STRUCTURAL ANALYSIS OF COLLUSION IN FIRST-PRICE AUCTIONS

LIUYI GUO

ABSTRACT. This paper provides the nonparametric identification results for asymmetric first-price auctions with unknown collusion schemes. Although collusion schemes take various forms in auctions, testing collusion crucially relies upon the assumed collusion scheme in the existing literature. This paper shows that regardless of the unknown collusion scheme, collusive bidders in a partial cartel can be identified from winning bids and identities of winners with auction-specific covariates satisfying an independence condition. Furthermore, the value distributions of collusive bidders can be identified under two types of cartels characterized by McAfee and McMillan (1992): strong cartels with efficient collusion mechanisms; and weak cartels with lottery mechanisms. Based on the identification results, a testing procedure is provided to recover the identities of collusive bidders. The test method is applied to the California highway procurement auctions. The test results suggest no statistical evidence of large-scale collusion among non-fringe firms in the sample.

1. INTRODUCTION

In the auction literature, collusion refers to a bidding ring or a cartel in which collusive bidders devise a mechanism, and one is selected as the representative of the cartel.[1] The competition between collusive bidders is successfully suppressed in the main auction. By McAfee and McMillan (1992), bidding rings in first-price auctions are classified into two categories: weak cartels in which side-payments are not allowed; and strong cartels in which side-payments are allowed. One of the main insights of McAfee and McMillan (1992) was that strong cartels could be efficient by implementing a knockout auction to decide the designated bidder with the highest private value among cartel members.[2] Though weak cartels can still be profitable, the impossibility of transfers prevents efficient collusion.

DEPARTMENT OF ECONOMICS, 2935 RESEARCH PARKWAY, TEXAS A&M UNIVERSITY, COLLEGE STATION, TX 77843-4228, USA.

E-mail address: lguo49@tamu.edu.

Date: November 1, 2022.

Key words and phrases. Collusion, first-price auctions.

The author thanks Yonghong An, Jackson Bunting, Zheng Fang, Matthew Gentry, Huiyi Guo, Fernando Luco, Qi Li, Matt Matsen, Paulo Somaini, and Silvana Krasteva for their helpful discussions.

[1] By Marshall and Marx (2007), there are two types of mechanisms. The first is the bid coordination mechanism, in which the cartel can arrange transfers and recommend bids to the ring members but cannot control the bids. The second is the bid submission mechanism, in which the cartel can control the bids. This paper focuses on the bid submission mechanism.

[2] A cartel is efficient if and only if the designated bidder has the highest value among ring members.

McAfee and McMillan asserted that the optimal collusion scheme for weak cartels is to allocate the object to cartel members with a constant probability, which implies that weak cartels are inefficient.[3]

Most empirical auction papers on collusion presumed that the cartel's designated bidder has the highest value among the ring members, e.g., Baldwin et al. (1997), Bajari and Ye (2003), Aryal and Gabrielli (2013), Marmer et al. (2016), and Schurter (2017). However, cartels are more likely to be inefficient from empirical and theoretical perspectives. First, side-payments are seldom observed in practice. Many real cases investigated by competition authorities do not involve any exchange of side-payments (Che et al., 2018). This means the cartels in those cases could be inefficient by McAfee and McMillan's theory. Second, implementing efficient cartels hinges on a pre-auction mechanism to reveal bidders' private values. On many occasions, the designated bidder is predetermined before the realization of private values. For an example of wheat auctions studied in Banerji and Meenakshi (2004), cartels may adopt a prearranged bid rotation scheme since wheat auctions are usually fast-paced and bidders are limited to exchanging information in this environment. Another canonical example is the "Electrical Conspiracy", in which 29 electrical equipment manufacturers took turns in submitting bids in procurement auctions (Durlauf and Blume, 2016). In the financial industry, a number of large companies, including Blackstone and Goldman Sachs, were accused of colluding by agreeing not to compete with each other in leverage buyout transactions between 2003 and 2007.[4] The total amount of settlement with investors is more than $590 million.

This paper examines two theoretical issues in the structural analysis of collusion in first-price auctions: (i) Can researchers identify collusive bidders with an unknown collusion scheme? (ii) Is it possible to distinguish different types of cartels if collusion is detected? This paper addresses these two issues in the asymmetric first-price auction model within the independent private value (IPV) paradigm. Existing methods of identifying collusive bidders typically exploit information on the structure of bidder markups with the assumption of efficient cartels. However, in first-price auctions, the non-collusive bidder's markup can be expressed as a function of the distribution of the highest competing bid, which depends upon the underlying collusion scheme. Thus, the test for collusion could be misleading if the structure of bidder markups is misspecified with the assumption of efficient cartels. It is desired to examine collusion in a unified framework for incorporating various collusion schemes. The second issue concerns whether different types of cartels can be distinguished if collusion is detected with empirical auction data. The information on the types of cartels is valuable to policymakers since the optimal response could depend

[3] Pesendorfer (2000) showed that random allocation is an optimal mechanism for weak cartels only in single-unit first-price auctions. When there are multiple items for sale, weak cartels can do better by employing a mechanism called "Ranking Mechanism".

[4] The case is Civil Action No. 07-12388-EFH.

upon the collusion scheme. For example, if the seller is a government concerned only with efficiency, the seller will be indifferent to the existence of efficient cartels. If the seller's objective is to maximize his surplus, the optimal response may differ between efficient and inefficient cartels. For example, Gruyer (2009) characterized an optimal auction mechanism for efficient cartels but not weak cartels.

To start with, Section 3 establishes an asymmetric first-price auction model with a partial cartel, accommodating various collusion schemes. The main assumption in the collusion model is that the cartels are formed ex ante, i.e., before bidders learn their private values. It is also assumed that the set of collusive bidders and the collusion scheme are unchanged across auctions. Following the literature on mechanism design theory, the direct collusion scheme is modeled as a map from the space of private value profiles to the set of distributions over the set of collusive bidders. By construction, the ex ante value distribution of the designated bidder is well-defined, and the designated bidder does not necessarily have the highest value among collusive bidders, reflecting an important feature of inefficient cartels. The characterization of equilibrium bidding strategies for serious bidders follows from the literature on asymmetric first-price auctions.

The objects to be identified are the primitives of the collusion model: the set of collusive bidders, the collusion scheme, and the bidders' value distributions. Section 4 studies identification of collusive bidders without assuming a particular collusion scheme. As shown later in Section 4, in terms of winning bids and identities of winners, there is an equivalence relation between the competitive model and the collusion model with efficient cartels. But this equivalence relation no longer holds in the collusion model with inefficient cartels. This leads to new challenges in identification. One cannot simply apply Theorem 3.2 in Athey and Haile (2007) to obtain identification of bid distributions whenever cartels are inefficient. This paper presents some novel results in the presence of inefficient cartels. The identification results are based on the observation in the collusion model that any winning bid must be submitted from either a designated bidder or a non-collusive bidder. A critical insight in this paper is that if a bidder is non-collusive, one can still correctly recover this bidder's bid distribution from winning bids and winners' identities despite the collusion scheme and the configuration of cartels being unknown. The recovered bid distribution, together with the distribution of winning bids, can be used to correctly infer the bidder's markup as well as value distribution if the bidder is non-collusive. This finding provides a basis for identifying identities of collusive bidders by an instrument approach credited to Schurter (2017). More precisely, if some auction-level instruments are independent of the collusion scheme and value distributions, one can construct value distributions that respond differently to auction-level instruments for collusive and non-collusive bidders. Accordingly, the set of collusive bidders can be identified from variations in instruments. As a result, non-collusive bidders' value distributions are nonparametrically identified.

Section 4 also derives a non-identification result for the collusion structure consisting of the collusion scheme and the profile of collusive bidders' value distributions. This non-identification result is obtained by observing that any collusion structure is equivalent to another collusion structure in the ex ante distribution of the designated bidder's private values and identities. This non-identification result suggests that collusive bidder's value distributions cannot be identified without further information on the collusion mechanism, even if the identities of collusive bidders are known. While Schurter (2017) achieved the identification results for efficient cartels, this paper shows identification of collusive bidders' value distributions in weak cartels with lottery mechanisms, which are the optimal collusion mechanisms for weak cartels described in McAfee and McMillan (1992). Therefore, the value distributions of collusive bidders can be identified under two types of cartels: strong cartels with efficient collusion mechanisms; and weak cartels with lottery mechanisms. Additional information on value distributions is required to distinguish the types of cartels. For example, suppose some non-collusive bidder is the same type as the collusive bidder. In that case, the non-collusive bidder's inferred value distribution can be exploited to distinguish the types of cartels.

This paper's main contribution is to characterize identification in the collusion model with general cartel mechanisms. In asymmetric IPV first-price auctions with an unknown cartel mechanism, this paper shows that identities of collusive bidders and value distributions of non-collusive bidders can be identified with auction-level covariates satisfying a regular independence condition. Compared to Schurter (2017), this paper does not assume an efficient collusion scheme was employed. Relaxing the assumption of efficient cartels is non-trivial for identification. The assumption of efficient cartels is crucial in Schurter (2017) since this assumption is necessary to invoke an identification result in Athey and Haile (2007) for recovering bid distributions and value distributions. Therefore, the identification arguments in Schurter (2017) are inapplicable to the case of inefficient cartels. Nonetheless, this paper shows that identities of collusive bidders can still be identified without the assumption of efficient cartels. In addition, this paper is the first to show identification of collusive bidders' value distributions in inefficient cartels and provide conditions for distinguishing different types of cartels in first-price auctions.

This paper also contributes to the empirical works on testing collusion in procurement auctions. Previous empirical works mainly focused on testing the implications of the competitive model, e.g., Bajari and Ye (2003) and Krasnokutskaya (2011). However, since collusive bidders can use simple strategies to mimic competitive bidding behavior, it is generally insufficient to detect collusion by only testing the necessary conditions of the competitive model. The identification results derived in Section 4 naturally provide a testable implication of collusion. This paper proposes a test statistic based on integrated quantile functions and applies the test method to the California highway procurement auctions. The test results suggest no statistical evidence of large-scale collusion among large firms. This paper confirms the finding of a previous study. Using the same dataset, Aryal and Gabrielli (2013) implemented a reduced-form bid analysis proposed by Bajari and Ye (2003) and found evidence of collusion between several groups of firms. However, Aryal and Gabrielli concluded no collusion based on a two-step structural analysis.

The paper proceeds as follows. Section 2 reviews the related empirical auction literature on collusion. Section 3 introduces the collusion model with a partial cartel and characterizes serious bidders' equilibrium bidding functions. Section 4 presents the main identification results. Section 5 describes test statistics, nonparametric estimation, and inference procedure. Section 6 illustrates the test with an empirical application to the California highway procurement auctions. Finally, Section 7 concludes. All proofs of theoretical results can be found in the appendix.

2. Related Literature

Collusion was documented by the literature at many different auctions for a long time, including highway construction projects, school milk delivery, and timber auctions, among others. Without prior knowledge of the existence of cartels, some econometric tests have been proposed to detect collusive bidding behavior (Porter and Zona, 1993; Baldwin et al., 1997; Price, 2008). Another strand of the literature on collusion examined the behaviors of collusive bidders with data from antitrust legislation, in which both identities of collusive bidders and the collusive agreement are accessible to researchers (Porter and Zona, 1999; Pesendorfer, 2000; Asker, 2010).

This paper is related to the literature on collusion in IPV first-price auctions under the framework of Guerre et al. (2000). Much of the existing empirical works presumed efficient cartels in a static collusion model. In this paper, the assumption of efficient cartels is relaxed. This paper is most closely related to Schurter (2017), who identified the efficient cartel with instrumental variables that affect bidding behavior but not value distributions. Bajari and Ye (2003) showed two necessary conditions of bids generated from a competitive model. Two conditions are conditional independence and exchangeability of bids, which can be used to test collusive behavior through a regression-based method. Aryal and Gabrielli (2013) proposed a two-step procedure to test collusion. First, some bidders are classified as potential collusive bidders if they fail the test proposed by Bajari and Ye (2003). Second, collusion implies the first-order stochastic dominance of the recovered cost distributions between the competition and collusion models. Athey et al. (2011) observed that the open auction prices are substantially lower than first-price auction prices in U.S. forest service (USFS) timber auctions, which may indicate potential collusion in open auctions.

Besides IPV first-price auctions, collusion is studied in other information structures and bidding formats. Hendricks et al. (2008) analyzed collusion in the environment of affiliated private value and common value. Conley and Decarolis (2016) proposed statistical tests to detect the bid coordination in average-bid public procurements. Kawai and Nakabayashi (2021) identified collusion by focusing on rebids bids after bids failed to meet the secret reserve price. Marmer et al. (2016) studied identification of efficient bidding rings in English auctions and derived a testable implication. Recently, some empirical works begin to examine collusion in dynamic models (Chassang and Ortner, 2019; Chassang et al., 2021; Kawai et al., 2021; Ortner et al., 2021; Chassang et al., 2022). For more detailed surveys of empirical works on collusion, see Harrington (2005), Hendricks and Porter (2007), Hortaçsu and Perrigne (2021), and Li and Zheng (2021).

3. The Collusion Model

3.1. **Setup.** The collusion model herein is a natural extension of McAfee and McMillan (1992). The first-price sealed-bid auctions with collusion are modeled as static Bayesian games. A non-strategic auctioneer sells one indivisible good to a set of ex ante asymmetric bidders $\mathcal{N}$ through first-price sealed-bid auctions. A proper subset of bidders $\mathcal{K} \subset \mathcal{N}$ form a cartel.[5] If the set of non-collusive bidders is non-empty (i.e., $\mathcal{N} \setminus \mathcal{K} \neq \emptyset$), then the cartel is referred to as a partial cartel. All bidders are risk-neutral and have independent private values drawn from absolutely continuous distribution functions $\{F_j(\cdot)\}_{j\in\mathcal{N}}$ on a common value interval $\mathcal{V}_j = [\underline{v}, \overline{v}]$.[6] The probability density functions $\{f_j(\cdot)\}_{j\in\mathcal{N}}$ are assumed to be bounded and strictly positive on the interval $(\underline{v}, \overline{v}]$.

Before the main auction, a designated bidder is selected from the cartel members. The selected bidder will bid as the representative of the cartel in the main auction. The problem facing the cartel is how to choose the designated bidder. By the revelation principle in mechanism design theory, it is without loss of generality to focus only on incentive-compatible, direct collusion mechanisms, in which each collusive bidder truthfully reports his private value before the main auction. The formal definition of a direct collusion scheme is as follows.

Definition 1. *A direct collusion scheme Γ is a map from the space of private value profiles to the set of distributions over the set of collusive bidders, i.e.,*

$$\Gamma : \prod_{k\in\mathcal{K}} \mathcal{V}_k \longrightarrow \Delta\left(\mathcal{K}\right),$$

where $\Delta(\mathcal{K})$ is a class of distributions over the set of collusive bidders $\mathcal{K}$.

[5] At least two collusive bidders form a cartel. When $\mathcal{K} = \emptyset$ or $|\mathcal{K}| = 1$, the collusion model naturally reduces to the competitive model. The benchmark model only considers a single cartel with fixed ring members. The equilibrium outcome can be easily generalized in the case of multiple cartels.

[6] As shown later, if a collusion model is misspecified as a competitive model, the common support of values implies the difference of the upper extremities of pseudo values between collusive bidders and non-collusive bidders, which may provide a testing implication. Implementing such a nonparametric test is probably hindered by boundary effects caused by kernel density estimators. Formal testing approaches based on this fact have not been developed.

A collusion mechanism may involve complicated rules related to enforcement, side-payments, and cover bids. The essence of collusion in auctions is the selection of the designated bidder, which dwarfs the competition. By Definition 1, any direct collusion scheme Γ specifies the probability mass function over the designated bidder's identity conditional on the realization of collusive bidders' private value profile, i.e., $\mathbb{P}\left\{I_{\mathcal{K}}=k|\boldsymbol{V}_{\mathcal{K}}=\boldsymbol{v}_{\mathcal{K}};\Gamma\right\}$, where $\boldsymbol{V}_{\mathcal{K}} \equiv (V_k)_{k\in\mathcal{K}}$ and $I_{\mathcal{K}}$ an indicator of the identity of the designated bidder. The collusion structure is a pair of the collusion scheme Γ and the profile of value distributions $\{F_k(\cdot)\}_{k\in\mathcal{K}}$. It can be shown that the ex ante distribution of the designated bidder's private values is well-defined given the collusion structure. To see this, let $V_{\mathcal{K}}$ denote the designated bidder's private values. For the direct collusion scheme Γ, the ex ante value distribution of the designated bidder is denoted by

$$F_{\mathcal{K}}(v;\Gamma) \equiv \mathbb{P}\left\{V_{\mathcal{K}} \le v;\Gamma\right\}.$$

By the law of total probability,

$$F_{\mathcal{K}}(v;\Gamma) = \sum_{k\in\mathcal{K}} \mathbb{P}\left\{V_{\mathcal{K}} \le v, I_{\mathcal{K}}=k;\Gamma\right\}.$$

When collusive bidder $k\in\mathcal{K}$ is selected as the designated bidder, it follows that

$$\mathbb{P}\left\{V_{\mathcal{K}} \le v, I_{\mathcal{K}}=k;\Gamma\right\} = \int_{D_k(v)} \mathbb{P}\left\{I_{\mathcal{K}}=k|\boldsymbol{V}_{\mathcal{K}}=\boldsymbol{v}_{\mathcal{K}};\Gamma\right\} f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}})\, d\boldsymbol{v}_{\mathcal{K}}, \tag{1}$$

where $D_k(v) \equiv \left\{\boldsymbol{v}_{\mathcal{K}} \in [\underline{v},\overline{v}]^K : \underline{v} \le v_k \le v, \underline{v} \le v_\ell \le \overline{v}, \ell \ne k, \ell \in \mathcal{K}\right\}$ with $K=|\mathcal{K}|$, and $f_{\boldsymbol{V}_{\mathcal{K}}}(\cdot)$ is the joint density of collusive bidders' private values.[7] Hence, given a collusion scheme Γ, the ex ante value distribution of the designated bidder can be expressed as

$$F_{\mathcal{K}}(v;\Gamma) = \sum_{k\in\mathcal{K}} \int_{D_k(v)} \mathbb{P}\left\{I_{\mathcal{K}}=k|\boldsymbol{V}_{\mathcal{K}}=\boldsymbol{v}_{\mathcal{K}};\Gamma\right\} f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}})\, d\boldsymbol{v}_{\mathcal{K}}, \tag{2}$$

The integrand on the right-hand side of equation (2) is the product of the conditional probability mass function over identities of the designated bidder (which is specified by the collusion scheme Γ in Definition 1) and the joint density function of collusive bidders' private values. The ex ante value distribution of the designated bidder is essential in characterizing equilibrium bidding behavior in the main auction. Equation (2) establishes a connection between the ex ante value distribution of the designated bidder and the collusion scheme. Two examples are presented below for illustrative purposes.

Example 1. *Suppose that the designated bidder is randomly selected with an equal probability of $1/K$ for any profile of private values, i.e.,*

$$\mathbb{P}\left\{I_{\mathcal{K}}=k|\boldsymbol{V}_{\mathcal{K}}=\boldsymbol{v}_{\mathcal{K}};\Gamma\right\} = \mathbb{P}\left\{I_{\mathcal{K}}=k;\Gamma\right\} = 1/K,$$

[7] The identity holds by observing that

$$\mathbb{P}\left\{V_{\mathcal{K}} \le v, I_{\mathcal{K}}=k;\Gamma\right\} = \mathbb{P}\left\{V_k \le v, I_{\mathcal{K}}=k;\Gamma\right\} = \int_{D_k(v)} f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}}|I_{\mathcal{K}}=k;\Gamma)\mathbb{P}\left\{I_{\mathcal{K}}=k;\Gamma\right\} d\boldsymbol{v}_{\mathcal{K}},$$

where $f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}}|I_{\mathcal{K}}=k;\Gamma)\mathbb{P}\left\{I_{\mathcal{K}}=k;\Gamma\right\} = \mathbb{P}\left\{I_{\mathcal{K}}=k|\boldsymbol{V}_{\mathcal{K}}=\boldsymbol{v}_{\mathcal{K}};\Gamma\right\} f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}})$.

for all $k \in \mathcal{K}$ and $\boldsymbol{v}_{\mathcal{K}} \in [\underline{v}, \overline{v}]^K$. Then, equation (2) reduces to

$$F_{\mathcal{K}}(v; \Gamma) = \sum_{k \in \mathcal{K}} \int_{D_k(v)} K^{-1} f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}})\, d\boldsymbol{v}_{\mathcal{K}} = K^{-1} \sum_{k \in \mathcal{K}} F_k(v).$$

The ex ante value distribution of the designated bidder is equal to a weighted sum of collusive bidders' value distributions if the designated bidder is randomly selected.

