_u} F_{V|S}(v \mid s) f_S(s) ds. && (2.5) \end{aligned}$$

The function $\Lambda(v \mid \bar{s})$ captures two events: the rival does not bid; and, if the rival bids, she draws a value less than v. The function $\Lambda(v \mid \bar{s})$ is independent of the reserve price r, as the effect of r cancels out between the two terms.

From the results in MSX, the bidding strategy for an active bidder is given by

$$B(v \mid \bar{s}, r) = v - \int_r^v \left(\frac{\Lambda(u \mid \bar{s})}{\Lambda(v \mid \bar{s})} \right)^{N-1} du \text{ for } v > r, \tag{2.6}$$

and $B(v \mid \bar{s}, r) = 0$ for $v \leq r$. The expression follows the same format as in the equilibrium bidding strategy for standard first-price auctions without the entry stage, see e.g. Krishna (2010, page 16). The only difference is that the function Λ is replaced by the CDF of valuations in the standard case. Note that equation (2.6) further implies that the expected payment by an active bidder with $V = v > r$ is given by

$$P(v \mid \bar{s}, r) = v \Lambda(v \mid \bar{s})^{N-1} - \int_r^v \Lambda(u \mid \bar{s})^{N-1} du. \tag{2.7}$$

In equilibrium, the marginal entrant with $S = \bar{s}$ must earn zero expected profit:

$$\Pi(\bar{s}, \kappa) = \int_{r}^{v_u} \left(v\Lambda(v \mid \bar{s})^{N-1} - P(v \mid \bar{s}, r) \right) f_{V|S}(v \mid \bar{s}) dv - \kappa. \tag{2.8}$$

Hence, the entry threshold $\bar{s}$ is determined by

$$\Pi(\bar{s}, \kappa) = 0. \tag{2.9}$$

2.2. Signal informativeness

This work develops from MSX and introduces a new channel that will affect the auction outcomes - the correlation between signals and values. To this end, we use a semi-parametric approach and model the joint distribution of values and signals by a single-parameter family of copulas. The copula's parameter captures the strength of the relationship between values and signals, which we refer to as signal informativeness. The following assumption parametrizes the dependence between values and signals.

Assumption 2.2.

(i) The joint CDF of the values and the signals satisfies $F(v, s) = C(F_V(v), F_S(s), \theta_0)$ for some $\theta_0 \in \Theta \subset \mathbb{R}$, where $C(\cdot, \cdot, \theta)$ is a known function up to the value of the parameter $\theta \in \Theta$.

(ii) $\partial C(x, y, \theta)/\partial\theta \geq 0$ for all $x, y \in [0, 1]$ and all $\theta \in \Theta$.

(iii) $C_2(x, y; \theta) = \partial C(x, y; \theta)/\partial y$ exists for all $x, y \in [0, 1]$ and all $\theta \in \Theta$.

(iv) $C_{22}(x, y; \theta) = \partial^2 C(x, y; \theta)/\partial y^2 \leq 0$.

(v) $\partial C_2(x, y, \theta)/\partial\theta$ exists for all $x, y \in [0, 1]$ and all $\theta \in \Theta$.

The family of copulas $\{C(\cdot, \cdot, \theta) : \theta \in \Theta\}$ is positively ordered by Assumption 2.2 (ii): for all $x, y \in [0, 1]$ and $\theta_1 \leq \theta_2$,

$$C(x, y, \theta_1) \leq C(x, y, \theta_2).$$

There are many families that satisfy the positive ordering assumption including the Gaussian copula and important members of the class of Archimedean copulas such as Ali-Mikhail-Haq, Clayton, Frank, Gumbel, and Joe. For all these copulas, a higher value of θ corresponds to a stronger association between values and signals, measured by statistics such as Kendall's τ or Spearman's ρ (Nelsen, 2007, Chapter 5). Hence, in our auction context, the positive ordering assumption ensures higher values of θ imply *more informative signals*.

The positive ordering has an important implication for the distribution of values conditional on entry. Let $P_\theta(\cdot)$ denote the probability function implied by $C(\cdot, \cdot, \theta)$, $F_V(\cdot)$, and

$F_S(\cdot)$. Under positive ordering, we have that for $\theta_1 \leq \theta_2$, all $\bar{s}$ and all $v \in [v_l, v_u]$,

$$P_{\theta_1}(V \leq v \mid S \geq \bar{s}) \geq P_{\theta_2}(V \leq v \mid S \geq \bar{s}), \tag{2.10}$$

i.e. the distribution of values conditional on entry is more stochastically dominant for a higher value of the copula parameter θ, i.e. for more informative signals.

Parts (iii) - (v) of Assumption 2.2 are relevant in our model because of the following property from copula:

$$F_{V|S}(v \mid \bar{s}) = C_2(F_V(v), F_S(\bar{s}); \theta_0).$$

The concavity assumption of the copula function with respect to its second argument is consistent with the "good news" condition in Assumption 2.1(iv).

It is also worth noting that for more informative signals, the conditional distribution of values given $S = s$ is more concentrated around $v = s$. Hence, there is no first-order stochastic dominance relationship between the conditional CDFs of values given $S = s$ corresponding to $\theta_1 \leq \theta_2$, i.e. the CDFs have a crossing at $v = s$.

2.3. Signals' informativeness and the seller's expected revenue

In this section, we present our main analytical results and introduce the information and cutoff effects of changes in the signals' informativeness on the seller's expected revenue. The following proposition shows that the seller's expected revenue can be expressed in terms of the $\Lambda(v \mid \bar{s})$ function.

Proposition 2.1. *Under Assumption 2.1, the equilibrium expected revenue of the seller given the entry threshold $\bar{s}$ is*

$$R(\bar{s}, \theta_0) = v_u - r\Lambda(r \mid \bar{s})^N - N \int_r^{v_u} \left(1 - \frac{N-1}{N}\Lambda(v \mid \bar{s})\right) \Lambda(v \mid \bar{s})^{N-1} dv. \tag{2.11}$$

where

$$\Lambda(v \mid \bar{s}) = F_S(\bar{s}) + F_V(v) - C\left(F_V(v), F_S(\bar{s}); \theta_0\right)$$

Since we are focusing on the information channel, the reservation price r plays no fundamental role in our analysis. Hence, in what follows we assume no reserve price, i.e. $r = v_l$ unless mentioned otherwise. In this case, $r\Lambda(r \mid \bar{s})^N = v_l F_S(\bar{s})^N$.

Suppose first that that $v_l = 0$. In this case, the effect of changes in θ on the seller's expected revenue works through $\Lambda(v|\bar{s})$. Let us denote the marginal changes on R and Λ by either policy tool as R' and Λ', respectively. We have

$$R'(\bar{s}, \theta_0) = -N(N-1) \int_{v_l}^{v_u} \left(1 - \Lambda(v \mid \bar{s})\right) \Lambda(v \mid \bar{s})^{N-2} \Lambda'(v \mid \bar{s}) dv. \tag{2.12}$$

That is, the impact on revenues is of the opposite sign than the marginal changes on Λ.

A change in signal informativeness has two effects on Λ:

$$\frac{d\Lambda(k,\theta_0)}{d\theta_0} = \underbrace{\frac{\partial\Lambda}{\partial\theta_0}}_{\text{information effect}} + \underbrace{\frac{\partial\Lambda}{\partial F_S(\bar{s})}\cdot\frac{\partial F_S(\bar{s})}{\partial\theta_0}}_{\text{cutoff effect}}$$

The first effect is a direct one: as the signal becomes more informative, it has an immediate implication on the bidder's winning probability. We refer to this as the information effect. However, there is another indirect effect. When the potential entrants' signals become more informative, it further changes the entry cutoff for bidding, therefore the marginal participant is different in the equilibrium. Such a cutoff effect will also have an impact on the seller's expected revenue.

