# Inference on Number of Potential Bidders in Selective Entry

Shuo Tian

November 28, 2021

**Abstract**

This paper studies inference on number of potential bidders in selective entry for first-price sealed-bid auctions. Information of potential bidders set is of great empirical importance in auctions. However, such a set is often constructed by ad hoc methods. I develop a method to test whether these methods are plausible. The test is based on the comparison of bidders' value distributions in two models. The first one is a selective entry model given information of potential bidders. The second one is another selective entry model with the number of potential bidders unknown and random. I apply the methods to procurement auctions from the Californian Department of Transportation (Caltrans). I find that recovered private value distributions fluctuate significantly when different construction methods are employed. I examine the widely used construction method that relies on the time length of a bidder's previous participation. The results show that the time length criterion for Caltrans data should be selected from 120 to 180 days.

# 1 Introduction

Entry and endogenous participation have been playing an important role in the study of markets. Entry represent that potential market entrants endogenously decide whether they enter the market. The participants of the market will behave differently from those who choose not to enter. The endogenous choice of potential entrants leads to a selection according to individuals' characteristics. In practical auctions, the phenomena of entry 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 (See Bulow and Klemperer, 2009). 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.

Entry in auctions has been widely studied in the literature. However, the practical auction data set often lacks the information of bidders' entry behaviors. A commonly used method is to assign potential entrants to each auction according to some criterion and analyze the problem under the constructed potential bidders set. However, construction methods are usually ad hoc. For example, for procurement auctions, a frequently used construction is to assign a bidder as a potential bidder of an auction if the bidder participates in at least one another auction in the same district and within a specific time length prior to the auction. This method is very common in the research of timber auctions, but the choice of time length differs in the literature (e.g., Li and Zhang, 2010; Li and Zheng, 2012; Li and Zhang, 2015). Besides, there are practical auction data sets contain some particular "proxy" information that can be used to construct potential bidders set. For example, in procurement auctions collected from Texas Department of Transportation (TDoT), Li and Zheng (2009) suggest contractors who submit an official bidding proposal on time are considered as potential bidders. In Caltrans data, Krasnokutskaya and Seim (2011) utilize plan holders information to form the potential bidders set. Construction methods based on "proxy" are usually ad hoc as other information in data such as qualification contractor list or project advertisement receiver information can also be used to select potential bidders. Other research may focus on the entry behavior for large firms, (e.g., Bajari et al., 2010; Gil and Marion, 2013). This construction imposes the assumption that large firms are always potential bidders for each auction.

Research on entry in auctions rely heavily on potential bidders' entry behaviors, but ad hoc methods usually give very different constructed potential bidders set. As shown in the application on Caltrans data, I follow the time length method described above, and the average number of potential bidders increases from 16.57 to over 30.10 when the time length is set equal to from 30 days to 360 days. The recovered private cost distribution for entry bidders could fluctuate much when different time length is utilized. For some particular bidders, the recovered private costs differ a lot for different time lengths. The lack of robustness and the uncertainty of recovered model primitives using ad hoc methods will lead to biased economic insights and misleading policy implications. Therefore, a natural research 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.

In this article, I propose a criterion to judge correctly whether a construction of potential bidders set is plausible 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 from both models can be identified and recovered respectively from data. If some ad hoc method of construction of potential bidders set is correct, the private value distributions recovered from the two models should be very close 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 introduce two selective entry models and show a characterization of equilibrium for them. Based on the two-stage selective entry model framework (e.g., Marmer et al., 2013; Gentry and Li, 2014; Gentry et al., 2017) I introduce the selective entry model with constructed potential bidders set (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 a participation cost. Potential bidders can observe the number of potential bidders in the first stage. In the second stage, entry bidders observe their private valuations but not the number of entrants and submit bids under a first-price auction mechanism. SEMCP follows largely on Gentry and Li (2014) framework, but the second stage is adjusted to fit in a first-price auction mechanism. I define and show equilibrium strategies, including the entry strategy and the bidding strategy, for a SEMCP model. Based on the SEMCP model, I further propose a selective entry model with an unknown number of potential bidders (SEMUP) where I assume the number of potential bidders is a random variable. In SEMUP models, there is a stage 0 prior to the entry stage and bidding stage. Nature chooses potential bidders set from the entire bidders' pool at this stage. The timing and auction mechanism for the entry stage and bidding stage 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, like 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 a weighted sum of different scenarios where the actual number of potential bidders picks every possible value on its support.

Second, I investigate the identification of the two models. In the SEMCP model, combined with Gentry and Li (2014) 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 under appropriate assumptions. In my paper, I additionally derive an equation that indicates the private valuation of an entry bidder is equal to the summation of his bid and a function of distribution functions of bids. The private value of an entry bidder can be directly identified from the idea of Guerre et al. (2000). As for a SEMUP model, to prove the identification of private values, I derive a similar result in which the private value of an entry bidder can be rewritten as a function of bid, the distribution of highest bid, and that of second highest bid. This equation circumvents the direct usage of the 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 from both SEMCP and SEMUP models. 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 closely related to the entry threshold, so the estimation result is influenced by the construction method of the 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 correctness of a particular construction of potential bidders set can be transformed into the problem of measuring the distance between the two recovered distributions of private values from the two models. For example, the criterion of time length for timber auctions described previously can be selected reasonably if the chosen time length can provide a very small distance between two distributions. As the problem is transformed to an equality problem of two distributions, I employ a Kolmogorov - Smirnov test (KS test) and apply the test result to shed light on whether a construction of potential bidders set is plausible.

Finally, I apply the entire procedure on construction contracts procurement auctions data from the California Department of Transportation (Caltrans). The data set consists of paving contracts collected by Caltrans from 1999 to 2005. I construct the potential bidders set for it using the time length 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. For each constructed potential bidders set, I structurally estimate the SEMUP model and SEMCP model. For Caltrans data, I pick the time length value from 30 days to 360 days for every multiple of ten days. The estimation results reveal that recovered private value distributions fluctuate much when different time lengths are employed. I calculate the distance between the two private costs distributions from the two models for each picked time length and depict a scatter plot of the distance and time length. The connection between the distance and time length forms a U - shape curve, and thus we can select an interval of the time length from 120 to 180 days, which gives lowest distance compared to other time length values. Besides, I conduct KS tests for each case. The test results show that time lengths from 120 to 180 days seem to be reasonable choices. Very small time length or very large time length are rejected by the KS test which indicates that the time length should be chosen very carefully.

This paper contributes to the literature of empirical auction research 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) investigate the effect of information on the bidding and survival of entrants in procurement auctions; Li and Zheng (2009) consider entry and competition effects in procurement auctions; Li and Zheng (2012) analyze bidders' entry and bidding decisions in Michigan timber sale auctions using Bayesian method; Li and Zhang (2010) study affiliation using bidders entry behaviors; Athey et al. (2011) discuss endogenous participation under sealed bid and open auctions with heterogeneous bidders; Krasnokutskaya and Seim (2011) study the bid preference program and paricipation decisions in California highway procurement auctions; Li and Zhang (2015) 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 addresses the ad hoc issue and can be used as a starting model to check 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 et al., 2013;Gentry and Li, 2014) and auction theory models of entry (e.g., Samuelson, 1985; Levin and Smith, 1994, Espín-Sánchez and Parra, 2018). 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.