Example 2. *Suppose that the cartel agreement Γ_e is efficient, i.e., the designated bidder has the highest value among all collusive bidders. Then for all $k \in \mathcal{K}$ and $\boldsymbol{v}_{\mathcal{K}} \in [\underline{v}, \overline{v}]^K$,*

$$\mathbb{P}\left\{I_{\mathcal{K}} = k | \boldsymbol{V}_{\mathcal{K}} = \boldsymbol{v}_{\mathcal{K}}; \Gamma_e\right\} = \mathbb{1}\left\{v_k \geq \boldsymbol{v}_{-k}\right\},$$

where $v_k \geq \boldsymbol{v}_{-k}$ if and only if $v_k \geq v_\ell$ for all $\ell \neq k$. Then, equation (2) reduces to

$$F_{\mathcal{K}}(v; \Gamma_e) = \sum_{k \in \mathcal{K}} \int_{D_k(v)} \mathbb{1}\left\{v_k \geq \boldsymbol{v}_{-k}\right\} f_{\boldsymbol{V}_{\mathcal{K}}}(\boldsymbol{v}_{\mathcal{K}})\, d\boldsymbol{v}_{\mathcal{K}} = \sum_{k \in \mathcal{K}} \mathbb{P}\left\{V_k \leq v, \max_{\ell \in \mathcal{K} \setminus \{k\}} V_\ell < V_k\right\}.$$

By the independence of private values, observe that for each $k \in \mathcal{K}$,

$$\mathbb{P}\left\{V_k \leq v, \max_{\ell \in \mathcal{K} \setminus \{k\}} V_\ell < V_k\right\} = \int_{\underline{v}}^{v} \prod_{\ell \in \mathcal{K} \setminus \{k\}} F_\ell(t)\, dF_k(t).$$

Hence,

$$F_{\mathcal{K}}(v; \Gamma_e) = \sum_{k \in \mathcal{K}} \int_{\underline{v}}^{v} \prod_{\ell \in \mathcal{K} \setminus \{k\}} F_\ell(t)\, dF_k(t) = \prod_{k \in \mathcal{K}} F_k(v).$$

The ex ante value distribution of the designated bidder is equal to the product of collusive bidders' value distributions if the cartel is efficient.

The following assumptions are imposed on the cartel throughout the paper.

Assumption 1. *A partial cartel is formed ex ante, i.e., before bidders learn their private values, and*
(i) the set of collusive bidders $\mathcal{K}$ and the collusion scheme Γ do not change across auctions;
(ii) the non-collusive bidders do not know the identity of the designated bidder;
(iii) $F_k(v|I_{\mathcal{K}} = k; \Gamma)$ is absolutely continuous in v for each collusive bidder $k \in \mathcal{K}$.
(iv) $\mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\}$ is strictly positive for each collusive bidder $k \in \mathcal{K}$.

With Assumption 1(i), this paper abstracts from the issues of cartel formation and enforcement. The cartel is merely assumed to be exogenously formed, and it is stable across auctions. Assumption 1(ii) concerns non-collusive bidders' information on the cartel, which is standard in the auction literature analyzing static collusion models. It distinguishes between the classic competitive and the collusion models. Assumption 1(iii) requires that every collusive bidder has an absolutely continuous value distribution conditional on being selected as the designated bidder. Assumption 1(iii) ensures that the value distribution of the designated bidder is differentiable for technical convenience. Assumption 1(iv) assumes that the ex ante probability of being selected as the designated bidder under the collusion scheme Γ is strictly positive for each collusive bidder. Assumption 1(iv) is necessary for the identification results in Section 4.

3.2. **Equilibrium Bidding Strategies.** The collusion model incorporates a competitive bidding stage as non-collusive bidders participate in the main auction. This subsection characterizes equilibrium bidding strategies for the serious bidders in the main auction.

A non-collusive bidder faces competition from an unknown designated bidder and other non-collusive bidders; meanwhile, the designated bidder only bids competitively against all non-collusive bidders. In first-price sealed-bid auctions, the designated bidder's bid is higher than complementary bids submitted from all other collusive bidders by design. In equilibrium, all non-collusive bidders form a correct belief of the designated bidder's value distribution.[8] Let $\beta_{\mathcal{K}}(\cdot)$ and $\{\beta_i(\cdot)\}_{i\in\mathcal{N}\setminus\mathcal{K}}$ denote equilibrium bidding strategies defined over the value interval $[\underline{v},\overline{v}]$ for the designated bidder and non-collusive bidders, respectively. Each serious bidder in the main auction seeks to optimize his expected surplus. The equilibrium bidding strategies of non-collusive bidders and the designated bidder can be characterized by the following optimization problems respectively:

$$\max_{b_i}\,(v_i - b_i)\,\mathbb{P}\left\{\beta_{\mathcal{K}}\left(V_{\mathcal{K}}\right) \leq b_i\right\} \prod_{j\in\mathcal{N}\setminus(\{i\}\cup\mathcal{K})} \mathbb{P}\left\{\beta_j\left(V_j\right) \leq b_i\right\}, \tag{3}$$

and

$$\max_{b_{\mathcal{K}}}\,(v_{\mathcal{K}} - b_{\mathcal{K}}) \prod_{j\in\mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_j\left(V_j\right) \leq b_{\mathcal{K}}\right\}. \tag{4}$$

In the optimization problem (3), the first term is non-collusive bidder i's surplus if he wins the auction. The second and third terms in the optimization problem (3) are the probability of bidder i's bid higher than an unknown designated bidder's bid and the probability of bidder i's bid higher than non-collusive competitors' bids, respectively. Similarly, in the optimization problem (4), the first term is the designated bidder's surplus if he is the winner, and the second term is the winning probability of the designated bidder. With Assumption 1 and the common support of private values, the following proposition formalizes the properties of Bayesian-Nash equilibrium (BNE) in the main auction.[9]

Proposition 1. *Under Assumption 1, there exists a unique Bayesian-Nash equilibrium of the main auction in pure strategies $(\beta_{\mathcal{K}}(\cdot), \{\beta_i(\cdot)\}_{i\in\mathcal{N}\setminus\mathcal{K}})$ that are strictly increasing and differentiable on the common value interval $(\underline{v},\overline{v}]$. Moreover, the inverses of equilibrium strategies $(\beta_{\mathcal{K}}(\cdot), \{\beta_i(\cdot)\}_{i\in\mathcal{N}\setminus\mathcal{K}})$ are the solutions to a system of first-order differential equations:*

$$\frac{d}{db}\sum_{j\in\mathcal{N}\setminus(\{i\}\cup\mathcal{K})} \log F_j(\beta_j^{-1}(b)) + \log F_{\mathcal{K}}(\beta_{\mathcal{K}}^{-1}(b)) = \frac{1}{\beta_i^{-1}(b) - b}, \tag{5}$$

[8] Graham and Marshall (1987) described a stylized fact that non-collusive bidders are unaware of the cartel, but collusive bidders perfectly know the ring members. In equilibrium, each bidder should know the distribution of the highest competing bid; otherwise, the solution concept cannot be the Bayesian-Nash equilibrium.

[9] It is possible to extend the collusion model without assuming the common support of values. Lebrun (2006) studied the existence and the uniqueness of the equilibrium in asymmetric first-price sealed-bid auctions in which supports of values are possibly different across bidders.

and

$$\frac{d}{db}\sum_{j\in\mathcal{N}\setminus\mathcal{K}}\log F_j(\beta_j^{-1}(b)) = \frac{1}{\beta_{\mathcal{K}}^{-1}(b)-b}, \tag{6}$$

with following boundary conditions in mandatory bidding: $\beta_j^{-1}(\eta)=\overline{v}$ *for all* $j\in\mathcal{N}\setminus\mathcal{K}$, $\beta_{\mathcal{K}}^{-1}(\eta)=\overline{v}$, $\beta_j^{-1}(\underline{v})=\lim_{v\to\underline{v}}\beta_j^{-1}(v)=\underline{v}=\underline{b}$ *for all* $j\in\mathcal{N}\setminus\mathcal{K}$, *and* $\beta_{\mathcal{K}}^{-1}(\underline{v})=\lim_{v\to\underline{v}}\beta_{\mathcal{K}}^{-1}(v)=\underline{v}=\underline{b}$, *where* $\overline{b}=\eta$ *is the common maximum bid such that* $\underline{v}<\eta<\overline{v}$ *and* $\underline{b}=\underline{v}$ *is the common minimum bid.*

Proof. See Theorem 1 in Lebrun (1997). □

In Proposition 1, equilibrium bidding strategies are strictly increasing and differentiable on the value interval $(\underline{v},\overline{v}]$, and satisfy a system of first-order conditions. Equilibrium bidding strategies share a common maximum bid at the upper extremity of the value interval. In mandatory bidding, every serious bidder's minimum bid equals the lower extremity of the value interval. The boundary condition differs between mandatory and voluntary bidding at the lower extremity of the value interval.[10] If voluntary bidding is assumed, then the boundary condition at $\underline{v}$ in Proposition 1 needs to be changed to one of the following conditions: (i) $\beta_{\mathcal{K}}^{-1}(\underline{v})=\beta_j^{-1}(\underline{v})=\underline{v}$ for all but at most one $j\in\mathcal{N}$; (ii) $\beta_{\mathcal{K}}^{-1}(\underline{v})>\beta_j^{-1}(\underline{v})=\underline{v}$ for all $j\in\mathcal{N}$; see Theorem 2 in Lebrun (1997).

4. Identification

This section is devoted to identification of the primitives in the collusion model from observables generated in a sequence of independent first-price auctions. In each auction indexed by t, econometricians can observe potential bidders $\mathcal{N}_t$, all bids $\boldsymbol{B}_t=(B_{tj})_{j\in\mathcal{N}_t}$, bidder identities $\boldsymbol{I}_t=(I_{tj})_{j\in\mathcal{N}_t}$, a vector of auction-specific characteristics $\boldsymbol{X}_t$, and auction-level instruments Z_t which are assumed to affect equilibrium bidding strategies but not the underlying model primitives. The first implication of the collusion model is that any winning bid is submitted from either a designated bidder or a non-collusive bidder. All identification results are based on the following assumption regarding the data generating process (DGP).

Assumption 2. *Given the model primitives* $(\{F_j(\cdot)\}_{j\in\mathcal{N}_t},\Gamma,\mathcal{K}_t)$, *potential bidders* $\mathcal{N}_t$, *and auction-level covariates* $(\boldsymbol{X}_t,Z_t)$, *the winning bid and the identity of the winning bidder in auction t are generated by the unique BNE strategies* $(\beta_{\mathcal{K}}(\cdot),\{\beta_i(\cdot)\}_{i\in\mathcal{N}_t\setminus\mathcal{K}_t})$, *that are strictly increasing and differentiable over the value interval* $(\underline{v},\overline{v}]$, *and satisfy the first-order conditions and boundary conditions in Proposition 1.*

[10] In voluntary bidding, bidders have the option not to bid; for example, a bidder will drop out if his private value is less than the reserve price. In mandatory bidding, any bidder must submit a bid greater than or equal to $\underline{v}$.

For simplicity of notation, indices of auctions are suppressed whenever it is well understood. All results are conditional on auction-specific covariates and the set of potential bidders. Though, in general, the system of differential equations (5) and (6) do not lead to closed-form solutions, the inverse of the equilibrium bidding strategy can be written as a function of the distribution of the bidder's highest competing bid. The bidder's value distribution has a quantile representation due to the strict monotonic relationship between bids and private values in equilibrium. Applying the indirect approach by Guerre et al. (2000), the inverse value distributions and the inverse bid distributions satisfy that for any non-collusive bidder $i \in \mathcal{N} \setminus \mathcal{K}$ and $\alpha \in (0, 1]$,

$$F_i^{-1}(\alpha) = \beta_i^{-1}(G_i^{-1}(\alpha)) = G_i^{-1}(\alpha) + \frac{1}{\sum_{j \in \mathcal{N} \setminus (\{i\} \cup \mathcal{K})} \frac{g_j(G_i^{-1}(\alpha))}{G_j(G_i^{-1}(\alpha))} + \frac{g_{\mathcal{K}}(G_i^{-1}(\alpha))}{G_{\mathcal{K}}(G_i^{-1}(\alpha))}}, \tag{7}$$

and for any collusive bidder $k \in \mathcal{K}$ and $\alpha \in (0, 1]$,

$$F_k^{-1}(\alpha) = \beta_{\mathcal{K}}^{-1}(G_k^{-1}(\alpha)) = G_k^{-1}(\alpha) + \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(G_k^{-1}(\alpha))}{G_j(G_k^{-1}(\alpha))}}, \tag{8}$$

where $G_k(\cdot) \equiv \mathbb{P}\{\beta_{\mathcal{K}}(V_k) \le \cdot\}$ is the distribution of collusive bidder k's serious bids $\beta_{\mathcal{K}}(V_k)$, $G_j(\cdot) \equiv \mathbb{P}\{\beta_j(V_j) \le \cdot\}$ is the distribution of non-collusive bidder j's bids $\beta_j(V_j)$, $G_{\mathcal{K}}(\cdot) \equiv \mathbb{P}\{\beta_{\mathcal{K}}(V_{\mathcal{K}}) \le \cdot\}$ is the distribution of the designated bidder's bids $\beta_{\mathcal{K}}(V_{\mathcal{K}})$, $g_j(\cdot)$ is the density function of $G_j(\cdot)$, $g_{\mathcal{K}}(\cdot)$ is the density function of $G_{\mathcal{K}}(\cdot)$, $F_j^{-1}(\alpha) \equiv \inf\{v : F_j(v) \ge \alpha\}$, and $G_j^{-1}(\alpha) \equiv \inf\{b : G_j(b) \ge \alpha\}$. By the boundary conditions in Proposition 1, $F_i^{-1}(0) = F_{\mathcal{K}}^{-1}(0) = \underline{v}$ for each non-collusive bidder $i \in \mathcal{N} \setminus \mathcal{K}$. This quantile representation of value distributions plays a critical role in identification. Equations (7) and (8) reveal the structure of bidder markups in the collusion model. Equation (7) implies that the non-collusive bidder's markup is a function of the designated bidder's value distribution, which depends on the underlying collusion scheme. Equation (8) can be viewed as a result in an extreme case where collusive bidder k is always the cartel's designated bidder. It is worth mentioning that the distribution of collusive bidder k's serious bids, $G_k(\cdot)$, is not directly observed from bidder k's all bids, even if bidder k is known as a collusive bidder to econometricians.

The first result of this paper shows that the bidder's bid distribution is nonparametrically identified from winning bids and identities of winning bidders whenever this bidder is non-collusive and the collusion scheme satisfies Assumption 1.

Lemma 1. *Under Assumption 1 and 2, if bidder i is non-collusive, then bidder i's bid distribution is nonparametrically identified from winning bids and identities of winning bidders. In particular, if bidder i is non-collusive, then for* $b \in [\underline{b}, \overline{b}]$,

$$G_i(b) = \exp\left(-\int_b^{\overline{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)} \, dH_i(t)\right),$$

where $H_j(\cdot)$ is bidder j's cumulative incidence function defined by

$$H_j(b) \equiv \mathbb{P}\left\{B_j = \max_{\ell \in \mathcal{N}} B_\ell, B_j \leq b\right\}.$$

Proof. See Appendix A. □

Lemma 1 applies to any unknown collusion scheme satisfying Assumption 1. The proof of Lemma 1 is closely related to identification in the competing risks model with independent non-identically distributed risks. Specifically, given a finite set of independent random variables $\{Y_n\}_{1 \leq n \leq N}$ with non-identical absolutely continuous distributions $\{F_{Y_n}(\cdot)\}_{1 \leq n \leq N}$, Berman (1963) proved that the cumulative incidence functions

$$H_{Y_n}(y) \equiv \mathbb{P}\left\{Y_n = \max_{1 \leq \ell \leq N} Y_\ell, Y_n \leq y\right\}$$

identify the marginal distribution $F_{Y_n}(\cdot)$ for each random variable Y_n.[11] Berman's result can be used to show identification of bid distributions in the classic competitive model of asymmetric first-price auctions; see Athey and Haile (2007). Assuming an efficient collusion scheme, Schurter (2017) demonstrated that serious bid distributions for both collusive bidders and non-collusive bidders can be recovered from winning bids and identities of winning bidders. This is because whenever the cartel is efficient, the winner's identity can be determined by the order of all bidders' serious bids.[12] For any collusive bidder $k \in \mathcal{K}$, the event of winning the auction is equivalent to that $B_k = \beta_{\mathcal{K}}(V_k) = \max_{\ell \in \mathcal{K}} \beta_{\mathcal{K}}(V_\ell)$ and $B_k = \beta_{\mathcal{K}}(V_k) \geq \max_{j \in \mathcal{N} \setminus \mathcal{K}} \beta_j(V_j)$; therefore, two events are equivalent within an efficient cartel for collusive bidder k:

$$\left\{B_k = \max_{\ell \in \mathcal{N}} B_\ell, B_k \leq b\right\} = \left\{\beta_{\mathcal{K}}(V_k) = \max_{\ell \in \mathcal{K}, j \in \mathcal{N} \setminus \mathcal{K}} \{\beta_{\mathcal{K}}(V_\ell), \beta_j(V_j)\}, \beta_{\mathcal{K}}(V_k) \leq b\right\}. \tag{9}$$

On the other hand, for any non-collusive bidder $i \in \mathcal{N} \setminus \mathcal{K}$, the event of winning the auction is equivalent to that $B_i = \beta_i(V_i) \geq \max_{\ell \in \mathcal{K}} \beta_{\mathcal{K}}(V_\ell)$ and $B_i = \beta_i(V_i) = \max_{j \in \mathcal{N} \setminus \mathcal{K}} \beta_j(V_j)$; therefore, two events are equivalent within an efficient cartel for non-collusive bidder i:

$$\left\{B_i = \max_{\ell \in \mathcal{N}} B_\ell, B_i \leq b\right\} = \left\{\beta_i(V_i) = \max_{\ell \in \mathcal{K}, j \in \mathcal{N} \setminus \mathcal{K}} \{\beta_{\mathcal{K}}(V_\ell), \beta_j(V_j)\}, \beta_i(V_i) \leq b\right\}. \tag{10}$$

Identification of the distributions of $\{\beta_{\mathcal{K}}(V_k)\}_{k \in \mathcal{K}}$ and $\{\beta_i(V_i)\}_{i \in \mathcal{N} \setminus \mathcal{K}}$ in the case of efficient cartels immediately follows from Berman (1963).

There are two comments on Lemma 1. First, it is stressed that the settings of the collusion model in Lemma 1 are substantially different from that of the collusion model with efficient cartels. The essential difference is that the designated bidder does not necessarily have the highest private value among the ring members. In the model with efficient cartels,

[11] The monotonic function $H_{Y_n}(\cdot)$ is known as the cumulative incidence function in the competing risks model; see Klein and Andersen (2005).