We next explain how the cutoff effect makes it unclear whether the seller should always prefer more informative or less informative signals. Indeed, it is straightforward to see how the information effect benefits the seller. For all $v \in [v_l, v_u]$,

$$\frac{\partial\Lambda}{\partial\theta_0} = -\frac{\partial C(F_V(v), F_S(\bar{s}), ;\theta_0)}{\partial\theta_0} < 0.$$

Hence, if there is only a direct effect, the seller would like to make the signal as informative as possible in order to maximize the expected revenue. This is rather intuitive. When only the information effect is at work, the entry is unaffected, and the pool of active bidders is fixed. The positive ordering assumption implies that these bidders tend to have higher values when the signals become more informative. Therefore, the seller now faces the bidders from a better distribution in the sense of stochastic dominance. This in turn translates into a higher auction revenue for the seller.

However, the presence of the cutoff effect complicates the situation. First, we note the following lemma.

Lemma 2.1. *Under Assumption 2.1,*

$$\frac{\partial\Lambda(v\mid\bar{s})}{\partial\bar{s}} = f_S(\bar{s})(1-F_{V|S}(v\mid\bar{s})) \geq 0. \tag{2.13}$$

According to the lemma, the seller would like to encourage entry by lowering the cutoff $\bar{s}$ in general. This term in the cutoff effect already has a different sign than the information effect. What makes it even more intriguing is the response of the equilibrium cutoff $\bar{s}$ to changes in θ is undetermined. To see this, recall the entry cutoff in equilibrium is determined by the following equation:

$$\Pi(\kappa,\bar{s};\theta_0) = \int_{v_l}^{v_u}\left[1-C_2(F_V(v),F_S(\bar{s});\theta_0)\right]\Lambda(v|\bar{s})^{N-1}dv - \kappa = 0.$$

Thus,

$$\frac{\partial F_S(\bar{s})}{\partial \theta_0} = -\frac{\partial \Pi(k, \bar{s}; \theta_0)/\partial \theta_0}{\partial \Pi(k, \bar{s}; \theta_0)/\partial F_S(\bar{s})}.$$

Assumption 2.2(iv) and the equation (4.4) are jointly sufficient to pin down the sign of the marginal effect $\partial \Pi(k, \bar{s}; \theta_0)/\partial F_S(\bar{s})$, which is always positive. However, the sign of $\partial \Pi(k, \bar{s}; \theta_0)/\partial \theta_0$ is ambiguous, because the auction model does not restrict how the conditional distribution of values given s changes with respect to the signal informativeness. More informative signals only make the distribution of values conditional on $S = s$ more concentrated around $v = s$. Hence, $\frac{\partial C_2(F_V(v), F_S(\bar{s}); \theta_0)}{\partial \theta_0}$ is left undetermined.

Relaxing the $v_l = 0$ assumption does not change the picture substantially as the cutoff effect is undetermined.

2.4. Illustration with a numerical example

In this section, we illustrate the information and cutoff effects using numerical examples. We use the Gaussian copula and Uniform$(0, 1)$ marginals for the values and signals. There is no reserve price. We consider two levels of entry cost: low ($\kappa = 0.015$) and high ($\kappa = 0.150$). We also consider markets with different numbers of potential bidders: few ($N = 4$) and many ($N = 10$). The figures in the Appendix provide detailed descriptions of the two effects under different scenarios. Table 1 provides a summary.

TABLE 1. The information and cutoff effects on the seller's expected revenues

Entry Cost	N	Information Effect	cutoff Effect	Dominating Effect	Informativeness
$\kappa = 0.150$					
	N=4	$+$	$-$	information	$\theta \uparrow$
	N=10	$+$	$-$	information	$\theta \uparrow$
$\kappa = 0.015$					
	N=4	$+$	$-$	cutoff	$\theta \downarrow$
	N=10	$+$	$-$	non-monotonic	$\min(\theta) = 0.4$

The information effects on revenues are positive as expected in all the four markets we consider. The cutoff effects are all negative, which means the entry cutoffs in equilibrium always increase with the signal informativeness in our specification. Since the two effects work in opposite directions, the next question is which of the two dominates.

When the entry cost is high ($k = 0.150$) entry is low regardless of the number of potential bidders: the entry cost acts as a selection mechanism resulting in few active bidders. In this scenario, the information effect dominates resulting in higher expected revenue for the seller. (One can see the relative magnitudes of the two effects in Figure ?? in the Appendix.) Thus, in situations with a high entry cost, the seller is likely to benefit from more informative signals.

In the case of a low entry cost ($k = 0.015$), one also has to take into account the number of potential bidders. In the small market with $N = 4$, the cutoff effect is more responsive to an increase in signals' informativeness than the information one. Overall, the more informative signals will reduce the seller's expected revenue. Thus, with few potential bidders and a low entry cost, the seller would prefer uninformed bidders. The intuition is that when the entry cost is low and there are few potential bidders, one can have full entry when signals are uninformative. However, under more informative signals, there is a higher probability of non-entry, which results in very few active bidders.

The situation is more complicated in the large market with $N = 10$: the cutoff effect dominates when $\theta < 0.4$, while the information effect is dominant for $\theta > 0.4$. One can see from Figure ?? that there is a high degree of participation under uninformative signals, but bidders are less likely to enter for higher values of θ. Nevertheless, because the number of potential bidders is large, there will be sufficiently many active bidders for the information effect to dominate. As a result, the effect of the signal's informativeness on the seller's expected revenue is non-monotone.

While one can describe the general tendencies of the two effects and how they depend on the entry cost and the number of potential bidders, in practice the overall effect is undetermined and depends on additional factors such as the marginal distribution of values. Hence, practical recommendations to sellers depend on specific applications.

3. Methodology

3.1. Semi-parametric identification

Part (i) of Assumption 2.2 parametrizes the dependence between values and signals. Note that the marginal distributions of signals and values remain nonparametric. Since signals are unobserved, one cannot expect nonparametric identification of their marginal CDF $F_S(\cdot)$. Moreover, with a parametric copula family, one cannot impose a normalization such as a uniform distribution of signals. However, since

$$F_{V|S}(v \mid \bar{s}) = C_2(F_V(v), F_S(\bar{s}); \theta_0),$$

the marginal distribution of signals only plays a role through the probability of non-entry $F_S(\bar{s})$, which can be treated as an unknown parameter. The knowledge of the CDF $F_S(\cdot)$ in general is not a must for our analysis.

3.1.1. **Identification of the primitives.** The starting point of our semi-parametric approach is similar to that outlined in GL, see footnote 18 therein, where they outline semi-parametric identification of the marginal distribution of bidders' values using a parametric specification of the copula function in the setting with no reserve price and with uniformly distributed signals. Note that with a parametrically specified copula, one cannot treat the assumption of uniformly distributed signals as a simple normalization. Nevertheless, parametric assumptions on the marginal distribution of signals are not required. This is possible because the marginal distribution of signals appears only in the probability of drawing a signal above the threshold, i.e. the probability of entry. We also note below that the semi-parametric identification approach can accommodate a binding reserve price, and the probability of entry can be identified together with the marginal distributions of values in the presence of a binding reserve price.

We assume that the econometrician observes data on L independent auctions, and the data generating process satisfies the following assumption.

Assumption 3.1.

(a) (V_{il}, S_{il}) and N_l are independent.

(b) $\{(V_{il}, S_{il}) : i = 1, \ldots, N_l; l = 1, \ldots, L\}$ are independent across the auctions and the bidders. $\{N_l : l = 1, \ldots, L\}$ are independent across the auctions.

(c) The unknown entry cost $\kappa > 0$ is constant across the auctions.

(d) The reserve price $r \geq v_l$ is constant across the auctions.

The entry threshold for signals in auction l is determined by equation (2.9) with $N = N_l$:

$$\bar{s}_l = \bar{s}(N_l). \tag{3.1}$$

Hence, there is variation in $\bar{s}$ across the auctions due to variation in the number of potential bidders N_l. The variation can be exploited for the purpose of identification.