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

# 2 Models, Identification and Estimation

In this section, 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)

The SEMCP model is largely from the selective entry model described in Gentry and Li (2014). 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 et al. (2013) and Gentry and Li (2014), a single and indivisible good is allocated to $N \geq 1$ 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 $s_i \in [0,1]$ of her private value $v_i \in [\underline{v}, \bar{v}]$. Entry bidders can observe their private values in the second stage. The private signal and private value follow a joint distribution $F(v,s)$ that is continuous in $(v,s)$ on its support $[\underline{v}, \bar{v}] \times [0,1]$. Each potential bidder simultaneously decide whether to enter the auction. If a bidder determine to enter, he incur an entry cost $k \in [\underline{k}, \bar{k}]$ which is related to preparation costs, opportunity costs, or uncertainty for risk. In the second stage, $n \geq 1$ entry bidders choose to enter. Each entrant observes his private value $v_i$ and proceeds with a first-price sealed-bid auction.

In SEMCP models, I maintain the assumption in Gentry and Li (2014) 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.

### 2.1.1 Equilibrium in a SEMCP Model

I assume that the joint cumulative distribution function follows,

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

1. *For each bidder $i$, the conditional distribution of $V_i$ is stochastically ordered in $S_i$, i.e., if for $\forall s' \geq s$, $F(v|s') \leq F(v|s)$.*

2. *The random pairs $(V_i, S_i)$ are independent across bidders: $(V_i, S_i) \perp (V_j, S_j)$ for $\forall j \neq i$.*

3. *Without loss of generality, 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), the model employs a weaker restriction on the relationship between the private value and private signal. They assume that the two variables follow stochastic ordering rather than affiliation relation like, Milgrom and Weber (1982), Ye (2007), and Marmer et al. (2013).

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

Besides the assumption 1, I follow Gentry and Li (2014) and impose the regularity condition on the relationship between values and signals. This condition provides some level of smoothness so that a well-behaved equilibrium exists.

To apply the result on the procurement auction of interest, I focus on the first-price sealed-bid auction as the mechanism in the second stage. The strategy of each bidder consist of two parts, an entry strategy, $\bar{s}$, and a bidding strategy, $\beta(v, \bar{s})$. The bidding strategy is an entry threshold where if a bidder has his private signal in the first stage $s_i \geq \bar{s}$, he chooses to enter the auction. The bidding strategy, $\beta(v; \bar{s})$ is the bid a bidder will propose when he has a private valuation $v$ in the second stage and the entry threshold in the first stage is $\bar{s}$.

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

**Definition 1 (Equilibrium of a SEMCP model)** *The symmetric pure strategy Bayesian Nash equilibrium of the two-stage auction game in a SEMCP model, $(\bar{s}^*, \beta^*)$, solve the following two equations.*
*(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 $v_i$,*

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

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,

$$\begin{aligned} G^*_{1:N-1}(b;\bar{s}) &= \text{Prob(winning with bid } b|\bar{s}) \\ &= [\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{N-1} \end{aligned} \tag{5}$$

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

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

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

$$\beta(v;\bar{s}) = v - \frac{1}{[\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1}} \int_0^{\bar{v}} [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1} dv \tag{7}$$

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

$$\pi^{cp}(v;\bar{s},N) = [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1}(v - \beta(v;\bar{s})) \tag{8}$$

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

$$\Pi^{cp}(s;\bar{s},N) = \int_0^{\overline{v}} \pi^{cp}(v;\bar{s}) f(v|s)dv \tag{9}$$

The entry threshold is characterized by,

$$\Pi^{cp}(\bar{s}^*;\bar{s}^*,N) = k \tag{10}$$

Equation (10) and Equation (7) can jointly specify the equilibrium strategy for a potential bidder. Following the Proposition 1 in Gentry and Li (2014), the equilibrium $(\bar{s}^*, \beta^*(\cdot))$ exists uniquely under Assumptions 1-2.

### 2.1.2 Identification of a SEMCP Model

The primitives of a SEMCP model include the entry strategy $\bar{s}$, the entry cost, $k$, the conditional distribution in the second stage, $F^*(v;\bar{s})$, the joint distribution of the private value and private signal, $F(v,s)$, the private value of entry bidders, $v$, and the private value distribution of entry bidders, $F^*(v)$.

In this paper, I mainly need the identification of private values and private value distribution of entry bidders[1]. From the maximization problem in the second stage, the first order condition leads to the following result as shown in the Appendix A,

$$v = b + \frac{1}{N-1}\frac{\bar{s} + (1-\bar{s})G^*(b;\bar{s})}{(1-\bar{s})g^*(b;\bar{s})} \tag{11}$$

where $G^*(\cdot)$ and $g^*(\cdot)$ represent the conditional distribution of bids in the bidding stage.

Therefore, under homogeneous auction setting, when the number of potential bidders, $N$, entry bidders' bids, $b_i$, and the entry decisions for each potential entrants are observed, the

[1]In a SEMUP model, only private values and corresponding distribution can be identified. The main objective is to compare the difference on the recovered private values from the two models. Therefore, in a SEMCP model, identification of other primitives is less important

model primitives of interest, both the entry threshold, $\bar{s}$, and private value of each bidder, $v_i$ are identified as follows.

The entry threshold can be identified from the information of entry frequency,

$$\bar{s} = 1 - \frac{\mathrm{E}(n|N)}{N} \tag{12}$$

The distribution terms, $G^*(b;\bar{s})$ and $g^*(b;\bar{s})$ can be identified directly from the bids distribution in the second stage.

The private value and corresponding private value distribution can be identified from the one-to-one mapping of Equation (11) by plugging in the bidder's bid, number of potential bidders, the identified entry threshold, and the values of conditional distributions of bids.

In fact, in a heterogeneous auction setting as described in Gentry and Li (2014), the observebles in a selective entry model includes the number of potential bidders for auction $j$, $N_j$, the number of entrants for each auction, $n_j$, a vector of submitted bids $b_j$, and a vector of auction-level instruments, $x_j$, which is a cost shifter that only shifts entry behaviors without affecting underlying distributions. Gentry and Li (2014) shows that under appropriate assumptions, the model primitives of a 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, $k$.

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

In practical auctions, 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 McMillan, 1987; Song, 2015).

### 2.2.1 Model Setup

At stage 0 (Nature Stage), auctioneer proposes a project and nature selects a set of bidders $A$, from the pool of all possible bidders $\mathcal{A}$. The set of bidders $A$ can be considered as the potential bidders set. The cardinality of set $A$ is $N$. The set of all possible bidders $\mathcal{A}$ for a specific data set can be considered as the pool of all bidders and has cardinality $\bar{N}$. The random variable $N$ can take its value on the support $\{1, 2, \cdots, \bar{N}\}$. The distribution of $N$ are denoted by a sequence of probabilities,

$$p_l = \text{Prob}(N = l) \tag{13}$$

which represents the probability that random variable $N = l$.

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

$$q_l^i = \text{Prob}(N = l | i \in A) \tag{14}$$

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

$$q_l = q_l^i \quad \text{for all } i \tag{15}$$

Each active bidder is informed a private signal $s_i$ which is correlated with his private value $v_i$ with a joint distribution $F(v, s)$. The joint distribtuion follows the same setting as in SEMCP models. Potential bidders choose whether to enter. If entry, the bidder pays a participation (preparation) cost $k$.