[12] The serious bid of a collusive bidder is the bid he would submit if he were the designated bidder.

if the winner of the main auction is a collusive bidder, then his serious bid must be higher than all other collusive bidders. In contrast, this is not true in the case of inefficient cartels. The winner's identity cannot be determined by the order of all bidders' serious bids once the cartel is inefficient. This complication suggests that both equivalence relations (9) and (10) break down in the case of inefficient cartels. Consequently, one cannot simply apply Berman (1963) or Athey and Haile (2007) to obtain identification of bid distributions without assuming the cartel is efficient. At first glance, it seems impossible to recover bid distributions from winning bids and identities of winners if the collusion scheme is unknown. However, Lemma 1 states that if a bidder is non-collusive, then his bid distribution can still be learned from winning bids and identities of winning bidders without assuming the cartel is efficient.

Second, Lemma 1 has implications for bidders' all bids. Suppose all bids (possibly including cover bids if collusion occurs) and bidder identities are observed. For each bidder $j \in \mathcal{N}$, define

$$M_j(b) \equiv \exp\left(-\int_b^{\overline{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_j(t)\right),$$

with the cumulative incidence function $H_\ell(\cdot)$ given in Lemma 1. Lemma 1 implies that, if bidder j is non-collusive, the distribution of bidder j's all bids coincide with the function $M_j(\cdot)$ that is induced from winning bids and identities of winners. This is equivalent to saying that if the distribution of bidder j's all bids does not coincide with the function $M_j(\cdot)$, then bidder j must be a collusive bidder.[13] This implication is consistent with the literature on collusion. In order to avoid an antitrust investigation, a successful cartel should be able to manipulate cover bids to mimic competition.

Lemma 1 is of great significance as it allows econometricians to correctly recover the distribution of the highest competing bid for any non-collusive bidder. Formally, let $G_{-j}(\cdot)$ denote the distribution of bidder j's highest competing bid. A non-collusive bidder bids competitively against the designated bidder and all other non-collusive bidders in the main auction; therefore, if bidder j is non-collusive,

$$G_{-j}(b) = \mathbb{P}\left\{\max_{\ell \in \mathcal{N}\setminus\{j\}} B_\ell \le b\right\} = \mathbb{P}\left\{\max_{i \in \mathcal{N}\setminus(\mathcal{K}\cup\{j\})} \{B_{\mathcal{K}}, B_i\} \le b\right\}.$$

By the independence of private values, the distribution of bidder j's highest competing bid $G_{-j}(b)$ can be written as

$$G_{-j}(b) = G_{\mathcal{K}}(b) \prod_{i \in \mathcal{N}\setminus(\mathcal{K}\cup\{j\})} G_i(b).$$

[13] Krasnokutskaya (2011) exploited this fact to test possible collusive behavior in the Michigan highway procurement auctions. The procedure is to recover the bid distributions of regular and fringe firms from winning bids in the first step and compare bid distributions to those estimated from the full samples. However, this method is insufficient for identifying collusive bidders. The equality of those two distributions is a necessary condition of a competitive model.

To further recover the distribution of bidder j's highest competing bid, it is unnecessary to know the distribution of the designated bidder's bids $G_{\mathcal{K}}(\cdot)$, which depends on the underlying collusion scheme and collusive bidders' value distributions. Notice that the winning bid is the maximum of the designated bidder's and non-collusive bidders' bids. The distribution of winning bids, denoted by $W(b)$, is the product of the designated bidder's bid distribution and all non-collusive bidders' bid distributions, i.e.,

$$W(b) = G_{\mathcal{K}}(b) \prod_{i \in \mathcal{N} \setminus \mathcal{K}} G_i(b) = G_{\mathcal{K}}(b) G_j(b) \prod_{i \in \mathcal{N} \setminus (\mathcal{K} \cup \{j\})} G_i(b).$$

Therefore, if bidder j is non-collusive, then the distribution of bidder j's highest competing bid can be written as the ratio of the distribution of winning bids to bidder j's bid distribution, i.e., $G_{-j}(b) = W(b)/G_j(b) = W(b)/M_j(b)$, where the second identity holds due to $G_j(b) = M_j(b)$ by Lemma 1. Furthermore, the first-order condition (5) implies that bidder j's inverse bidding strategy, $\beta_j^{-1}(b) = b + G_{-j}(b)/g_{-j}(b)$ with $g_{-j}(b) \equiv d/dbG_{-j}(b)$, is correctly recovered from winning bids and identities of winning bidders. By equation (7), non-collusive bidder j's inverse value distribution is identified if the collusion scheme satisfies Assumption 1. This result is formalized in the following proposition.

Proposition 2. *Under Assumption 1 and 2, if bidder i is non-collusive, then bidder i's value distribution is nonparametrically identified from winning bids and identities of winning bidders. In particular, if bidder i is non-collusive, then* $\underline{v} = F_i^{-1}(0) = M_i^{-1}(0) = \underline{b}$, *and when* $\alpha \in (0, 1]$,

$$F_i^{-1}(\alpha) = M_i^{-1}(\alpha) + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(M_i^{-1}(\alpha))}{\sum_{\ell \in \mathcal{N} \setminus \{i\}} H'_\ell(M_i^{-1}(\alpha))},$$

where

$$M_i(b) \equiv \exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)} \, dH_i(t) \right),$$

and $H_j(\cdot)$ *is bidder j's cumulative incidence function defined by*

$$H_j(b) \equiv \mathbb{P}\left\{ B_j = \max_{\ell \in \mathcal{N}} B_\ell, B_j \leq b \right\}.$$

Proof. See Appendix A. □

Turning to identification of collusive bidders, one can construct the inverse pseudo-value distribution for each bidder $j \in \mathcal{N}$:

$$U_j^{-1}(\alpha) \equiv \begin{cases} M_j^{-1}(\alpha) + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(M_j^{-1}(\alpha))}{\sum_{\ell \in \mathcal{N} \setminus \{j\}} H'_\ell(M_j^{-1}(\alpha))} & \text{for } \alpha \in (0, 1]; \\ \underline{b} & \text{for } \alpha = 0, \end{cases} \tag{11}$$

where the functions $M_j(\cdot)$ and $H_\ell(\cdot)$ are given in Proposition 2.[14] The inverse pseudo-value distribution can be nonparametrically estimated as all quantities on the right-hand side of equation (11) are directly observed from winning bids and identities of winning bidders. Moreover, the construction of inverse pseudo-value distributions in equation (11) depends neither on the configuration of the cartel nor on the collusion scheme. Proposition 2 suggests that despite the unknown collusion scheme, bidder j's inverse pseudo-value distribution coincides with the inverse value distribution if bidder j is non-collusive. On the contrary, if bidder j is collusive, it is observed that the inverse pseudo-value distribution does not equal the inverse value distribution since bidder j's inverse bidding strategy is incorrectly inferred.[15] This result forms a basis for identifying collusive bidders with proper auction-level instruments that affect equilibrium bidding strategies but not the underlying model primitives. In particular, the inverse pseudo-value distribution will respond differently to auction-level instruments for collusive and non-collusive bidders.

Assumption 3. *There exist auction-level instruments Z_t such that*
(i) each bidder's equilibrium bidding strategy depends non-trivially on instruments Z_t;
(ii) the collusion scheme and bidder value distributions are invariant to instruments Z_t.

Assumption 3(i) assumes that auction-level instruments can provide sufficient variation in equilibrium bidding strategies. Assumption 3(ii) is the independence condition, which requires that auction-level instruments do not affect the underlying collusion scheme and value distributions. To satisfy the independence condition, the information on instruments should ideally be revealed to bidders only after the realization of private values and the selection of the designated bidder. Examples of instruments in the empirical auction literature include cost shifters (Gentry and Li, 2014; Somaini, 2020), and exogenous participation of bidders (Haile et al., 2003; Guerre et al., 2009; Marmer et al., 2013; Aryal et al., 2018).

Additional information like the independence condition of instruments is indispensable for detecting collusion in first-price auctions. As pointed out by Bajari and Ye (2003), bids generated from a collusion model with partial cartels can be rationalized by a competitive model as long as bids can satisfy some necessary conditions of the competitive model. Unsurprisingly, it is generally insufficient to identify collusive bidders only with information on distributions of all bids. Nevertheless, with proper auction-level instruments in hand,

[14] Notice that U_j is exactly the same as bidder j's true value distribution in a competitive model. Thus, U_j is also called the competitively rationalizing valuation distribution in Schurter (2017).

[15] To see this, using the fact that $m_j(\cdot)/M_j(\cdot) = H_j'(\cdot)/\sum_{\ell\in\mathcal{N}} H_\ell(\cdot)$ for all $j \in \mathcal{N}$, for collusive bidder $k \in \mathcal{K}$,

$$U_k^{-1}(\alpha) = b + \frac{1}{\sum_{j\in\mathcal{N}\setminus\mathcal{K}} \frac{g_j(b)}{G_j(b)} + \sum_{\ell\in\mathcal{K}\setminus\{k\}} \frac{m_\ell(b)}{M_\ell(b)}}\Bigg|_{b=M_k^{-1}(\alpha)} < b + \frac{1}{\sum_{j\in\mathcal{N}\setminus\mathcal{K}} \frac{g_j(b)}{G_j(b)}}\Bigg|_{b=M_k^{-1}(\alpha)} = \beta_{\mathcal{K}}^{-1}(M_k^{-1}(\alpha)).$$

Assuming the common support of private values, then $G_k^{-1}(1) = M_k^{-1}(1) = \overline{b}$ the boundary, which implies $U_k^{-1}(1) < \beta_{\mathcal{K}}^{-1}(G_k^{-1}(1)) = \beta_{\mathcal{K}}^{-1}(\overline{b}) = \overline{v}$.

identities of collusive bidders are identified from observables regardless of the unknown collusion mechanism.

Theorem 1. *Under Assumption 1, 2, and 3, identities of collusive bidders and value distributions of non-collusive bidders are identified from winning bids, identities of winning bidders, and auction-level instruments.*

Proof. See Appendix A. □

Theorem 1 applies to a broad class of collusion schemes satisfying Assumption 1. Identification of collusive bidders solely relies on winning bids, identities of winners, and auction-level instruments. Once identities of collusive bidders are identified, value distributions of non-collusive bidders are immediately identified by Proposition 2. Interestingly, the identification results in Theorem 1 can also be achieved within a more complex information structure, in which collusive bidders may have interdependent private values but non-collusive bidders have independent private values.

The detailed proof of Theorem 1 can be found in Appendix A. The proof idea is outlined as follows. First, the inverse pseudo-value distribution is invariant to instruments for each non-collusive bidder by Proposition 2 and Assumption 3(ii). Heuristically, identification of collusive bidders can be proved by showing that the inverse pseudo-value distribution is a non-trivial function of instruments for every collusive bidder. That is to say, for any collusive bidder $k \in \mathcal{K}$, there exists a strictly positive Lebesgue measure $\mathcal{A} \subset (0,1]$, $z, z' \in \mathcal{Z} \subset \mathbb{R}^{d_z}$, and $z \neq z'$ such that

$$U_k^{-1}(\alpha; z) \neq U_k^{-1}(\alpha; z')$$

for $\alpha \in \mathcal{A}$. For collusive bidder $k \in \mathcal{K}$, the inverse pseudo-value distribution conditional on $Z_t = z$ is given by

$$U_k^{-1}(\alpha; z) \equiv M_k^{-1}(\alpha; z) + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(M_k^{-1}(\alpha; z); z)}{\sum_{\ell \in \mathcal{N} \setminus \{k\}} H'_\ell(M_k^{-1}(\alpha; z); z)}. \tag{12}$$

for $\alpha \in (0,1]$. Denote $m_j(b) \equiv d/db M_j(b)$. Note that $m_j(b)/M_j(b) = H'_j(b)/\sum_{\ell \in \mathcal{N}} H_\ell(b)$ for all $j \in \mathcal{N}$, equation (12) reduces to

$$U_k^{-1}(\alpha; z) = M_k^{-1}(\alpha; z) + \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)} + \sum_{\ell \in \mathcal{K} \setminus \{k\}} \frac{m_\ell(M_k^{-1}(\alpha;z);z)}{M_\ell(M_k^{-1}(\alpha;z);z)}},$$

where $M_j(\alpha; z) = G_j(\alpha; z)$ and $m_j(\alpha; z) = g_j(\alpha; z)$ for each non-collusive bidder $j \in \mathcal{N} \setminus \mathcal{K}$ by Lemma 1. The challenge of the proof is that the inverse pseudo-value distribution $U_k^{-1}(\alpha; z)$, as a function of instruments, is analytically intractable since $M_k(\alpha; z)$ is a complicated function of instruments.

To address this complication, notice that the inverse pseudo-value distribution can be expressed as a summation of two quantities $L_{k,1}(\alpha; z)$ and $L_{k,2}(\alpha; z)$, i.e.,

$$U_k^{-1}(\alpha; z) = L_{k,1}(\alpha; z) + L_{k,2}(\alpha; z),$$

where

$$L_{k,1}(\alpha; z) \equiv M_k^{-1}(\alpha; z) + \frac{1}{\sum_{j\in\mathcal{N}\setminus\mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)}},$$

$$L_{k,2}(\alpha; z) \equiv \frac{1}{\sum_{j\in\mathcal{N}\setminus\mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)} + \sum_{\ell\in\mathcal{K}\setminus\{k\}} \frac{m_\ell(M_k^{-1}(\alpha;z);z)}{M_\ell(M_k^{-1}(\alpha;z);z)}} - \frac{1}{\sum_{j\in\mathcal{N}\setminus\mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)}}.$$

The proof of identification proceeds in two steps. First, $L_{k,1}(\alpha; z)$ is invariant to instruments. Second, $L_{k,2}(\alpha; z)$ can be expressed in terms of the designated bidder's equilibrium bidding strategy that depends non-trivially on instruments. If the cartel is efficient, it is not hard to see that $L_{k,1}(\alpha; z)$ is invariant to instruments by the independence condition. This is because $M_k(\alpha; z)$ equals the distribution of collusive bidder k's serious bids in an efficient cartel, that is, $M_k(\alpha; z) = G_k(\alpha; z)$. By equation (8), $L_{k,1}(\alpha; z)$ is exactly the α-th quantile of collusive bidder k's value distribution in an efficient cartel. However, the distribution of the collusive bidder's serious bids cannot be recovered in an inefficient cartel. It is somewhat unexpected that $L_{k,1}(\alpha; z)$ is also invariant to instruments in the case of inefficient cartels.

An important intermediate result is that, for any collusion structure $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ satisfying Assumption 1, there exists a unique profile of value distributions $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$ such that two collusion structures $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ and $(\{F_k^*(\cdot)\}_{k\in\mathcal{K}}, \Gamma_e)$ are equivalent in the joint distribution of the designated bidder's private value and identities; see Lemma A1. This equivalence result provides an implicit connection between $L_{k,1}(\alpha; z)$ and the model primitives. By the strict monotonicity of the designated bidder's inverse bidding strategy, Proposition A1 shows that $L_{k,1}(\alpha; z)$ is the α-th quantile of the distribution $F_k^*(\cdot)$ for each collusive bidder $k \in \mathcal{K}$, which is invariant to instruments by the independence condition. On the other hand, for any fixed $\alpha \in (0, 1]$ and $z \in \mathcal{Z}$, $L_{k,2}(\alpha; z)$ can be expressed in terms of the designated bidder's equilibrium bidding strategy. Since the designated bidder's equilibrium bidding strategy depends non-trivially on instruments, $L_{k,2}(\alpha; z)$ cannot be invariant to instruments. Combining these results, it follows that the inverse pseudo-value distribution

$$U_k^{-1}(\alpha; z) = L_{k,1}(\alpha; z) + L_{k,2}(\alpha; z)$$

is a non-trivial function of instruments for each collusive bidder $k \in \mathcal{K}$ under Assumption 3. This completes the proof.

Remark. *If there are multiple cartels, say $\mathcal{K}_1$ and $\mathcal{K}_2$, it is obvious that results pertain to Lemma 1 and Proposition 2 still hold. The inverse pseudo-value distribution coincides with the inverse value distribution for each non-collusive bidder under multiple cartels. Similar to the case of a single* *cartel, identification of collusive bidders $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2$ can still be obtained by showing that the pseudo-value distribution is a non-trivial function of instruments for every collusive bidder in $\mathcal{K}$.*

Given the model primitives $(\{F_j(\cdot)\}_{j\in\mathcal{N}}, \Gamma, \mathcal{K})$, Theorem 1 states that the configuration of the cartel $\mathcal{K}$ is uniquely determined by observables with the independence condition of instruments. The value distributions of non-collusive bidders $\{F_i(\cdot)\}_{i\in\mathcal{N}\setminus\mathcal{K}}$ are identified from Proposition 2. However, without further information, it is impossible to infer collusive bidders' value distributions and the collusion scheme. Given the non-collusive bidders' value distributions, the collusion structure $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ is not identified if there exists another collusion structure $(\{\tilde{F}_k(\cdot)\}_{k\in\mathcal{K}}, \tilde{\Gamma})$ that leads to the same cumulative incidence functions $\{H_j(\cdot)\}_{j\in\mathcal{N}}$ in equilibrium. A simple numerical example is provided below for illustrative purposes.

Example 3. *Suppose two collusive bidders $\mathcal{K} = \{1, 2\}$ with uniform value distributions form a weak cartel and adopt a lottery mechanism Γ, in which the designated bidder is randomly chosen with an equal probability $1/2$. Then, for each $k \in \mathcal{K} = \{1, 2\}$,*

$$p_k(v;\Gamma) \equiv \mathbb{P}\{V_k \le v, I_{\mathcal{K}} = k; \Gamma\} = \mathbb{P}\{V_k \le v | I_{\mathcal{K}} = k; \Gamma\}\, \mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\} = \frac{v}{2},$$

and the designated bidder's value distribution is

$$\mathbb{P}\{V_{\mathcal{K}} \le v; \Gamma\} = \sum_{k\in\mathcal{K}} \mathbb{P}\{V_k \le v, I_{\mathcal{K}} = k; \Gamma\} = v.$$

There exists a unique set of distribution functions $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$ such that

$$F_k^*(v) = \exp\left(-\int_v^{\overline{v}} \frac{p_k'(t;\Gamma)}{F_{\mathcal{K}}(t;\Gamma)}\, dt\right) = \exp\left(-\int_v^{\overline{v}} \frac{1}{2t}\, dt\right) = \exp\left(\frac{1}{2}\log v\right) = \sqrt{v}.$$

Let Γ_e be an efficient collusion mechanism. Under the collusion structure $(\{F_k^*(v) = \sqrt{v}\}_{k\in\mathcal{K}}, \Gamma_e)$, for each $k \in \mathcal{K}$ and $v \in [\underline{v}, \overline{v}]$,

$$\mathbb{P}^*\{V_k \le v, I_{\mathcal{K}} = k; \Gamma_e\} = \int_{\underline{v}}^{v} \sqrt{t}\, d\sqrt{t} = \frac{v}{2} = \mathbb{P}\{V_k \le v, I_{\mathcal{K}} = k; \Gamma\},$$

where $\mathbb{P}^*$ is the probability measure under distribution functions $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$. Two collusion structures $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ and $(\{F_k^*(\cdot)\}_{k\in\mathcal{K}}, \Gamma_e)$ induce the same value distribution of the designated bidder, i.e.,

$$\mathbb{P}^*\{V_{\mathcal{K}} \le v; \Gamma_e\} = \mathbb{P}\{V_{\mathcal{K}} \le v; \Gamma\} = v.$$

Given the value distributions of non-collusive bidders, equilibrium bidding functions are the same under two collusion structures $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ and $(\{F_k^*(\cdot)\}_{k\in\mathcal{K}}, \Gamma_e)$. It can be shown that the cumulative incidence functions $\{H_j(\cdot)\}_{j\in\mathcal{N}}$ are also the same under two collusion structures.