When there is a binding reserve price, one has to distinguish between the events of entry and bidding:

$$\text{Entry:} \qquad S_{il} \geq \bar{s}(N_l), \tag{3.2}$$

$$\text{Bidding:} \qquad S_{il} \geq \bar{s}(N_l) \text{ and } V_{il} \geq r. \tag{3.3}$$

Bidders that submitted bids above the reserve price are active bidders. The econometrician observes submitted bids and the number of potential bidders $\{(B_{il}, N_l) : i = 1, \ldots, N_l; l = 1, \ldots, L\}$, where $B_{il} = 0$ for non-active bidders with $V_{il} < r$ or $S_{il} < \bar{s}(N_l)$. Using only data for active bidders, the conditional CDF and PDF of bids conditional on bidding can be estimated using standard nonparametric techniques. Similarly, since non-active status is observed, the econometrician can estimate the probability of bidding

$$p(\bar{s}(N), r) = \Pr(S_{il} > \bar{s}(N), V_{il} > r). \tag{3.4}$$

Note that the event of entry without bidding, $S_{il} \geq \bar{s}(N_l)$ but $V_{il} < r$, is in general not directly observable. Hence, the probability of entry, $1 - F_S(\bar{s}(N))$ cannot be estimated directly from data as a proportion of bidders who paid the entry cost but did not bid when the reserve price is binding. Note further that due to the semi-parametric copula specification, the probability of entry plays a central role in identification.

From the results in MSX, the values conditional on bidding, i.e. conditional on $V_{il} > r$ and $S_{il} > \bar{s}(N_l)$, are identified from the conditional distribution of the bids and the probability of bidding. Let $\xi(b \mid N)$ denote the inverse bidding strategy in an auction with N potential bidders. By Proposition 3 in MSX,

$$\xi(b \mid N) = b + \frac{1}{N-1}\left(\frac{G(b \mid N)}{g(b \mid N)} + \frac{1 - p(\bar{s}(N), r)}{p(\bar{s}(N), r)}\frac{1}{g(b \mid N)}\right), \quad \text{for } b \geq r, \tag{3.5}$$

where $G(\cdot \mid N)$ is the CDF of the distribution of bids of active bidders in auctions with N potential bidders, $g(\cdot \mid N)$ is the corresponding PDF. The values of active bidders are identified as

$$V_{il} = \xi(B_{il} \mid N_l), \quad \text{for } V_{il} \geq r, S_{il} \geq \bar{s}(N_l). \tag{3.6}$$

Using the identified values for the active bidders, the econometrician can now estimate the CDF of values conditional on bidding: for $v \geq r$,

$$F^*(v \mid \bar{s}(N_l), r), \tag{3.7}$$

which can now be treated as known for the purpose of identification for all $v \geq r$.

Following the approach of GL and correcting for the presence of the reserve price r and the unknown distribution of signals F_S, we can express $F^*(v \mid \bar{s}, r)$ as

$$\begin{aligned} p(\bar{s}, r)F^*(v \mid \bar{s}, r) &= \int_{\bar{s}}^{s_u} \left(C_2(F_V(v), F_S(s); \theta_0) - C_2(F_V(r), F_S(s); \theta_0)\right) dF_S(s) \\ &= F_V(v) - C(F_V(v), F_S(\bar{s}); \theta_0) - (F_V(r) - C(F_V(r), F_S(\bar{s}); \theta_0)), \end{aligned} \tag{3.8}$$

where we used the properties $F_S(s_u) = 1$ and $C(F_V(v), 1) = F_V(v)$, and (**??**). Note that for every $v \geq r$, the expression on the left-hand side of (3.8) can be estimated from the data, which forms the basis of identification of the unknown terms on the right-hand side. The expression depends on the probability of non-entry $F_S(\bar{s})$ (which cannot be directly estimated from the data) and the probability of bidding $p(\bar{s}, r)$. However the two probabilities are related by

$$F_S(\bar{s}(N)) = 1 - p(\bar{s}(N), r) - (F_V(r) - C(F_V(r), F_S(\bar{s}(N)); \theta_0)), \tag{3.9}$$

which allows us to write

$$\begin{aligned} p(\bar{s}(N), r)F^*(v \mid \bar{s}(N), r) &= F_V(v) - C(F_V(v), F_S(\bar{s}(N)); \theta_0) \\ &\quad -1 + p(\bar{s}(N), r) + F_S(\bar{s}(N)). \end{aligned} \tag{3.10}$$

The model's primitives can now be identified from the system of equations in (3.9)–(3.10): for a given $v \geq r$, the system has $2 \times |\mathcal{N}|$ equations with $|\mathcal{N}| + 3$ unknowns: $F_S(\bar{s}(N))$ for $N \in \mathcal{N}$, $F_V(v)$, $F_V(r)$ and θ_0. Hence, without imposing any further restrictions, the model's primitives can be identified if $|\mathcal{N}| \geq 3$. In the absence of the reserve price, $F_S(\bar{s}(N))$ can be estimated directly from the data as the probability of non-entry and $|\mathcal{N}| = 2$ is sufficient for identification.

The concavity of $C(y, \cdot; \theta)$ assumed in Assumption 3.1**??** allows one to invert equation (3.9) to obtain unique expressions $F_S(\bar{s}(N)) = \psi_N(F_V(r), \theta_0)$, where the ψ_N function is known. Similarly using the concavity of $C(\cdot, y; \theta)$, one can invert (3.10) to obtain unique expressions $F_S(\bar{s}(N)) = \varphi_N(F_V(v), \theta_0)$ with a known function φ_N. By setting $\psi_N(F_V(r), \theta_0) = \varphi_N(F_V(v), \theta_0)$, one obtains $|\mathcal{N}|$ equations with three unknowns: $F_V(v)$, $F_V(r)$ and θ_0.

More equations can be added by using the cross v's restrictions. Consider the system in (3.9)–(3.10) for $v \in v_1, \ldots, v_M$ with all $v_j \geq r$. In this case, the system has $(M+1) \times |\mathcal{N}|$ restrictions with $M + |\mathcal{N}| + 2$ unknowns. The extra restrictions can be used, for example, to test the parametric specification of the copula.

3.1.2. **Identification of counterfactuals.** The practitioner can be interested in several counterfactual experiments: changes in the entry cost κ or introduction of entry fees, changes in the informativeness of the signals, and their effects on the seller's expected revenue. Since the seller's expected revenue is determined by the function $\Lambda(v \mid \bar{s})$ and the probability of bidding $p(\bar{s}, r)$, one can construct their counterfactual versions to perform counterfactual experiments.

Using the semi-parametric copula assumption in Assumption **??** and (2.5), the function $\Lambda(v \mid \bar{s})$ can now be conveniently written as

$$\Lambda(v \mid \bar{s}) = F_S(\bar{s}) + \int_{\bar{s}}^{s_u} C_2(F_V(v), F_S(s); \theta_0) dF_S(s) = F_S(\bar{s}) + F_V(v) - C(F_V(v), F_S(\bar{s}); \theta_0). \tag{3.11}$$

3.2. Estimation

In this section, we focus on the case with no reserve price. Write $p_N = 1 - F_S(\bar{s}_N)$. The following equation is the basis for the estimation:

$$p_N \cdot F^*(v \mid \bar{s}_N) = \tilde{C}(F_V(v), 1 - p_N, \theta_0), \tag{3.12}$$

where we define

$$\tilde{C}(u, v, \theta) = u - C(u, v, \theta).$$

Further, let $\tilde{C}^{-1}(\cdot, v, \theta)$ denote the inverse function of $\tilde{C}(\cdot, v, \theta)$. Now, the model's restriction can be written as

$$F_V(v) = \tilde{C}^{-1}(p_N \cdot F^*(v \mid \bar{s}_N), 1 - p_N, \theta_0),$$

and it holds for all $v \in [v_l, v_u]$ and $N \in \mathcal{N}$.