At stage 2 (Bidding Stage), entrants observe their private values, $v_i$. They propose their bids based on their information. I assume that potential bidders can only observe the distribution of the number of potential bidders $N$, but not the actual realization of the number of potential bidders. In addtion, entrants cannot observe the number of entrants $n$ in the second stage. The auction is assumed to be a first-price sealed-bid auction.

### 2.2.2 Equilibrium for a SEMUP Model

Similar to the setting in a SEMCP model, strategies of a potential bidder consists of two parts, an entry strategy $\bar{s}$, that is an entry threshold, and a bidding strategy $\beta(v; \bar{s})$. The symmetric pure strategy Bayesian Nash equalibrium of the two-stage auction game $(\bar{s}^*, \beta^*)$ in a SEMUP model can be characterized as,

**Definition 2 (Equilibrium of a SEMUP model)** *The symmetric pure strategy Bayesian Nash equilibrium of the two-stage auction game in a SEMUP model, $(\bar{s}^*, \beta^*)$, solve the following two equations.*
*(a) the entry threshold strategy $\bar{s}^*$ can be characterized*

$$\Pi^{up}(\bar{s}|\bar{s}) = k \tag{16}$$

*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 $v_i$,*

$$\beta^*(v_i; \bar{s}^*) = \arg\max_b Prob(\max_{j \neq i} \beta^*(v_j; \bar{s}^*) \leq b)(v_i - b) \tag{17}$$

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

$$F^*(v;\bar{s}) \equiv \frac{1}{1-\bar{s}} \int_{\bar{s}}^{1} F(v|t)dt \tag{18}$$

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

$$F^*_{1:N-1}(v;\bar{s}) = [\bar{s} + (1-\bar{s})F^*(v;\bar{s})]^{N-1} \tag{19}$$

Similar to the SEMCP models, suppose to 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,

$$\begin{aligned} G^*_{1:N-1}(b;\bar{s}) &= \text{Prob(winning with bid } b|\bar{s}) \\ &= [\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{N-1} \end{aligned} \tag{20}$$

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

$$\sum_{l=1}^{\bar{N}} q_l[\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{l-1} \tag{21}$$

The expected payoff therefore equals,

$$\left\{ \sum_{l=1}^{\bar{N}} q_l[\bar{s} + (1-\bar{s})F^*(\beta^{-1}(b);\bar{s})]^{l-1} \right\} (v-b) \tag{22}$$

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

$$\beta(v;\bar{s}) = \sum_{l=1}^{\bar{N}} \frac{q_l[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}}{\sum_{l=1}^{\bar{N}} q_l[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}} \left(v - \frac{1}{[\bar{s}+(1-\bar{s})F^*(v;\bar{s})]^{l-1}} \int_0^v [\bar{s}+(1-\bar{s})F^*(u;\bar{s})]^{l-1} du\right) \tag{23}$$

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

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

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

$$\Pi^{up}(s;\bar{s}) = \int_0^{\overline{v}} \pi^{up}(v;\bar{s}) f(v|s) dv \tag{25}$$

The entry threshold is characterized by,

$$\Pi^{up}(\bar{s}^*;\bar{s}^*) = k \tag{26}$$

After $\bar{s}^*$ is specified, the bidding strategy at the second stage can be characterized by plugging in $\bar{s}^*$ to Equation (23), i.e., $\beta(v;\bar{s}^*)$.

### 2.2.3 Identification for Private Values in a SEMUP Model

In a SEMCP model, observables include the vector of bids for each auction. Model primitives include entry strategy $\bar{s}^*$, entry cost, $k$, the conditional private value distribution in the second stage, $F^*(v;\bar{s})$, the joint distribution of private value and private signal, $F(v,s)$, distributions of the number of potential bidders, $p_l$ and $q_l$, the private value of entry bidders $v$, and the corresponding private value distribution for entry bidders, $F^*(v)$.

Unlike SEMCP models, econometrician and potential bidders in a SEMUP model do not observe the information of potential bidders set nor actual realization of the number of potential bidders. The lack of observation causes that most of the model primitives cannot be identified. However, private values and private value distribution in a SEMUP model can be identified by employing the fact that private values can be written as a function of bids, distribution of highest bids and that of second highest bids.

From the equation of maximization problem (22), the first order condition suggests that,

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

In the appendix, one can show that, $q_l = lp_l/N^*$, where $N^*$ is a constant defined as the expected number of potential bidders in the entire bidders' pool $\mathcal{A}$.

Therefore we have,

$$v = b + \frac{\sum_{l=1}^{\bar{N}} p_l l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}}{\sum_{l=2}^{\bar{N}} p_l l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-2}(l-1)(1-\bar{s})g^*(b;\bar{s})} \tag{27}$$

If I consider the distribution of winning bids[2], it can be characterized as (see proof in Appendix F),

$$G^{(1)}(b;\bar{s}) = \frac{\sum_{l=1}^{\bar{N}} p_l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l} - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l}{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l} \tag{28}$$

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

$$G^{(2)}(b;\bar{s}) = \frac{(1-\bar{s})(1-G^*(b;\bar{s}))\sum_{l=1}^{\bar{N}} p_l l[\bar{s} + (1-\bar{s})G^*(b;\bar{s})]^{l-1}}{1 - \sum_{l=1}^{\bar{N}} p_l \bar{s}^l} + G^{(1)}(b) \tag{29}$$

Private values can be identified from equation (27) from the fact that,

$$v = b + \frac{G^{(2)}(b;\bar{s}) - G^{(1)}(b;\bar{s})}{g^{(2)}(b;\bar{s})} \tag{30}$$

[2]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. Under homogeneous auction setting, distribution terms, $G^{(1)}(b;\bar{s})$ and $G^{(2)}(b;\bar{s})$ can be identified directly from the observation of entry 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. The private value distribution of entry bidders thereafter can be identified from the private values.

Since no information of entry frequence can be observed, entry strategy $\bar{s}^*$ is not observed. The fact that the entry threshold is not identified will cause distribution in the first stage, $F(v,s)$ and the entry cost, $k$, are not identified. The distribution of the number of potential bidders, $p_l$ and $q_l$ are not identified neither because of the lack of information of entry behaviors.

### 2.2.4 Intuition of Identification Results

In this section, I try to explain the intuition behind the identification of private values and private value distribution in SEMUP models.

Suppose $X_1, X_2, \cdots, X_n$ are $n$ independent draws from a distribution $F(x)$. Suppose we rank the $n$ random variables and pick the $i$th highest value, then we get the order statistic $X^{(i:n)}$. One can prove that the distribution of $X^{(i:n)}$ follows,

$$G^{(i:n)}(x) = \frac{n!}{(i-1)!(n-i)!}\int_0^{F(x)} t^{n-i}(1-t)^{i-1}dt$$

When the number of $n$ is known, the relation above can show that, the distribution of order statistics can be used to identify that of its parent distribution, $F(x)$. Based on this insight, Athey and Haile (2002) show that in the symmetric IPV model, one order statistic of a bidder's valuation nonparametrically identifies the bidders' private value distribution, when the number of bidders is known.