Theorem 2. *Under Assumption 1, 2, and 3, the collusion structure $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ is not identified from winning bids, identities of winning bidders, and auction-level instruments.*

Proof. See Appendix A. □

The proof of Theorem 2 shows that the cumulative incidence functions $\{H_j(\cdot)\}_{j\in\mathcal{N}}$ are the same under two different collusion structures in equilibrium. This non-identification result arises from the equivalence between collusion structures in the joint distribution of the designated bidder's private values and identities. Theorem 2 implies that value distributions of collusive bidders cannot be identified without additional assumptions on the collusion mechanism. A natural question is under which collusion schemes can value distributions of collusive bidders be identified. From Schurter (2017), the value distributions of collusive bidders can be identified in an efficient cartel. For a broad class of weak cartels in single-object first-price auctions, the theory literature claimed that the optimal collusion mechanism is the lottery mechanism, i.e., a random selection of the designated bidder. It follows from Theorem 1 that if the weak cartel adopts a lottery mechanism, the value distributions of collusive bidders can be identified from winning bids, identities of winning bidders, and auction-level instruments.

Corollary 1. *Suppose that the cartel's designated bidder is randomly selected. Under Assumption 1, 2, and 3, all bidders' value distributions are nonparametrically identified from winning bids, identities of winning bidders, and auction-level instruments.*

Proof. See Appendix A. □

It is already known that collusive bidders' identities and value distributions can be identified under either efficient cartels or inefficient cartels using lottery mechanisms. An important question for policymakers is whether those two different types of cartels can be distinguished with empirical auction data. Unfortunately, the answer to this question is negative without additional information since bids generated from an inefficient cartel satisfying Assumption 1 could be rationalized by an efficient cartel. Further information on value distributions is required to distinguish the types of cartels. This is because two collusion models, namely strong cartels and weak cartels, will imply different collusive bidders' value distributions. If econometricians have prior value distributions for collusive bidders, a goodness-of-fit test is helpful to determine the types of cartels.[16] Or, if there exists some non-collusive bidder who is the same type as the collusive bidder, then the inferred value distribution of the non-collusive bidder can be used to distinguish the types of cartels.

5. Testable Implication, Estimation, and inference

The identification results in Section 4 are constructive and provide a testable implication of collusion. Only the non-collusive bidder's pseudo-value distribution is invariant to auction-level instruments for any configuration of the cartel and the collusion scheme. To

[16] In an empirical analysis of collusion in procurement auctions, Bajari and Ye (2003) collected prior information on cost distributions from knowledgeable industry sources and computed the likelihood of the collusion model given the bid data.

test collusive bidding behavior, a family of hypotheses $\{\mathrm{H}_0^j\}_{j=1}^J$ can be formulated for each bidder $j \in \mathcal{N}$,

$$\mathrm{H}_0^j : U_j(v; z, \boldsymbol{x}, \mathcal{N}) = U_j(v; z', \boldsymbol{x}, \mathcal{N}) \quad \text{for any } z \neq z',$$

where $U_j(v; z, \boldsymbol{x}, \mathcal{N})$ is bidder j's pseudo-value distribution conditional on auction-level covariates $Z_t = z$, $\boldsymbol{X}_t = \boldsymbol{x}$, and $\mathcal{N}_t = \mathcal{N}$. Instead of joint hypothesis testing, this is a multiple testing problem as econometricians wish to know the statistical decision for the individual hypothesis. The statistical decision for the individual hypothesis helps to infer the identities of collusive and non-collusive bidders. This section describes the test statistic, nonparametric estimation, and simultaneous inference procedures. A Monte Carlo study is presented at the end of the section.

5.1. **Test Statistic.** This subsection introduces the test statistic for the individual hypothesis H_0^j. It is well-known that two distribution functions are the same if and only if their corresponding integrated quantile functions are the same. For each bidder $j \in \mathcal{N}$, the testing problem is equivalent to the following individual hypothesis in terms of integrated quantile functions:

$$\mathrm{H}_0^j : Q_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) = Q_j(\alpha; z', \boldsymbol{x}, \mathcal{N}) \quad \text{for any } z \neq z',$$

where $Q_j(\alpha; z, \boldsymbol{x}, \mathcal{N})$ is the integrated quantile function defined as

$$Q_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) = \int_0^\alpha U_j^{-1}(t; z, \boldsymbol{x}, \mathcal{N})\, dt.$$

The integrated quantile function in auction models is well-studied in the recent literature (Liu and Vuong, 2013; Barrett et al., 2014; Liu and Luo, 2017; Luo and Wan, 2018). The advantages of constructing the test statistic with integrated quantile functions are twofold. First, it avoids the construction of pseudo valuations in comparing pseudo-value distributions, especially when only winning bids and identities of winning bidders are observed. Second, the estimated quantile-trimmed mean of pseudo-values may converge faster than the estimated quantile of pseudo-values; see Theorem 2 and Theorem 3 in Haile et al. (2003).

This paper proposes to construct the test statistic based on integrated quantile functions. When the instrumental variable Z_t takes values on a finite support $\mathcal{Z}$, the test statistic is constructed based on a general statistical functional ϕ applied to the difference between two estimated integrated quantile functions $\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N})$, i.e.,

$$t_j = \sum_{z, z' \in \mathcal{Z}} \phi\left(\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})\right),$$

Two common test statistics are the Kolmogorov-Smirnov (KS) type statistic with the supremum norm:

$$t_j = \sum_{z, z' \in \mathcal{Z}} \sup_{\alpha \in [0,1]} |\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})|,$$

and the Cramér von-Mises type statistic with the L_1 norm:

$$t_j = \sum_{z,z' \in \mathcal{Z}} \int_0^1 |\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})|\, d\alpha.$$

When the instrumental variable is continuous, the test statistic is modified to

$$t_j = \int \phi\left(\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})\right) dP(z) dP(z').$$

It is possible to discretize the support of continuous instrumental variables for computational convenience. Alternatively, Schurter (2017) proposed a test based on the correlation of pseudo private values when the instrumental variable is continuous. This paper mainly focuses on the case of the discrete instrumental variable.

5.2. **Estimation.** This subsection discusses nonparametric estimation of inverse pseudo-value distributions and integrated quantile functions conditional on auction-level covariates under the family of null hypotheses $\{\mathrm{H}_0^j\}_{j=1}^J$.

Let B_{tj} denote the bid submitted from bidder j at auction t. From a sequence of independent first-price sealed-bid auctions indexed by $t \in \{1, \cdots, T\}$, econometricians can observe potential bidders $\mathcal{N}_t$, winning bids $W_t = \max_{\ell \in \mathcal{N}_t} B_{t\ell}$, identities of winning bidders, auction-level characteristics $\boldsymbol{X}_t \in \mathcal{X}^d \subset \mathbb{R}^d$, and auction-level instruments $Z_t \in \mathcal{Z} \subset \mathbb{R}$. For illustrative purposes, it is assumed that the instrumental variable Z_t is discrete while auction-level characteristics $\boldsymbol{X}_t$ may contain both continuous and discrete variables.

Under the family of null hypotheses $\{\mathrm{H}_0^j\}_{j=1}^J$, conditional inverse pseudo-value distributions are estimated in two steps. First, estimate bidder j's inverse bidding strategy conditional on auction-level covariates,

$$\beta_j^{-1}(b; z, \boldsymbol{x}, \mathcal{N}) = b + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(b; z, \boldsymbol{x}, \mathcal{N})}{\sum_{\ell \in \mathcal{N} \setminus \{j\}} H'_\ell(b; z, \boldsymbol{x}, \mathcal{N})}.$$

where $H_j(b; z, \boldsymbol{x}, \mathcal{N})$ is the conditional cumulative incidence function defined as

$$H_j(b; z, \boldsymbol{x}, \mathcal{N}) \equiv \mathbb{P}\left\{B_{tj} = W_t, B_{tj} \le b | Z_t = z, \boldsymbol{X}_t = \boldsymbol{x}, \mathcal{N}_t = \mathcal{N}\right\}.$$

The nonparametric estimator for inverse bidding strategies is based on the kernel method. Second, estimate bidder j's conditional bid distribution

$$M_j(b; z, \boldsymbol{x}, \mathcal{N}) \equiv \exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t; z, \boldsymbol{x}, \mathcal{N})} dH_j(t; z, \boldsymbol{x}, \mathcal{N})\right)$$

from winning bids and identities of winning bidders.

The estimation of bidder inverse bidding strategies is as follows. For bidder $j \in \mathcal{N}$, the probability of winning the auction and his bid lower than b conditional on auction-level covariates $Z_t = z$, $\boldsymbol{X}_t = \boldsymbol{x}$, and $\mathcal{N}_t = \mathcal{N}$ can be empirically estimated by

$$\hat{H}_j(b; z, \boldsymbol{x}, \mathcal{N}) = \frac{\sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{B_{tj} = W_t, B_{tj} \le b\right\} \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}}{\sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}},$$

where $K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x})$ is a multivariate kernel function defined as a product of univariate kernel functions $K(\cdot)$ with smoothing parameters $h_x = (h_s)_{s=1}^d$:

$$K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) = \prod_{s=1}^{d} K\left(\frac{X_{ts} - x_{ts}}{h_s}\right).$$

When auction-level characteristics $\boldsymbol{X}_t$ contain both continuous and discrete variables, say $\boldsymbol{X}_t^C$ and $\boldsymbol{X}_t^D$ respectively, $K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x})$ is the product of kernel functions such that

$$K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) = K_{h_x^C}(\boldsymbol{X}_t^C, \boldsymbol{x}^C) K_{h_x^D}(\boldsymbol{X}_t^D, \boldsymbol{x}^D),$$

where $h_x = (h_x^C, h_x^D)$, $K_{h_x^C}(\cdot,\cdot)$ is a kernel function for continuous variables, and $K_{h_x^D}(\cdot,\cdot)$ is a kernel function smoothing categorical variables (see Li and Racine (2007)). Note that the sum of all bidders' cumulative incidence functions equals the distribution of winning bids. The estimator of the sum of conditional cumulative incidence functions is given by

$$\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(b; z, \boldsymbol{x}, \mathcal{N}) = \frac{\sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{W_t \le b\right\} \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}}{\sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}}.$$

An estimator of the derivative of $H_j(b; z, \boldsymbol{x}, \mathcal{N})$ takes the form

$$\hat{H}_j'(b; z, \boldsymbol{x}, \mathcal{N}) = \frac{\sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) K_{h_b}(B_{tj}, b) \mathbb{1}\left\{B_{tj} = W_t\right\} \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}}{\sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}},$$

where h_x and h_b are smoothing parameters. The Silverman rule-of-thumb is used to get the optimal bandwidth h_b^* for the univariate kernel function $K_{h_b}(\cdot,\cdot)$, i.e.,

$$h_b^* = c_K \sigma T^{-1/5},$$

where σ is the standard deviation of winning bids, and c_K is a kernel-specific constant; see Silverman (2018). The optimal bandwidth h_x^* for the multivariate kernel function $K_{h_x}(\cdot,\cdot)$ can be obtained by minimizing the Mean Square Error (MSE) associated with $K_{h_x}(\cdot,\cdot)$. Following the analogy principle and replacing unknown quantities with sample counterparts, the estimation of bidder j's inverse bidding function conditional on auction-level covariates is

$$\hat{\beta}_j^{-1}(b; z, \boldsymbol{x}, \mathcal{N}) = b + \frac{\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(b; z, \boldsymbol{x}, \mathcal{N})}{\sum_{\ell \in \mathcal{N} \setminus \{j\}} \hat{H}_\ell'(b; z, \boldsymbol{x}, \mathcal{N})}.$$

The second step is to estimate bidder j's conditional bid distribution $M_j(b; z, \boldsymbol{x}, \mathcal{N})$ from winning bids and identities of winning bidders. A computational advantage is that the conditional bid distribution $M_j(b; z, \boldsymbol{x}, \mathcal{N})$ can be consistently estimated using the empirical distribution of $H_\ell(\cdot; z, \boldsymbol{x}, \mathcal{N})$ where $\ell \in \mathcal{N}$. To see this, note that by analogy principle,

$$\begin{aligned}
\hat{M}_j(b; z, \boldsymbol{x}, \mathcal{N}) &= \exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(t; z, \boldsymbol{x}, \mathcal{N})} \, d\hat{H}_j(t; z, \boldsymbol{x}, \mathcal{N})\right) \\
&= \exp\left(-\sum_{t=1}^{T} \frac{\mathbb{1}\{B_{tj} = W_t\} \mathbb{1}\{B_{tj} \ge b\} K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}}{\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(B_{tj}; z, \boldsymbol{x}, \mathcal{N}) \sum_{t=1}^T K_{h_x}(\boldsymbol{X}_t, \boldsymbol{x}) \mathbb{1}\left\{\mathcal{N}_t = \mathcal{N}, Z_t = z\right\}}\right).
\end{aligned}$$

Alternatively, one can adapt the importance sampling method to obtain a smoothed estimator of $M_j(b; z, \boldsymbol{x}, \mathcal{N})$ by

$$\tilde{M}_j(b; z, \boldsymbol{x}, \mathcal{N}) = \exp\left(- \sum_{l=1}^{L} \frac{\mathbb{1}\{S_r \geq b\} \hat{H}'_j(S_r; z, \boldsymbol{x}, \mathcal{N})}{\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(S_r; z, \boldsymbol{x}, \mathcal{N})} \frac{1}{\mu(S_l)} \right),$$

where $(S_l)_{l=1}^{L}$ is a sequence of random variables drawn from some known density function μ (e.g., the density function of the standard normal distribution). But this method is done at the cost of estimating the derivative $H'_j(\cdot; z, \boldsymbol{x}, \mathcal{N})$.

With estimates $\hat{M}_j(b; z, \boldsymbol{x}, \mathcal{N})$ and $\hat{\beta}_j^{-1}(b; z, \boldsymbol{x}, \mathcal{N})$ in hand, bidder j's conditional inverse pseudo-value distribution can be consistently estimated by

$$\hat{U}_j^{-1}(\alpha; z, \boldsymbol{x}, \mathcal{N}) = \hat{M}_j^{-1}(\alpha; z, \boldsymbol{x}, \mathcal{N}) + \frac{\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(\hat{M}_j^{-1}(\alpha; z, \boldsymbol{x}, \mathcal{N}); z, \boldsymbol{x}, \mathcal{N})}{\sum_{\ell \in \mathcal{N} \setminus \{j\}} \hat{H}'_\ell(\hat{M}_j^{-1}(\alpha; z, \boldsymbol{x}, \mathcal{N}); z, \boldsymbol{x}, \mathcal{N})},$$

where

$$\hat{M}_j^{-1}(\alpha; z, \boldsymbol{x}, \mathcal{N}) = \inf\left\{ b : \hat{M}_j(b; z, \boldsymbol{x}, \mathcal{N}) \geq \alpha \right\}.$$

Under the family of null hypotheses $\{\mathrm{H}_0^j\}_{j=1}^J$, Lemma 1 implies that the distribution of bidder j's all bids coincides with the distribution $M_j(b; z, \boldsymbol{x}, \mathcal{N})$ induced from winning bids and identities of winners. Hence, if all bids are observed, one can also estimate bidder j's conditional inverse pseudo-value distribution by

$$\hat{U}_j^{-1}(\alpha; z, \boldsymbol{x}, \mathcal{N}) = \hat{B}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) + \frac{\sum_{\ell \in \mathcal{N}} \hat{H}_\ell(\hat{B}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}); z, \boldsymbol{x}, \mathcal{N})}{\sum_{\ell \in \mathcal{N} \setminus \{j\}} \hat{H}'_\ell(\hat{B}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}); z, \boldsymbol{x}, \mathcal{N})},$$

where $\hat{B}_j(\alpha; z, \boldsymbol{x}, \mathcal{N})$ is an estimate of α-quantile of bidder j's all bids across auctions conditional on auction-level covariates. The α-quantile of bidder j's all bids can be estimated easily by sorting bidder j's observed bids.

The conditional integrated quantile function can be estimated as follows. Let a vector of uniform knots be denoted by $(\alpha_0, \alpha_1, \cdots, \alpha_S)$ in the domain $[0, \alpha)$ with $\alpha_0 = 0$ and $\alpha \in (0, 1]$. Since the quantile process is a right-continuous step function, the estimated conditional inverse pseudo-value distribution is a piecewise linear function. Thanks to this property, a consistent estimator of bidder j's conditional integrated quantile function is written as

$$\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) = \int_0^\alpha \hat{U}_j^{-1}(t; z, \boldsymbol{x}, \mathcal{N})\, dt = (S-1)^{-1} \sum_{s=0}^{S} \hat{U}_j^{-1}(\alpha_s; z, \boldsymbol{x}, \mathcal{N}).$$

Two empirical issues are discussed here. First, the kernel estimator is inconsistent at the boundary of the support, which is known as boundary effects. The estimation of inverse bidding functions is biased near the boundary of bids. Boundary correction methods have been extensively studied in the literature; see Zhang et al. (1999) for a survey on boundary correction methods. Hickman and Hubbard (2015) used a transformation and reflection method to address boundary effects in estimating density functions of private values in first-price auctions. Trimming near the boundaries is another approach to avoid boundary effects in the literature; see Guerre et al. (2000). To better illustrate this notion, let b_τ denote the τ-th quantile of $M_j(b; z, \boldsymbol{x}, \mathcal{N})$, that is, $b_\tau = M_j^{-1}(\tau; z, \boldsymbol{x}, \mathcal{N})$. Choose $\tau \in (0,1)$ such that $b_\tau \geq \underline{b} + h_b^*$ and $b_{1-\tau} \leq \overline{b} - h_b^*$, where h_b^* is the optimal bandwidth for the univariate kernel function $K_{h_b}(\cdot, \cdot)$. The testing hypothesis is modified to

$$\mathrm{H}_0^j : Q_{j,\tau}(\alpha; z, \boldsymbol{x}, \mathcal{N}) = Q_{j,\tau}(\alpha; z', \boldsymbol{x}, \mathcal{N}) \quad \text{for any } z \neq z' \text{ and } \alpha \in [\tau, 1-\tau],$$

for each $j \in \mathcal{N}$, where $Q_{j,\tau}(\alpha; z, \boldsymbol{x}, \mathcal{N})$ is the trimmed integrated quantile function defined by

$$Q_{j,\tau}(\alpha; z, \boldsymbol{x}, \mathcal{N}) = \int_\tau^\alpha U_j^{-1}(t; z, \boldsymbol{x}, \mathcal{N})\, dt.$$

Second, though nonparametric estimation is attractive in principle, it may suffer from the curse of dimensionality in the finite sample. Most empirical auction papers exploit the approach in Haile et al. (2003) to reduce the dimensions of auction-level covariates. The method in Haile et al. (2003) assumes the additive (or multiplicative) separability structure of private values and covariates. Under this structure, a simple regression can eliminate the dimensions of auction covariates. One may alternatively adopt the single index semi-parametric model to control auction heterogeneity; see Kong (2021).

Remark. *Imposing the monotonicity constraint in estimating inverse bidding functions is of great interest in the literature. For instance, Henderson et al. (2012) proposed a constrained reweighting approach to achieve the monotonicity of the estimator. Ma et al. (2021) developed a smooth rearrangement approach and investigated the asymptotic property of this monotonicity-constrained estimator. In the spirit of Luo and Wan (2018), the monotonicity constraint can be imposed by taking the greatest convex minorant (GCM) of the estimated integrated quantile function.*[17] *However, constructing critical value in the testing problem with shape enforcing operators becomes complicated when the parameter is not $\sqrt{n}$-estimable; see Fang (2021). It is not pursued in this paper.*

5.3. **Simultaneous Inference.** Recall that the test statistic for the individual null hypothesis H_0^j is

$$t_j = \sum_{z, z' \in \mathcal{Z}} \phi\left(\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})\right).$$

The individual null hypothesis H_0^j is rejected for a large value of the test statistic t_j. In the problem of simultaneously testing finite hypotheses, it is desired to control the family-wise error rate (FWER), which is the probability of one or more false rejections for all possible configurations of true and false hypotheses. This paper proposes to perform simultaneous inference by the Holm procedure, which is one of a class of step-down procedures. Inference methods that control the FWER are typically described by the p-values. Let the significance level be $\alpha \in (0,1)$. The implementation of the Holm procedure is as follows.

