Inference on Number of Potential Bidders in Selective Entry

Shuo Tian

November 23, 2021

Abstract

This paper studies the inference on number of potential bidders in selective entry for first price sealed bid auctions. The number of potential bidders and information of potential bidders set is of great importance in auction entry problems, however, existing literature often employs ad hoc methods to construct the potential bidders set. I develop a selective entry model allowing number of potential bidders unknown and use it as a benchmark model. By comparing it and the tranditional selective entry model with constructed potential bidders set, the difference between distributions recovered from the two models can reveal whether the construction method of potential bidders set is plausible or optimal. Based on that, I propose a test to distinguish whether any construction of potential bidders set is plausible and reasonable in terms of bidders' private values and beliefs. Finally, I apply these methods on procurement auction data from the Californian Department of Transportation.

1 Introduction

Entry and endogenous participation has been playing an important role in the study of markets. When it comes to auctions, the phenomena of entry that a set of potential entrants endogenously make entry decisions is very common. For auctioneers or policy makers, a major concern of practical design is to attract bidders since participation with too few bidders inevitably leads to an unprofitable and inefficient outcome[5]. Moreover, when bidders endogenously make their entry decision, the occurrence of selection on bidders makes properties of entry bidders less representative of the entire population. This gives particular importance to consideration of entry and endogenous participation in terms of auction research.

Auction entry has been widely studied in the literature, however, when it comes to an empirical topic of a practical auction, the data set often lacks the information of bidders' entry behaviors. To take entry into account, a commonly used method is to assign potential entrants for each auction in the data according to some criterion, and analyze the problem under the constructed potential bidders set. For example, a commonly used method to assign potential bidders is to identify a bidder in the entire data set as a potential bidder of an auction if the bidder participated in at least one another auction in the same district and within a specific time length prior to the bidding date of the auction. This method is frequently used in Timber auctions. (e.g., Li and Zhang, 2010[13]; Athey, Levin, and Seira, 2011[2]; Li and Zheng, 2012[16]; Li and Zhang, 2015[14]). Besides, some of the practical auction data set contains some "proxy" information that can be used to construct potential bidders set. For example, in the highway mowing auction data collected from Texas Department of Transportation (TDoT) (Li and Zheng, 2009[15]), bidders who have requested an official bidding proposal for the project no later than 21 days prior to the letting date are considered as potential bidders. Other research may focus on the entry behavior for large firms, or assume large firms are always potential bidders for each auction in the data set (e.g., Bajari, Hong, and Ryan, 2010[3]; Gil and Marion, 2013[9]).

The information of entry behaviors is of most importance in auction entry related research, however, the methods of constructing potential bidders set are ususally ad hoc. In timber auctions example, the choices of time length differ in the literature. In TDoT auctions, other information like qualification of bidders, advertisement annoucement of the project can also be used as a "proxy" of entry. Therefore, a natural question is how to select the criterion to best fit the data, and how to evaluate whether some construction of potential bidders set is plausible.

To handle this research question, in this article, I propose a criterion to tell whether a construction of potential bidders set is reasonable and consistent in terms of bidders' private values and beliefs. The fundamental idea is to propose a test that is based on two comparable selective entry models. One of the models depends on the information of potential bidders set, and the other one does not. The private values of entry bidders of both models can be identified and recovered respectively from data. If some ad hoc method of construction of potential bidders set is correct, it is believed that the private value distributions recovered from the two models should be very closed to each other. The rationality of a specific construction of potential bidders set therefore can be transformed to the problem of measuring the distance between distributions of entry bidders private valuation recovered from the two models.

Motivated by this idea, I first employ the two-stage selective entry model framework (Marmer, shneyerov, and Xu, 2013[17]; Gentry and Li, 2014[8]) to deal with the constructed potential bidders set with some specific construction method (I denote the selective entry model with constructed potential bidders set by SEMCP). In the first stage, each potential bidder observes a private signal that is correlated with his private value for the object and chooses whether to enter the auction. If entry, the bidder pays an participation cost. Potential bidders can observe the number of potential bidders in the first stage. In stage 2, entry bidders observe their private valuations but not the number of entrants, and submit bids under a first price auction mechanism. The SEMCP model follows largely on Gentry and Li (2014)[8] framework, but the second stage is adjusted to first price auction. I derive the equilibrium strategies and the identification equation for pseudo value estimation. Combined with Gentry and Li (2014)[8] result, the entry threshold, private value of entry bidders, private value distribution in the second stage, participation cost, and the private value distribution in the first stage are all identified.

Second, based on the selective entry model, I further develop a model called selective entry model with an unknown number of potential bidders (denoted by SEMUP) where I assume randomness on the number of potential bidders. The timing and auction mechanism are the same as the SEMCP model, but the information structure changes in SEMUP models. Potential bidders in the first stage cannot observe the actual number of potential bidders, but the distribution of the number of potential bidders. In the second stage, similar to SEMCP models, entry bidders observe their valuations but not the number of entrants. I derive the equilibrium strategies for SEMUP models. The bidding strategy under this setting shows a form of weighted sum of different scenarios where the actual number of potential bidders changes. I then derive a similar equation for identification result in which the private value of entry bidders can be written as the sum of bid and a ratio of two weighted sum of distributions. I follow Song (2015)[22], and derive a similar result in which the private value of entry bidders can be rewritten as a function of bid, distribution of highest bid, and the distribution of second highest bid. This equation circumvent the direct usage of entry threshold, so it makes it possible to estimate the private values without information of potential bidders set.

In the third step, for each entry bidder, I recover his private value for both SEMCP and SEMUP model. Based on the identification result of private values of entry bidders, I develop an estimation procedure to recover the pseudo values of private valuation of entry bidders. From the identification result, the SEMCP model is closed related to the entry threshold, so the estimation result is influenced by the construction method of potential bidders set. On the other hand, the estimation of private values in the SEMUP model has nothing to do with entry information if we control auction level characteristics. Therefore, the rationality of a specific construction of potential bidders set can be transformed to the problem of measuring the distance between the two entry bidders private value distributions recovered from the two models. For example, the criterion of time length for timber auctions described previously can be selected in terms of minimum distance of two distributions. As the problem is transformed to an equality problem of two distributions, I use a Kolmogorov - Smirnov test (KS test), and apply the test result to tell whether a construction of potential bidders set is plausible.

Finally, I apply the entire procedure on construction contracts procurement auctions data from the California Deparment of Transportation (Caltrans). I study the paving contracts by Caltrans from 1999 to 2005. I construct potential bidders set using the commonly used method described previously, i.e., assigning a bidder as a potential bidder if the bidder attends another auction in the same district and within a specific time length prior to the auction. I examine different time length changing from 30 days to 360 days. I structurally estimate the SEMUP model and the SEMCP model for each constructed potential bidders set. I calculate the distance between the two private costs distributions from the two models for every ten days, and depict a scatter plot of the distance and time length. I observe a U - shape curve, and thus I can select an interval of the time length, $[120, 180]$ days, which gives lowest distance between the two models. Besides, I conduct KS tests for each case, and get corresponding bootstrap p values. The p values shows that days around $[120, 180]$ seem to be reasoable choice, since p values there cannot be rejected. Very small time length or very large time length are rejected by the KS test, and this means that the time length should be chosen very carefully.

This paper contributes to the literature of empirical auction research with consideration of endogenous participation. To the best of my knowledge, this paper is the first one that provides a rationale to examine and test the constructed potential bidders set. For example, De Silva et al. (2009)[6] investigate the effect of information on the bidding and survival of entrants in procurement auctions; Li and Zheng (2009)[15] consider entry and competition effects in procurement auctions; Li and Zheng (2010)[16] analyze bidders' entry and bidding decisions in Michigan timber sale auctions using Bayesian method; Li and Zhang (2010)[13] study affiliation using bidders entry behaviors; Athey, Levin, and Seira (2011)[2] discuss endogenous participation under sealed bid and open auctions with heterogeneous bidders; Krasnokutskaya and Seim (2011)[11] study the bid preference program and paricipation decisions in California highway procurement auctions; Li and Zhang (2015)[14] examine the affiliation and entry effects with heterogeneous bidders. In all of these articles, the potential bidders set is either constructed in an ad hoc way, or defined by using some "proxy" information. The article answers the ad hoc question and can be used as a starting model to check out the validity of construction of potential bidders set.

This paper is also closely related to the literature of the identification of auction entry models. The article is based on the selective entry model literature(e.g., Marmer, Shneyerov, and Xu ,2013[17];Gentry and Li, 2014[8]) and auction theory models of entry problem (e.g., Samuelson ,1985[21]; Levin and Smith, 1994[12], Espín-Sánchez and Parra, 2018[7]). The identification of these models depends on the information of bidders entry behaviors. However, the developed SEMUP model in this paper extends the selective entry model in order that the number of potential bidders is random and unobserved. This paper reveals the identification of private values and private value distributions of entry bidders even the potential bidders set is not observed.

This article is organized as follows. Section 2 introduces the two selective entry models and gives the identification of private values of entry bidders. Section 3 describes the estimation details of SEMCP and SEMUP models. Section 4 proposes a method to deal with construction of potential bidders set with a specific continuous variable as a criterion. It also introduces the KS test to tell whether a specific construction is appropriate and plausible. In Section 5, I apply the method and the test on Caltrans data. Section 6 concludes.

2 Models and Identification

In this section, based on the selective entry model, I introduce two entry models - the selective entry model with constructed potential bidders set (SEMCP), and the selective entry model with an unknown number of potential bidders (SEMUP).

2.1 Selective Entry Model with Constructed Potential Bidders Set (SEMCP), Gentry and Li, 2014

The SEMCP model is largely from the selective entry model described in Gentry and Li (2014)[8]. The main change is that the constructed potential bidders set is being used as the potential bidders set and the second stage of the model is adjusted as a first price auction.

In a two-stage selective entry model as developed in Marmer, Shneyerov, and Xu, (2013)[17] and Gentry and Li, (2014)[8], a single and indivisible good is allocated to N symmetric potential bidders via a two-stage auction game, where bidders are assumed to have independent private values for the good. In the first stage, each potential bidder i observes a private signal s_i of her private value v_i. Entry bidders can observe their private values in the second stage. The private signal and private value follow a joint distribution $F(v, s)$. Each potential bidder simultaneously decide whether to enter the auction. If a bidder determine to enter, he will pay an entry cost k which is related to preparation costs, opportunity costs, or uncertainty for risk. In the second stage, n entry bidders observe their private values v_i and proceed with a sealed bid first price auction.

In SEMCP models, I maintain the assumption as mentioned in Gentry and Li (2014)[8] that bidders could observe the number of potential bidders N prior to entry, but do not observe the number of entrants n until the auction concludes. SEMCP models basically follows the selective entry model in the literature, though the information of potential bidders are observed or obtainable in terms of some constructed potential bidders set.

2.1.1 Equilibrium in a SEMCP Model

As I mentioned previously, I denote the joint distribution between each bidder's private value and private signal by $F(v, s)$. I make the following assumption on this joint cumulative distribution function.