In addition to this result, when the number of bidders is unknown, the value distribution cannot be directly identified from one order statistic. Song (2015) extends the idea, and shows that a pair of ranked statistics (refers to order statistic when the total sample size is not observed) can be used to identify the parent distribution, i.e., two ranked statistics can be used to identify bidder's value distribution even if the number of bidders is unknown. As an example, Song (2006) considers the identification and estimation of a first-price auction model with an uncertain number of bidders. It shows that both the private value distribution and the number of bidders are identified from the first and second-highest bids.

In this paper, based on the setting in Song (2006), I further consider the endogenous participation by combining the selective entry model and auction models with an unknown number of bidders. Therefore, the SEMUP model contains a new model primitive, the entry threshold, $\bar{s}$. By incorporating endogenous participation, the number of potential bidders is no longer identified. However, the identification result on private values holds. This comes from the fact that both of the two order statistics rely on the entry threshold in a entry framework. The private values and private value distribution should depend on the information of bid distribution and that of entry threshold. By converting the method to the usage of a pair of two order statistics, entry threshold information are circumvented by the entry information carried by the two order statistics. As shown in the equation previously, all the entry threshold terms are cancelled out by using the Equation (38).

In fact, one could use other two order statistics of bids to get similar results, so we do not have to focus on highest and second highest bids. The rank of some order statistics is very important information. For example, research can use the second highest and third highest bids to identify the model, but the observation of the first bid is still important, at least the information that there exists a first bid greater than the second and third one.

## 2.3 Estimation

The estimation of a SEMCP and SEMUP model is based on the identification equations (11) and (38). The idea is to estimate every component in the identification equation, and plug in the estimated terms to get the pseudo private value for each entrant in each auction. The private value distribution can be obtained from the empirical CDF of these pseudo values.

To make SEMCP and SEMUP comparable, I assume that for both SEMCP and SEMUP model, there is a cost shifter $x$ with dimension $\rho$ that brings about the hetereogeneity on the auction level.

In a SEMCP model, for each auction, the number of potential bidders, $n_j$, the number of potential bidders, $N_j$, and the bid $b_{ji}$ are directly observed. The entry strategy in the identification is,

$$\bar{s}_N(x) = 1 - \frac{\mathrm{E}(n|x,N)}{N} \tag{31}$$

where $\mathrm{E}(n|x,N)$ can be estimated nonparametrically such as by the Nadaraya-Watson estimator.

For distribtution terms, $G^*(b;\bar{s})$ and $g^*(b;\bar{s})$, if we assume that $\bar{s}$ is affected completely by the cost shifter $x$ and the number of potential bidders $N_j$, the two distributions can be estimated by,

$$\hat{G}(b|\bar{s}) = \hat{G}(b,x,N) = \frac{1}{Jh_G^\rho}\sum_{j=1}^{J}\frac{1}{N_j}\sum_{i=1}^{N_j}1\left(b_{ji}\le b\right)K_G\left(\frac{x-X_j}{h_G},\frac{N-N_j}{h_{GN}}\right) \tag{32}$$

and

$$\hat{g}(b|\bar{s}) = \hat{G}(b,x,N) = \frac{1}{Jh_g^{\rho+1}}\sum_{j=1}^{J}\frac{1}{N_j}\sum_{i=1}^{N_j}K_g\left(\frac{b-b_{ji}}{h_g},\frac{x-X_j}{h_g},\frac{N-N_j}{h_{gN}}\right) \tag{33}$$

where $h_G$, $h_{GN}$, $h_g$ and $h_{gN}$ are some bandwidths, and $K_G$ and $K_g$ are kernels with bounded supports.

The pseudo private values can be estimated by,

$$\hat{v}_{ji} = b_{ji} + \frac{1}{N_j-1}\frac{\hat{\bar{s}} + (1-\bar{s})\hat{G}^*(b;\bar{s})}{(N_j-\hat{\bar{s}})\hat{g}^*(b;\bar{s})} \tag{34}$$

As for a SEMUP model, we only need to characterize the estimation of distribution of highest bid and second highest bid. The entry threshold for a SEMUP model is not a function of $N$, so the distribution terms can be estimated by,

$$\hat{G}^{(2)}(b;\bar{s}) = \hat{G}^{(2)}(b,x) = \frac{1}{Jh_{2G}^{\rho}} \sum_{j=1}^{J} 1(b_j^{(2)} \leq b) K_{2G}\left(\frac{x - X_j}{h_{2G}}\right) \tag{35}$$

$$\hat{G}^{(1)}(b;\bar{s}) = \hat{G}^{(1)}(b,x) = \frac{1}{Jh_{1G}^{\rho}} \sum_{j=1}^{J} 1(b_j^{(1)} \leq b) K_{1G}\left(\frac{x - X_j}{h_{1G}}\right) \tag{36}$$

and

$$\hat{g}^{(2)}(b;\bar{s}) = \hat{g}^{(2)}(b,x) = \frac{1}{Jh_{2g}^{\rho+1}} \sum_{j=1}^{J} K_{2g}\left(\frac{b - b_j^{(2)}}{h_{2g}}, \frac{x - X_j}{h_{2g}}\right) \tag{37}$$

where $h_{1G}$, $h_{2G}$, and $h_{2g}$ are some bandwidths, and $K_{1G}$, $K_{2G}$, and $k_{2g}$ are kernels with bounded supports.

The pseudo private values can be estimated by,

$$\hat{v}_{ji} = b_{ji} + \frac{\hat{G}^{(2)}(b;\bar{s}) - \hat{G}^{(1)}(b;\bar{s})}{\hat{g}^{(2)}(b;\bar{s})} \tag{38}$$

# 3 Inference on the Number of Potential Bidders

In this section, I propose a rationale to deal with construction of potential bidders set, i.e., distributions of a SEMCP and SEMUP model should be very close to each other. For construction methods with a continuous variable as a criterion, I suggest selecting the one that can give minimum distance between the two distributions. To make it more stringent, a test of the equality of two distributions can be employed. I propose a Kolmogorov-Smirnov test to indicate whether a construction method is acceptable. The test can also be employed for any given construction of potential bidders set such as construction with "proxy" information.

## 3.1 Minimum Distance and Test

The construction of potential bidders set sometimes involves the selection of a continuous variable. For example, in timber auctions, the potential bidders set is usually formed by adding all actual bidders in the data set who participated in at least one another auction in the same district and during a specific time length before the bidding date of the auction. In Li and Zheng (2012), the time length is set equal to 30 days. Li and Zhang (2010) restrict this value to 90 days, and in Li and Zhang (2015), it is 365 days. One can consider these articles are taking a similar method 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 should we select the criterion. When focusing on the time length construction method, the question is how many days is the most appropriate option to construct a potential bidders set.

From the previous section, the SEMUP model is considered a more general model than every SEMCP model. From the model setting, SEMUP model is invariant to different construction of potential bidders set. Meanwhile, SEMCP models will give different private valuation when the construction method changes. Thus, the difference between private values of the SEMUP model and the SEMCP model can be used to indicate whether the construction method of potential bidders set is plausible in terms of equilibrium strategies and beliefs. If I choose the distance between the distribution of private values of entry bidders from the SEMUP model and that from the SEMCP model as a rationale, the option of interest should be the SEMCP model that gives the least distance. In the construction method in terms of a continuous variable, this selection is equivalent to find a specific value (or interval) that gives least distance between the two distributions.