[17] The greatest convex minorant of a real-valued function f defined on $\mathcal{X}$ is the maximal convex function g such that $g(x) \leq f(x)$ for all $x \in \mathcal{X}$.

STEP 1: Nonparametrically estimate the test statistic for bidder $j \in \{1, \cdots, J\}$ from the original sample $(\boldsymbol{B}_t, Z_t, \boldsymbol{X}_t, \mathcal{N}_t)_{t=1}^T$:

$$t_j = \sum_{z,z' \in \mathcal{Z}} \phi\left(\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})\right).$$

STEP 2: Generate the bootstrap algorithm samples $\{(\boldsymbol{B}_t^*(r), Z_t^*(r), \boldsymbol{X}_t^*(r), \mathcal{N}_t^*(r))_{t=1}^T\}_{r=1}^R$ from the original sample for a fixed number of bootstrap replications R.[18] For each replication $r \in \{1, \cdots, R\}$ and bidder $j \in \{1, \cdots, J\}$, compute the bootstrap statistic by

$$t_j^*(r) = \sum_{z,z' \in \mathcal{Z}} \phi\left([\hat{Q}_{j,b}^*(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_{j,b}^*(\alpha; z', \boldsymbol{x}, \mathcal{N})] - [\hat{Q}_j(\alpha; z, \boldsymbol{x}, \mathcal{N}) - \hat{Q}_j(\alpha; z', \boldsymbol{x}, \mathcal{N})]\right),$$

where $\hat{Q}_{j,b}^*(\alpha; z, \boldsymbol{x}, \mathcal{N})$ is a bootstrap analogue of bidder j's integrated quantile function computed from the bootstrap sample $(\boldsymbol{B}_t^*(r), Z_t^*(r), \boldsymbol{X}_t^*(r), \mathcal{N}_t^*(r))_{t=1}^T$. The bootstrap p-values $\{\hat{p}_j\}_{j=1}^J$ can be computed from

$$\hat{p}_j = R^{-1} \sum_{r=1}^R \mathbb{1}\left\{|t_j^*(r)| > |t_j|\right\}.$$

STEP 3: Order bootstrap p-values by $\hat{p}_{(1)} \leq \hat{p}_{(2)} \leq \cdots \leq \hat{p}_{(J)}$, and denote the associated hypotheses by $\mathrm{H}_0^{(1)}, \mathrm{H}_0^{(2)}, \cdots, \mathrm{H}_0^{(J)}$. If $\hat{p}_{(1)} \geq \alpha/J$, accept all hypotheses $\mathrm{H}_0^{(1)}, \mathrm{H}_0^{(2)}, \cdots, \mathrm{H}_0^{(J)}$.

STEP 4: If $\hat{p}_{(1)} < \alpha/J$, reject $\mathrm{H}_0^{(1)}$ and test the remaining $J-1$ hypotheses at the significance level $\alpha/(J-1)$.

STEP 5: If $\hat{p}_{(2)} \geq \alpha/(J-1)$, accept all hypotheses $\mathrm{H}_0^{(2)}, \mathrm{H}_0^{(3)}, \cdots, \mathrm{H}_0^{(J)}$; otherwise, reject $\mathrm{H}_0^{(2)}$ and test the remaining $J-2$ hypotheses at the significance level $\alpha/(J-2)$.

STEP 6: Continue this procedure to reject or accept hypotheses $\mathrm{H}_0^{(3)}, \mathrm{H}_0^{(4)}, \cdots, \mathrm{H}_0^{(J)}$.

Under the Holm procedure (Step 1-6 above), the family-wise error rate (FWER) of any false rejection is less or equal to α (Lehmann et al., 2005).[19]

5.4. **Simulations.** Two Monte Carlo experiments are conducted to investigate the finite sample performance of the test. Only winning bids and identities of winners are observed in a finite sequence of first-price reverse auctions. For the convenience of DGP, let every bidder have a uniform cost distribution $F(c) = c$ on the support $[0, 1]$.[20] The number of bootstrap repetitions is $R = 200$, and the number of DGP repetitions is 400 due to the computational expense. Test statistics are constructed with the supremum norm.

---

[18] The bootstrap method is Efron's bootstrap.

[19] One potential limitation of the Holm procedure is the conservativeness which may result in lower power. To improve the power properties of the test, Romano and Wolf (2005) provided a stepwise multiple testing method that captures the joint dependence structure of the test statistics.

[20] In first-price reverse auctions with N symmetric bidders, the equilibrium bidding strategy is

$$\beta(c) = c + (1 - F(c))^{-(N-1)} \int_c^{\bar{c}} (1 - F(x))^{N-1}\, dx.$$

When every bidder has a uniform cost distribution $F(c) = c$ on the interval $[0, 1]$, $\beta(c) = \frac{1+(N-1)c}{N}$.

The first experiment studies the empirical size of the test, i.e., the probability that the test rejects the null hypothesis when no collusion occurs. Suppose there is a set of three competitive bidders $\mathcal{S}_1$. The number of additional competitive bidders serves as the instrumental variable. The test statistic for each bidder in $\mathcal{S}_1$ is constructed from the paired samples with different numbers of additional competitive bidders $N' \in \{0, 2, 3\}$. Table 1 summarizes the empirical sizes under the nominal levels $\alpha \in \{0.01, 0.05, 0.1\}$. In all scenarios, the empirical sizes are close to the nominal levels as the sample size increases.

The second experiment studies the empirical local power of the test, i.e., the probability that the test rejects false null hypotheses. Suppose a set of two collusive bidders $\mathcal{S}_2$ form a weak cartel. The designated bidder is randomly selected with an equal probability. The test statistic for each bidder in $\mathcal{S}_2$ is constructed from the paired samples with different numbers of additional competitive bidders $N' \in \{1, 2, 3\}$. There are various ways to define power in multiple testing, e.g., minimal power, complete power, individual power, and average power (Chen et al., 2011). The comparison of different definitions is beyond the scope of this paper. In this simulation study, the power is defined as the average proportion of rejected false null hypotheses, i.e., average power. Table 2 summarizes the empirical local powers. For each nominal level α, the empirical local powers are increasing in the sample size.

Table 1. Empirical Sizes

	$N' \in \{0, 2\}$			$N' \in \{0, 3\}$		
T	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
250	0.020	0.076	0.130	0.020	0.078	0.132
500	0.018	0.055	0.083	0.015	0.075	0.128
1000	0.010	0.055	0.100	0.018	0.060	0.110
1500	0.015	0.045	0.090	0.017	0.058	0.115

T is the total number of auctions in each scenario.

Table 2. Empirical Local Powers

	$N' \in \{1, 2\}$			$N' \in \{1, 3\}$		
T	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
250	0.034	0.109	0.178	0.154	0.300	0.384
500	0.048	0.144	0.225	0.205	0.386	0.504
1000	0.084	0.250	0.354	0.326	0.548	0.660
1500	0.243	0.463	0.561	0.596	0.793	0.854

T is the total number of auctions in each scenario.

6. Detect Collusion in Procurement Auctions

Many prosecution cases have shown that collusion is pervasive in public procurement auctions. In school milk procurement auctions, more than 35 bid-rigging cases have been documented in Texas and Florida (Pesendorfer, 2000). Thirteen dairies were charged with collusion in school milk procurement auctions from 1980 to 1990 in Ohio (Porter and Zona, 1999). In addition to the school milk market, a large proportion of bid-rigging cases were filed in public procurement auctions for construction projects. In New York, several large firms were convicted in federal court of rigging bids in Long Island highway construction projects (Porter and Zona, 1993). According to court documents, one former employee in the California Department of Transportation (Caltrans) was sentenced to 6.5 years for bid rigging. Caltrans incurred a loss of more than $1.2 million.

There is a well-recognized difficulty in identifying collusive bidders in auctions because cartels can mimic competitive bidding. An empirical challenge is to develop a statistical test that can discriminate between collusive bidders and non-collusive bidders in public procurement auctions. Based on the testable implication of collusion, this section applies the statistical method in Section 5 to detect collusion in the California highway procurement auctions. All identification results in previous sections can accommodate procurement auction settings. Model assumptions and empirical issues are to be discussed in the following subsections.

6.1. **Data Description.** The raw data were collected by Bajari et al. (2014) from the California Department of Transportation. The raw data contain information on the California highway construction contracts awarded through public procurement auctions from 1999 to 2008. For each project, the data include the bidding date, total bids, bidder identities, the location of the project, the number of bidders, the engineer's estimate of the project's cost, the distance of the firm to the project, the utilization rate, the job type, and the number of days to complete the project. For each bidder, the utilization rate is defined as the ratio of backlog to capacity.[21] Following An and Tang (2019), the job type is a dummy variable, which equals one if the contract involves major construction or rebuilding based on the project description. Using a classification criterion similar to Jofre-Bonet and Pesendorfer (2003), all firms are divided into two types: non-fringe firms, which have a market share above 5% with at least 60 bids submitted in the market; and fringe firms, which are the remaining firms.[22]

A small portion of procurement auctions in which the contract was not awarded to the firm with the lowest bid is dropped. Auctions with bids lower than 1.5 percentile or higher than 98.5 percentile are excluded from the sample. The full sample includes 1,488 contracts

[21] Backlog is defined as the sum of the dollar values of contracts won but not yet completed. Capacity is the maximum backlog during the sample period.

[22] In this empirical application, several neighboring districts are grouped as one market.

issued by Caltrans, with 7,026 bids submitted by 455 contractors. Table 3 summarizes the descriptive statistics of the sample. There is noticeable auction-specific heterogeneity. The number of participating firms ranges from 2 to 9 in most auctions. On average, four to five firms submit bids in a standard auction, among which three are fringe firms. The median winning bid is around $969,000, which is approximately 10.4% less than the median engineer's estimate of the project's cost. The project requires 96 days to complete on average.

Table 3. Descriptive Statistics

Variables	Auctions	Mean	Standard deviation	P5	Median	P95
Winning bid ($1000)	1,488	3,691	10,333	205	969	13,343
Average bid ($1000)	1,488	4,021	10,996	238	1,082	14,325
Engineer's estimate ($1000)	1,488	3,911	10,257	229	1,057	13,442
Distance to project (100 miles)	1,488	1.096	0.906	0.184	0.891	2.801
Utilization rate	1,488	0.128	0.132	0	0.088	0.380
Job type	1,488	0.480	0.500	0	0	1
Working days	1,488	96	117	18	60	293
Number of bidders	1,488	4.662	2.343	2	4	9
Number of fringe bidders	1,488	3.418	2.367	1	3	8

Non-fringe firms and fringe firms are substantially different in bidding behavior. Table 4 compares firm bidding behavior. Table 4 shows that non-fringe firms participate more frequently than fringe firms. The larger capacity of non-fringe firms can explain this difference in participation. On average, a non-fringe firm wins 66 contracts, while a fringe firm only wins two contracts in the sample. The win rate of non-fringe firms is also much higher than that of fringe firms. Around 40% procurement auctions are won by nine non-fringe firms.

Table 4. Comparison Between Firms

Variables	Non-fringe firms	Fringe firms
Number of firms	9	446
Number of wins	591	897
Average wins per firm	65.67	2.01
Number of bids submitted	1,808	5,254
Average bids submitted	200.89	11.78

The identification arguments in Section 4 only apply to the same set of potential bidders. Thus, the estimation of inverse pseudo-cost distributions must be undertaken by fixing the set of potential bidders. A frequently used method in the literature is to assign a bidder as a potential bidder in one auction if this bidder participates in at least one another auction in the same district and within a specific time length prior to this auction. Considering collusive bidding behavior, this paper groups neighboring districts where firms have similar participation behavior to generate the set of potential bidders for convenience of estimation and inference.[23]

There are twelve administrative districts in the California highway procurement auctions. The firm participation behavior can largely be explained by geography. It is unlikely that those firms active in distant districts will collude. Since firms are typically active in several neighboring districts, it is more reasonable to pool all data in those neighboring districts to test potential collusive bidding behavior. To determine which neighboring districts should be grouped as one region, the participation behavior of top local firms is compared in each district. Top local firms have a market share higher than 5% and a win rate higher than 1% in the district.[24] In each district $d \in \{1, \cdots, 12\}$, firm j's local market share in district d is defined as

$$s_{j,d} = \frac{\text{total of firm } j\text{'s winning bids in district } d}{\text{total of winning bids in district } d}.$$

By construction, there is a set of top local firms corresponding to each district. Notice that top local firms in one district may also be active in others. The idea is to group districts where top local firms have similar participation behavior. Finally, there are four regions in the sample: districts 1 and 2 are region 1; districts 4 and 5 are region 2; districts 3, 6, 9, and 10 are region 3; and districts 7, 8, 11, and 12 are region 4. The pattern of participation behavior is evident by observing the Caltrans district map (see Figure 2). It is interesting to notice that neighboring districts are grouped as one region. The details can be found in Appendix B.

Because collusive bidders may refrain from bidding, it is infeasible to correctly construct the set of potential bidders for each auction in a specific region. As an approximation, it is assumed that non-fringe firms in each region are always the potential bidders and the participation of fringe firms is exogenous. This assumption is also maintained by Bajari et al. (2010) in an empirical analysis of equilibrium selection with the Caltrans procurement auctions. In summary, for each auction, the set of potential bidders includes all non-fringe firms in this region and fringe firms who submit bids.

6.2. **Collusion Between Potential Bidders.** More than 450 firms submit at least one bid in the sample. Subject to the finite sample size, it is infeasible to perform the test for each firm. A practical issue is to consider which potential bidders may form a cartel. Firms

[23] Grouping different districts as one region is common in the analysis of auction data; for example, Li and Zhang (2010), Athey et al. (2011), and Schurter (2017).

[24] The definition of top local firms is different from that of non-fringe firms in Bajari et al. (2014). In Bajari et al. (2014), the non-fringe firm has a market share higher than 1% in one unified market.

participating in the Caltrans procurement auctions are divided into two types by market share criterion. Figure 1 lists the top firms with a market share above 2% in each region. An important feature is that market shares in each region are quite skewed. For example, among 75 firms bidding in region 1, the two largest firms have a market share of around 40%. This pattern of market shares suggests that collusion between those non-fringe firms would be practitioners' primary concern.

In the industrial organization literature, it is common to consider that a cartel faces numerous competitive fringe firms (Montero and Guzman, 2010; De Roos and Smirnov, 2021). This empirical application focuses on collusion between non-fringe firms. Following Aryal and Gabrielli (2013), it is assumed that non-fringe firms have no incentive to collude with fringe firms and all fringe firms are competitive. There are two main reasons for this consideration. First, fringe firms have limited influence on winning bids. While there are 455 contractors in the sample, the market share of the ten largest firms is almost 50%. Non-fringe firms win around 40% procurement auctions in the sample. Second, many fringe bidders only bid a few times in the sample. Table 4 shows that fringe firms submit bids only around 12 times on average. The number of bid observations is too few to infer the pseudo-cost distributions of fringe firms.

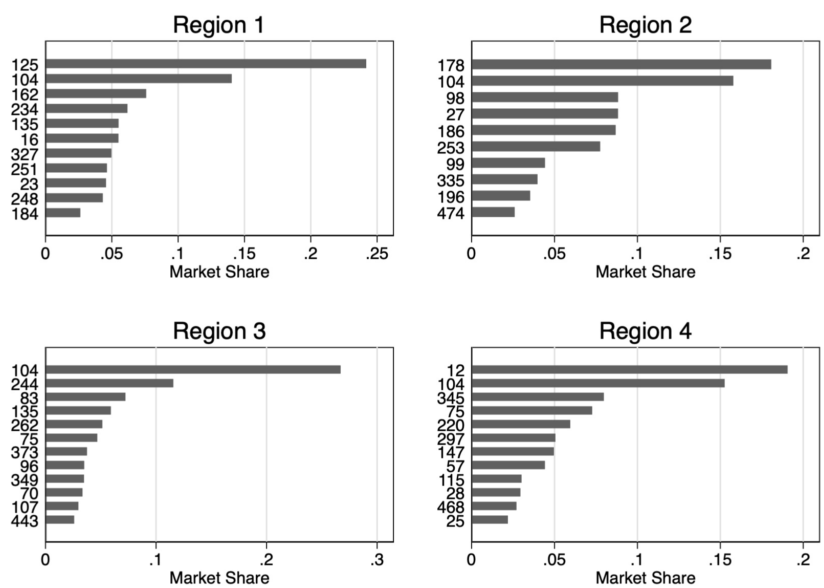

FIGURE 1. —This figure lists the firms with a market share above 2% in each region. Identities of firms are labeled on the vertical axis.

6.3. **Auction-Level Instruments.** Identification of collusive bidders in Theorem 1 requires econometricians to observe auction-level instruments satisfying Assumption 3. Examples of instruments in the empirical auction literature include reserve prices, the distance of the bidder to the project (cost shifters), and sets of potential bidders. However, the California Department of Transportation does not formally set public reserve prices. In an environment with asymmetric bidders, auction-level instruments like cost shifters should include every bidder's rivals, which may suffer from the curse of dimensionality. This paper exploits the variation in bidders to identify possible bidding rings in the California highway procurement auctions.

In procurement auctions, the excludable variation in bidders is supported by the local nature of participating firms. A few non-fringe firms repeatedly compete with numerous competitive fringe firms across auctions in the same administrative district. If the cartel formed by non-fringe firms can exclude competitive fringe firms, then the participation of fringe firms can serve as instruments to identify collusive non-fringe firms. To fix this idea, consider a set of non-fringe firms $\mathcal{S}$ in a region and assume that competitive fringe firms are homogeneous in cost distributions. The instrumental variable Z_t is the number of competitive fringe firms participating in auction t. If non-fringe firm $j \in \mathcal{S}$ does not collude, the exclusion restriction requires firm j's pseudo-cost distribution to satisfy

$$U_j(v; z, \boldsymbol{x}, \mathcal{S}) = U_j(v; z', \boldsymbol{x}, \mathcal{S}).$$

Under the null hypothesis of no collusion, researchers can test the equality of non-fringe firms' pseudo-cost distributions across auctions with different numbers of competitive fringe firms.

If the cartel is stable between non-fringe firms by Assumption 1, it is expected that the infrequent participation of fringe firms will not change the collusion mechanism. It is also assumed that the participation of fringe firms does not affect collusive bidders' cost distributions. Table 5 shows that both the average winning bid and the average bid are decreasing with the number of fringe firms after controlling the auction heterogeneity with the approach in Haile et al. (2003). This observation supports the assumption of exogenous participation of fringe firms.

6.4. **Implementation.** This paper uses the approach in Haile et al. (2003) to reduce the dimensions of auction-level characteristics. Specifically, assume that the bidder's valuations have a multiplicative separable structure such that

$$V_{jt} = \delta(\boldsymbol{X}_t) A_{jt},$$

where $\delta(\boldsymbol{X}_t) = \exp(\boldsymbol{X}_t \gamma)$ and the bidder-specific private information A_{jt} is independent of auction-level covariates $(\boldsymbol{X}_t, \mathcal{N}_t)$. By Haile et al. (2003), the equilibrium bidding function also preserves the multiplicative separability, that is,

$$\beta_j(V_{jt}; \boldsymbol{X}_t, \mathcal{N}_t) = \delta(\boldsymbol{X}_t) \beta_j(A_{jt}; \boldsymbol{X}_0, \mathcal{N}_t)$$

TABLE 5. NUMBER OF FRINGE FIRMS AND HOMOGENIZED BIDS

		Average winning bid		Average bid	
Number of fringe firms	Auctions	Mean	SE	Mean	SE
1	251	0.989	0.209	1.058	0.203
2	294	0.943	0.175	1.033	0.178
3	261	0.903	0.163	1.016	0.166
4	200	0.863	0.152	0.982	0.161
5	150	0.817	0.151	0.935	0.157
6	105	0.800	0.140	0.926	0.150
7	62	0.797	0.152	0.939	0.155
8	40	0.751	0.129	0.885	0.137

with $\delta(\boldsymbol{X}_0) = 1$. The multiplicative separability of the equilibrium bidding function implies that bids can be homogenized by regressing the logarithm of bids on auction-level covariates, including the logarithm of the engineer's estimate of cost, the average distance to the project, the average utilization rate, the job type, and the logarithm of working days. Table 6 displays the results of regression analysis. The regression results show that the engineer's estimate of the project's cost is a very powerful explanatory variable. An R-square of 97.4% in regression (1) indicates that the auction heterogeneity is captured well by the engineer's estimate of the project's cost. The bidder's cost increases in the distance to the project and working days. Regression (4) is used to homogenize bids before estimating pseudo-cost distributions. The homogenized bid B_{jt}^h, which is bidder j would have submitted in auction t if auction-level characteristics are $\boldsymbol{X}_0$, is the exponential of estimated residuals:

$$B_{jt}^h = \exp(\log B_{jt} - \boldsymbol{X}_t \hat{\gamma}), \tag{13}$$

where $\hat{\gamma}$ is a vector of estimated coefficients by ordinary least squares regression (OLS).