Assumption 1 *The value-signal pairs (V_i, S_i) are drawn symmetrically across bidders from a joint distribution such that,*

1. *The support of the random variable V_i is a bounded interval $\mathcal{V} = [0, \bar{v}]$, and the joint CDF $F(v, s)$ is continuous in (v, s).*
2. *For each bidder i, the conditional distribution of V_i is stochastically ordered in S_i, i.e., if for $\forall s' \geq s$, $F(v|s') \leq F(v|s)$.*
3. *The random pairs (V_i, S_i) are independent across bidders: $(V_i, S_i) \perp (V_j, S_j)$ for $\forall j \neq i$.*
4. *First stage signals S_i is normalized to have a uniform marginal distribution on $[0, 1]$: $S_i \sim U[0, 1]$.*

As mentioned in Gentry and Li (2014)[8], their model employs a weaker restriction on the connection of private value and private signal. They assume that the two variables follow restriction of stochastic ordering rather than affiliation restriction like, Milgrom and Weber (1982)[19], Ye (2007)[23], and Marmer, Shneyerov, and Xu (2013)[17].

Beyond the assumption 1, I impose an additional regularity condition on the relationship between values and signals as in Gentry and Li (2014)[8]. This condition provides some level of smoothness so that the equilibrium exists.

Assumption 2 *The conditional expectation $E[V_i|S_i = s]$ is continuous s on $[0, 1]$.*

To apply the result on the procurement auction of interest, I focus on the sealed bid first price auction as the mechanism in the second stage.

The strategy of each bidder consist of two parts, the entry strategy, $\bar{s}$ where bidders choose to entry if their private values $s_i \geq \bar{s}^*$, and the bidding strategy, $\beta(v; \bar{s})$ where bidders make their bids $\beta(v; \bar{s})$ under a symmetric entry threshold $\bar{s}$.

The symmetric pure strategy Bayesian Nash equalibrium of the two-stage auction game $(\bar{s}^*, \beta^*)$ in this model can be characterized as,

(a) the entry threshold strategy $\bar{s}^*$ can be characterized

$$\Pi^{cp}(\bar{s}|\bar{s}, N) = k \tag{1}$$

where $\Pi^{cp}(s_i|\bar{s}, N)$ is the expected utility when bidder i with a private signal s_i decides to enter and all other bidders holds symmetric entry strategy $\bar{s}$.

(b) for all v_i,

$$\beta^*(v_i; \bar{s}^*) = \arg\max_b \text{Prob}(\max_{j \neq i} \beta^*(v_j; \bar{s}^*) \leq b)(v_i - b) \tag{2}$$

The equilibrium can be solved as the following procedure. The distribution of values among entrants with entry strategy $\bar{s}$ can be derived as,

$$F^*(v; \bar{s}) \equiv \frac{1}{1 - \bar{s}} \int_{\bar{s}}^{1} F(v|t)dt \tag{3}$$

For an active bidder with a private value v at bidding stage, the probability that he considers himself to win against $N - 1$ potential rivals is,

$$F^*_{1:N-1}(v; \bar{s}) = [\bar{s} + (1 - \bar{s})F^*(v; \bar{s})]^{N-1} \tag{4}$$

Suppose I consider a symmetric bidding strategy, $\beta(\cdot)$, which is an increasing continuous function. The probability for an active entry bidder with bid b and $N - 1$ potential rivals to win is,

$$G^*_{1:N-1}(b;\bar{s}) = \text{Prob}(\text{winning with bid } b|\bar{s}) \tag{5}$$

$$= [\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{N-1} \tag{6}$$

In this situation, bidders have certain belief of the potential number of bidders, N, so the expected payoff therefore equals,

$$[\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{N-1}(v-b) \tag{7}$$

From the first order condition, I can characterize $\beta(\cdot)$ as,

$$\beta(v;\bar{s}) = v - \frac{1}{[\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1}} \int_0^{\bar{v}} [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1} dv \tag{8}$$

The expected payoff of a bidder with private value v in the bidding stage is,

$$\pi^{cp}(v;\bar{s},N) = [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1}(v - \beta(v;\bar{s})) \tag{9}$$

The expected payoff of a bidder with private signal s in the entry stage is,

$$\Pi^{cp}(s;\bar{s},N) = \int_0^{\bar{v}} \pi^{cp}(v;\bar{s}) f(v|s) dv \tag{10}$$

The entry threshold is characterized by,

$$\Pi^{cp}(\bar{s}^*;\bar{s}^*,N) = k \tag{11}$$

Likewise, Equation (11) and Equation (8) can jointly specified the equilibrium strategy for a potential bidder. Following the Proposition 1 in Gentry and Li (2014)[8], the equilibrium $(\bar{s}^*, \beta^*(\cdot))$ exists uniquely under Assumptions 1-2.

2.1.2 Identification for a SEMCP Model

The primitives of selective entry model are identified on the support of entry threshold, including joint distribution of private value and private signal pairs, $F(v, s)$, entry strategy, $\bar{s}^*$, and entry cost, k, if extra assumptions on a cost shifter are imposed (Gentry and Li, 2014)[8].

In this section, I mainly present that private values of entry bidders for a SEMCP model are identified under the second stage being a first price sealed bid auction.

For the entry bidders, they will choose their b to maximize their expected payoff given entry threshold $\bar{s}$,

$$\max_{b} \; [\bar{s} + (1 - \bar{s})G^*(b; \bar{s})]^{N-1}(v - b) \tag{12}$$

The first order condition lead to the following result as shown in the Appendix A,

$$v = b + \frac{1}{N-1} \frac{\bar{s} + (1 - \bar{s})G^*(b; \bar{s})}{(1 - \bar{s})g^*(b; \bar{s})} \tag{13}$$

The entry threshold $\bar{s}$ is identified from the data from the conclusion of Gentry and Li (2014)[8]. The bid distribution $G^*(\cdot)$ and $g^*(\cdot)$ are identified from the entry bids in the data. The number of potential bidder and each bidder's bid is directly observed in the data. Therefore, the private value and the private value distribution for the entry bidders are identified.

2.2 Selective Entry Model with an Unknown Number of Potential Bidders (SEMUP)

In practical auction, as the number of potential bidders is often uncertain, I develop an extension to the traditional selective entry model. Rather than considering that all bidders can observe the number of potential bidders N, bidders are assumed to have a belief on the distribution of the number of potential bidders. I denote this extended model by selective entry model with an unknown number of potential bidders (SEMUP). SEMCP models can be considered as special cases of the SEMUP model where the discrete distribution of N collapse to a point mass probability on some specific value of N.

To model the stochastic number of potential bidders, I combine the selective entry model with the feature of auction models with an unknown number of bidders (e.g., McAfee and MacMillan, 1987[18], Song, 2015[22]).

2.2.1 Model Setup

At time 0 (Nature Stage), auctioneer proposes a project and nature selects a pool of active bidders (potential bidders set), A, with number of elements, N. N can take a value up to $\bar{N}$. The distribution of N are denoted by a sequence of probabilities,

$$p_l = \text{Prob}(N = l) \tag{14}$$

which represents the probability that $N = l$.

At time 1 (Entry Stage), potential bidders are informed that they are active. They cannot observe the true value of N, but have a belief related to the number of competitors, where

$$q_l^i = \text{Prob}(N = l | i \in A) \tag{15}$$

which means an active bidder i's belief of the number of potential bidders is l. I assume symmetric beliefs, so I define

$$q_l = q_l^i \quad \text{for all } i \tag{16}$$

Each active bidder is informed a private signal s_i which is correlated with his private value v_i with a joint distribution $F(v_i, s_i)$. Potential bidders choose whether to enter. If entry, the bidder pays a participation (preparation) cost k. The entry strategy is characterized as a threshold strategy, $\bar{s}$, i.e., if $s \geq \bar{s}$, the bidder will choose to enter.

At time 2 (Bidding Stage), entry bidders observe their private values, v_i. They make their bids based on their information. I assume that bidders only observe the distribution of the number of potential bidders N, but cannot observe the number of entrants n before making their bids. The auction is assumed to be a first price sealed bid auction. The bidder with the highest bid wins.

2.2.2 Equilibrium for a SEMUP Model

The symmetric pure strategy Bayesian Nash equalibrium of the two-stage auction game $(\bar{s}^*, \beta^*)$ in a SEMUP model can be characterized as,

(a) the entry threshold strategy $\bar{s}^*$ can be characterized

$$\Pi^{up}(\bar{s}|\bar{s}) = k \tag{17}$$

where $\Pi(s_i|\bar{s})$ is the expected utility when bidder i with a private signal s_i decides to enter and all other bidders holds symmetric entry strategy $\bar{s}$.

(b) for all v_i,

$$\beta^*(v_i; \bar{s}^*) = \arg\max_b \text{Prob}(\max_{j \neq i} \beta^*(v_j; \bar{s}^*) \leq b)(v_i - b) \tag{18}$$

The equilibrium of SEMUP models can be solved similarly. The distribution of values among entrants with entry strategy $\bar{s}$ can also be derived as since the distribution is only influenced by the entry strategy rather than the uncertainty of the number of potential bidders,

$$F^*(v; \bar{s}) \equiv \frac{1}{1-\bar{s}} \int_{\bar{s}}^{1} F(v|t)dt \tag{19}$$

If an active bidder believes that the number of potential bidders is N, an active bidder with a private value v at bidding stage considers the probability of winning the auction against $N-1$ potential rivals is,

$$F^*_{1:N-1}(v;\bar{s}) = [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1} \tag{20}$$

Similar to the SEMCP models, suppose I consider a symmetric bidding strategy, $\beta(\cdot)$, which is an increasing continuous function. Under the belief of N potential bidders, the probability of winning for a bidder with bid b and $N-1$ potential rivals is,

$$\begin{aligned} G^*_{1:N-1}(b;\bar{s}) &= \text{Prob}(\text{winning with bid } b|\bar{s}) && (21)\\ &= [\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{N-1} && (22) \end{aligned}$$

With consideration of all possible number of potential bidders, the probability that one believes he will win is,

$$\sum_{l=1}^{\bar{N}} q_l[\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{l-1} \tag{23}$$

The expected payoff therefore equals,

$$\left\{\sum_{l=1}^{\bar{N}} q_l[\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{l-1}\right\}(v-b) \tag{24}$$

From the first order condition, I can characterize $\beta(\cdot)$ as (See details in the Appendix D),

$$\beta(v;\bar{s}) = \sum_{l=1}^{\bar{N}} \frac{q_l[\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{l-1}}{\sum_{l=1}^{\bar{N}} q_l[\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{l-1}} \left(v - \frac{1}{[\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{l-1}} \int_0^v [\bar{s} + (1-\bar{s})F^*(u;\bar{s})]^{l-1} du\right) \tag{25}$$

The expected payoff of a bidder with private value v in the bidding stage is,

$$\pi^{up}(v;\bar{s}) = \left\{ \sum_{l=1}^{\bar{N}} q_l [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{l-1} \right\} (v - \beta(v;\bar{s})) \tag{26}$$

The expected payoff of a bidder with private signal s in the entry stage is,

$$\Pi^{up}(s;\bar{s}) = \int_0^{\overline{v}} \pi^{up}(v;\bar{s}) f(v|s) dv \tag{27}$$

The entry threshold is characterized by,

$$\Pi^{up}(\bar{s}^*;\bar{s}^*) = k \tag{28}$$

After $\bar{s}^*$ is specified, the bidding strategy at the second stage can be characterized by plugging in $\bar{s}^*$ to Equation (25), i.e., $\beta(v;\bar{s}^*)$.