Suppose the construction method of potential bidders set is based on some rules in terms of a continuous variable. I denote this variable by $\xi$ with domain $\Xi$. 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,

$$d(\xi) = \sup_{v\in[0,\bar{v}]} |F_{up}(v) - F_{cp}(v;\xi)| \tag{39}$$

From this setting, the minimizer $\xi^*$ of the function $d(\xi)$ on its support is of most interest. In Proposition 1, it reveals 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 $\xi$ for any $v \in [0,\bar{v}]$, and that $\xi$ has a compact support $\Xi$. The function $d(\xi)$ is bounded below and attains its infimum on the support.*

Proposition 1 is important because it allows the comparison between two distributions from the SEMCP and SEMUP model will work when the construction method involves a continuous variable. Despite this good property of the distance function, the uniqueness of minimizer does not necessarily hold. A simple numerical simulation in the next section can illustrate this lack of uniqueness.

After we are able to estimate the private values from the two distributions, the estimated distance between a SEMCP and a SEMUP model can be obtained from the two empirical cumulative distribution functions. Therefore, I can get the estimate of distance from,

$$\hat{d}(\xi) = \sup_{v\in[0,\bar{v}]} \left| \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{v}^j_{up,i} \leq v\} - \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{1}^{n_j} 1\{\hat{v}^j_{cp,i} \leq v\} \right| \tag{40}$$

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 variable to choose in order to construct the potential bidders set.

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 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.

I formulate the KS test to indicate 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 using the recovered private values,

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

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

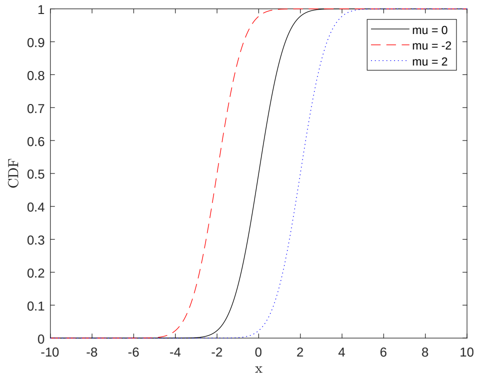
## 3.2 A Numerical Illustration of Minimum Distance

Proposition 1 provides the existence of the minimizer $\xi$. 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 $F_{up}(v)$.

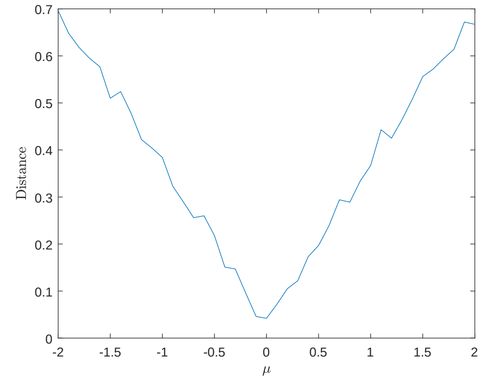
For example, in Figure 1(a), the black solid line represents a standard normal distribution function, i.e., the mean $\mu = 0$, and the standard deviation $\sigma = 1$. The red dashed line and the blue dotted line refers to the CDF distribution with different mean $\mu = -2$ and $\mu = 2$ with identical $\sigma = 1$. 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 $\mu$ 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 $\mu$ as the mean of distribution and standard deviation $\sigma = 1$. Based on the random draws, for each value of $\mu$ 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 $\mu$, I can compute the distance between the empirical CDF from random draws of specific $\mu$ and the empirical CDF from the initial standard normal distribution random draws. Figure 1(b) shows it works quite good as the lowest distance occurs around $\mu = 0$.

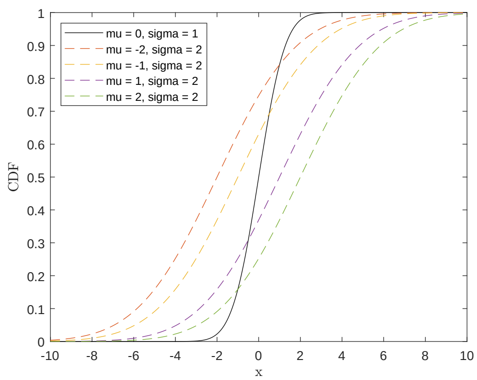
However, when the shape of distributions, $F_{cp}(v; \xi)$ and $F_{up}(v)$ are not completely the same, the minimum of the distance function of two distributions is hard to analyze and even not unique. For example, in Figure 1(c), the solid line refers to the CDF of normal distribution with $\mu = 0, \sigma = 1$, and dashed lines represent different CDF of normal distributions with different values of $\mu$ and $\sigma = 2$. When the value of $\mu$ 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 1(d). 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 - $\mu$ curve.



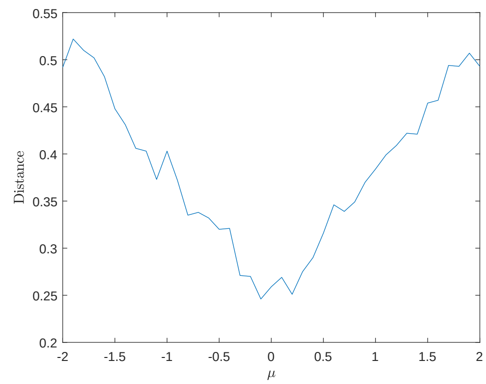

(a) Different $\mu$ with Identifcal $\sigma$




(b) Simulated Distance under Different $\mu$ with Identifcal $\sigma$




(c) Different $\mu$ with Different $\sigma$




(d) Simulated Distance under Different $\mu$ with Different $\sigma$

Figure 1: Normal Distribution Example

# 4 Empirical Application: Caltrans Procurement Auctions

In this section, I examine the frequently used method in literature that uses time information as a criterion to assign 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.

## 4.1 Data

I employ the data set of Caltrans procurement auctions studied in Bajari et al. (2014). 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 et al. (2014), 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 |

## 4.2 Estimation with Auction Level Characteristics

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 Gentry and Li (2014), 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.

### 4.2.1 Estimation Procedure of a SEMCP model

Based on the identification results and the assumptions above, the estimation equation for a procurement auction in a SEMCP model can be written as,

$$\hat{c}_{ji} = b_{ji} - \frac{1}{N_j - 1} \frac{1 - \hat{\hat{s}}(z_j'\beta, x_j)\hat{G}^*(b_{ji}|z_j, x_j)}{\hat{\hat{s}}(z_j'\beta, x_j)\hat{g}^*(b_{ji}|z_j, x_j)} \tag{42}$$

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.

To deal with the estimation of single index and distribution of bids, 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,

$$\log(b_{ji}) = \beta_0 + z_j'\beta + x_j'\zeta + e_{ji} \tag{43}$$

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,

$$G_{B|Z,X}(b_{ji}|z_j, x_j) = Prob(\alpha_0 + z_j'\beta + x_j'\zeta + e_{ji} \leq \log(b_{ji})) = F_e(\log(b_{ji}) - \alpha_0 - z_j'\beta - x_j'\zeta) \tag{44}$$

and

$$g_{B|Z,X}(b_{ji}|z_j, x_j) = \frac{\partial}{\partial b} F_e(\log(b_{ji}) - \alpha_0 - z_j'\beta - x_j'\zeta) = f_e(\log(b_{ji}) - z_j'\beta - x_j'\zeta)/b_{ji} \tag{45}$$

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 (44) and equation (45).