The estimation procedure in Section 5 can be adapted for procurement auctions. Under the null hypothesis of no collusion, the inverse pseudo-cost distribution of bidder $j \in \mathcal{S}$ is estimated by

$$\hat{U}_j^{-1}(\alpha; z, \mathcal{S}) = \hat{B}_j^h(\alpha; z, \mathcal{S}) - \frac{1 - \sum_\ell \hat{H}_\ell(\hat{B}_j^h(\alpha; z, \mathcal{S}); z, \mathcal{S})}{\sum_{\ell \neq j} \hat{H}'_\ell(\hat{B}_j^h(\alpha; z, \mathcal{S}); z, \mathcal{S})}, \tag{14}$$

where the conditional cumulative incidence function and its derivative are estimated using homogenized bids, and $\hat{B}_j^h(\alpha; z, \mathcal{N})$ is an estimate of α-quantile of bidder j's homogenized bids across auctions conditional on $Z_t = z$. The estimation of the derivative of the cumulative incidence functions adopts the Gaussian kernel function and the Silverman rule-of-thumb bandwidth.

The test procedure with homogenized bids involves the following steps.

TABLE 6. REGRESSION RESULTS

Variables	(1) log(total bid)	(2) log(total bid)	(3) log(total bid)	(4) log(total bid)
log(engineer's estimate)	0.988***	0.983***	0.985***	0.971***
	(0.00190)	(0.00189)	(0.00193)	(0.00251)
Average distance to project		0.0142***	0.0148***	0.0179***
		(0.00365)	(0.00368)	(0.00377)
Average utilization rate			-0.0965***	-0.0708***
			(0.0221)	(0.0220)
Job type				0.0235***
				(0.00492)
log(working days)				0.0281***
				(0.00344)
Constant	0.150***	0.256***	0.249***	0.316***
	(0.0270)	(0.0333)	(0.0336)	(0.0346)
Competition FE		YES	YES	YES
Observations	7062	7062	7062	7062
R-squared	0.974	0.976	0.976	0.976

Standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

<u>STEP 1</u>: Given the sample $(\boldsymbol{B}_t, Z_t, \boldsymbol{X}_t, \mathcal{S}_t)_{t=1}^T$, control the observed auction-specific heterogeneity with the method in Haile et al. (2003). Specifically, run the regression (4) in Table 6 and obtain the sample of homogenized bids $(\boldsymbol{B}_t^h, Z_t, \mathcal{S}_t)_{t=1}^T$ by equation (13).

<u>STEP 2</u>: For each configuration of non-fringe firms $\mathcal{S}$ and the number of fringe firms z, compute the trimmed integrated quantile function from the sample of homogenized bids $(\boldsymbol{B}_t^h, Z_t, \mathcal{S}_t)_{t=1}^T$:

$$\hat{Q}_{j,\tau}(\alpha; z, \mathcal{S}) = \int_\tau^\alpha \hat{U}_j^{-1}(t; z, \mathcal{S})\, dt,$$

where $\tau = 5\%$ and $\alpha \in [\tau, 1-\tau]$ and $\hat{U}_j^{-1}(t; z, \mathcal{S})$ is given by equation (14). The test statistic for bidder $j \in \mathcal{S}$ is

$$t_j = \sum_{z, z' \in \mathcal{Z}} \sup_{\alpha \in [\tau, 1-\tau]} |\hat{Q}_j(\alpha; z, \mathcal{S}) - \hat{Q}_j(\alpha; z', \mathcal{S})|.$$

<u>STEP 3</u>: Generate the bootstrap samples $\{(\boldsymbol{B}_t^*(r), Z_t^*(r), \boldsymbol{X}_t^*(r), \mathcal{S}_t^*(r))_{t=1}^T\}_{r=1}^R$ from the original sample $(\boldsymbol{B}_t, Z_t, \boldsymbol{X}_t, \mathcal{S}_t)_{t=1}^T$ for bootstrap replications $R = 500$. Next, run the regression (4) in Table 6 and homogenize the bootstrap samples of bids by equation (13). The

bootstrap samples of homogenized bids are denoted by $\{(\boldsymbol{B}_t^{*h}(r), Z_t^*(r), \mathcal{S}_t^*(r))_{t=1}^T\}_{r=1}^R$. For each replication r, compute the bootstrap test statistic for bidder $j \in \mathcal{S}$:

$$t_j^*(r) = \sum_{z,z' \in \mathcal{Z}} \sup_{\alpha \in [\tau, 1-\tau]} |[\hat{Q}_j^*(\alpha; z, \mathcal{S}) - \hat{Q}_j^*(\alpha; z', \mathcal{S})] - [\hat{Q}_j(\alpha; z, \mathcal{S}) - \hat{Q}_j(\alpha; z', \mathcal{S})]|,$$

where the trimmed integrated quantile function is computed from the bootstrap samples of homogenized bids $\{(\boldsymbol{B}_t^{*h}(r), Z_t^*(r), \mathcal{S}_t^*(r))_{t=1}^T\}_{r=1}^R$.

<u>Step 4</u>: Perform the simultaneous inference in Section 5 for each configuration $\mathcal{S}$.

6.5. **Results.** Since the number of fringe firms serves as the instrumental variable, the test statistics are computed from the paired samples. The test only uses the paired samples in which the number of fringe firms is less than or equal to five and non-fringe firms have more than 20 bids. Table 7 reports bootstrap p-values of test statistics. For each non-fringe firm in different regions, Table 7 reports bootstrap p-values computed from two paired samples with different numbers of fringe firms. The bootstrap p-values in the last column of Table 7 are calculated by aggregating the test statistics for all paired samples. The bootstrap p-values in the last column are all higher than 10%. Under the model assumptions, the test results suggest no statistical evidence of large-scale collusion among non-fringe firms in the sample after controlling for the observed auction-specific heterogeneity. This finding supports Aryal and Gabrielli (2013).

Table 7. Bootstrap p-Values

Regions	Firms	Pairwise										Aggregate
		(1,2)	(1,3)	(1,4)	(1,5)	(2,3)	(2,4)	(2,5)	(3,4)	(3,5)	(4,5)	
1	104	.460	.156			.586						.370
	125	.318		.204			.480					.288
	162	.772										.772
2	98								.198	.108	.962	.394
	104	.894	.082	.094		.054	.072		834			.114
	186	.180		.874	.586		.216	.470			.564	.494
	253	.952										.952
3	104	.088	.072	.380	.002	.976	.834	.180	.690	.216	.128	.120
	244	.170	.024			.748						.152
	262	.790	.294			.608						.602
4	12					.502	.944	.984	.580	.496	.820	.910
	104					.586	.920	.456	.696	.144	.322	.582

7. Conclusion

This paper provides identification results in asymmetric first-price sealed-bid auctions with general collusion mechanisms. In a static first-price auction model with partial cartels, identities of collusive bidders can be identified with auction-level instruments for a broad class of collusion schemes. The key insight is that bidder's value distribution can be correctly inferred from winning bids and identities of winners if this bidder is non-collusive. The collusion structure consisting of collusive bidders' value distribution and the collusion scheme is not identified. All bidders' value distributions are nonparametrically identified if the cartel is efficient or the designated bidder is randomly selected. This paper proposes a test statistic based on identification results and applies it to detect potential collusion in the California highway procurement auctions. The test results suggest no statistical evidence of large-scale collusion among non-fringe firms in the sample.

References

An, Yonghong and Xun Tang. 2019. *Specification and negotiation in incomplete contracts*, The RAND Journal of Economics **50**, no. 1, 57–92.

Aryal, Gaurab and Maria F. Gabrielli. 2013. *Testing for collusion in asymmetric first-price auctions*, International Journal of Industrial Organization **31**, no. 1, 26–35.

Aryal, Gaurab, Serafin Grundl, Dong-Hyuk Kim, and Yu Zhu. 2018. *Empirical relevance of ambiguity in first-price auctions*, Journal of Econometrics **204**, no. 2, 189–206.

Asker, John. 2010. *A study of the internal organization of a bidding cartel*, The American Economic Review **100**, no. 3, 724–62.

Athey, Susan and Philip A Haile. 2007. *Nonparametric approaches to auctions*, Handbook of econometrics **6**, 3847–3965.

Athey, Susan, Jonathan Levin, and Enrique Seira. 2011. *Comparing open and sealed bid auctions: Evidence from timber auctions*, The Quarterly Journal of Economics **126**, no. 1, 207–257.

Bajari, Patrick, Han Hong, and Stephen P Ryan. 2010. *Identification and estimation of a discrete game of complete information*, Econometrica **78**, no. 5, 1529–1568.

Bajari, Patrick, Stephanie Houghton, and Steven Tadelis. 2014. *Bidding for incomplete contracts: An empirical analysis of adaptation costs*, American Economic Review **104**, no. 4, 1288–1319.

Bajari, Patrick and Lixin Ye. 2003. *Deciding between competition and collusion*, Review of Economics and statistics **85**, no. 4, 971–989.

Baldwin, Laura H, Robert C Marshall, and Jean-Francois Richard. 1997. *Bidder collusion at forest service timber sales*, Journal of Political Economy **105**, no. 4, 657–699.

Banerji, Abhijit and JV Meenakshi. 2004. *Buyer collusion and efficiency of government intervention in wheat markets in northern india: An asymmetric structural auctions analysis*, American journal of agricultural economics, 236–253.

Barrett, Garry F, Stephen G Donald, and Debopam Bhattacharya. 2014. *Consistent nonparametric tests for lorenz dominance*, Journal of Business & Economic Statistics **32**, no. 1, 1–13.

Berman, Simeon M. 1963. *Note on extreme values, competing risks and semi-markov processes*, The Annals of Mathematical Statistics **34**, no. 3, 1104–1106.

Chassang, Sylvain, Kei Kawai, Jun Nakabayashi, and Juan Ortner. 2022. *Robust screens for noncompetitive bidding in procurement auctions*, Econometrica **90**, no. 1, 315–346.

Chassang, Sylvain, Kei Kawai, Jun Nakabayashi, and Juan M Ortner. 2021. *Data driven regulation: Theory and application to missing bids*, National Bureau of Economic Research.

Chassang, Sylvain and Juan Ortner. 2019. *Collusion in auctions with constrained bids: Theory and evidence from public procurement*, Journal of Political Economy **127**, no. 5, 2269–2300.

Che, Yeon-Koo, Daniele Condorelli, and Jinwoo Kim. 2018. *Weak cartels and collusion-proof auctions*, Journal of Economic Theory **178**, 398–435.

Chen, Jie, Jianfeng Luo, Kenneth Liu, and Devan V Mehrotra. 2011. *On power and sample size computation for multiple testing procedures*, Computational Statistics & Data Analysis **55**, no. 1, 110–122.

Conley, Timothy G and Francesco Decarolis. 2016. *Detecting bidders groups in collusive auctions*, American Economic Journal: Microeconomics **8**, no. 2, 1–38.

De Roos, Nicolas and Vladimir Smirnov. 2021. *Collusion, price dispersion, and fringe competition*, European Economic Review **132**, 103640.

Durlauf, Steven and Lawrence E Blume. 2016. *The new palgrave dictionary of economics*, Springer.

Fang, Zheng. 2021. *A unifying framework for testing shape restrictions*, Working Paper.

Gentry, Matthew and Tong Li. 2014. *Identification in auctions with selective entry*, Econometrica **82**, no. 1, 315–344.

Graham, Daniel A and Robert C Marshall. 1987. *Collusive bidder behavior at single-object second-price and english auctions*, Journal of Political Economy **95**, no. 6, 1217–1239.

Gruyer, Nicolas. 2009. *Optimal auctions when a seller is bound to sell to collusive bidders*, The Journal of Industrial Economics **57**, no. 4, 835–850.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong. 2000. *Optimal nonparametric estimation of first-price auctions*, Econometrica **68**, no. 3, 525–574.

______. 2009. *Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions*, Econometrica **77**, no. 4, 1193–1227.

Haile, Philip, Han Hong, and Matthew Shum. 2003. *Nonparametric tests for common values at first-price sealed-bid auctions*.

Harrington, Joseph E. 2005. *Detecting cartels*, Working paper.

Henderson, Daniel J, John A List, Daniel L Millimet, Christopher F Parmeter, and Michael K Price. 2012. *Empirical implementation of nonparametric first-price auction models*, Journal of Econometrics **168**, no. 1, 17–28.

Hendricks, Ken, Robert Porter, and Guofu Tan. 2008. *Bidding rings and the winner's curse*, The RAND Journal of Economics **39**, no. 4, 1018–1041.

Hendricks, Ken and Robert H Porter. 2007. *An empirical perspective on auctions*, Handbook of Industrial Organization **3**, 2073–2143.

Hickman, Brent R and Timothy P Hubbard. 2015. *Replacing sample trimming with boundary correction in nonparametric estimation of first-price auctions*, Journal of Applied Econometrics **30**, no. 5, 739–762.

Hortaçsu, Ali and Isabelle Perrigne. 2021. *Empirical perspectives on auctions*, Handbook of Industrial Organization **5**, no. 1, 81–175.

Jofre-Bonet, Mireia and Martin Pesendorfer. 2003. *Estimation of a dynamic auction game*, Econometrica **71**, no. 5, 1443–1489.

Kawai, Kei and Jun Nakabayashi. 2021. *Detecting large-scale collusion in procurement auctions*, Journal of Political Economy, forthcoming.

Kawai, Kei, Jun Nakabayashi, and Juan M Ortner. 2021. *The value of privacy in cartels: An analysis of the inner workings of a bidding ring*, National Bureau of Economic Research.

Klein, John P and Per Kragh Andersen. 2005. *Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function*, Biometrics **61**, no. 1, 223–229.

Kong, Yunmi. 2021. *Sequential auctions with synergy and affiliation across auctions*, Journal of Political Economy **129**, no. 1, 148–181.

Krasnokutskaya, Elena. 2011. *Identification and estimation of auction models with unobserved heterogeneity*, The Review of Economic Studies **78**, no. 1, 293–327.

Lebrun, Bernard. 1997. *First price auctions in the asymmetric n bidder case*, Working Paper.

———. 2006. *Uniqueness of the equilibrium in first-price auctions*, Games and Economic Behavior **55**, no. 1, 131–151.

Lehmann, Erich Leo, Joseph P Romano, and George Casella. 2005. *Testing statistical hypotheses*, Vol. 3, Springer.

Li, Qi and Jeffrey Scott Racine. 2007. *Nonparametric econometrics: theory and practice*, Princeton University Press.

Li, Tong and Bingyu Zhang. 2010. *Testing for affiliation in first-price auctions using entry behavior*, International Economic Review **51**, no. 3, 837–850.

Li, Tong and Xiaoyong Zheng. 2021. *Econometric analyses of auctions: a selective review*, Handbook of Research Methods and Applications in Empirical Microeconomics.

Liu, N and Q Vuong. 2013. *Nonparametric test of monotonicity of bidding strategy in first price auctions*, New York University.

Liu, Nianqing and Yao Luo. 2017. *A nonparametric test for comparing valuation distributions in first-price auctions*, International Economic Review **58**, no. 3, 857–888.

Luo, Yao and Yuanyuan Wan. 2018. *Integrated-quantile-based estimation for first-price auction models*, Journal of Business & Economic Statistics **36**, no. 1, 173–180.

Ma, Jun, Vadim Marmer, Artyom Shneyerov, and Pai Xu. 2021. *Monotonicity-constrained nonparametric estimation and inference for first-price auctions*, Econometric Reviews **40**, no. 10, 944–982.

Marmer, Vadim, Artyom Shneyerov, and Pai Xu. 2013. *What model for entry in first-price auctions? a nonparametric approach*, Journal of Econometrics **176**, no. 1, 46–58.

Marmer, Vadim, Artyom A Shneyerov, and Uma Kaplan. 2016. *Identifying collusion in english auctions*, Available at SSRN 2738789.

Marshall, Robert C and Leslie M Marx. 2007. *Bidder collusion*, Journal of Economic Theory **133**, no. 1, 374–402.

McAfee, R. Preston and John McMillan. 1992. *Bidding rings*, The American Economic Review **82**, no. 3, 579–599.

Montero, Juan-Pablo and Juan Ignacio Guzman. 2010. *Output-expanding collusion in the presence of a competitive fringe*, The Journal of Industrial Economics **58**, no. 1, 106–126.

Ortner, Juan, Kei Kawai, Sylvain Chassang, and Jun Nakabayashi. 2021. *Using bid rotation and incumbency to detect collusion: a regression discontinuity approach*, Review of Economic Studies, forthcoming.

Pesendorfer, Martin. 2000. *A study of collusion in first-price auctions*, The Review of Economic Studies **67**, no. 3, 381–411.

Porter, Robert H. and J. Douglas Zona. 1993. *Detection of bid rigging in procurement auctions*, Journal of Political Economy **101**, no. 3, 518–538.

———. 1999. *Ohio school milk markets: An analysis of bidding*, The RAND Journal of Economics **30**, no. 2, 263–288.

Price, Michael K. 2008. *Using the spatial distribution of bidders to detect collusion in the marketplace: evidence from timber auctions*, Journal of Regional Science **48**, no. 2, 399–417.

Romano, Joseph P and Michael Wolf. 2005. *Stepwise multiple testing as formalized data snooping*, Econometrica **73**, no. 4, 1237–1282.

Schurter, Karl. 2017. *Identification and inference in first-price auctions with collusion*, Penn State Economics Working Paper.

Silverman, Bernard W. 2018. *Density estimation for statistics and data analysis*, Routledge.

Somaini, Paulo. 2020. *Identification in auction models with interdependent costs*, Journal of Political Economy **128**, no. 10, 3820–3871.

Zhang, Shunpu, Rohana J Karunamuni, and M Chris Jones. 1999. *An improved estimator of the density function at the boundary*, Journal of the American Statistical Association **94**, no. 448, 1231–1240.

Appendix A. Proofs of Identification Results

This Appendix presents detailed proofs of arguments in Section 4.