2.2.3 Identification for Private Values

Unlike SEMCP models, econometrician and potential bidders do not observe the actual realization of the number of potential bidders. Not all of the model primitives can be identified. In this section, I show that the private values and private values distribution can be identified by employing the fact that the pseudo private values can be written as a function of bids, distribution of highest bids and second highest bids.

From the equation of payoff maximization, we have,

$$\left\{ \sum_{l=1}^{\bar{N}} q_l [\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1} \right\} (v - b) \tag{29}$$

where $G^*(b;\bar{s})$ can be considered as the bid distribution at bidding stage conditional on entry threshold $\bar{s}$.

FOC suggests that,

$$(v-b)\left\{\sum_{l=2}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-2}(l-1)(1-\bar{s})g^*(b;\bar{s})\right\} = \sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}$$

$$v = b + \frac{\sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}}{\sum_{l=2}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-2}(l-1)(1-\bar{s})g^*(b;\bar{s})}$$

From the fact that,

$$q_l = lp_l/N^*$$

where N^* is defined as the expected number of potential bidders.

Therefore we have,

$$v = b + \frac{\sum_{l=1}^{\bar{N}} p_l l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}}{\sum_{l=2}^{\bar{N}} p_l l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-2}(l-1)(1-\bar{s})g^*(b;\bar{s})} \tag{30}$$

If I consider the distribution of winning bids[1], it can be characterized as (see proof in Appendix F),

$$G^{(1)}(b;\bar{s}) = \frac{\sum_{l=1}^{\bar{N}} p_l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l} - \sum_{l=1}^{\bar{N}} p_l\bar{s}^l}{1-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l} \tag{31}$$

The distribution of the second highest bid can be expressed in a form of the distribution of the highest bid,

$$G^{(2)}(b;\bar{s}) = \frac{(1-\bar{s})(1-G^*(b;\bar{s}))\sum_{l=1}^{\bar{N}} p_l l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}}{1-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l} + G^{(1)}(b) \tag{32}$$

Private values can be identified from equation (30) from the fact that,

$$v = b + \frac{G^{(2)}(b;\bar{s}) - G^{(1)}(b;\bar{s})}{g^{(2)}(b;\bar{s})} \tag{33}$$

[1] If there is no winning bid, I assume that the winning bid is 0

where $g^{(2)}(b;\bar{s})$ is the PDF of the second highest bids. $G^{(1)}(b;\bar{s})$ and $G^{(2)}(b;\bar{s})$ can be identified directly from bids. Therefore, the private values of the entry bidders can be directly identified from bidders' bids, the distribution of highest entry bid and that of second highest entry bid.

3 Structural Estimation

Since I will deal with procurement auction data in the application section, this structural estimation part will be written in a private costs framework. The estimation procedure for a SEMCP model is based on the formula derived in the identification part but private values here are replaced by private costs in the situation of procurement auctions,

$$c_i = b_i - \frac{1}{N-1}\frac{1-\bar{s}G^*(b;\bar{s})}{\bar{s}g^*(b;\bar{s})} \tag{34}$$

where $\bar{s}$ refers to the entry threshold that bidders choose to enter if their private signal in the first stage is less than $\bar{s}$.

Based on the method in Guerre, Perrigne, and Vuong (2000)[10], pseudo values of private costs of entry bidders can be estimated by plugging in their bids, number of potential bidders, entry threshold and bid distribution in the bidding stage. The private cost distribution of entry bidders can then be estimated through the empirical cumulative distribution function derived from the pseudo private costs.

There are several auction level characteristics that makes the estimation difficult to handle. To account for the auction level characteristics and avoid the curse of dimentionality, I assume that, the distribution in the first stage, $F(v|s)$, is influenced by a single index of auction level characteristics, which is denoted by $F(v|s,z'\beta)$. In the application section, the variable z is a vector contains $[job \quad ave.fri \quad ave.dis \quad log(est.cost)]$ where job refers to a dummy variable to indicate whether the project is major construction, $ave.fri$ is the average distance between the locations of contractors and the project, $ave.fri$ is the average of a dummy variable $fringe$ which indicates whether a firm is small firm and $log(est.cost)$ refers to the engineer estimate of the project calculated by the government. The $ave.fri$ and $ave.dis$ are calculated from entry bidders.

In a selective entry model, as mentioned in Gentry and Li (2014)[8], the identification relies on a participation cost shifter which is assumed to have effects on participation cost, but will not shift the private cost distribution. I denote this variable by x in the model. In the application, the variable $ave.uti$ is employed as the cost shifter in the estimation procedure. $ave.uti$ is the average utilization rate of entry bidders which indicate the average level of workload of contractors at that period of time. If the variable is large, bidders are quite busy so that the cost of preparing a new project will be increased. On the other hand, this variable is not closely related to the working cost of projects, thus it is plausible to consider that the private cost distribution is not affected by the variable. Therefore, it might be a plausible option to pick $ave.uti$ as the cost shifter.

Based on the assumptions above, the equation (34) can be rewritten as,

$$\hat{c}_{ji} = b_{ji} - \frac{1}{N_j - 1} \frac{1 - \hat{\bar{s}}(z_j'\beta, x_j)\hat{G}^*(b_{ji}|z_j'\beta, x_j)}{\hat{\bar{s}}(z_j'\beta, x_j)\hat{g}^*(b_{ji}|z_j'\beta, x_j)} \tag{35}$$

From this setting, I estimate the SEMCP model with the following procedure,

Step 1, Run a linear regression of the natural logarithm of bids on auction level characteristics z_j and cost shifter x_j.

Step 2, From the estimated coefficients of auction level characteristics, get the value of single index for each auction.

Step 3, From the error terms of linear regression, estimate the distribution of the error terms. Based on the connection between the distribution of bids and the distribution of error terms, calculate the estiamted distribution terms.

Step 4, Based on the calculated single index, and cost shifter variable, estimate the entry threshold using a binomial logistic regression.

Step 5, The pseudo values of private costs can be estimated by plugging the bids, the number of potential bidders, the estimated entry threshold, and the estimated distribution terms.

3.1 Estimation of Single Index and Distribution of Bids in the Second Stage

I employ a single index approach so that all the characteristics in the initial distribution $F(v|s, z'\beta)$ influence the model primitives only through the scalar $z'\beta$. From this setting, I assume contractors' bidding behaviors follow the form,

$$\log(b_{ji}) = \beta_0 + z_j'\beta + x_j'\zeta + e_{ji} \tag{36}$$

Meanwhile, the distribution in the second stage $G^*(\cdot)$ and $g^*(\cdot)$ can be estimated through the distriubtion of error terms. The distributions are connected through the following relation,

$$G_{B|Z,H}(b_{ji}|z_j, h_j) = Prob(\alpha_0 + z_j'\beta + x_j'\zeta + e_{ji} \leq \log(b_{ji})) = F_e(\log(b_{ji}) - \alpha_0 - z_j'\beta - x_j'\zeta) \tag{37}$$

and

$$g_{B|Z,H}(b_{ji}|z_j, h_j) = \frac{\partial}{\partial b} F_e(\log(b_{ji}) - \alpha_0 - z_j'\beta - x_j'\zeta) = f_e(\log(b_{ji}) - z_j'\beta - x_j'\zeta)/b_{ji} \tag{38}$$

I adopt kernel density estimation to acquire the CDF and PDF of error terms e_{ji}, and finally get the value of distribution of entry bids through the transformation in equation (37) and equation (38).

3.2 Estimation of Entry Thresholds

I employ a binomial logistic regression with $n_j \sim \mathcal{B}\left(N_j, \frac{\exp(\gamma m_j)}{1+\exp(\gamma m_j)}\right)$ to estimate the entry thresholds $\bar{s}(z_j\beta, x_j)$. The variable m_j contains constant term, the combination of the single index estimated in the last section, and the cost shifter x_j.

After the estimation of entry threshold and distributions in the second stage, I can recover the pseudo values of private costs of entry bidders by plugging in the estimates into equation (35).

3.3 Estimation of the SEMUP model

Similar to the procedure of the estimation of a SEMCP model, the estimation of a SEMUP model is based on the equation below,

$$\hat{c}_{ji} = b_{ji} - \frac{\hat{G}^{(L1)}(b_{ji}; z_j'\beta, x_j) - \hat{G}^{(L2)}(b_{ji}; z_j'\beta, x_j)}{\hat{g}^{(L2)}(b_{ji}; z_j'\beta, x_j)} \tag{39}$$

To get distributions of the lowest and second lowest bids, I follow the linear regression in equation (36) and get the error terms for each bidder. The lowest bid distribution is estimated through a kernel density using the error terms of lowest bids and the second lowest bid distribution comes from the error terms of second lowest bidders. Based on the estiamted distributions, I predict the value of distributions for each error term from the kernel density function. After the estimation of distribution values, I can recover the pseudo values of private costs by plugging in estimates into equation (39).

4 Test on Construction of Potential Bidders Set

In this section, I first consider methods of constructing potential bidders set when a continuous variable is applied as a rule to distinguish potential bidders from all bidders. For example, in Li and Zhang(2010[13], 2015[14]) and Li and Zheng (2012)[16], the potential bidders set is formed by adding all actual bidders in the data set who participated in at least one auction in the same district and during a specific time length before the bidding date of the auction. In Li and Zhang (2010)[13], this time length is set equal to 90 days (a quarter) and in Li and Zhang (2015)[14], it is set equal to 365 days (a year). Meanwhile, in Li and Zheng (2012), this value is set equal to 30 days (a month). I can consider these articles are taking similar methods to construct potential bidders set, i.e., a continuous variable, the time length prior to auctions, is utilized as a criterion to decide a potential bidder. A natural question for this procedure is, how many days is the most appropriate to construct a potential bidders set.

As I mentioned in the last section, the SEMUP model is a more general model than every SEMCP model. Meanwhile, from Equation (33) and Equation (13), based on the estimation method in Guerre, Perrigne, and Vuong (2000)[10], I can recover pseudo values of private values for the two models. From models setting, SEMUP model is invariant to different methods of constructing potential bidders set. However, SEMCP models will give different results of pseudo private values. Thus, the difference between pseudo private values of the SEMUP model and the SEMCP model can be used to tell whether the constructing method of potential bidders set is consistent to the recovered bidders strategies and beliefs. In this article, I choose the distance between the distribution of private values of entry bidders from SEMUP model and SEMCP model as a rationale to tell which SEMCP model gives the most close estimates of private values.