Let $\hat{F}^*(\cdot \mid \bar{s}_N)$ and $\hat{p}_N$ denote the estimators of $F^*(\cdot \mid \bar{s}_N)$ and p_N respectively. For a given θ and N, the implied CDF $F_V(v)$ is $\tilde{C}^{-1}(\hat{p}_N \cdot \hat{F}^*(v \mid \bar{s}_N), 1 - \hat{p}_N, \theta)$. Since the distribution of values is independent of N, we can use the average to estimate the implied CDF $F_V(v)$:

$$\hat{F}_V(v \mid \theta) = \frac{1}{|\mathcal{N}|} \sum_{N \in \mathcal{N}} \tilde{C}^{-1}(\hat{p}_N \cdot \hat{F}^*(v \mid \bar{s}_N), 1 - \hat{p}_N, \theta).$$

We can now estimate θ by solving

$$\hat{\theta} = \arg\min_{\theta} \int_{\underline{v}}^{\overline{v}} \left(\tilde{C}^{-1}(\hat{p}_N \cdot \hat{F}^*(v \mid \bar{s}_N), 1 - \hat{p}_N, \theta) - \hat{F}_V(v \mid \theta) \right)^2 w(v) dv, \tag{3.13}$$

where $w(\cdot)$ is a weight function chosen by the econometrician, and $\underline{v}, \overline{v}$ are chosen so that $v_l \leq \underline{v} < \overline{v} \leq v_u$. The CDF of values $F_V(\cdot)$ then can be estimated as

$$\hat{F}_V(v) = \hat{F}_V(v \mid \hat{\theta}).$$

The optimization problem in (3.13) is one-dimensional and can be solved fast.

Once $\hat{\theta}$ and $\hat{F}(\cdot)$ are computed, we can proceed to the estimation of the entry cost and the seller's expected revenue. To estimate the entry cost κ, we can use

$$\hat{\kappa}(N) = \int_{v_l}^{v_u} (1 - C_2(\hat{F}_V(v), \hat{F}_S(\bar{s}_N), \hat{\theta})) \hat{\Lambda}(v \mid \hat{F}_S(\bar{s}_N))^{N-1} dv,$$

where

$$\hat{\Lambda}(v \mid \hat{F}_S(\bar{s}_N)) = \hat{F}_S(\bar{s}_N) + \hat{F}_V(v) - C(\hat{F}_V(v), \hat{F}_S(\bar{s}), \hat{\theta}).$$

We can estimate the seller's expected revenue by

$$\hat{R}(N) = v_u - N \int_{v_l}^{v_u} \left(1 - \frac{N-1}{N} \hat{\Lambda}(v \mid \hat{F}_S(\bar{s}_N)) \right) \hat{\Lambda}(v \mid \hat{F}_S(\bar{s}_N))^{N-1} dv.$$

4. Empirical Application

4.1. Data

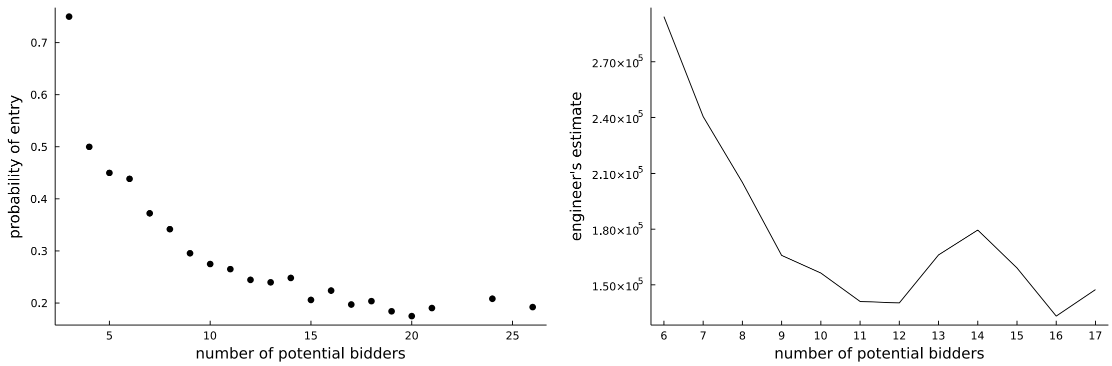
In this section, we revisit the application in Li and Zheng (2009) Li and Zheng (2009) while focusing on the role of the signals' informativeness. We are interested in estimating the signals' informativeness and its optimal level that maximizes the seller's revenue, or in this case, minimizes procurement costs.

The dataset has 540 auctions of "mowing highway right-of-way" maintenance jobs held by Texas Department of Transportation (TDoT) between January 2001 and December 2003. The data contain information not only on the submitted bids and bidder's identities but also on the project characteristics. The latter includes the engineer's estimates, contract length (the number of working days), number of items in the bidding proposal, the acreage of full-width mowing, whether it is a state job, and whether the job is on an interstate highway. Li and Zheng (2009) provides a detailed description of the dataset.

Next, we describe the timeline and the information flow of an auction in this application. TDoT first announces the project 3-6 weeks before the letting date. The advertisement includes a brief summary regarding the project - location, expected completion time, and engineer's estimate. Interested bidders have to request project plans and bidding proposal documents from TDoT no later than 21 days prior to letting. These documents provide some specific information about the project such as the item schedule etc. TDoT maintains and publishes the list of planholders for any project prior to the bid submission deadline. A planholder submits her bid in a sealed envelope by the bid opening time, if she chooses to do so.

From the viewpoint of a typical bidder, her involvement in the auction evolves as follows. She first comes to know the project under auction from an advertisement and makes a request for the bidding plan from TDoT. By doing so, the bidder becomes a potential bidder for the project. At this stage, she has a rough idea of how costly the project could be from reading both the advertisement and bidding documents. We assume that from this step the bidder receives a "signal" regarding her valuation of the project as the information can be obtained free of charge. The randomness of signals across bidders can come from their different abilities to interpret job complexity, flexibility in planning, and capacity constraints from the backlog of existing workload.

Once the bidder believes the project is worth further exploring, ie, her signal is sufficiently favorable, she will decide to learn more about the project and prepare for the bidding. By choosing to enter the auction, the bidder becomes an actual bidder. Only these bidders will incur the cost κ to update the information at hand and learn their



(A) Probability of Entry (B) Average Engineer's Estimates

FIGURE 1. Entry Probability and Average Project Size by Number of Potential Bidders

"valuations" for the project. In practice, this entry cost should entail examining the job complexity on the site, coordinating with suppliers, allocating human resources, updating the budget plan, and any other efforts spent towards the bid preparation and submission process.

Figure 1 shows two data patterns relevant to our study. First, not all the potential bidders chose to submit a bid. The probabilities of entry are shown in sub-figure (A). Indeed, except for auctions with three potential bidders, only less than a 50% of planholders submit bids. This probability declines with auctions attracting more potential interest from bidders.

Second, the project sizes also decrease along with the number of potential bidders. The sub-figure (B) plots the pattern of average engineer's estimates in this application. This seems to suggest that auctions with more potential bidders are associated with smaller and possibly easier jobs to accomplish. Hence, one can expect some variation in signals' informativeness with the number of potential bidders.

4.2. Estimation and Results

To account for observed heterogeneity, we follow the homogenization approach in Haile, Hong, and Shum (2003). Let X_ℓ denote a collection of observed auction-specific characteristics in auction ℓ. We assume the valuation for bidder i in auction ℓ follows is given by

$$V_{i,\ell} = g(X_\ell) \cdot V^*_{i,\ell} \tag{4.1}$$

where $V^*_{i,\ell}$ is bidder i's idiosyncratic value. The function $g(\cdot)$ shifts the mean of the distribution of bidders' idiosyncratic values. As in Haile, Hong, and Shum (2003), we also assume X and V^* are independent. Consequently, one can show that this structure in

(4.1) is preserved after imposing the equilibrium bidding strategy $b(\cdot)$.

$$b(V_{i,\ell} \mid \bar{s}_{N_\ell}) = g(X_\ell) \cdot b(V^*_{i,\ell} \mid \bar{s}_{N_\ell}) \tag{4.2}$$

We follow the common practice in this literature and use the logarithm of bids and other observed variables in our data analysis.