Proof of Lemma 1. Let $B_{\mathcal{K}}$ denote the designated bidder's bid. Since the designated bidder's bid is the highest in the cartel with any collusion scheme, i.e., $B_{\mathcal{K}} = \max_{k\in\mathcal{K}} B_k$, then

$$\sum_{\ell\in\mathcal{N}} H_\ell(b) = \mathbb{P}\left\{\max_{\ell\in\mathcal{N}} B_\ell \le b\right\} = \mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus\mathcal{K}} \{B_{\mathcal{K}}, B_j\} \le b\right\}.$$

By the independence of the designated bidder's bid and non-collusive bidders' bids, for any $b \in [\underline{b}, \overline{b}]$,

$$\sum_{\ell\in\mathcal{N}} H_\ell(b) = \mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus\mathcal{K}} \{B_{\mathcal{K}}, B_j\} \le b\right\} = \mathbb{P}\left\{B_{\mathcal{K}} \le b\right\}\mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus\mathcal{K}} B_j \le b\right\}.$$

By definition, the distribution of the designated bidder's bids is $G_{\mathcal{K}}(b) \equiv \mathbb{P}\left\{\beta_{\mathcal{K}}(V_{\mathcal{K}}) \le b\right\}$, and the distribution of non-collusive bidder's bids is $G_j(b) \equiv \mathbb{P}\left\{\beta_j(V_j) \le b\right\}$ for $j \in \mathcal{N}\setminus\mathcal{K}$. It follows that

$$\sum_{\ell\in\mathcal{N}} H_\ell(b) = \mathbb{P}\left\{\beta_{\mathcal{K}}(V_{\mathcal{K}}) \le b\right\}\mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus\mathcal{K}} \beta_j(V_j) \le b\right\} = G_{\mathcal{K}}(b)\prod_{j\in\mathcal{N}\setminus\mathcal{K}} G_j(b).$$

A non-collusive bidder bids against other non-collusive bidders and the designated bidder in the main auction; therefore, for any $i \in \mathcal{N}\setminus\mathcal{K}$,

$$\begin{aligned} H_i(b) = \mathbb{P}\left\{\max_{\ell\in\mathcal{N}} B_\ell \le B_i, B_i \le b\right\} &= \mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus(\mathcal{K}\cup\{i\})} \{B_j, B_{\mathcal{K}}\} \le B_i, B_i \le b\right\} \\ &= \mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus(\mathcal{K}\cup\{i\})} \{\beta_j(V_j), \beta_{\mathcal{K}}(V_{\mathcal{K}})\} \le \beta_i(V_i), \beta_i(V_i) \le b\right\}, \end{aligned}$$

where the designated bidder's bid is $B_{\mathcal{K}} = \beta_{\mathcal{K}}(V_{\mathcal{K}})$ and the non-collusive bidder j's bid is $B_j = \beta_j(V_j)$. Next, it can be shown that

$$\begin{aligned} \mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus(\mathcal{K}\cup\{i\})} \{\beta_j(V_j), \beta_{\mathcal{K}}(V_{\mathcal{K}})\} \le \beta_i(V_i), \beta_i(V_i) \le b\right\} &= \int_{\underline{b}}^{b} G_{\mathcal{K}}(t)\prod_{j\in\mathcal{N}\setminus(\mathcal{K}\cup\{i\})} G_j(t)\, dG_i(t) \\ &= \int_{\underline{b}}^{b} G_{\mathcal{K}}(t)\prod_{j\in\mathcal{N}\setminus\mathcal{K}} G_j(t)\, d\log G_i(t). \end{aligned}$$

Since $\sum_{\ell\in\mathcal{N}} H_\ell(b) = G_{\mathcal{K}}(b)\prod_{j\in\mathcal{N}\setminus\mathcal{K}} G_j(b)$, then for each non-collusive bidder $i \in \mathcal{N}\setminus\mathcal{K}$ and $b \in [\underline{b}, \overline{b}]$,

$$H_i(b) = \int_{\underline{b}}^{b} \sum_{\ell\in\mathcal{N}} H_\ell(t)\, d\log G_i(t).$$

Taking derivatives on both sides yields

$$dH_i(b) = \sum_{\ell\in\mathcal{N}} H_\ell(b)\, d\log G_i(b),$$

or equivalently

$$\frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(b)} dH_i(b) = d \log G_i(b).$$

With the boundary condition $G_i(\bar{b}) = 1$, integrating on both sides gives

$$\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_i(t) = \int_b^{\bar{b}} d \log G_i(t) = -\log G_i(b).$$

That is, for any non-collusive bidder $i \in \mathcal{N} \setminus \mathcal{K}$,

$$G_i(b) = \exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_i(t) \right).$$

□

Proof of Proposition 2. Let $b \in [\underline{b}, \bar{b}]$. By Leibniz integral rule, for any bidder $j \in \mathcal{N}$, define

$$m_j(b) \equiv \frac{d}{db} M_j(b) = \frac{H_j'(b)}{\sum_{\ell \in \mathcal{N}} H_\ell(b)} \exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_j(t) \right),$$

where $H_j'(b) \equiv d/db H_j(b)$. It can be proved that, for any collusion scheme,

$$\frac{g_{\mathcal{K}}(b)}{G_{\mathcal{K}}(b)} = \sum_{k \in \mathcal{K}} \frac{m_k(b)}{M_k(b)}.$$

To see this, notice that the distribution of the highest bid is equal to the product of distributions of the designated bidder's bids and distributions of all other non-collusive competitors' bids, i.e.,

$$\mathbb{P}\left\{ \max_{\ell \in \mathcal{N}} B_\ell \le b \right\} = \sum_{\ell \in \mathcal{N}} H_\ell(b) = G_{\mathcal{K}}(b) \prod_{j \in \mathcal{N} \setminus \mathcal{K}} G_j(b).$$

It follows that

(A1)
$$\frac{\sum_{\ell \in \mathcal{N}} H_\ell'(b)}{\sum_{\ell \in \mathcal{N}} H_\ell(b)} = \frac{\frac{d}{db} G_{\mathcal{K}}(b) \prod_{j \in \mathcal{N} \setminus \mathcal{K}} G_j(b)}{G_{\mathcal{K}}(b) \prod_{j \in \mathcal{N} \setminus \mathcal{K}} G_j(b)} = \frac{g_{\mathcal{K}}(b)}{G_{\mathcal{K}}(b)} + \sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(b)}{G_j(b)} = \frac{g_{\mathcal{K}}(b)}{G_{\mathcal{K}}(b)} + \sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{m_j(b)}{M_j(b)},$$

where $G_j(\cdot) = M_j(\cdot)$ and $g_j(\cdot) = m_j(\cdot)$ for $j \in \mathcal{N} \setminus \mathcal{K}$ by Lemma 1. On the other hand,

$$\sum_{j \in \mathcal{N}} \frac{m_j(b)}{M_j(b)} = \sum_{j \in \mathcal{N}} \frac{\frac{H_j'(b)}{\sum_{\ell \in \mathcal{N}} H_\ell(b)} \exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_j(t) \right)}{\exp\left(-\int_b^{\bar{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_j(t) \right)} = \frac{\sum_{\ell \in \mathcal{N}} H_\ell'(b)}{\sum_{\ell \in \mathcal{N}} H_\ell(b)}. \tag{A2}$$

Equalities (A1) and (A2) imply

$$\sum_{j \in \mathcal{N}} \frac{m_j(b)}{M_j(b)} = \frac{g_{\mathcal{K}}(b)}{G_{\mathcal{K}}(b)} + \sum_{i \in \mathcal{N} \setminus \mathcal{K}} \frac{m_i(b)}{M_i(b)},$$

or equivalently

$$\frac{g_{\mathcal{K}}(b)}{G_{\mathcal{K}}(b)} = \sum_{k \in \mathcal{K}} \frac{m_k(b)}{M_k(b)}. \tag{A3}$$

For non-collusive bidder $i \in \mathcal{N} \setminus \mathcal{K}$, the inverse bidding strategy can be written as

$$\begin{aligned}\beta_i^{-1}(b) &\equiv b + \frac{1}{\sum_{j \in \mathcal{N} \setminus (\{i\} \cup \mathcal{K})} \frac{g_j(b)}{G_j(b)} + \frac{g_{\mathcal{K}}(b)}{G_{\mathcal{K}}(b)}} \\ &= b + \frac{1}{\sum_{j \in \mathcal{N} \setminus (\{i\} \cup \mathcal{K})} \frac{m_j(b)}{M_j(b)} + \sum_{k \in \mathcal{K}} \frac{m_k(b)}{M_k(b)}},\end{aligned}$$

where the second equality holds by Lemma 1 and equation (A3). Using the fact that $m_j(\cdot)/M_j(\cdot) = H_j'(\cdot)/\sum_{\ell \in \mathcal{N}} H_\ell(\cdot)$ for all $j \in \mathcal{N}$, simple algebra yields

$$\beta_i^{-1}(b) = b + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(b)}{\sum_{\ell \in \mathcal{N} \setminus \{i\}} H_\ell'(b)},$$

which proves the first statement. Identification of the non-collusive bidder's inverse value distribution immediately follows from equation 7. From the strict monotonicity of inverse bidding strategy $\beta_i^{-1}(\cdot)$ and Lemma 1, for $i \in \mathcal{N} \setminus \mathcal{K}$ and $\alpha \in (0, 1]$,

$$F_i^{-1}(\alpha) = \beta_i^{-1}(G_i^{-1}(\alpha)) = M_i^{-1}(\alpha) + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(M_i^{-1}(\alpha))}{\sum_{\ell \in \mathcal{N} \setminus \{i\}} H_\ell'(M_i^{-1}(\alpha))}.$$

When $\alpha = 0$, $F_i^{-1}(0) = \underline{v} = \underline{b} = M_i^{-1}(0)$ by the boundary conditions. □

Before the formal proof of Theorem 1, two intermediate results are presented.

Lemma A1. *Suppose a direct collusion scheme Γ with absolutely continuous value distribution functions $\{F_k(\cdot)\}_{k \in \mathcal{K}}$ satisfies Assumption 1. There exists a unique profile of absolutely continuous value distribution functions $\{F_k^*(\cdot)\}_{k \in \mathcal{K}}$ such that under an efficient collusion scheme Γ_e, for all $k \in \mathcal{K}$ and $v \in [\underline{v}, \overline{v}]$,*

$$\mathbb{P}^* \{V_k \leq v, I_{\mathcal{K}} = k; \Gamma_e\} = \mathbb{P} \{V_k \leq v, I_{\mathcal{K}} = k; \Gamma\},$$

where $\mathbb{P}^*$ and $\mathbb{P}$ are probability measures under distribution functions $\{F_k^*(\cdot)\}_{k \in \mathcal{K}}$ and $\{F_k(\cdot)\}_{k \in \mathcal{K}}$ respectively.

Proof of Lemma A1. Given a direct collusion scheme Γ and a profile of collusive bidders' value distributions $\{F_k(\cdot)\}_{k \in \mathcal{K}}$, recall that value distribution of the designated bidder is given by

$$F_{\mathcal{K}}(v; \Gamma) \equiv \mathbb{P} \{V_{\mathcal{K}} \leq v; \Gamma\} = \sum_{k \in \mathcal{K}} \mathbb{P} \{V_{\mathcal{K}} \leq v, I_{\mathcal{K}} = k; \Gamma\} = \sum_{k \in \mathcal{K}} \mathbb{P} \{V_k \leq v, I_{\mathcal{K}} = k; \Gamma\}.$$

Under an efficient collusion scheme Γ_e, by the independence of private value,

$$\mathbb{P}^* \{V_k \leq v, I_{\mathcal{K}} = k; \Gamma_e\} = \mathbb{P}^* \left\{V_k \leq v, \max_{\ell \in \mathcal{K} \setminus \{k\}} V_\ell < V_k\right\} = \int_{\underline{v}}^{v} \prod_{\ell \in \mathcal{K} \setminus \{k\}} F_\ell^*(t)\, dF_k^*(t).$$

for any $k \in \mathcal{K}$ and $v \in (\underline{v}, \overline{v}]$. Set

$$p_k(v; \Gamma) \equiv \mathbb{P} \{V_k \leq v, I_{\mathcal{K}} = k; \Gamma\} = \mathbb{P}^* \{V_k \leq v, I_{\mathcal{K}} = k; \Gamma_e\}.$$

Differentiating on both sides of above equality with respect to v leads to a K-dimensional system of non-autonomous first-order differential equations:

$$\begin{aligned} f_1^*(v) \prod_{\ell \in \mathcal{K} \setminus \{1\}} F_\ell^*(v) &= p_1'(v;\Gamma), \\ f_2^*(v) \prod_{\ell \in \mathcal{K} \setminus \{2\}} F_\ell^*(v) &= p_2'(v;\Gamma), \\ &\vdots \\ f_K^*(v) \prod_{\ell \in \mathcal{K} \setminus \{K\}} F_\ell^*(v) &= p_K'(v;\Gamma). \end{aligned} \tag{A4}$$

Note that

$$F_{\mathcal{K}}'(v;\Gamma) = \sum_{k \in \mathcal{K}} p_k'(v;\Gamma) = \sum_{k \in \mathcal{K}} \left(\prod_{\ell \in \mathcal{K} \setminus \{k\}} F_\ell^*(v) f_k^*(v) \right) = \left(\prod_{k \in \mathcal{K}} F_k^*(v) \right)',$$

or equivalently

$$F_{\mathcal{K}}(v;\Gamma) = \prod_{k \in \mathcal{K}} F_k^*(v). \tag{A5}$$

The system of equations (A4) and (A5) imply that for any $k \in \mathcal{K}$,

$$\frac{f_k^*(v)}{F_k^*(v)} = \frac{p_k'(v;\Gamma)}{F_{\mathcal{K}}(v;\Gamma)}.$$

Set the initial value condition $F_k^*(\overline{v}) = 1$. Integrating on both sides from v to $\overline{v}$ yields

$$\int_v^{\overline{v}} \frac{p_k'(t;\Gamma)}{F_{\mathcal{K}}(t;\Gamma)}\, dt = \int_v^{\overline{v}} \frac{f_k^*(t)}{F_k^*(t)}\, dt = -\log F_k^*(v).$$

Solving for F_k^* gives

$$F_k^*(v) = \exp\left(- \int_v^{\overline{v}} \frac{p_k'(t;\Gamma)}{F_{\mathcal{K}}(t;\Gamma)}\, dt \right).$$

Hence, the unique set of value distribution functions $\{F_k^*(\cdot)\}_{k \in \mathcal{K}}$ satisfy the desired property. □

Lemma A1 leads to the following proposition.

Proposition A1. *Under Assumption 1 and 2, if a set of collusive bidders $\mathcal{K}$ with absolutely continuous private value distribution functions $\{F_k(\cdot)\}_{k \in \mathcal{K}}$ form a cartel and adopt a collusion scheme Γ, then for all $k \in \mathcal{K}$ and $b \in [\underline{b}, \overline{b}]$,*

$$M_k(b) \equiv \exp\left(- \int_b^{\overline{b}} \frac{1}{\sum_{\ell \in \mathcal{N}} H_\ell(t)}\, dH_k(t) \right) = F_k^*(\beta_{\mathcal{K}}^{-1}(b)),$$

where $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$ is distribution functions induced from value distribution functions $\{F_k(\cdot)\}_{k\in\mathcal{K}}$ and the collusion scheme Γ by Lemma A1, $\beta_{\mathcal{K}}^{-1}(\cdot)$ is the inverse bidding strategy of the designated bidder, and $H_j(\cdot)$ is a function defined with the winning bid $B_j = \max_{\ell\in\mathcal{N}} B_\ell$ such that for $j \in \mathcal{N}$,

$$H_j(b) \equiv \mathbb{P}\left\{B_j = \max_{\ell\in\mathcal{N}} B_\ell, B_j \leq b\right\}.$$

Proof of Proposition A1. For any collusive bidder $k \in \mathcal{K}$, the event of winning the auction is equivalent to the event of being selected as the designated bidder and submitting a bid higher than that of all non-collusive competitors; thus, for each $k \in \mathcal{K}$ and $b \in [\underline{b}, \overline{b}]$,

$$H_k(b) = \mathbb{P}\left\{\max_{j\in\mathcal{N}\setminus\{k\}} B_j \leq B_k, B_k \leq b\right\} = \mathbb{P}\left\{\max_{i\in\mathcal{N}\setminus\mathcal{K}} B_i \leq B_k, I_{\mathcal{K}} = k, B_k \leq b\right\}.$$

Note that $B_i = \beta_i(V_i)$ for $i \in \mathcal{N}\setminus\mathcal{K}$ and $B_k = \beta_{\mathcal{K}}(V_k)$ for $I_{\mathcal{K}} = k \in \mathcal{K}$. It follows from the absolute continuity of the probability measure $\mathbb{P}\{V_k \leq v | I_{\mathcal{K}} = k\}$ that

$$\begin{aligned}
H_k(b) &= \mathbb{P}\left\{\max_{i\in\mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), I_{\mathcal{K}} = k, \beta_{\mathcal{K}}(V_k) \leq b\right\} \\
&= \mathbb{P}\{I_{\mathcal{K}} = k\}\,\mathbb{P}\left\{\max_{i\in\mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), \beta_{\mathcal{K}}(V_k) \leq b | I_{\mathcal{K}} = k\right\} \\
&= \mathbb{P}\{I_{\mathcal{K}} = k\}\,\mathbb{P}\left\{\max_{i\in\mathcal{N}\setminus\mathcal{K}} \beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq V_k, V_k \leq \beta_{\mathcal{K}}^{-1}(b) | I_{\mathcal{K}} = k\right\} \\
&= \mathbb{P}\{I_{\mathcal{K}} = k\} \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i\in\mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k\right\} dF_k(t | I_{\mathcal{K}} = k).
\end{aligned}$$

Since the non-collusive bidder's private value is independent of the collusive scheme, it follows that

$$H_k(b) = \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i\in\mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t\right\} \frac{d}{dt}\mathbb{P}\{V_k \leq t, I_{\mathcal{K}} = k\}\, dt.$$

By Lemma A1, there exists a unique set of distribution functions $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$ such that

$$\frac{d}{dt}\mathbb{P}\{V_k \leq t, I_{\mathcal{K}} = k\} = \prod_{\ell\in\mathcal{K}\setminus\{k\}} F_\ell^*(t) f_k^*(t),$$

for each $k \in \mathcal{K}$. Therefore,

$$H_k(b) = \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i\in\mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t\right\} \prod_{\ell\in\mathcal{K}\setminus\{k\}} F_\ell^*(t) f_k^*(t)\, dt.$$

Let $s = \beta_{\mathcal{K}}(t)$; thus, $ds = \beta'_{\mathcal{K}}(t)dt$. The change of variables leads to

$$\begin{aligned}
H_k(b) &= \int_{\underline{b}}^{b} \prod_{i\in\mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq \beta_{\mathcal{K}}^{-1}(s)\right\} \prod_{\ell\in\mathcal{K}\setminus\{k\}} F_\ell^*(\beta_{\mathcal{K}}^{-1}(s)) \frac{f_k^*(\beta_{\mathcal{K}}^{-1}(s))}{\beta'_{\mathcal{K}}(\beta_{\mathcal{K}}^{-1}(s))}\, ds \\
&= \int_{\underline{b}}^{b} \prod_{i\in\mathcal{N}\setminus\mathcal{K}} \mathbb{P}\{\beta_i(V_i) \leq s\} \prod_{\ell\in\mathcal{K}\setminus\{k\}} F_\ell^*(\beta_{\mathcal{K}}^{-1}(s))\, dF_k^*(\beta_{\mathcal{K}}^{-1}(s)).
\end{aligned}$$

Let $G_k^*(\cdot) \equiv F_k^*(\beta_{\mathcal{K}}^{-1}(\cdot)) = \mathbb{P}^*(V_k \leq \beta_{\mathcal{K}}^{-1}(\cdot))$; hence,

$$\begin{aligned} H_k(b) &= \int_{\underline{b}}^{b} \prod_{i\in\mathcal{N}\setminus\mathcal{K}} G_i(s) \prod_{\ell\in\mathcal{K}\setminus\{k\}} G_\ell^*(s)\, dG_k^*(s) \\ &= \int_{\underline{b}}^{b} \prod_{k\in\mathcal{K}} G_k^*(s) \prod_{i\in\mathcal{N}\setminus\mathcal{K}} G_i(s)\, d\log G_k^*(s) \\ &= \int_{\underline{b}}^{b} G_{\mathcal{K}}(s) \prod_{i\in\mathcal{N}\setminus\mathcal{K}} G_i(s)\, d\log G_k^*(s) \\ &= \int_{\underline{b}}^{b} \sum_{\ell\in\mathcal{N}} H_\ell(s)\, d\log G_k^*(s), \end{aligned}$$

where $G_i(\cdot) \equiv \mathbb{P}\{\beta_i(V_i) \leq \cdot\}$ for any non-collusive bidder $i \in \mathcal{N}\setminus\mathcal{K}$, and $G_{\mathcal{K}}(\cdot) = \prod_{k\in\mathcal{K}} G_k^*(\cdot)$ holds in the third equality by equation (A5). Solving for G_k^* gives the desired result: for each $k \in \mathcal{K}$ and $b \in [\underline{b}, \overline{b}]$,

$$G_k^*(b) = \exp\left(-\int_b^{\overline{b}} \frac{1}{\sum_{\ell\in\mathcal{N}} H_\ell(t)}\, dH_k(t)\right) = M_k(b).$$