To estimate the entry threshold, 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 (42).

### 4.2.2 Estimation Procedure of a 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,

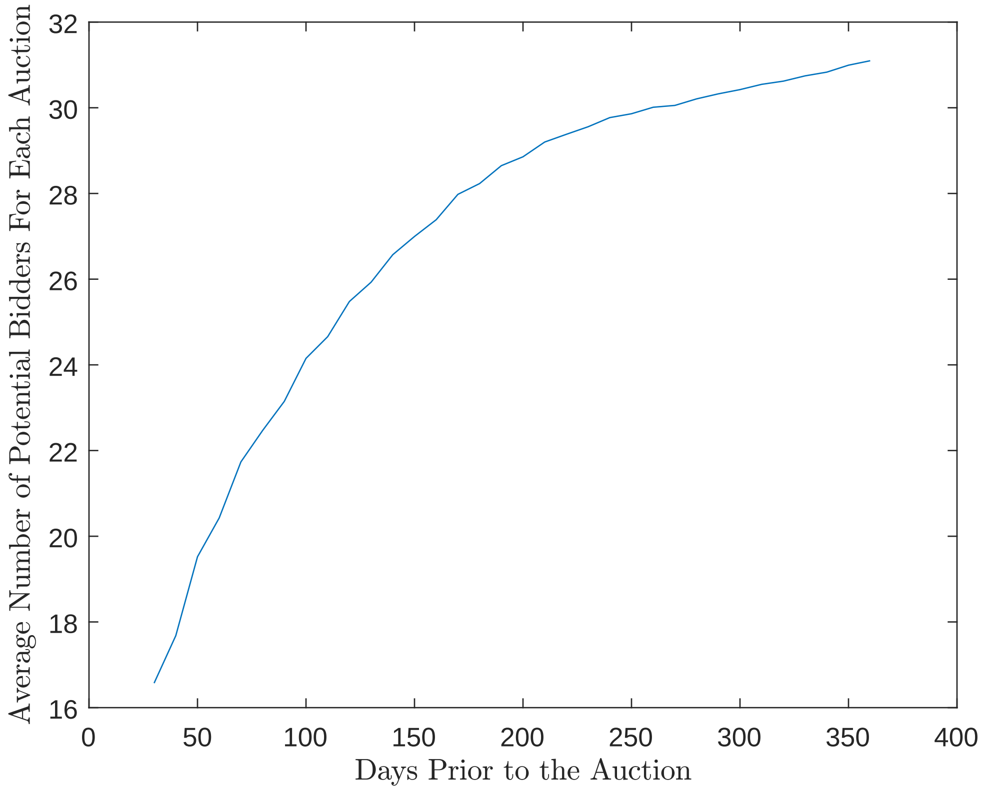
$$\hat{c}_{ji} = b_{ji} - \frac{\hat{G}^{(L1)}(b_{ji}; z_j, x_j) - \hat{G}^{(L2)}(b_{ji}; z_j, x_j)}{\hat{g}^{(L2)}(b_{ji}; x_j)} \tag{46}$$

To get distributions of the lowest and second lowest bids, I follow the linear regression in equation (43) and get the error terms for each bidder. The lowest bid distribution is estimated through a kernel density using the error terms calculated from 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. The distribution of bids can be similarly estimated through the transformation in Equation (44) and (45). After the estimation of distribution values, I can recover the pseudo values of private costs by plugging estimates into equation (46).

## 4.3 Application Result: Time as a Criterion

For different value of time length, the number of potential bidders for each auction changes. Figure 2 shows the relation between the average number of potential bidders of each auction and the selection of time length. From the figure, the 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

Figure 2: Average Number of Potential Bidders for Each Auction




potential bidders since the number of auctions and bidders is limited. The average number of potential bidders for each auction approximately doubles when the time length value increases from 30 days to 360 days. It reveals that in practical auction data, the potential bidders set can be very different when the criterion differs.

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. For the estimated private costs, I run a simple OLS regression between the two contruction methods. The regression result reveals an estimated slope of 0.901 which to some extent represents a 10% difference between the two construction methods overall. However, the worse situation is that some of the cost pairs indicate a very large difference on estimation from different constructions of potential bidders set. The 3 depicts the histogram of ratios of recovered private costs from 360 days to 30 days. The vertical red line represents that the ratio equals 1. However, from the histogram, we can observe that the ratios have a large value spread. Some of the extreme points can have very implausible values, such as less than 0.5 or greater than 2.

I similarly do the same thing for private costs recovered from potential bidders sets with different time length criterion. For different time lengths, private costs vary a lot. Figure 4 plots the histogram of ratios of recovered private costs. 4(a) and 4(b) have distribution surrounding the vertical line ratio = 1. Ratios in 4(b) seems to cluster more around 1. Meanwhile, the histograms in 4(c) and 4(d) locate more or less away from 1. Thus, as for Caltrans data, it is important to check the method of constructing potential bidders set for the research of entry behaviors. Furthermore, there is not a clear pattern of estimated slopes and private cost ratios. 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.

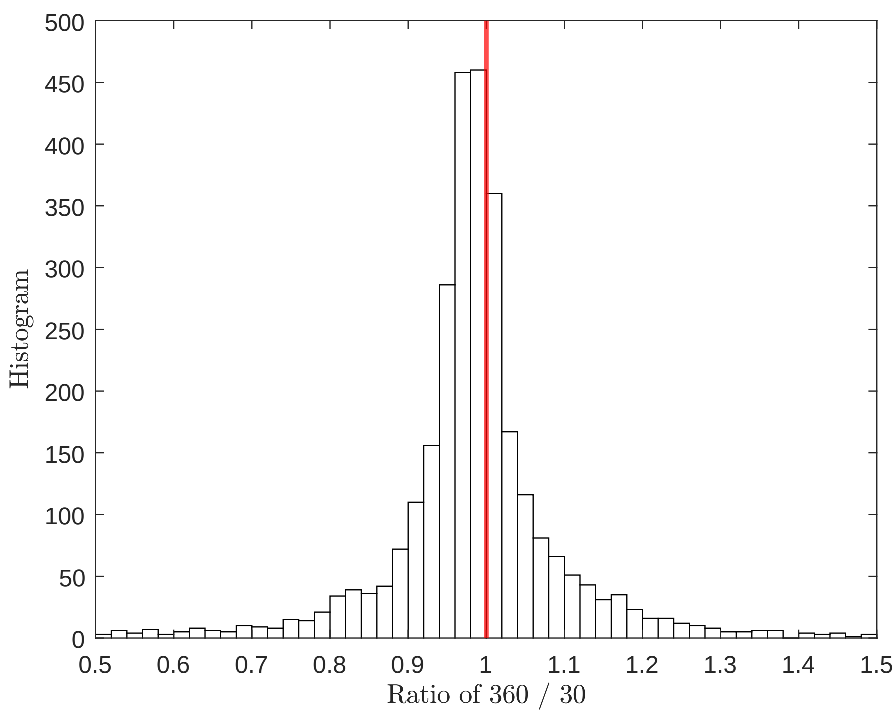
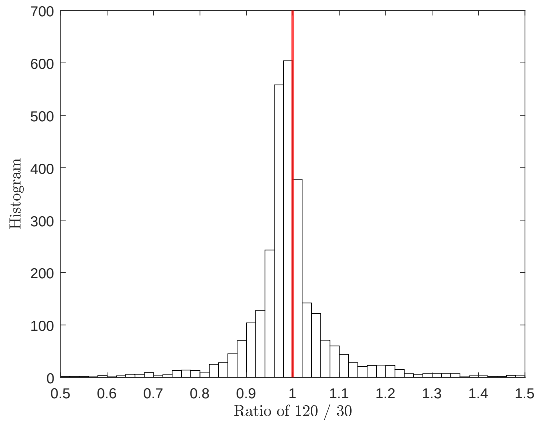
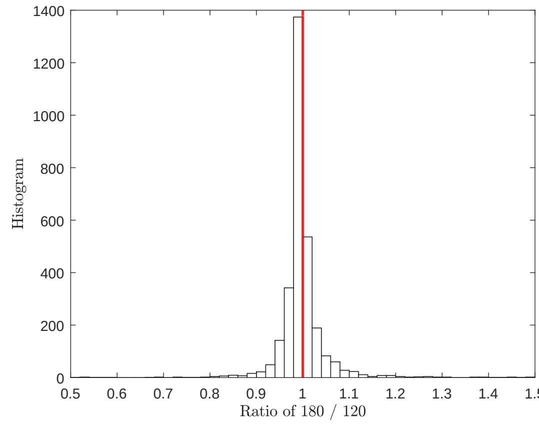