We first regress all observed bids on the covariates X_ℓ and a set of dummy variables for each value of N. The sum of each residual and the corresponding intercept estimate provides an estimate of each homogenized bid. These estimated bids are what bidders would have submitted in the equilibrium in a generic N-bidder auction with no heterogeneity.

Following Li and Zheng (2009), we first include several auction-specific characteristics into X: the engineer's estimates, number of working days, acreage of full-width mowing, number of items, whether it is a state job, and whether it is on an interstate highway. However, we found that all covariates other than the engineer's estimates were statistically insignificant and their contributions to R^2 were marginal. Therefore, we use only the engineer's estimates to account for the auction heterogeneity.[2] In this case, (4.2) becomes

$$\log(B_{i,\ell}) = \beta_1 \log(EngEst_\ell) + \log(B^*_{i,\ell})$$

where the homogenized bid B^* includes both the intercept estimate for N and the residual. The estimated parameter β_1 is 0.998 with a 95% confidence interval at $(0.985, 1.012)$. We thus decided to normalize bids simply as follows:

$$B^*_{i,\ell} = \frac{B_{i,\ell}}{EngEst_\ell}.$$

Hence, the valuations and bids in our analysis should be interpreted as fractions of their corresponding engineer's estimates.

To adjust for the low-bid procurement auctions, we transform homogenized bids as

$$\tilde{B}^*_{i,\ell} = \max\{B^*_{i,\ell}\} - B^*_{i,\ell}$$

where the $\max$ is taken over all i, ℓ, N. In other words, we find the maximum in our bid sample as the upper bound and compute, for each bid, its distance to the upper bound. This effectively transforms low-bid auctions into high-bid auctions. The $\max\{B^*_{i,\ell}\}$ in our sample is 3.1. We work with the sample of transformed bids $\{\tilde{B}^*_{i,\ell}\}$ for our subsequent empirical analysis.

Since the distributions of values conditional on entry must be nonparametrically estimated for each number of potential bidders, we exclude N's with fewer than 50 actual bids. As a result, our sample includes auctions with $N \in \{6, 7, .., 16, 17\}$.

[2] A similar observation was made with Michigan highway procurement data in Xu (2013) as well.

It is plausible to expect that the information content of signals varies with N. Therefore we consider three groups of N's: 6–10, 11–14, and 15–17. The groupings were determined to satisfy the following criteria: the estimated value of θ is not on the boundary, and the algorithm returns reasonable estimates for CDF F_V, i.e., the lower and upper bounds of the distribution function approach to zero and one, respectively.

TABLE 2. The estimated copula parameter $\hat{\theta}$, its corresponding Spearman correlation $\hat{\rho}$, and the estimated entry cost $\hat{\kappa}$ for different numbers of potential bidders

number of potential bidders	$\hat{\theta}$	$\hat{\rho}$	$\hat{\kappa}$
6 - 10	3.859	0.543	0.0557
	(3.115, 4.462)	(0.462, 0.600)	(0.0472, 0.0669)
11 - 14	2.390	0.371	0.0579
	(2.010, 2.754)	(0.318, 0.419)	(0.0521, 0.0641)
15 - 17	9.683	0.853	0.0294
	(9.598, 9.714)	(0.851, 0.854)	(0.0214, 0.0403)

Note: The 95% nonparametric percentile bootstrap confidence intervals are in parentheses.

We estimate the model using the Frank copula; the results are reported in Table 2. We report the estimates of θ for each group and their corresponding estimated Spearman rank correlation coefficients ρ. It is interesting to observe that the signals' informativeness does not correlate with the number of potential bidders in a monotone fashion. The signals are least informative for auctions with $N = 11 - 14$ with $\hat{\rho} = 0.371$, and they are most informative for $N = 15 - 17$ with $\hat{\rho} = 0.853$; in auctions with $N = 6 - -10$, $\hat{\rho} = 0.543$.

We estimate the entry cost κ for each N in the sample and reported the average of the costs for each group of auctions in Table 2. The entry cost varies between 3-6%. While the estimated entry cost demonstrates a non-monotone relationship with the number of potential bidders, overall it tends to be lower for larger N's.

4.3. Counterfactual experiments: Optimizing the seller's expected revenue

In this section, we discuss our counterfactual experiments and estimate the optimal levels of signal informativeness in terms of the Spearman rank correlation ρ^* that maximize the seller's expected revenue. We consider $N = 10, 14, 17$ (one from each group). The results are reported in Table 3.

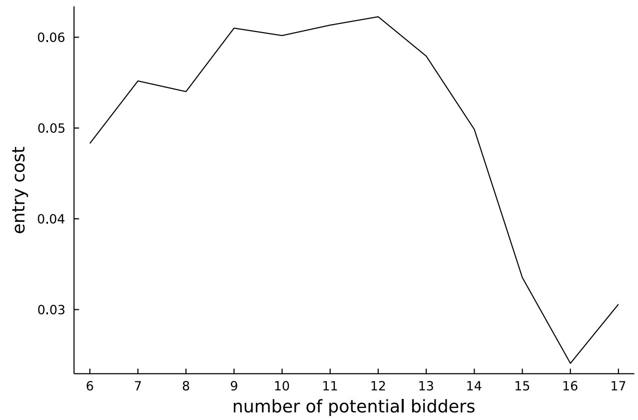


FIGURE 2. Estimated Entry Costs

TABLE 3. The estimated optimal signal informativeness in terms of the Spearman correlation $\hat{\rho}^*$, corresponding revenue increase, estimated equivalent entry cost $\hat{\kappa}^*$, and total expected entry subsidy

number of potential bidders	ρ^*	revenue increase	κ^*	subsidy
10	0.86	0.0185	0.0474	0.0257
	(0.78, 0.92)	(0.0057, 0.0423)	(0.0423, 0.0541)	(0.0069, 0.0608)
14	0.91	0.0651	0.0357	0.0946
	(0.88, 0.93)	(0.0472, 0.0838)	(0.0323, 0.0415)	(0.0613, 0.1247)
17	0.92	0.0024	0.0283	0.0038
	(0.87, 0.93)	(0.0001, 0.005)	(0.0213, 0.0382)	(0.0002, 0.0077)

Note: The 95% nonparametric percentile bootstrap confidence intervals are in parentheses.

The estimates of ρ^* range from 0.86 for $N = 10$ to 0.92 for $N = 17$. This suggests that TDoT should prefer well-informed bidders. This is consistent with our previous arguments that in auctions with many potential bidders, there will be sufficiently many active bidders for the seller to profit from the information effect. Nevertheless, the relationship between ρ and the seller's expected revenue is non-monotone: while well-informed bidders are preferred by the auctioneer, they should not be perfectly informed.

We further computed the resulting expected revenue increase by changing ρ from its estimates in the data to the estimated ρ^*. Since the estimated ρ is very close to the optimal for $N = 17$, we do not observe noticeable improvements. However, in the case of $N = 10$ and $N = 14$, we see improvements equal to 1.85 and 6.51 percent of the engineer's estimates respectively. The improvements are most significant for $N = 14$ because they come from the largest change of the signals' informativeness.

We plotted the relations between the seller's expected revenue and Spearman rank correlation coefficient ρ in Figure 3, where plots (A) -(C) are for $N = 10, 14, 17$, respectively. There is a concave relation for the three auctions, that is, the seller's revenue first increases with the signal's informativeness, and then decreases after reaching the peak at ρ^*. The results indicate that the seller's benefits from changing the signal's informativeness increase with the distance between ρ and ρ^*.

We plotted the probability of non-entry for N=14 in figure (D) in Figure 3. It shows that the cut-off signal $\bar{s}$ increases with the Spearman rank correlation ρ. Hence the estimated cutoff effect on the seller's expected revenue is always negative. Therefore, in this application, the information effect and cutoff effect work against each other. When signals are not sufficiently informative, the information effect dominates, and more informative signals improve the seller's expected revenue. However, after the peak at ρ^*, the cutoff effect takes over, and more information hurts the seller's expected revenue.