This completes the proof. □

Proof of Theorem 1. Conditional on $Z_t = z$, construct the inverse pseudo-value distribution for each bidder $j \in \mathcal{N}$ by

$$U_j^{-1}(\alpha; z) \equiv \begin{cases} M_j^{-1}(\alpha; z) + \frac{\sum_{\ell\in\mathcal{N}} H_\ell(M_j^{-1}(\alpha;z);z)}{\sum_{\ell\in\mathcal{N}\setminus\{j\}} H_\ell'(M_j^{-1}(\alpha;z);z)} & \text{for } \alpha \in (0,1]; \\ \underline{b} & \text{for } \alpha = 0, \end{cases}$$

where

$$M_j(b; z) \equiv \exp\left(-\int_b^{\overline{b}} \frac{1}{\sum_{\ell\in\mathcal{N}} H_\ell(t; z)}\, dH_j(t; z)\right).$$

It suffices to show that only non-collusive bidder's pseudo-value distributions are independent of instruments. That is, for any $i \in \mathcal{N}\setminus\mathcal{K}$, $\alpha \in (0,1]$ and $z \neq z'$,

$$U_i^{-1}(\alpha; z) = U_i^{-1}(\alpha; z');$$

but for every collusive bidder $k \in \mathcal{K}$, there exists $\mathcal{A} \subset (0,1]$ with a strictly positive Lebesgue measure, and $z \neq z'$ such that

$$U_k^{-1}(\alpha; z) \neq U_k^{-1}(\alpha; z')$$

for all $\alpha \in \mathcal{A}$. By Proposition 2 and Assumption 3(ii), non-collusive bidder's inverse pseudo-value distributions are independent of instruments since for $i \in \mathcal{N}\setminus\mathcal{K}$,

$$U_i^{-1}(\alpha; z) = F_i^{-1}(\alpha; z) = F_i^{-1}(\alpha; z') = U_i^{-1}(\alpha; z').$$

Conditional on $Z_t = z$, the inverse pseudo-value distribution for collusive bidder $k \in \mathcal{K}$ can be written as

$$U_k^{-1}(\alpha; z) \equiv M_k^{-1}(\alpha; z) + \frac{\sum_{\ell \in \mathcal{N}} H_\ell(M_k^{-1}(\alpha; z); z)}{\sum_{\ell \in \mathcal{N} \setminus \{k\}} H'_\ell(M_k^{-1}(\alpha; z); z)}$$
$$= M_k^{-1}(\alpha; z) + \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)} + \sum_{\ell \in \mathcal{K} \setminus \{k\}} \frac{m_\ell(M_k^{-1}(\alpha;z);z)}{M_\ell(M_k^{-1}(\alpha;z);z)}},$$

The inverse pseudo-value distribution can be expressed as a summation of two quantities $L_{k,1}(\alpha; Z)$ and $L_{k,2}(\alpha; z)$, i.e.,

$$U_k^{-1}(\alpha; z) = L_{k,1}(\alpha; z) + L_{k,2}(\alpha; z), \tag{A6}$$

where

$$L_{k,1}(\alpha; z) \equiv M_k^{-1}(\alpha; z) + \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)}},$$
$$L_{k,2}(\alpha; z) \equiv \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)} + \sum_{\ell \in \mathcal{K} \setminus \{k\}} \frac{m_\ell(M_k^{-1}(\alpha;z);z)}{M_\ell(M_k^{-1}(\alpha;z);z)}} - \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)}}.$$

The proof consists of two steps: first, $L_{k,1}(\alpha; z)$ is invariant to instruments; second, $L_{k,2}(\alpha; z)$ depends non-trivially on instruments. Let $V_k^*(\alpha; z) \equiv \beta_{\mathcal{K}}^{-1}(M_k^{-1}(\alpha; z); z)$ for any collusive bidder $k \in \mathcal{K}$. The increasing monotonicity of $\beta_{\mathcal{K}}^{-1}(\cdot; z)$ and $M_k^{-1}(\cdot; z)$ implies that V_k^* is the inverse of CDF F_k^*. To see this, by Proposition A1, for $\alpha \in (0, 1]$,

$$\alpha = M_k(M_k^{-1}(\alpha)) = F_k^*(\beta_{\mathcal{K}}^{-1}(M_k^{-1}(\alpha; z))) = F_k^*(V_k^*(\alpha; z)).$$

By definition,

$$V_k^*(\alpha; z) = \beta_{\mathcal{K}}^{-1}(M_k^{-1}(\alpha; z); z) = M_k^{-1}(\alpha; z) + \frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)}} = L_{k,1}(\alpha; z), \tag{A7}$$

which is invariant to instruments by Proposition A1 and Assumption 3(ii). Hence, equalities (A6) and (A7) imply that

$$U_k^{-1}(\alpha; z) = L_{k,1}(\alpha; z) + L_{k,2}(\alpha; z) = V_k^*(\alpha; z) + L_{k,2}(\alpha; z).$$

Next, it is observed that

$$\frac{1}{\sum_{j \in \mathcal{N} \setminus \mathcal{K}} \frac{g_j(M_k^{-1}(\alpha;z);z)}{G_j(M_k^{-1}(\alpha;z);z)}} = V_k^*(\alpha; z) - M_k^{-1}(\alpha; z) = V_k^*(\alpha; z) - \beta_{\mathcal{K}}(V_k^*(\alpha; z); z) > 0,$$

and for $\ell \in \mathcal{K}$,

$$m_\ell(b; z) = \frac{d}{db} M_\ell(b; z) = \frac{d}{db} F_\ell^*(\beta_{\mathcal{K}}^{-1}(b; z); z) = \frac{1}{\beta'_{\mathcal{K}}(\beta_{\mathcal{K}}^{-1}(b; z); z)} f_\ell^*(\beta_{\mathcal{K}}^{-1}(b; z); z).$$

It follows that

$$U_k^{-1}(\alpha;z) = V_k^*(\alpha;z) + \frac{1}{\left[V_k^*(\alpha;z) - \beta_{\mathcal{K}}(V_k^*(\alpha;z);z)\right]^{-1} + \frac{1}{\beta'_{\mathcal{K}}(V_k^*(\alpha;z);z)} \sum_{\ell \in \mathcal{K}\setminus\{k\}} \frac{f_\ell^*(V_k^*(\alpha;z);z)}{F_\ell^*(V_k^*(\alpha;z);z)}} - \frac{1}{\left[V_k^*(\alpha;z) - \beta_{\mathcal{K}}(V_k^*(\alpha;z);z)\right]^{-1}}.$$

Instruments do not affect the collusion scheme and collusive bidders' value distribution by Assumption 3(i), which implies $V_k^*(\alpha;z) = V_k^*(\alpha;z')$ for all $z \neq z'$. The difference between the inverse pseudo-value distributions under different levels of instruments is

$$\begin{aligned} & U_k^{-1}(\alpha;z) - U_k^{-1}(\alpha;z') \\ &= \left\{ \frac{1}{\left[V_k^*(\alpha) - \beta_{\mathcal{K}}(V_k^*(\alpha);z)\right]^{-1} + \frac{1}{\beta'_{\mathcal{K}}(V_k^*(\alpha);z)} \sum_{\ell \in \mathcal{K}\setminus\{k\}} \frac{f_\ell^*(V_k^*(\alpha))}{F_\ell^*(V_k^*(\alpha))}} - \frac{1}{\left[V_k^*(\alpha) - \beta_{\mathcal{K}}(V_k^*(\alpha);z)\right]^{-1}} \right\} \\ &- \left\{ \frac{1}{\left[V_k^*(\alpha) - \beta_{\mathcal{K}}(V_k^*(\alpha);z')\right]^{-1} + \frac{1}{\beta'_{\mathcal{K}}(V_k^*(\alpha);z')} \sum_{\ell \in \mathcal{K}\setminus\{k\}} \frac{f_\ell^*(V_k^*(\alpha))}{F_\ell^*(V_k^*(\alpha))}} - \frac{1}{\left[V_k^*(\alpha) - \beta_{\mathcal{K}}(V_k^*(\alpha);z')\right]^{-1}} \right\}. \end{aligned}$$

An important feature of $U_k^{-1}(\alpha;z)$ is that, for a fixed α, $U_k^{-1}(\alpha;z)$ is increasing in both $\beta_{\mathcal{K}}(V_k^*(\alpha);z)$ and $\beta'_{\mathcal{K}}(V_k^*(\alpha);z)$. Since $V_k^*(0) = \beta_{\mathcal{K}}^{-1}(M_k^{-1}(0)) = \beta_{\mathcal{K}}^{-1}(\underline{b}) = \underline{v}$, $\beta_{\mathcal{K}}(V_k^*(0);z') = \beta_{\mathcal{K}}(V_k^*(0);z)$. Since BNE strategies depend non-trivially on instruments Z_t by Assumption 3(ii), there must exist some α such that $\beta_{\mathcal{K}}(V_k^*(\alpha);z) > \beta_{\mathcal{K}}(V_k^*(\alpha);z')$ and $\beta'_{\mathcal{K}}(V_k^*(\alpha);z) > \beta'_{\mathcal{K}}(V_k^*(\alpha);z')$ for some $z \neq z'$; therefore, $U_k^{-1}(\alpha;z) \neq U_k^{-1}(\alpha;z')$, which completes the proof. □

Proof of Theorem 2. By Lemma A1, there exists a unique profile of absolutely continuous value distribution functions $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$ such that under an efficient collusion scheme Γ_e, for all $k \in \mathcal{K}$ and $v \in [\underline{v}, \overline{v}]$,

$$\mathbb{P}^*\{V_k \le v, I_{\mathcal{K}} = k; \Gamma_e\} = \mathbb{P}\{V_k \le v, I_{\mathcal{K}} = k; \Gamma\},$$

where $\mathbb{P}^*$ and $\mathbb{P}$ are probability measures under distribution functions $\{F_k^*(\cdot)\}_{k\in\mathcal{K}}$ and $\{F_k(\cdot)\}_{k\in\mathcal{K}}$ respectively. Let $v = \overline{v}$, then

$$\mathbb{P}^*\{I_{\mathcal{K}} = k; \Gamma_e\} = \mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\}.$$

which means each collusive bidder's probability of being selected as the designated bidder is the same under two collusion structures. It suffices to show that given value distributions of non-collusive bidders, two collusion structures $(\{F_k^*(\cdot)\}_{k\in\mathcal{K}}, \Gamma_e)$ and $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ induce the same cumulative incidence functions in equilibrium, i.e., for every $j \in \mathcal{N}$,

$$H_j(b;\Gamma) = H_j^*(b;\Gamma_e).$$

Since two structures $(\{F_k(\cdot)\}_{k\in\mathcal{K}}, \Gamma)$ and $(\{F_k^*(\cdot)\}_{k\in\mathcal{K}}, \Gamma_e)$ have the same value distribution of the designated bidder, the equilibrium bidding strategies of both non-collusive bidders and the designated bidder are the same under these two collusion structures. Thus, for non-collusive bidder $i \in \mathcal{N} \setminus \mathcal{K}$,

$$
\begin{aligned}
H_i(b;\Gamma) &= \mathbb{P}\left\{\max_{\ell \in \mathcal{N}} B_\ell \leq B_i, B_i \leq b; \Gamma\right\} \\
&= \mathbb{P}\left\{\max_{j \in \mathcal{N}\setminus(\mathcal{K}\cup\{i\})} \{B_j, B_{\mathcal{K}}\} \leq B_i, B_i \leq b; \Gamma\right\} \\
&= \mathbb{P}\left\{\max_{j \in \mathcal{N}\setminus(\mathcal{K}\cup\{i\})} \{\beta_j(V_j), \beta_{\mathcal{K}}(V_{\mathcal{K}})\} \leq \beta_i(V_i), \beta_i(V_i) \leq b; \Gamma\right\}, \\
&= \mathbb{P}^*\left\{\max_{j \in \mathcal{N}\setminus(\mathcal{K}\cup\{i\})} \{\beta_j(V_j), \beta_{\mathcal{K}}(V_{\mathcal{K}})\} \leq \beta_i(V_i), \beta_i(V_i) \leq b; \Gamma_e\right\} \\
&= H_i^*(b;\Gamma_e),
\end{aligned}
$$

where the second last equality holds since the equilibrium bidding strategies of both non-collusive bidders and the designated bidder are the same under two collusion structures. For each collusive bidder $k \in \mathcal{K}$,

$$
\begin{aligned}
H_k(b;\Gamma) &= \mathbb{P}\left\{\max_{j \in \mathcal{N}\setminus\{k\}} B_j \leq B_k, B_k \leq b; \Gamma\right\} \\
&= \mathbb{P}\left\{\max_{i \in \mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), I_{\mathcal{K}} = k, \beta_{\mathcal{K}}(V_k) \leq b; \Gamma\right\} \\
&= \mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\}\, \mathbb{P}\left\{\max_{i \in \mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), \beta_{\mathcal{K}}(V_k) \leq b | I_{\mathcal{K}} = k; \Gamma\right\} \\
&= \mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\} \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i \in \mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k; \Gamma\right\} dF_k(t | I_{\mathcal{K}} = k; \Gamma).
\end{aligned}
$$

Since non-collusive bidders' value distributions are independent of the collusion scheme, then for $i \in \mathcal{N} \setminus \mathcal{K}$,

$$
\mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k; \Gamma\right\} = \mathbb{P}^*\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k; \Gamma_e\right\}.
$$

Using the fact that $\mathbb{P}^*\{V_k \leq v, I_{\mathcal{K}} = k; \Gamma_e\} = \mathbb{P}\{V_k \leq v, I_{\mathcal{K}} = k; \Gamma\}$ and $\mathbb{P}^*\{I_{\mathcal{K}} = k; \Gamma_e\} = \mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\}$, for any collusive bidder $k \in \mathcal{K}$,

$$
\begin{aligned}
H_k(b;\Gamma) &= \mathbb{P}\{I_{\mathcal{K}} = k; \Gamma\} \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i \in \mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k; \Gamma\right\} dF_k(t | I_{\mathcal{K}} = k; \Gamma) \\
&= \mathbb{P}^*\{I_{\mathcal{K}} = k; \Gamma_e\} \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i \in \mathcal{N}\setminus\mathcal{K}} \mathbb{P}^*\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k; \Gamma_e\right\} dF_k^*(t | I_{\mathcal{K}} = k; \Gamma_e) \\
&= \mathbb{P}^*\left\{\max_{i \in \mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), I_{\mathcal{K}} = k, \beta_{\mathcal{K}}(V_k) \leq b; \Gamma_e\right\} \\
&= H_k^*(b;\Gamma_e),
\end{aligned}
$$

which implies that the cumulative incidence functions of collusive bidders are the same under two collusion structures. □

Proof of Corollary 1. By Theorem 1, identities of collusive bidders are identified from winning bids, identities of winners, and auction-level instruments. The value distributions of non-collusive bidders are identified by Proposition 2. It is left to show that the value distributions of collusive bidders are nonparametrically identified. It is observed that for any $k \in \mathcal{K}$,

$$\begin{aligned}
H_k(b) &= \mathbb{P}\left\{\max_{i \in \mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), I_{\mathcal{K}} = k, \beta_{\mathcal{K}}(V_k) \leq b\right\} \\
&= \mathbb{P}\left\{I_{\mathcal{K}} = k\right\} \mathbb{P}\left\{\max_{i \in \mathcal{N}\setminus\mathcal{K}} \beta_i(V_i) \leq \beta_{\mathcal{K}}(V_k), \beta_{\mathcal{K}}(V_k) \leq b | I_{\mathcal{K}} = k\right\} \\
&= \mathbb{P}\left\{I_{\mathcal{K}} = k\right\} \int_{\underline{v}}^{\beta_{\mathcal{K}}^{-1}(b)} \prod_{i \in \mathcal{N}\setminus\mathcal{K}} \mathbb{P}\left\{\beta_{\mathcal{K}}^{-1}(\beta_i(V_i)) \leq t | I_{\mathcal{K}} = k\right\} dF_k(t | I_{\mathcal{K}} = k).
\end{aligned}$$

Since the noncollusive bidder's private value is independent of the collusive scheme, for any $b \in [\underline{b}, \overline{b}]$,

$$H_k(b) = \mathbb{P}\left\{I_{\mathcal{K}} = k\right\} \int_{\underline{b}}^{b} \prod_{i \in \mathcal{N}\setminus\mathcal{K}} G_i(t)\, dG_k(t | I_{\mathcal{K}} = k).$$

With the initial condition $G_k(\underline{b} | I_{\mathcal{K}} = k) = 0$, it follows that

$$G_k(b | I_{\mathcal{K}} = k) = \frac{1}{\mathbb{P}\left\{I_{\mathcal{K}} = k\right\}} \int_{\underline{b}}^{b} \frac{1}{\prod_{i \in \mathcal{N}\setminus\mathcal{K}} G_i(b)} dH_k(b).$$

Note that non-collusive bidders' bid distributions $\{G_i(\cdot)\}_{i \in \mathcal{N}\setminus\mathcal{K}}$ are also identified, and the probability of collusive bidder k being selected as the designated bidder is identified from

$$\mathbb{P}\left\{I_{\mathcal{K}} = k\right\} = \int_{\underline{v}}^{\overline{v}} \prod_{\ell \in \mathcal{K}\setminus\{k\}} F_\ell^*(t)\, dF_k^*(t),$$

where $F_\ell^{*-1}(\alpha) = \beta_{\mathcal{K}}^{-1}(M_k^{-1}(\alpha))$. If the designated bidder is randomly selected, $\mathbb{P}\left\{V_k \leq v\right\} = \mathbb{P}\left\{V_k \leq v | I_{\mathcal{K}} = k\right\}$. It follows that for any $k \in \mathcal{K}$,

$$G_k(b) = \mathbb{P}\left\{V_k \leq \beta_{\mathcal{K}}^{-1}(b)\right\} = \mathbb{P}\left\{V_k \leq \beta_{\mathcal{K}}^{-1}(b) | I_{\mathcal{K}} = k\right\} = G_k(b | I_{\mathcal{K}} = k).$$

Identification of collusive bidders' value distributions follows from equation (8). □

Appendix B. Caltrans

Figure 2. California Department of Transportation District Map.

Neighboring districts are grouped as one region as follows. A top local firm has a market share higher than 5% and a win rate higher than 1% in the district. The set of top local firms in district d is denoted by $j \in \mathcal{J}_d$. Top local firms in one district may also be active in other districts. For a top local firm $j \in \mathcal{J}_d$ in district d, the set of active districts is denoted by $D_{j,d}$. To characterize the participation behavior of top local firms in a district, define

$$D_d = \left\{ d' : | \bigcup_{j \in \mathcal{J}_d, d' \in D_{j,d}} j | / |\mathcal{J}_d| > 1/2 \right\}.$$

D_d is a set of districts in which most top local firms in district d are active.

Then group two neighboring districts d_1 and d_2 if

$$\text{dist}(D_{d_1}, D_{d_2}) \equiv \inf \left\{ d' - d : d' \in D_{d_1}, d \in D_{d_2} \right\} = 0.$$

Define $D_{d_1,d_2} = D_{d_1} \cup D_{d_2}$ if districts d_1 and d_2 are grouped. Continuous to group other neighboring district with D_{d_1,d_2}. Finally, there are four regions in the sample: districts 1 and 2 are region 1; districts 4 and 5 are region 2; districts 3, 6, 9, and 10 are region 3; and districts 7, 8, 11, and 12 are region 4.