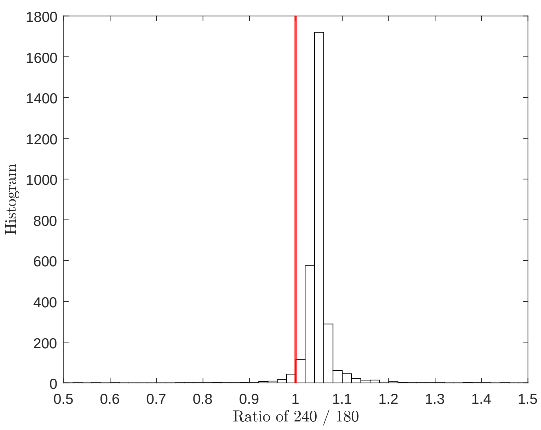

Figure 3: Comparison of Pseudo Private Costs: 30 days vs 360 days



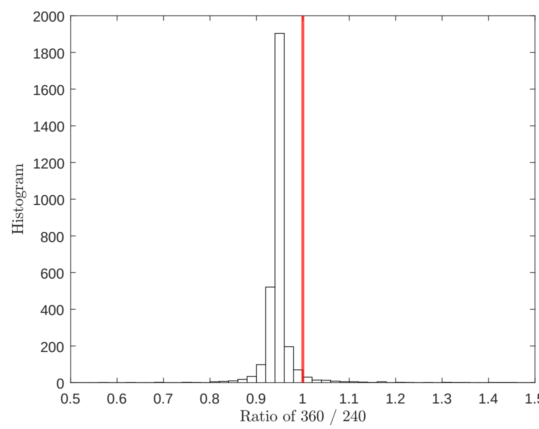

(a) Comparison of Pseudo Private Costs 30 vs 120




(b) Comparison of Pseudo Private Costs 120 vs 180




(c) Comparison of Pseudo Private Costs 180 vs 240




(d) Comparison of Pseudo Private Costs 240 vs 360

Figure 4: Histogram of ratios of Private Costs

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,

$$\hat{d}(\xi) = \sup_{c \in [0,\bar{c}]} \left| \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{1}^{n_j} 1\{\hat{c}^j_{up,i} \leq c\} - \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{1}^{n_j} 1\{\hat{c}^j_{cp,i} \leq c\} \right| \tag{47}$$

and the KS statistic is formed as,

$$KS = \sup_{c \in [0,\bar{c}]} \left| \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{1}^{n_j} 1\{\hat{c}^j_{up,i} \leq c\} - \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{1}^{n_j} 1\{\hat{c}^j_{cp,i} \leq c\} \right| \tag{48}$$

Figure 6(a) represents the difference between SEMUP and each SEMCP model. The two dotted lines represent the [5%, 95%] 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.

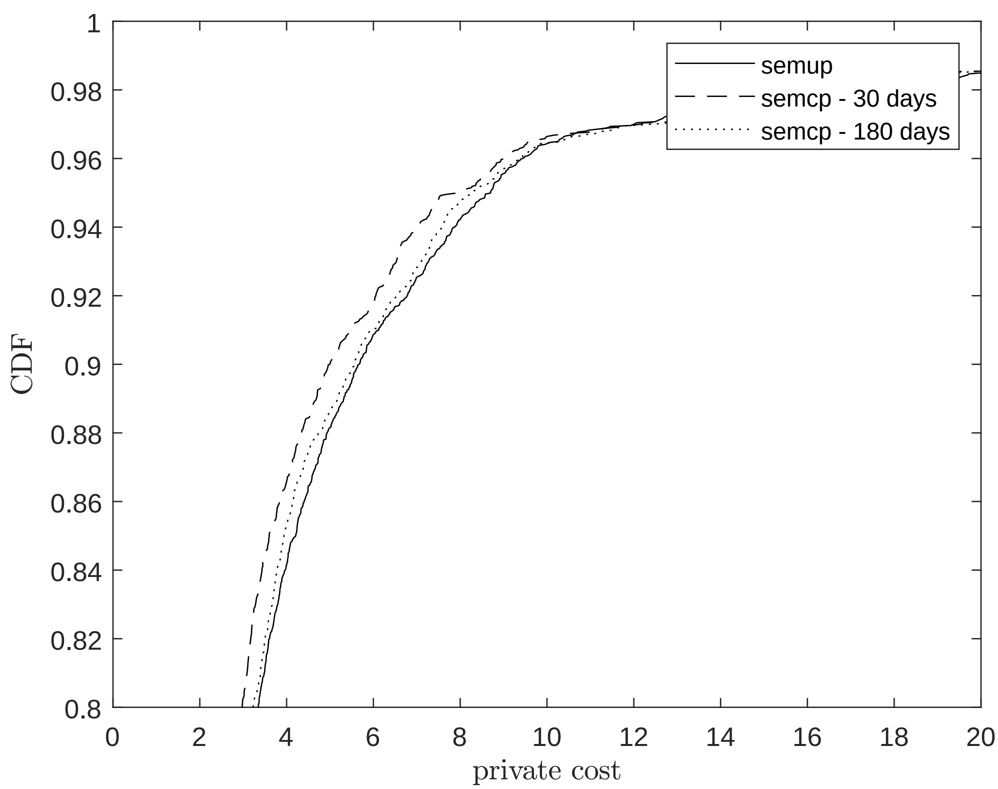
When we look into the empirical CDF recovered from SEMCP and SEMUP models, Figure 5 shows the difference of empirical CDF of two SEMCP models compared to a SEMUP model. As for time length being 180 days, the dotted line reveals a relative small distance in the figure. The dashed line indicates the empirical CDF has more deviation when the time length for construction is 30 days.