Suppose the constructing method of potential bidders set is based on some rules in terms of a continuous variable. I denote this variable by ξ with domain Ξ. The model primitives of interest are the distribution of private values for entry bidders. I denote the private values distribution of entry bidders in a SEMUP model as $F_{up}(v)$, and that in a SEMCP model as $F_{cp}(v;\xi)$. The distance between the two distribution is defined as,

$$d(\xi) = \sup_{v \in [0,\bar{v}]} |F_{up}(v) - F_{cp}(v;\xi)| \tag{40}$$

From this setting, the minimizer ξ^* of the function $d(\xi)$ on its support is of most interest. I prove in Proposition 1 that, under proper conditions, the minimizer of function $d(\xi)$ exists.

Proposition 1 (Existence of Minimum Distance) *Assume that the class of function $F_{up}(v;\xi)$ is continuous with respect to ξ for any $v \in [0, \bar{v}]$, and that ξ has a compact support Ξ. The function $d(\xi)$ is bounded below and attains its infimum on the support.*

In empirical work, we can estimate the pseudo private values of entry bidders, and estimate the empirical CDF for these pseudo values. Therefore, I can get the estimate of distance from,

$$\hat{d}(\xi) = \sup_{v\in[0,\bar{v}]} \left| \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{v}^j_{up,i} \le v\} - \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{v}^j_{cp,i} \le v\} \right| \tag{41}$$

where $\hat{v}^j_{up,i}$ and $\hat{v}^j_{cp,i}$ represent the estimated pseudo values of private values of bidder i in auction j from a SEMUP model and a SEMCP model respectively. To get a SEMCP model that is closest to the SEMUP model, I can choose $\hat{\xi}^*$ to minimize the estimated distance $\hat{d}(\xi)$ which is the continuous measure to choose to construct the potential bidders set.

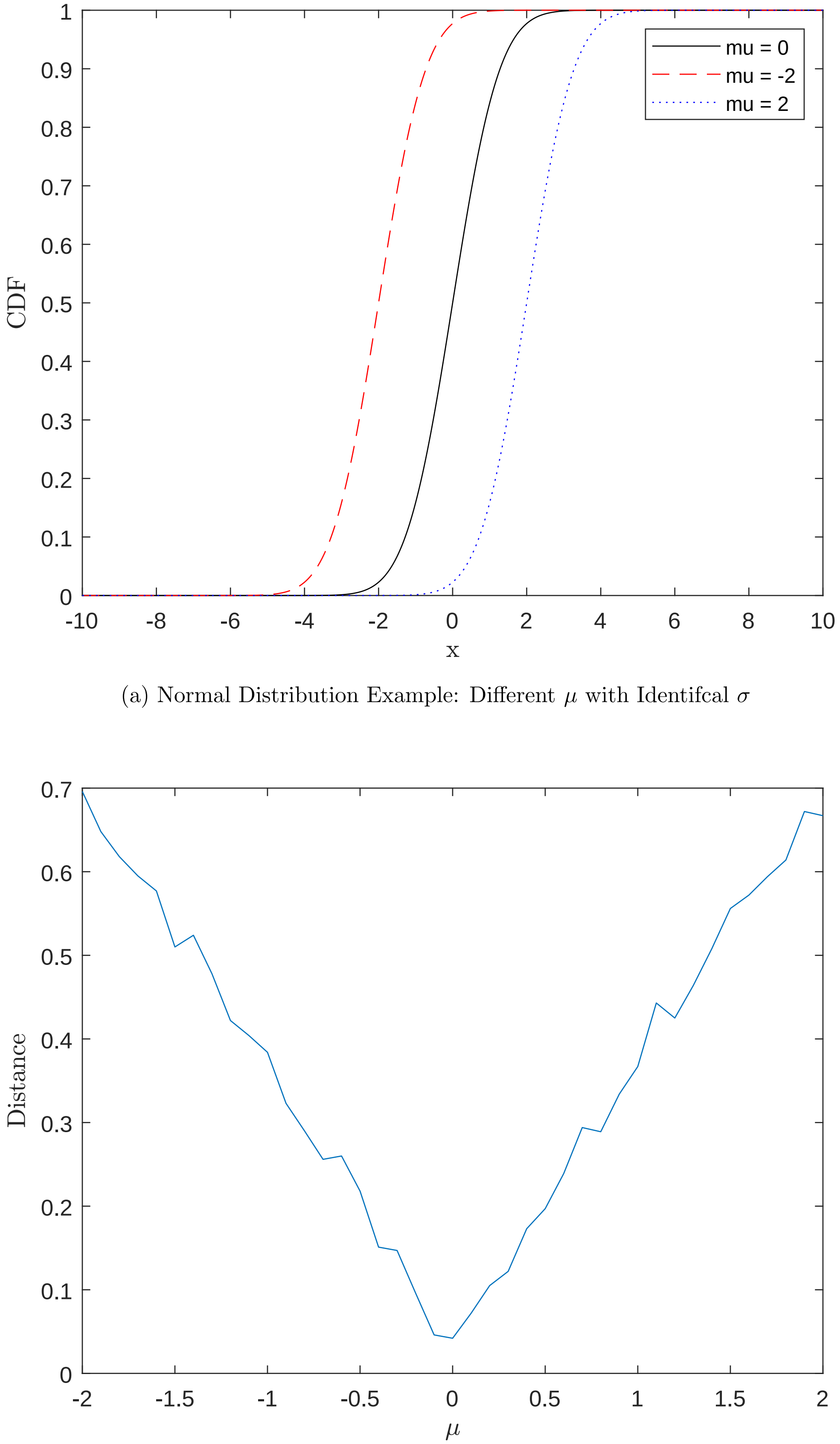
Proposition 1 provides the existence of the minimizer ξ. However, the uniqueness of the minimizer is not necessary. In fact, the property of the distance function $d(\xi)$ depends on the shape of the two distribution function, $F_{cp}(v;\xi)$, and $F_{up}(v)$.

For example, in Figure 1(a), the black solid line represents a standard normal distribution function, i.e., the mean $\mu = 0$, and the standard deviation $\sigma = 1$. The red dashed line and the blue dotted line refers to the CDF distribution with different mean $\mu = -2$ and $\mu = 2$ with identical $\sigma = 1$. If the black line is assumed to be the distribution of a SEMUP model, the process of selecting SEMCP models is similar to find the closest distribution of a SEMCP model by choosing μ on its support. When SEMCP models shift from the left (red dashed line) to the right (blue dotted line), the distance of the two models is decreasing at first, attains minimum distance 0 (completely overlapping), and then increasing afterwards.

In Figure 1(b), I make a simple simulation based on the previous setting. I first generate random draws from standard normal distribution as pseudo values of a SEMUP model. For each different value of $\mu \in [-2, 2]$, I generate a group of random draws based on the value of μ as the mean of distribution and standard deviation $\sigma = 1$. Based on the random draws, for each value of μ on the interval, I can construct the empirical CDF from the random draws. From the random draws of the initial standard normal distribution, I can similarly recover the empirical CDF. Therefore, for each value of μ, I can compute the distance between the empirical CDF from random draws of specific μ and the empirical CDF from the initial standard normal distribution random draws. Figure **??** shows it works quite good as the lowest distance occurs around $\mu = 0$.

However, when the shape of distributions, $F_{cp}(v; \xi)$ and $F_{up}(v)$ are not completely the same, the minimum of the distance function of two distributions is hard to analyze and even not unique. For example, in Figure 2(a), the solid line refers to the CDF of normal distribution with $\mu = 0, \sigma = 1$, and dashed lines represent different CDF of normal distributions with different values of μ and $\sigma = 2$. When the value of μ changes, different functions of dashed line shift from the left to right, it is not clear which one keeps the lowest distance especially when in the real case, more volatility exists in the shape of functions. Under this setting, a similar simple simulation is done, and the result is shown in Figure 2(b). Unlike a clear minimum in the previous example, a short interval with small distance appears around $\mu = 0$, and at the same time, more fluctuation occurs in the shape of distance - μ curve.

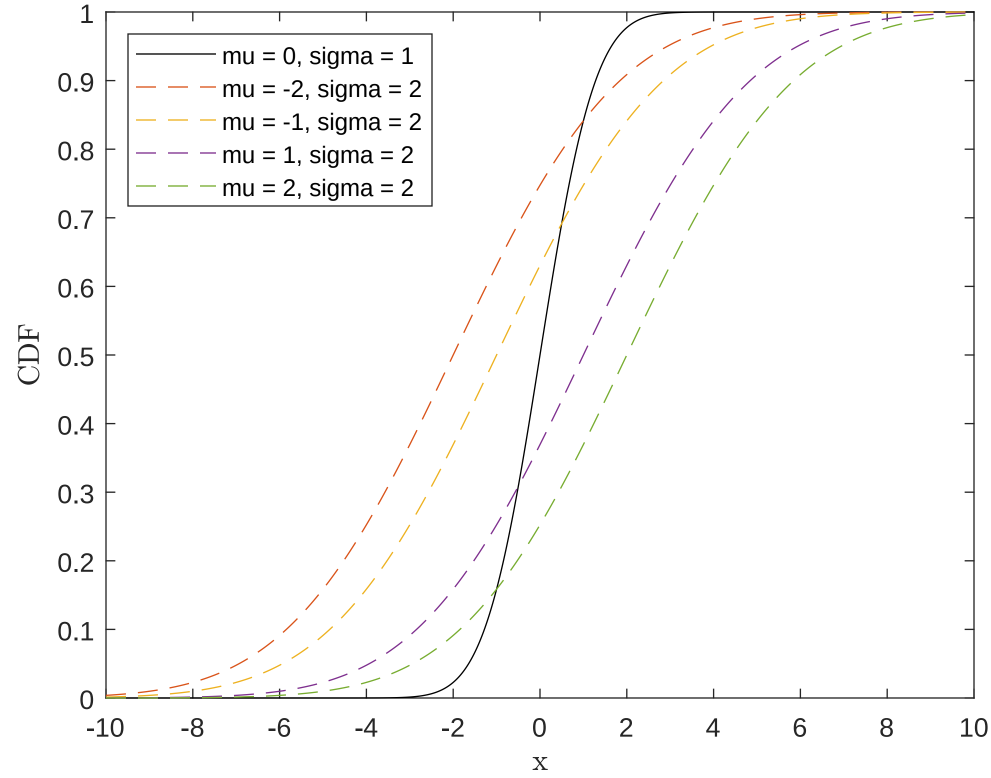
To make the criterion more stringent, I suggest that a Kolmogorov-Smirnov (KS) test can be employed to handle the problem. In empirical work, sometimes an econometrician only cares for some specific potential bidders set, like all large firms, or all bidders in the data set as potential bidders. In another case, some researcher may use some information as a "proxy" to distinguish potential bidders. In these situations, econometricians already have information on constructed potential bidders set, but how do we know whether this potential bidders set is plausible or not. Furthermore, as stated previously, when one assigns