4.4. Changing the entry cost

An alternative mechanism to changing the signal's informativeness is for the seller to subsidize entry. In this section, we estimate the cost of this approach. More specifically, we estimate the subsidy required to achieve the same seller's expected revenue as that for the optimal signal's informativeness level.

To this end, we first discuss how the seller's expected revenue reacts to changes in the entry cost. For any level of the entry cost κ, equation (2.9) defines the entry threshold $\bar{s}$ and thus the marginal participant:

$$\bar{s} = \psi(\kappa). \tag{4.3}$$

MSX showed that the function $\psi(\cdot)$ is monotone increasing and one-to-one in the range where $0 < p(\bar{s}) < 1$. Hence, we can analyze the changes in $\bar{s}$ in order to see the effect of changing the entry cost κ.

Recall from Lemma 2.1 that

$$\frac{\partial \Lambda(v \mid \bar{s})}{\partial \bar{s}} = f_S(\bar{s})(1 - F_{V|S}(v \mid \bar{s})) \geq 0. \tag{4.4}$$

Equation (2.12) then implies:

$$\frac{\partial R(\bar{s}, \theta_0)}{\partial \bar{s}} \leq 0. \tag{4.5}$$

An increase in $\bar{s}$ has two opposite effects on $\Lambda(v \mid \bar{s})$: the positive effect $f_S(\bar{s})$ and the negative effect $-f_S(\bar{s})F_{V|S}(v \mid \bar{s})$. The positive effect comes directly from the non-entry event: An active bidder with the signal S substantially above $\bar{s}$ faces fewer competitors at the bidding stage under a larger entry threshold. The negative effect is due to the good news condition in Assumption 2.1(iv): the values V and signals S are positively associated in the sense of the stochastic dominance relationship $F_{V|S}(v \mid \bar{s}_1) \geq F_{V|S}(v \mid \bar{s}_2)$ for $\bar{s}_1 \leq \bar{s}_2$ and all v. However, the magnitude of the cutoff effect depends on the CDF $F_{V|S}(v \mid \bar{s}) \leq 1$, and as a result, the positive effect of $\bar{s}$ on $\Lambda(v \mid \bar{s})$ is stronger.

Lemma 2.1 shows the stochastic dominance relationship for $\Lambda(v \mid \bar{s})$ is the opposite of that for $F_{V|S}(v \mid \bar{s})$:

$$\Lambda(v \mid \bar{s}_1) \leq \Lambda(v \mid \bar{s}_2) \text{ for all } \bar{s}_1 \leq \bar{s}_2 \text{ and all } v. \tag{4.6}$$

Since the bidders do not observe each other's entry decisions, the equilibrium bidding function is determined by $\Lambda(v \mid \bar{s})$ instead of $F_{V|S}(v \mid \bar{s})$. An increase in the entry threshold $\bar{s}$ leads to less aggressive bidding by the active bidders:

$$\frac{\partial B(v \mid \bar{s})}{\partial \bar{s}} \leq 0.$$

The effect of changes in the entry threshold $\bar{s}$ on the seller's revenue is unambiguous: a higher entry threshold results in fewer active bidders and less aggressive bidding. As a result, when everything else is equal, a higher entry threshold results in a lower revenue for the seller.

We now discuss equivalent changes in the entry cost required to attain the same seller's expected revenue as that under the optimal signals' informativeness. We use κ^* to denote the equivalent entry cost. Table 3 reports the results.

The estimated equivalent entry costs are 4.7%, 3.6%, and 2.8% of the engineer's estimates for $N = 10, 14, 17$, respectively. The numbers are all below the estimated entry costs in the data, which indicates that the seller can attain the information-optimal expected revenue by subsidizing entry. The column of Table 3 shows the total expected subsidy. In each case, we find that the required subsidy exceeds the benefits. Hence, entry-cost subsidies are not a viable option in our application.

To see the magnitude of potential benefits and costs, consider for example a typical auction with $N = 14$. The average engineer's estimate for such auctions is $155,259 USD. By making signals more informative, for example by releasing more detailed information and reducing the signals' variance, TDoT could increase its expected revenue by up to $10,103 USD. The total entry subsidy required to attain the same expected revenue is $14,682 USD and exceeds the potential benefits.