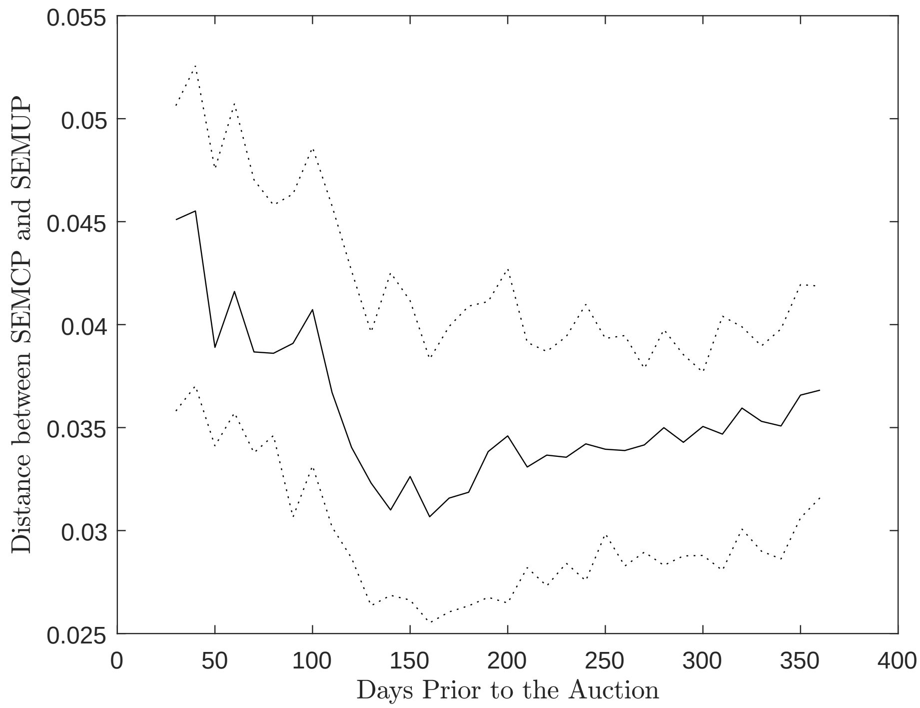
Figure 6(b) represents corresponding bootstrap p-values from the test defined in previous sections. On the 5% significant level, p-values within the interval $[120, 180]$ are relatively high. When the value of time length gets very small or very large, the equality of private cost distributions between SEMUP and SEMCP is rejected on the 5% significant level. This result behaves in the same way as the distance method.

# 5 Conclusion

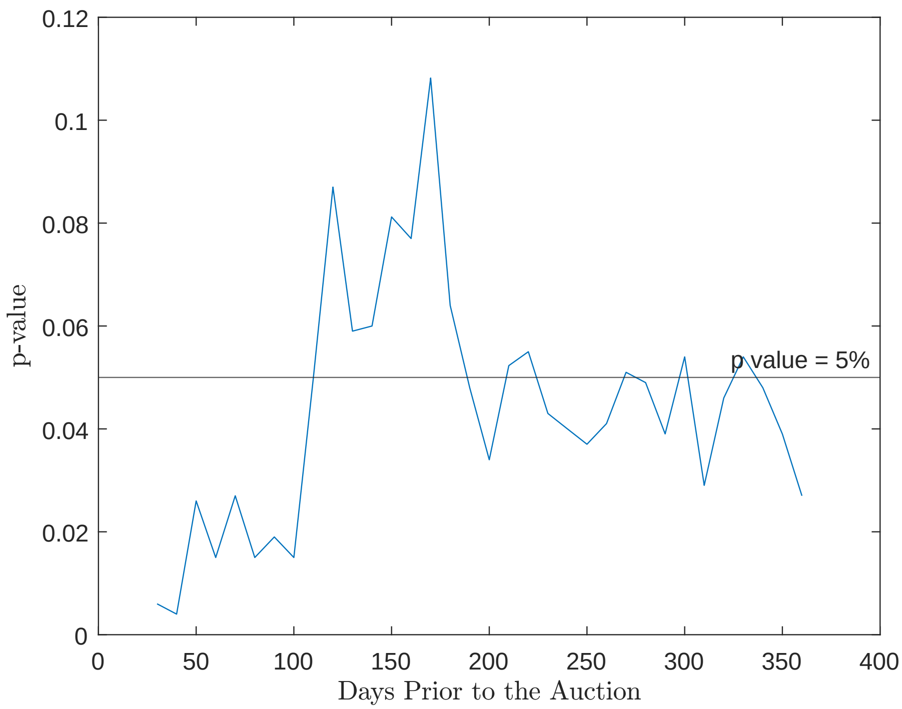
This paper proposed a criterion to distinguish potential bidders for practical auctions data. I developed the selective entry model with an unknown number of potential bidders. I showed the equilibrium result and identification of primitives for the model. The model generalizes the widely used selective 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

Figure 5: Distance between Private Cost Distributions of SEMCP and SEMUP models for Different Time Length






(a) Distance between Private Cost Distributions of SEMCP and SEMUP models for Different Time Length




(b) p Values for Different Time Length

Figure 6: Distance and 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 appropriate.

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 within 120 days to 180 days. The bootstrap p - values confirmed 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.

# Appendices

## 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 $\Delta$ refers to the single index described in the paper, $\Delta = Z'\beta$.

The distribution of private cost can be derived as,

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

## Appendix C: Connection between $p_l$ and $q_l$

We follow McAfee and McMillan (1987) to deal with the connection between the two kinds of probabilities.

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

$$\sum_{l=1}^{\bar{N}} \sum_{A,|A|=l} \beta_A = 1$$

Then the probability that $l$ bidders are present is,

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

The expected number of bidders is,

$$N^* = \sum_{l=1}^{\bar{N}} l p_l$$

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

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

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

$$N^* q_l = l p_l$$

## 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 (2012) p.27 Proposition 1.5.12, the following theorem shows the property of supremum of semicontinuous functions.

**Theorem 3** *If $(X,\tau)$ 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 $\xi$ with index $v$. From the assumption, $F_{cp}(v;\xi)$ is a continuous function of $\xi$ for any $v$ on its support. Therefore, $\mathcal{F}$ is a subset of $LSC(\xi)$.

I have the following result,

$$d(\xi) = \sup_{v \in [0,\bar{v}]} |F_{up}(v) - F_{cp}(v;\xi)| = \sup_{f \in \mathcal{F}} f_v(\xi) \tag{49}$$

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 $\xi$, $\Xi$ 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,

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

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,

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

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,

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

The PDF of it is,

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

Therefore, we have,

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

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

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

S. Athey and P. A. Haile. Identification of standard auction models. *Econometrica*, 70(6): 2107–2140, 2002.

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.

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

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.

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

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.

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

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

M. Gentry, T. Li, and J. Lu. Auctions with selective entry. *Games and Economic Behavior*, 105:104–111, 2017.

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.

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

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

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

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

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.

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.

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.

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.

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

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

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

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

U. Song. Nonparametric identification and estimation of a first-price auction model with an uncertain number of bidders. Technical report, Working paper, University of British Columbia, 2006.

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

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