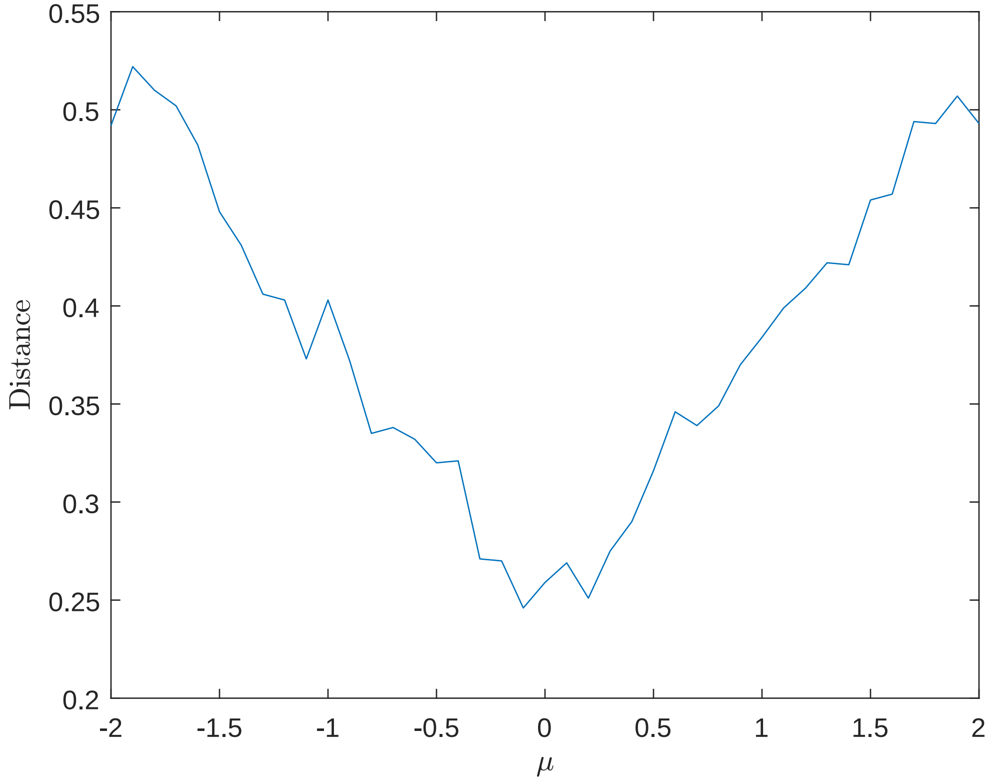
(a) Normal Distribution Example: Different μ with Identifcal σ

(b) Normal Distribution Example: Simulated Distance under Different μ with Identifcal σ

Figure 1: Normal Distribution Example with the Same Shape



(a) Normal Distribution Example: Different μ with Different σ



(b) Normal Distribution Example: Simulated Distance under Different μ with Different σ

Figure 2: Normal Distribution Example with Different Shapes

a specific number of time length, does it have to give a good performance on the recovery of potential bidders set, especially when multiple minimum exist from estimation? To resolve this problem, I extend the previous approach to form a test to determine whether some specific constructed potential bidders set is appropriate or not.

For a specific method of constructing potential bidders set, both SEMCP and SEMUP can be estimated. It is natural to consider that if the two models represent the same equilibrium, they will show the same private values distribution. Therefore, I can formulate the KS test to tell whether the two value distributions recovered from the data are the same.

$$H_0 : \forall v \in [0, \bar{v}] \quad F_{cp}(v) = F_{up}(v);$$
$$H_1 : \exists v \in [0, \bar{v}] \quad F_{cp}(v) \neq F_{up}(v).$$

The test statistic for the model can be characterized as,

$$KS = \sup_{v \in [0,\bar{v}]} \left| \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{1}^{n_j} 1\{\hat{v}^j_{up,i} \leq v\} - \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{1}^{n_j} 1\{\hat{v}^j_{cp,i} \leq v\} \right| \tag{42}$$

I can bootstrap the density of the test under the null to compute the critical value t_α at $\alpha\%$ significance level as introduced in Abadie (2002)[1].

5 Application to the Caltrans Procurement Data

In this section, I examine the frequently used method in literature that uses time information as a criterion to distinguish potential bidders, i.e., the potential bidders set is organized by all actual bidders in the data set who ever joined in at least one auction in the same district and during a specific period of time prior to the bidding date of the auction. I follow the method mentioned in the previous section and regard the time length as a continuous variable.

5.1 CalTrans Auctions: Data

I employ the data set of Caltrans procurement auctions collected by Bajari, Houghton, and Tadelis (2013)[4]. I study the paving contracts by Caltrans from 1999 to 2005. The data sample of interest consists of 819 projects and 3661 bids by 349 general contractors. Table 1 and Table 2 decribe some summary statistics of observed characteristics on auction level and contractor level.

In the table, *engineering estimate* variable is the estimated cost of each project prepared by government engineers. job variable includes the information of project construction type: $job = 1$ represents major construction and $job = 0$ for minor construction. $fringe$ is defined as a dummy variable for contractors. When a firm wins less than 1% of the value of the contracts, $fringe$ is set to equal to 1.

The utilization rate variable, *util*, is a variable that describes a contractor's production capacity and backlog situation. Following Bajari, Houghton, and Tadelis(2014)[4], I use the data of winning bids, bidding dates and contract days in the sample data set to construct this variable. First, I assuming each project is achived at a constant speed, a *backlog* variable for a contractor at a specific auction is defined as the summation of the remaining value of the projects on which the contractor is working at the bidding date of the auction. Second, the variable *capacity* is defined as the maximum *backlog* of the contractor. Finally, the utilization rate *util* is defined as *backlog/capacity* which measures the production capability and opportunity cost of a contractor.

dist is the distance between the potential bidder and the project. The sample data set gives addresses of contractors and locations of projects. I employ the location information, and geographically calculate the distance between any possible combinations of a bidder and a project/footnote I use the Python package geopy to transform the locations into a pair of latitude and longitude. I then use function *geodesic* to calculate the direct distance between the two pairs of latitude and longitude. The distance is measured in miles. . As the potential bidders are of my most interest, the distance between potential bidders and projects are also calculated. This variable measures the geographic cost advantage. The contractors with shorter distance is expected to have a lower transportation cost which might be considered as influential to the entry behaviors.

Table 1: Auction level summary statistics

Variables	mean	standard deviation	median
number of entry bidders	4.47	2.15	4
engineering estimate (1m dollars)	2.88	7.25	0.95
job	0.40	0.49	0
ave.fringe	0.56	0.29	0.60
ave.distance (miles)	102.89	84.55	81.73
ave.util	0.11	0.12	0.07

Table 2: Contractor level summary statistics

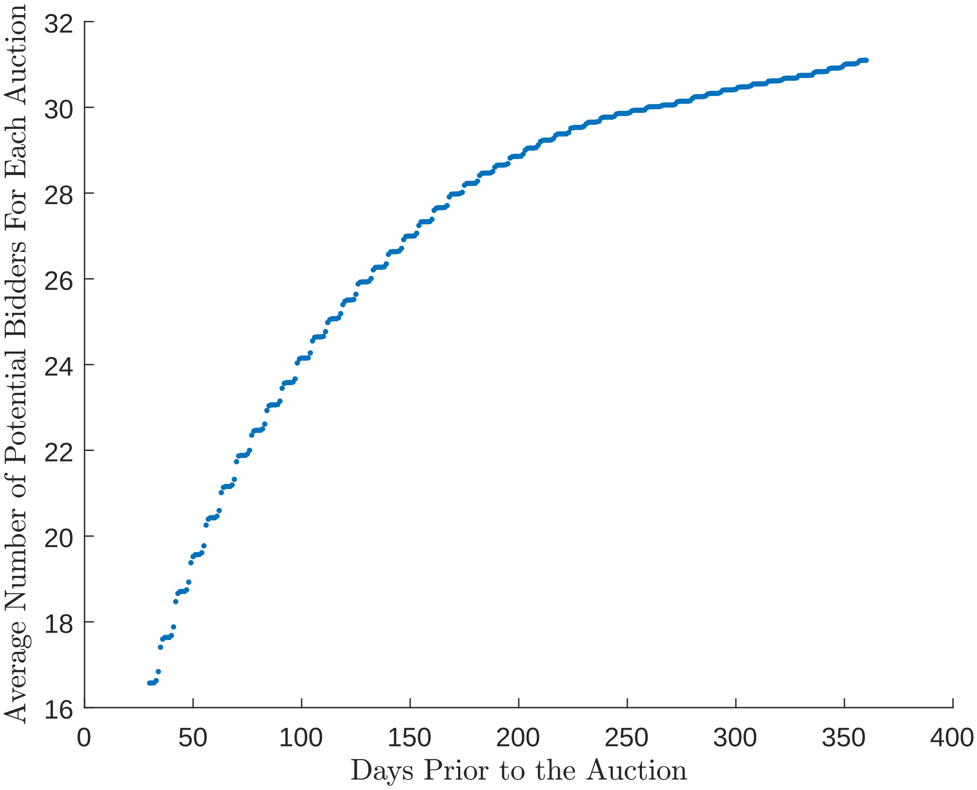
Variables	mean	standard deviation	median
fringe	0.94	0.23	1
distance (miles)	98.07	167.54	53.22
util	0.04	0.08	0
job	0.46	0.38	0.43
engineering estimate (1m dollars)	3.21	8.95	0.93

5.2 Application Result: Time as a Criterion

In this section, I employ the private cost distribution of the entry bidders. When it comes to auction level hetereogeneity, the distribution in this method can be regarded as a weighted sum of conditional distribution of private costs conditional on auction level characteristics. This is proved in the Appendix B. In fact, a similar result can be found if auction level characteristics is excluded in advance.

I consider the method of creating potential bidders set that is very frequently used in the literature (For example, Li and Zhang, 2010[13], 2015[14]; Li and Zheng, 2012[16]). I recover the set of potential bidders by adding all actual bidders who participated in at least one auction in the same district and during a specific time length prior to the bidding date of the auction.