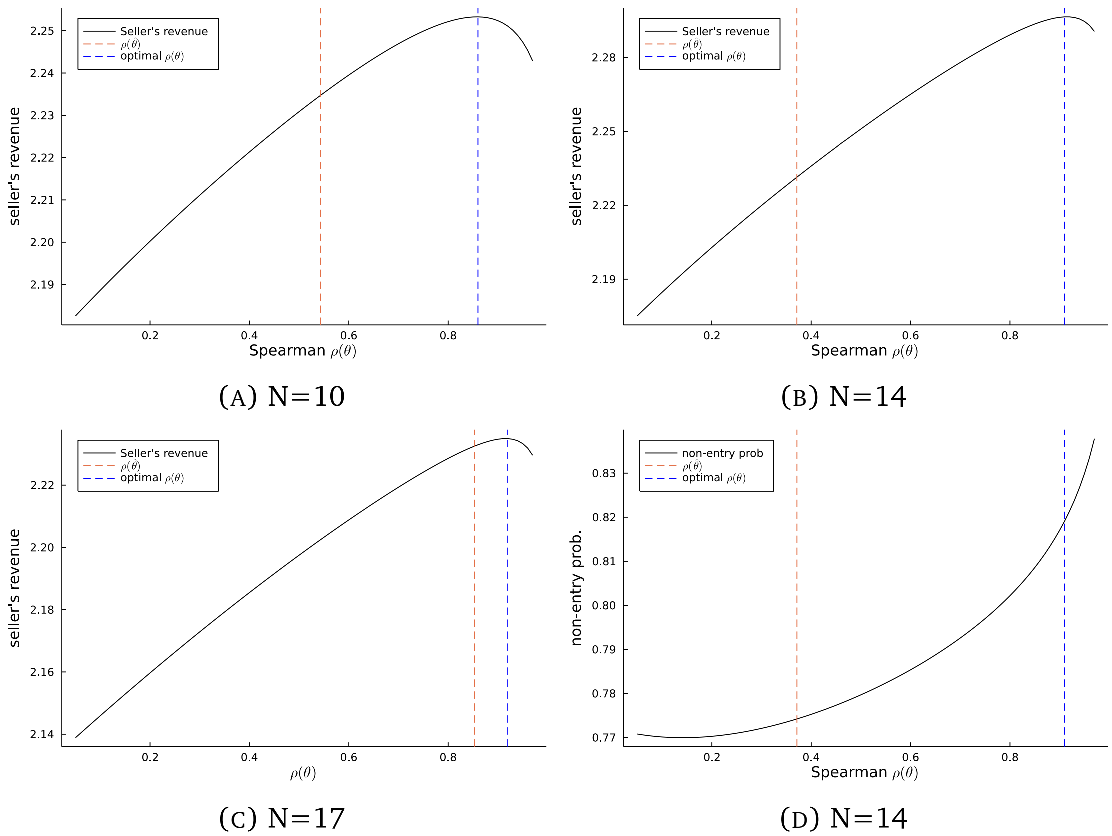


FIGURE 3. Seller's revenues with Spearman's rank

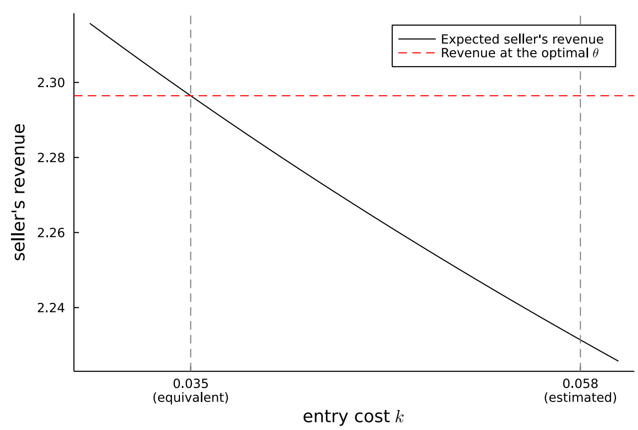


FIGURE 4. Equivalent Changes in κ for Optimal Revenues, N=14

Appendix A. Proofs

Proof of Proposition 2.1. The seller's expected revenue is

$$
\begin{aligned}
R(\bar{s}, r) &= N \int_{\bar{s}}^{s_u} \int_r^{v_u} P(v \mid \bar{s}, r) f_{V|S}(v \mid s) dv f_S(s) ds && \text{(A.1)} \\
&= N \int_r^{v_u} P(v \mid \bar{s}, r) \left(\int_{\bar{s}}^{s_u} f_{V|S}(v \mid s) f_S(s) ds \right) dv && \text{(A.2)} \\
&= p(\bar{s}, r) N \int_r^{v_u} P(v \mid \bar{s}, r) f^*(v \mid \bar{s}, r) dv, && \text{(A.3)}
\end{aligned}
$$

where

$$
f^*(v \mid \bar{s}, r) = \frac{1}{p(\bar{s}, r)} \int_{\bar{s}}^{s_u} f_{V|S}(v \mid s) f_S(s) ds \tag{A.4}
$$

is the PDF of values conditional on entry and bidding. By the definition of Λ in (2.4),

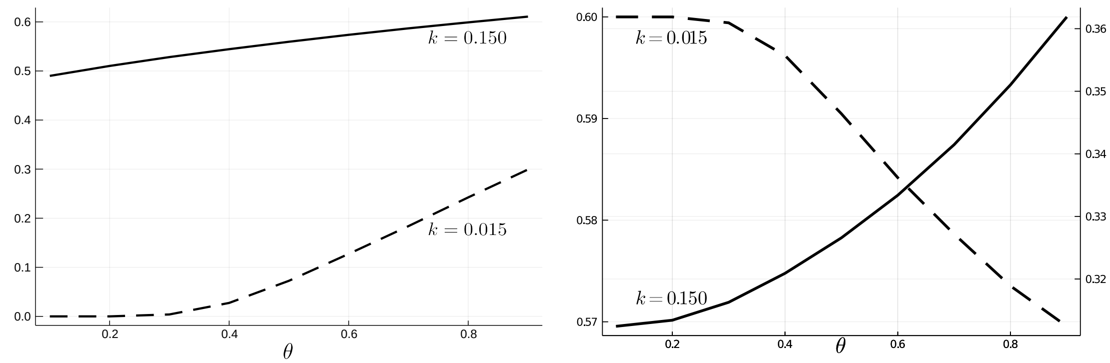
$$
\lambda(v \mid \bar{s}) = \frac{\partial \Lambda(v \mid \bar{s})}{\partial v} = p(\bar{s}, r) f^*(v \mid \bar{s}), \tag{A.5}
$$

and by (2.7), $R(\bar{s}, r)$ becomes

$$
\begin{aligned}
& N \int_r^{v_u} \left(v \Lambda(v \mid \bar{s})^{N-1} - \int_r^v \Lambda(u \mid \bar{s})^{N-1} du \right) \lambda(v \mid \bar{s}) dv && \text{(A.6)} \\
=\ & N \int_r^{v_u} v \Lambda(v \mid \bar{s})^{N-1} \lambda(v \mid \bar{s}) dv - N \int_r^{v_u} \left(\int_r^v \Lambda(u \mid \bar{s})^{N-1} du \right) d\Lambda(v \mid \bar{s}) && \text{(A.7)} \\
=\ & N \int_r^{v_u} v \Lambda(v \mid \bar{s})^{N-1} \lambda(v \mid \bar{s}) dv - N \int_r^{v_u} \Lambda(v \mid \bar{s})^{N-1} dv + N \int_r^{v_u} \Lambda(v \mid \bar{s})^N dv && \text{(A.8)} \\
=\ & \int_r^{v_u} v d\Lambda(v \mid \bar{s})^N + N \int_r^{v_u} \left(\Lambda(v \mid \bar{s}) - 1 \right) \Lambda(v \mid \bar{s})^{N-1} dv && \text{(A.9)} \\
=\ & v_u - r \Lambda(r \mid \bar{s})^N + \int_r^{v_u} \left((N-1) \Lambda(v \mid \bar{s})^N - N \Lambda(v \mid \bar{s})^{N-1} \right) dv, && \text{(A.10)}
\end{aligned}
$$

where the results hold by integration by parts and because $\Lambda(v_u \mid \bar{s}) = 1$.

□



(A) Probability of non-entry

(B) The seller's revenue (left axis for $k = 0.015$, right axis for $k = 0.150$)

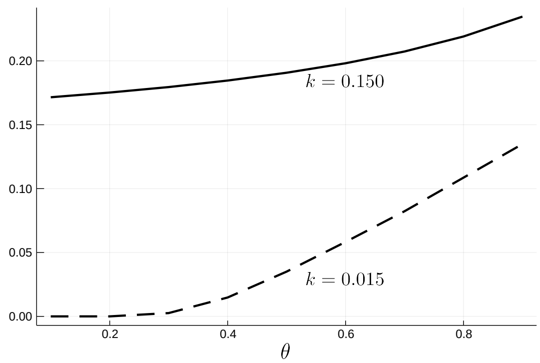
FIGURE 5. The number of potential bidders $N = 4$: The probability of non-entry and the seller's total revenue as a function of the informativeness parameter θ for different values of the entry cost k: dashed lines for $k = 0.015$, and solid lines for $k = 0.150$

Appendix B. Appendix: Numerical Examples

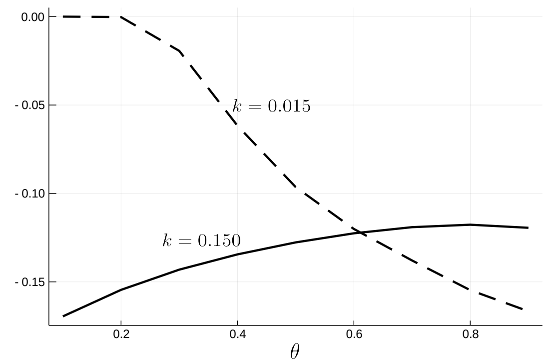
First, consider an auction with a relatively small number of potential bidders: $N = 4$. Figure 5 plots the probability of non-entry and the seller's total revenue as functions of the informativeness parameter θ. in the case of a low entry cost, there is full entry when the signals are almost uninformative. This is also the case that results in the largest revenue for the seller. Increasing the signals' informativeness starts affecting the probability of non-entry at around $\theta \approx 0.3$. The probability of non-entry increases from 0.0 to 0.3. Increasing the informativeness of the signals raises the entry threshold and negatively affects the seller's revenue.

In the case of a high entry cost, increased informativeness of the signals also increases the entry threshold, however, it happens to a lesser extent: the probability of non-entry increases from just under 0.5 to just over 0.6. At the same time and unlike in the low-cost case, the seller's revenue increases with the signals' informativeness.

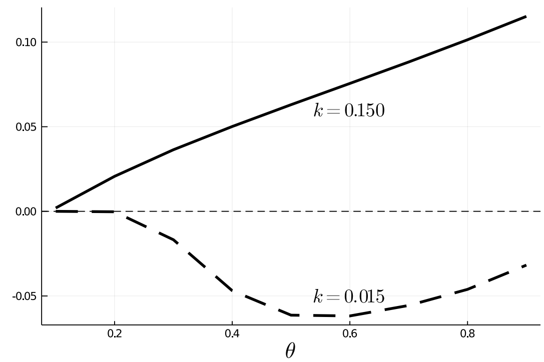
To understand the difference between the low and high entry cost k cases when the number of potential bidders is small ($N = 4$), we break down the total marginal effect of θ on the seller's revenue into the information and cutoff effects. Figure 6(A) shows that in both cases and as expected, the direct information effect is positive. On the other hand, the indirect cutoff effect is negative for the low entry cost case, and positive for the high entry cost case. Moreover, in the case of a low entry cost the cutoff effect dominates the information effect resulting in a total negative marginal effect on the seller's revenue.