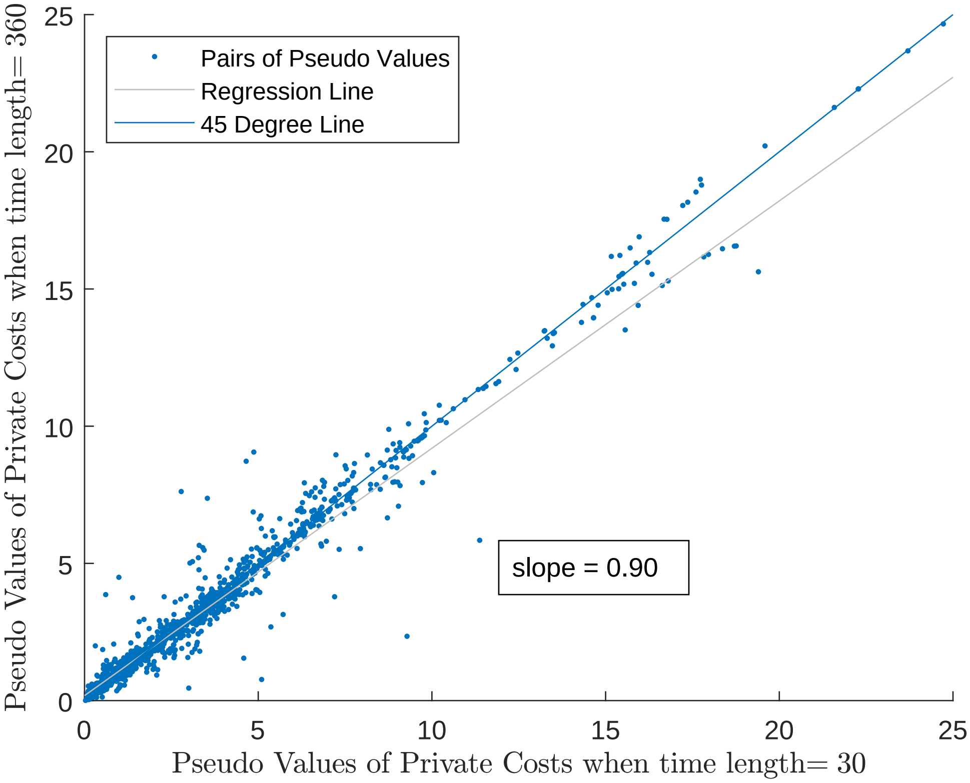
Figure 3: Average Number of Potential Bidders for Each Auction



For different value of time length, the number of potential bidders for each auction changes. Figure 3 shows the relation between the average number of potential bidders of each auction and the value of time length. From the figure, average number of potential bidders for each auction is non decreasing as more bidders are assigned as potential bidders when the time length gets longer. The curve increase quite rapidly at small value of time length, then the increasing rate slows down for higher time length values. This comes from the fact that when the time length grows very large, there will be a maximum for the average number of potential bidders since the number of auctions and bidders is limited.

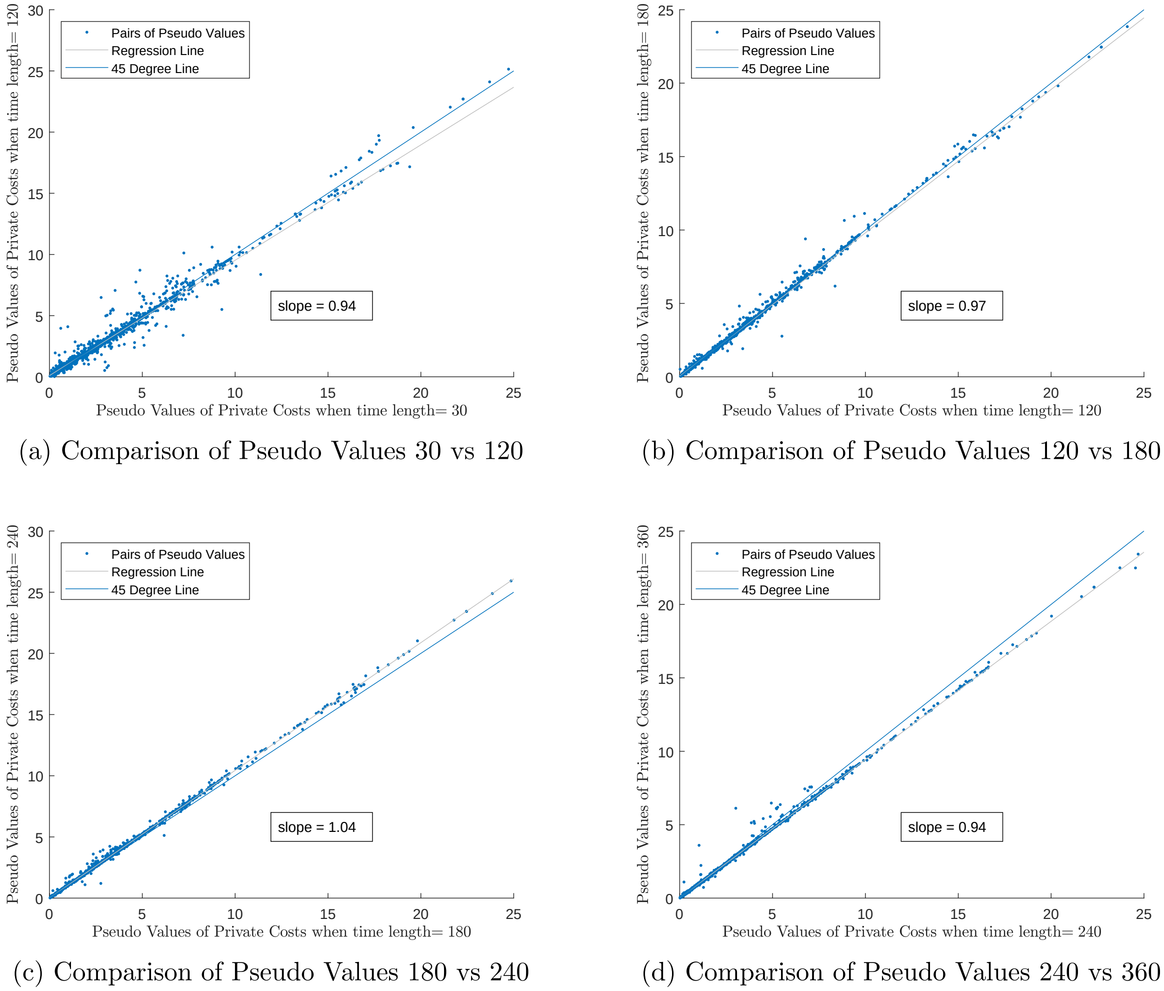
Following the estimation procedure defined in the previous section, I estimate the SEMUP model, and get pseudo private costs for each entry bidder. For different values of time length, I can estimate the SEMCP model under the corresponding constructed potential bidders set. Figure 4 shows the scatter plot of pairs estimated pseudo private costs of entry bidders under

Figure 4: Comparison of Pseudo Values $m = 30$ and $m = 360$



time length equals 30 days and time length equals 360 days. The figure depicts a fitted line of pseudo values and a 45° degree line. The regression result reveals an estimated slope of 0.901 which to some extent represents a 10% difference between the two construction methods.

In Figure 5, I similarly draw the scatterplot of private costs recovered from potential bidders sets with different time length criterion. When it comes to different criteria of time length, private costs vary a lot. In fact, when working on more different values of time length, results indicate that different time length makes estimated private costs very different. Thus, as for Caltrans data, it is important to check the method of constructing potential bidders set when it comes to the research of entry behaviors. On the other hand, there is not a clear pattern of estimated slopes. I suggest that it might be because the regression method filters out too much information, and therefore a comparison between costs distribution is very necessary.



(a) Comparison of Pseudo Values 30 vs 120

(b) Comparison of Pseudo Values 120 vs 180

(c) Comparison of Pseudo Values 180 vs 240

(d) Comparison of Pseudo Values 240 vs 360

Figure 5: First group of subfigures.

I estimate SEMUP models and SEMCP models for every ten days. Similar to the case of private valuation, the distance for procurement auctions can be anagously defined as,

$$\hat{d}(\xi) = \sup_{c\in[0,\bar{c}]} \left| \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{c}^j_{up,i} \leq c\} - \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{c}^j_{cp,i} \leq c\} \right| \tag{43}$$

and the KS statistic is formed as,

$$KS = \sup_{c\in[0,\bar{c}]} \left| \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{c}^j_{up,i} \leq c\} - \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{c}^j_{cp,i} \leq c\} \right| \tag{44}$$

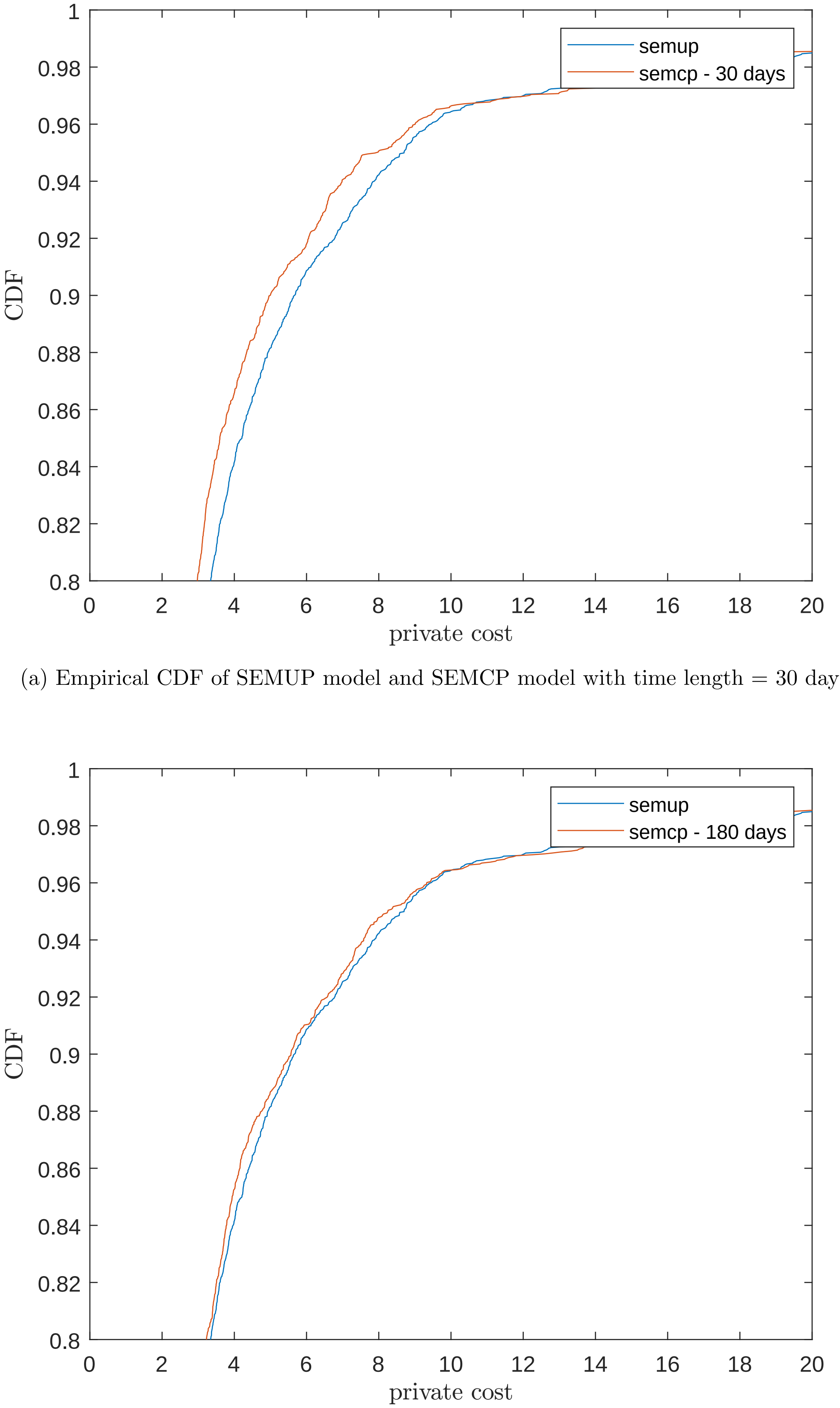
Figure 7 represents the difference between SEMUP and each SEMCP model. The two dotted lines represent the 90% confidence interval. The occurence of the U-shape pattern makes any time length within the interval $[120, 180]$ seem plausible to be a good choice of criterion time length. From this curve, we can observe that for very small or very large value of time length, the distance tends to be larger. This implicitly indicates that consideration with no entry (or very small amount of potential bidders) or with all bidders as potential bidders (or very large amount of potential bidders) will lead to distortion on recovery of bidders' private costs or beliefs.

When we look into the empirical CDF recovered from SEMCP and SEMUP models, Figure 6 shows the difference of empirical CDF of two SEMCP models compared to the same SEMUP models. As for time length being 180 days, it reveals a relative small distance in the previous figure. The empirical CDF indicates more deviation for the case that time length is 30 days.

Figure 8 represents corresponding bootstrap p - values from the test defined in previous sections. On the 5% significant level, p -values within the interval $[120, 180]$ are relatively high. When the value of time length gets very small or very large, the equality of private cost distributions between SEMUP and SEMCP is rejected on the 5% significant level. This result behaves in the same way as the distance method.

6 Conclusion

This paper gave a criterion to distinguish potential bidders for practical auctions data. I developed the selective entry model with an unknown number of potential bidders. I showed the equilibrium result and identification of primitives for the model. The model generalizes the traditional selective entry model by further assuming the number of potential bidders is random. By comparing it to the selective entry model with constructed potential bidders set, I proposed an idea on how to evaluate a method of creating potential bidders set for



(a) Empirical CDF of SEMUP model and SEMCP model with time length = 30 days

(b) Empirical CDF of SEMUP model and SEMCP model with time length = 180 days

Figure 6: Normal Distribution Example with the Same Shape

Figure 7: Distance between Private Cost Distributions of SEMCP and SEMUP models for Different Time Length

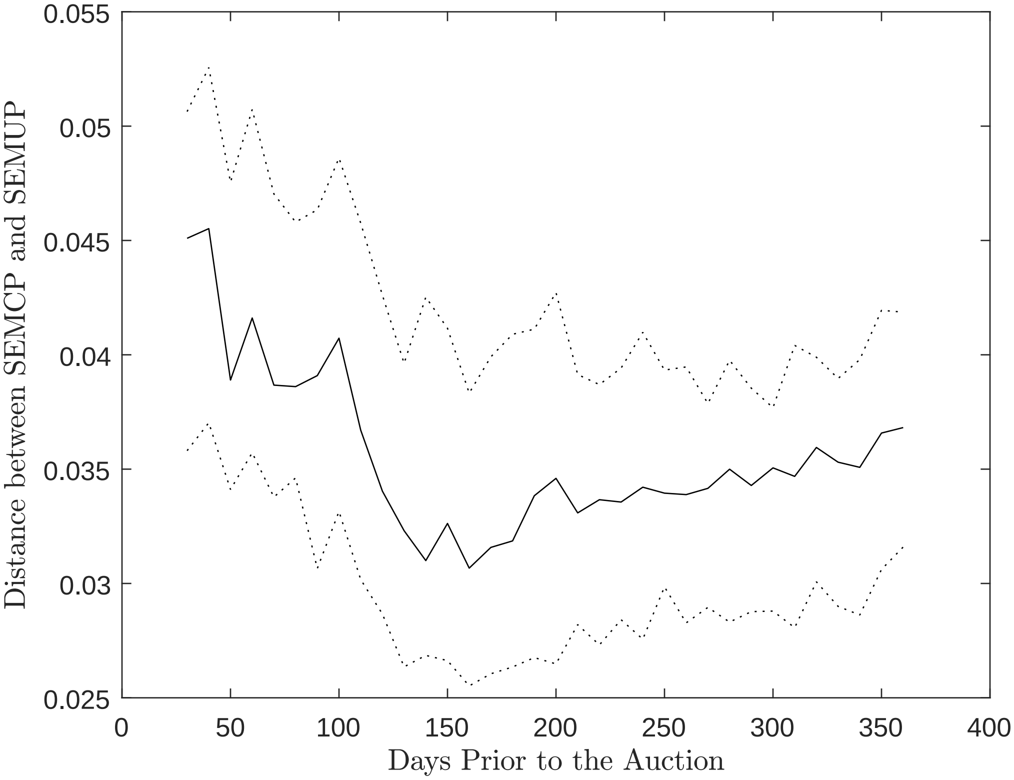
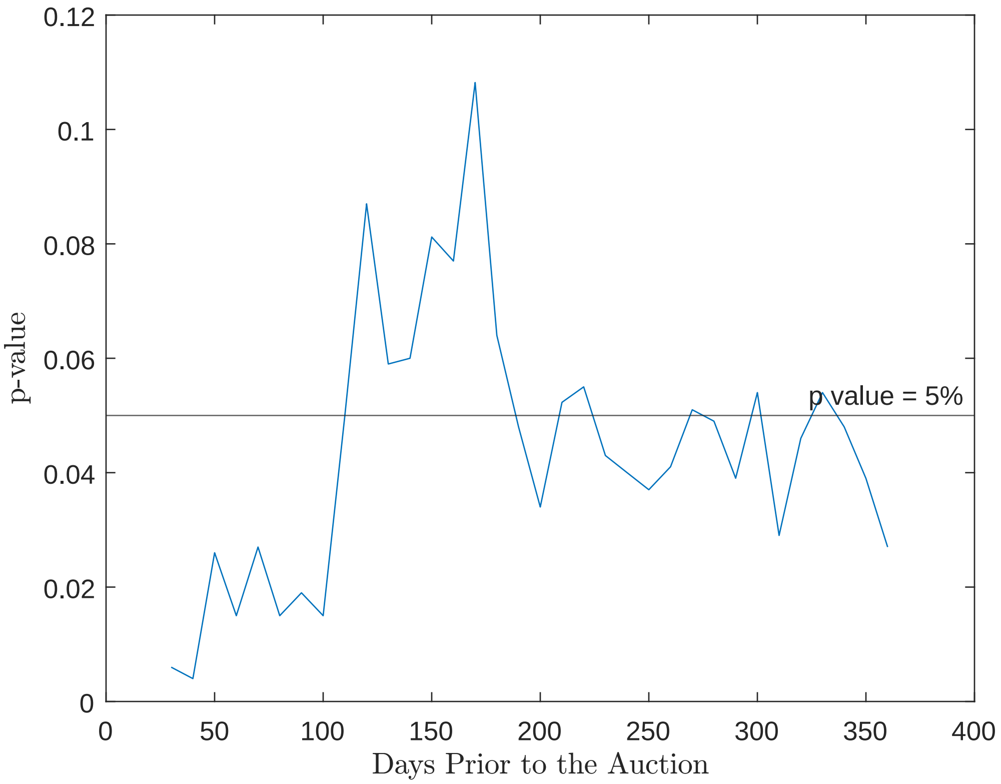


Figure 8: p Values for Different Time Length



auction data, i.e., the closer the distance between the two recovered estimated private values distributions is, the more plausible the method is. To measure this closeness, I further utilized KS test to indicate whether one construction method is plausible.

The paper utilized the procedure on an empirical application using Caltrans procurement auction data. I structurally estimated the models and made comparison among different models. I examined the commonly used method of constructing potential bidders set that assigns a bidder as a potential bidder for an auction if the bidder enters another auction in the same district and within a specific time length prior to the auction. As for Caltrans data, I find that a plausible choice of time length might be 120 days to 180 days. The bootstrap p - values comfirmed this finding, and additionally showed that too large or too small time length would be rejected. The method can be used as a starting point for practical auction research on entry. Policy makers or government can use it to distinguish potential bidders and measure the degree of entry.

Appendix A: Identification of Private Values in SEMCP

From the maximization problem, the maximizer is derived from the first order condition,

$$
\begin{aligned}
(N-1)[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{N-2}(1-\bar{s})g^*(b;\bar{s})(v-b) - [\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{N-1} &= 0 \\
v-b &= \frac{[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{N-1}}{(N-1)[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{N-2}(1-\bar{s})g^*(b;\bar{s})} \\
v-b &= \frac{\bar{s}+(1-\bar{s})G^*(b;\bar{s})}{(N-1)(1-\bar{s})g^*(b;\bar{s})} \\
v &= b+\frac{1}{N-1}\frac{\bar{s}+(1-\bar{s})G^*(b;\bar{s})}{(1-\bar{s})g^*(b;\bar{s})}
\end{aligned}
$$

Appendix B: Distribution and Conditional Distribution of Private Cost

Suppose the conditional distribution is denoted by $F_{C|\Delta}(c|\delta)$ where Δ refers to the single index described in the paper, $\Delta = Z'\beta$.

The distribution of private cost can be derived as,

$$
\begin{aligned}
F_C(c) &= \text{Prob}(C \le c) && (45) \\
&= \text{Prob}(C \le c, -\infty \le \delta \le \infty) && (46) \\
&= \int_{-\infty}^{\infty}\int_0^c f(t,\delta)dtd\delta && (47) \\
&= \int_{-\infty}^{\infty}\int_0^c f_{C|\Delta}(t|\delta)f_\Delta(\delta)dtd\delta && (48) \\
&= \int_{-\infty}^{\infty} F_{C|\Delta}(c|\delta)f_\Delta(\delta)d\delta && (49)
\end{aligned}
$$

Appendix C: Connection between p_l and q_l

We follow McAfee and MacMillan (1987)[18] to deal with the connection between the two kinds of probabilities.