(A) Information effect (marginal, direct)



(B) Cutoff effect (marginal, indirect)

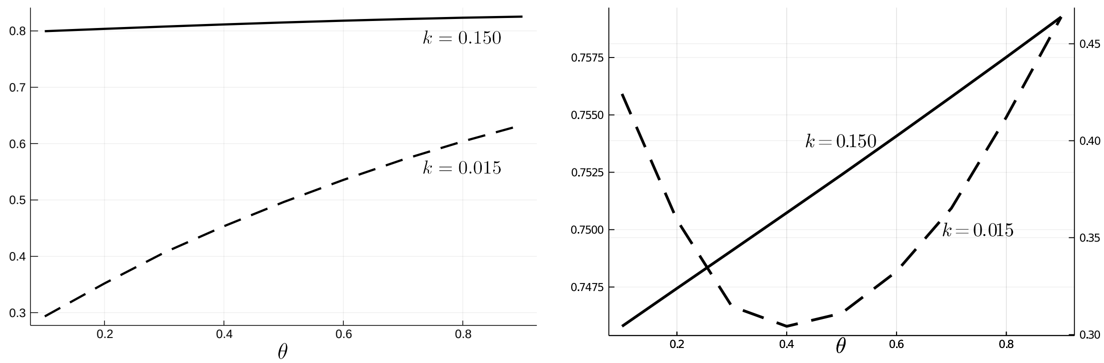


(C) Total marginal effect

FIGURE 6. The number of potential bidders $N = 4$: A decomposition of the total marginal effect of signals' informativeness θ on the seller's revenue into the information and cutoff effects for different values of the entry cost k: dashed lines for $k = 0.015$, and solid lines for $k = 0.150$

To understand these patterns, note that in the low entry cost case, there is a full entry and no selection when signals are not informative enough. Note that in this case, a signal provides little to no information about the true value, and even potential bidders with low signals can expect to win the auction with a reasonable probability. When θ increases and signals become sufficiently informative, potential bidders with low signals stop entering as now such the signals predict with enough accuracy that they are unlikely to win. The negative cutoff effect is stronger than the positive direct effect, and as a result, the seller's revenue decreases.

Note that in the case of a high entry cost, entry is already highly selective even when signals are uninformative: the probability of non-entry is close to 50% for the near-zero values of θ. Changing signals' informativeness leads only to small increases in the probability of non-entry, and while the cutoff effect is negative, its magnitude is smaller than that of the positive direct effect. Figure 6(C) shows that the positive direct information effect dominates when the entry cost is high.



(A) Probability of non-entry

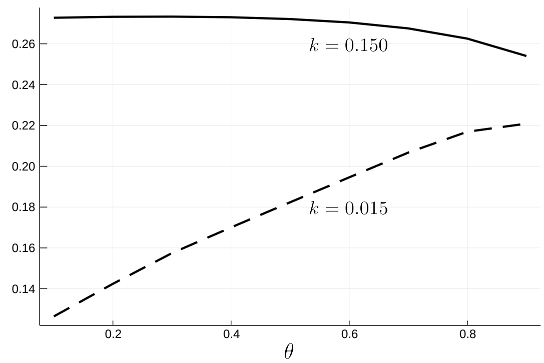
(B) The seller's revenue (left axis for $k = 0.015$, right axis for $k = 0.150$)

FIGURE 7. The number of potential bidders $N = 10$: The probability of non-entry and the seller's total revenue as a function of the informativeness parameter θ for different values of the entry cost k: dashed lines for $k = 0.015$, and solid lines for $k = 0.150$

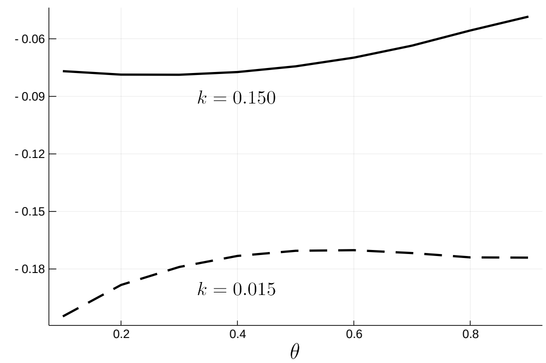
Overall, Figure 5(A) indicates that when the number of potential bidders is relatively small, from the seller's perspective it is preferred to have a low entry cost and uninformative signals.

Figures 7 and 8 plot the same functions for auctions with a larger number of potential bidders: $N = 10$. There is no full entry even when the entry cost is low ($k = 0.015$), and signals are non-informative: $F_S(\bar{s}) \approx 0.30$ for $\theta \approx 0.0$. Also, in the case with a low entry cost, the magnitudes of the positive information effect and the negative cutoff effects are very similar. The total marginal effect switches from negative to positive at $\theta \approx 0.4$, as shown in Figure 8(C). As a result, the effect of the signals' informativeness on the seller's revenue is now non-monotone, as shown in Figure 7(B): the seller's revenue is decreasing for $\theta < 0.4$, and it is increasing for $\theta > 0.4$. Hence, small changes in the signals' informativeness can have opposite effects depending on θ.

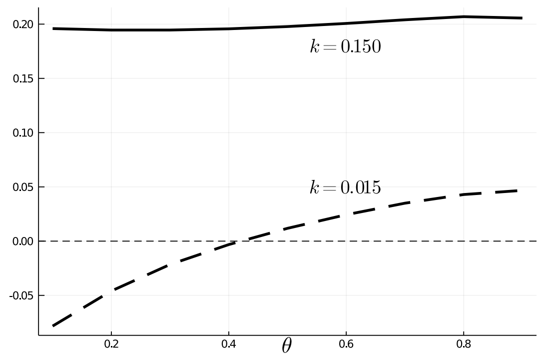
Moreover, Figure 7(B) shows that from the seller's perspective, having a low entry cost and as informative signals as possible are preferable. However, the latter result depends on the marginal distribution of values. For example, when the marginal distribution of values is Negative Squared, the seller's revenue is maximized when signals are least informative.



(A) Information effect (marginal, direct)



(B) cutoff effect (marginal, indirect)



(C) Total marginal effect

FIGURE 8. The number of potential bidders $N = 10$: A decomposition of the total marginal effect of signals' informativeness θ on the seller's revenue into the information and cutoff effects for different values of the entry cost k: dashed lines for $k = 0.015$, and solid lines for $k = 0.150$

References

CHEN, X., M. L. GENTRY, T. LI, AND J. LU (2020): "Identification and inference in first-price auctions with risk averse bidders and selective entry," *Available at SSRN 3681530*.

GENTRY, M., AND T. LI (2014): "Identification in Auctions with Selective Entry," *Econometrica*, 82(1), 315–344.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal nonparametric estimation of first-price auctions," *Econometrica*, 68(3), 525–574.

HAILE, P., H. HONG, AND M. SHUM (2003): "Nonparametric tests for common values at first-price sealed-bid auctions," .

KRISHNA, V. (2010): *Auction Theory*. Academic Press, second edn.

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., A. SHNEYEROV, AND P. XU (2013): "What Model for Entry in First-Price Auctions? A Nonparametric Approach," *Journal of Econometrics*, 176(1), 46–58.

NELSEN, R. B. (2007): *An introduction to copulas*. Springer Science & Business Media.

SAMUELSON, W. F. (1985): "Competitive bidding with entry costs," *Economics letters*, 17(1-2), 53–57.

XU, P. (2013): "Nonparametric Estimation Of Entry Cost In First-Price Procurement Auctions," *Journal of Applied Econometrics*, 28(6), 1046–1065.