Suppose we index the bidders in the total bidders set with natural numbers $\mathcal{A} = \{1, 2, 3, ..., N_{max}\}$. For any finite set $A \subseteq \mathcal{A}$, let β_A represent the probability that A is the set of active bidders. The probabilities should naturally satisfy,

$$\sum_{l=1}^{N_{max}} \sum_{A,|A|=l} \beta_A = 1 \tag{50}$$

Then the probability that l bidders are present is,

$$p_l = \sum_{A,|A|=l} \beta_A \tag{51}$$

The expected number of bidders is,

$$N^* = \sum_{l=1}^{N_{max}} l p_l \tag{52}$$

When bidder k is selected, he will update his probability of the number of bidders by,

$$q_n^k = \sum_{\substack{A \\ |A|=n \\ k \in A}} \beta_A \bigg/ \sum_{\substack{A \\ k \in A}} \beta_A \tag{53}$$

From Lemma 1 of McAfee and McMillan (1987), we have

$$N^* q_l = l p_l \tag{54}$$

Appendix D: Derivation of Bidding Strategy Under SEMUP Models

From the first order condition,

$$\left\{\sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{l-1}\right\}+$$
$$(b-v)\sum_{l=1}^{\bar{N}} q_l(1-\bar{s})[\bar{s}+(1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{l-2}(l-1)f^*(\beta^{-1}(b);\bar{s})\frac{d\beta^{-1}(b)}{db}=0$$

Under symmetric bidding strategy,

$$\beta'(v)\left\{\sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}\right\}+\beta(v)\sum_{l=1}^{\bar{N}} q_l(1-\bar{s})[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-2}(l-1)f^*(v;\bar{s})=$$
$$v\sum_{l=1}^{\bar{N}} q_l(1-\bar{s})[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-2}(l-1)f^*(v;\bar{s})$$

Therefore we have,

$$\left\{\beta(v)\left\{\sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}\right\}\right\}'=v\sum_{l=1}^{\bar{N}} q_l(1-\bar{s})[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-2}(l-1)f^*(v;\bar{s})$$

Finally we get,

$$\beta(v;\bar{s})=\sum_{l=1}^{\bar{N}}\frac{q_l[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}}{\sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}}\left(v-\frac{1}{[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}}\int_{\underline{v}}^{v}[\bar{s}+(1-\bar{s})F^*(u;\bar{s})]^{l-1}du\right)$$

Appendix E: Proof of Existence of Minimum Distance

According to Pedersen (2021)[20] p.27 Proposition 1.5.12, the following theorem shows the property of supremum of semicontinuous functions.

Theorem 3 *If (X, τ) is a topological space and if a set $\mathcal{F} \subset LSC(X)$, then $g : X \to \bar{R}$ defined by*

$$g(x) = \sup_{f \in \mathcal{F}} f(x), x \in X$$

is also contained in $LSC(X)$ where $LSC(X)$ refers to the set of all lower semicontinuous functions $X \to \bar{R}$.

I define $\mathcal{F} = \{f_v(\xi) = |F_{up}(v) - F_{cp}(v; \xi)| : \forall v \in [0, \bar{v}]\}$ as the set of functions of ξ with index v. From the assumption, $F_{cp}(v; \xi)$ is a continuous function of ξ for any v on its support. Therefore, $\mathcal{F}$ is a subset of $LSC(\xi)$.

I have the following result,

$$d(\xi) = \sup_{v \in [0, \bar{v}]} |F_{up}(v) - F_{cp}(v; \xi)| = \sup_{f \in \mathcal{F}} f_v(\xi) \tag{55}$$

Therefore, from Theorem 3, $d(\xi)$ is also lower semicontinuous with respect to $\xi \in \Xi$. From an extension of Weistrass extreme value theorem, if the function $d(\xi)$ is lower semicontinuous and the support of ξ, Ξ is compact, the function $d(\xi)$ is bounded below and attains its infimum.

Appendix F: Derivation of distribution of winning bids and second highest bids

Assume a bidder with valuation v will bid a potential bid $b = \beta(v; \bar{s})$. In our entry model, there is always a very small probability that auction passes, $\sum_{l=1}^{\bar{N}} p_l \bar{s}^l$.

With consideration of the possibility of auction pass, the distribution of the highest bid we observe is,

$$\begin{aligned}
G^{(1)}(b) &= \text{Prob}(B_{max} \leq b|\text{auction observed}) \\
&= \frac{\text{Prob}(B_{max} \leq b)}{\text{Prob(auction observed)}} \\
&= \frac{\text{Prob}(B_{max} \leq b \text{ or auction pass}) - \text{Prob(auction pass)}}{\text{Prob(auction observed)}} \\
&= \frac{\sum_{l=1}^{\bar{N}} p_l \cdot \text{Prob}(B_{max} \leq b \text{ or auction pass}|N = l) - \text{Prob(auction pass)}}{\text{Prob(auction observed)}} \\
&= \frac{\sum_{l=1}^{\bar{N}} p_l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^l - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}
\end{aligned}$$

As for the second highest bid, we define second highest bid equals 0 if the auction contains only one bidder.

The distribution of the second highest bid is,

$$\begin{aligned}
&G^{(2)}(b) \\
&= \text{Prob}(B_{2nd} \leq b|\text{auction observed}) \\
&= \frac{\text{Prob}(B_{2nd} \leq b)}{\text{Prob(auction observed)}} \\
&= \frac{\text{Prob}(B_{2nd} \leq b \text{ or auction pass}) - \text{Prob(auction pass)}}{\text{Prob(auction observed)}} \\
&= \frac{p_1 + \sum_{l=2}^{\bar{N}} p_l \cdot \left\{[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^l + l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}[(1-\bar{s})(1-G^*(b;\bar{s}))]\right\} - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}
\end{aligned}$$

This distribution can also be represented in a form,

$$
\begin{aligned}
&G^{(2)}(b)\\
&=\frac{p_1+\sum_{l=2}^{\bar{N}} p_l\cdot\left\{[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^l+l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}[(1-\bar{s})(1-G^*(b;\bar{s}))]\right\}-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l}{1-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l}\\
&=\frac{(1-\bar{s})(1-G^*(b;\bar{s}))\sum_{l=1}^{\bar{N}} p_l l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}+\sum_{l=1}^{\bar{N}} p_l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^l-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l}{1-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l}\\
&=\frac{(1-\bar{s})(1-G^*(b;\bar{s}))\sum_{l=1}^{\bar{N}} p_l l[\bar{s}+(1-\bar{s})G^*(b;\bar{s})]^{l-1}}{1-\sum_{l=1}^{\bar{N}} p_l\bar{s}^l}+G^{(1)}(b)
\end{aligned}
$$

Appendix G: Identification of Private Values

From previous sections, we can derived the distribution of the second highest bid as,

$$G^{(2)}(b)$$

$$= \frac{p_1 + \sum_{l=2}^{\bar{N}} p_l \cdot \left\{[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^l + l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}[(1-\bar{s})(1-G^*(b;\bar{s}))]\right\} - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}$$

The PDF of it is,

$$\begin{aligned} g^{(2)}(b) &= \sum_{l=2}^{\bar{N}} p_l \left\{ l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}(1-\bar{s})g^*(b;\bar{s}) + l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}(1-\bar{s})(-g^*(b;\bar{s})) \right. \\ &\qquad \left. + l(l-1)[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(1-\bar{s})(1-G^*(b;\bar{s}))(1-\bar{s})g^*(b;\bar{s}) \right\} \Big/ \{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l\} \\ &= \sum_{l=2}^{\bar{N}} p_l l(l-1)[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(1-\bar{s})\left(1-G^*(b;\bar{s})\right)(1-\bar{s})g^*(b;\bar{s}) \Big/ \{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l\} \\ &= (1-\bar{s})\left(1-G^*(b;\bar{s})\right) \sum_{l=2}^{\bar{N}} p_l l(l-1)[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(1-\bar{s})g^*(b;\bar{s}) \Big/ \{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l\} \end{aligned}$$

Therefore, we have,

$$\frac{G^{(2)}(b) - G^{(1)}(b)}{g^{(2)}(b)}$$

$$= \frac{(1-\bar{s})(1-G^*(b;\bar{s})) \sum_{l=1}^{\bar{N}} p_l l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1} \Big/ \{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l\}}{(1-\bar{s})\left(1-G^*(b;\bar{s})\right) \sum_{l=2}^{\bar{N}} p_l l(l-1)[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(1-\bar{s})g^*(b;\bar{s}) \Big/ \{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l\}}$$

$$= \frac{\sum_{l=1}^{\bar{N}} p_l l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}}{\sum_{l=2}^{\bar{N}} p_l l(l-1)[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(1-\bar{s})g^*(b;\bar{s})}$$

and,

$$\begin{aligned} v &= b + \frac{\sum_{l=1}^{\bar{N}} p_l l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}}{\sum_{l=2}^{\bar{N}} p_l l(l-1)[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(1-\bar{s})g^*(b;\bar{s})} \\ &= b + \frac{G^{(2)}(b) - G^{(1)}(b)}{g^{(2)}(b)} \end{aligned}$$

References

[1] A. Abadie. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American statistical Association*, 97(457):284–292, 2002.

[2] S. Athey, J. Levin, and E. Seira. Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, 126(1):207–257, 2011.

[3] P. Bajari, H. Hong, and S. P. Ryan. Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568, 2010.

[4] P. Bajari, S. Houghton, and S. Tadelis. Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review*, 104(4):1288–1319, 2014.

[5] J. Bulow and P. Klemperer. Why do sellers (usually) prefer auctions? *American Economic Review*, 99(4):1544–75, 2009.

[6] D. G. De Silva, G. Kosmopoulou, and C. Lamarche. The effect of information on the bidding and survival of entrants in procurement auctions. *Journal of Public Economics*, 93(1-2):56–72, 2009.

[7] J.-A. Espín-Sánchez and A. Parra. Entry games under private information. 2018.

[8] M. Gentry and T. Li. Identification in auctions with selective entry. *Econometrica*, 82(1):315–344, 2014.

[9] R. Gil and J. Marion. Self-enforcing agreements and relational contracting: Evidence from california highway procurement. *The Journal of Law, Economics, & Organization*, 29(2):239–277, 2013.

[10] E. Guerre, I. Perrigne, and Q. Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000.

[11] E. Krasnokutskaya and K. Seim. Bid preference programs and participation in highway procurement auctions. *American Economic Review*, 101(6):2653–86, 2011.

[12] D. Levin and J. L. Smith. Equilibrium in auctions with entry. *The American Economic Review*, pages 585–599, 1994.

[13] T. Li and B. Zhang. Testing for affiliation in first-price auctions using entry behavior. *International Economic Review*, 51(3):837–850, 2010.

[14] T. Li and B. Zhang. Affiliation and entry in first-price auctions with heterogeneous bidders: An analysis of merger effects. *American Economic Journal: Microeconomics*, 7(2):188–214, 2015.

[15] T. Li and X. Zheng. Entry and competition effects in first-price auctions: theory and evidence from procurement auctions. *The Review of Economic Studies*, 76(4):1397–1429, 2009.

[16] T. Li and X. Zheng. Information acquisition and/or bid preparation: A structural analysis of entry and bidding in timber sale auctions. *Journal of Econometrics*, 168(1):29–46, 2012.

[17] V. Marmer, A. Shneyerov, and P. Xu. What model for entry in first-price auctions? a nonparametric approach. *Journal of Econometrics*, 176(1):46–58, 2013.

[18] R. P. McAfee and J. McMillan. Auctions with a stochastic number of bidders. *Journal of economic theory*, 43(1):1–19, 1987.

[19] P. R. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, pages 1089–1122, 1982.

[20] G. K. Pedersen. *Analysis now*, volume 118. Springer Science & Business Media, 2012.

[21] W. F. Samuelson. Competitive bidding with entry costs. *Economics letters*, 17(1-2):53–57, 1985.

[22] U. Song. Identification of auction models with an unknown number of bidders, 2015.

[23] L. Ye. Indicative bidding and a theory of two-stage auctions. *Games and Economic Behavior*, 58(1):181–207, 2007.