Testing for Collusion in Asymmetric First-Price Auctions

Gaurab Aryal and Maria F. Gabrielli

ANU Working Papers in Economics and Econometrics
No. 564

November, 2011 JEL: C1, C4, C7, D4, L4.

ISBN: 086831 564 8

TESTING FOR COLLUSION IN ASYMMETRIC FIRST-PRICE AUCTIONS[1]

Gaurab Aryal and Maria F. Gabrielli[†,‡]

† The Australian National University, gaurab.aryal@anu.edu.au

‡ CONICET, Universidad Nacional de Cuyo, (corresponding author): florgabrielli@gmail.com

[1]The paper supersedes the paper "Detecting Collusion on Highway Procurement," by Gabrielli. Gabrielli thanks Joris Pinske for insights and comments on the earlier draft. Authors also thank the two referees and the audiences in ANU seminar, for their help. The usual caveats apply.

Testing for Collusion in Asymmetric First-Price Auctions

GAURAB ARYAL AND MARIA F. GABRIELLI

ABSTRACT. This paper proposes fully nonparametric tests to detect possible collusion in first-price procurement (auctions). The aim of the tests is to detect possible collusion before knowing whether or not bidders are colluding. Thus we do not rely on data on anti-competitive hearing, and in that sense is 'ex-ante'. We propose a two steps (model selection) procedure: First, we use a reduced form test of independence and symmetry to short-list bidders whose bidding behavior is at-odds with competitive bidding, and Second, the recovered (latent) cost for these bidders must be higher under collusion than under competition, because collusion dwarfs competition, hence detecting collusion boils down to testing if the estimated cost distribution under collusion first order stochastically dominates that under competition. We propose rank based and Kolmogorov-Smirnov (K-S) tests. We implement the tests for Highway Procurement data in California and conclude that there is no evidence of collusion even though the reduced form test supports collusion.

Keywords: Asymmetric Auctions; Collusion; Nonparametric Testing.
JEL: C1, C4, C7, D4, L4.

1. INTRODUCTION

Auction is the most widely used selling mechanism for both private and public goods. For example, federal government is the biggest auctioneer in the U.S., it sells the offshore oil leases, timber from national forests and construction/ highway projects through auction. The assets of bankrupt businesses are usually liquidated by means of an auction. However, auctions are susceptive to bid rigging where bidders collude to dwarf the competition, thereby hurting the taxpayers. Bid rigging is pervasive and has been studied in the literature; some of the examples include public procurement and construction (Porter and Zona [1993]; Bajari [2001]) and (Bajari and Ye [2003]) (henceforth, B&Y), right to supply milk to public schools (Porter and

Date: November 18, 2011.

Zona [1999] and Pesendorfer [2000]) and trading of stamps (Asker [2008]). Most of these papers study the market where collusion has already been proven in the civil court (except B&Y) and estimate the effect of collusion on welfare using reduced form approach (except Asker [2008]).[1] However, it is desirable to have a test that can be used to detect possible collusion without knowing if bidders are colluding. In this paper, we estimate two structural models– one with (non-inclusive) collusion and one with competition– and use rank and K-S tests that to test the null hypothesis that the cost distribution under competition explains the data better than that under collusion. For the purpose of the paper, we assume that the bidding ring can control the members bid and suppress collusion. This is the most favorable condition for collusion and the failure to detect collusion means that it is even more unlikely to detect in other cases; for more on collusion in first-price auction see Marshall and Marx [2007]. The aim of this paper is to contribute to our understanding of collusion. Clearly, as with any test, our procedure is not full proof and the test cannot replace wiretapping and thorough criminal investigation but can be used as a first step in assessing the likelihood of bid rigging. In the recent years, criminal enforcement of the antitrust laws has deterred price–fixing in some market settings, but not bidder collusion (Marshall and Meurer [2001]) and hence the social value of any test to detect collusion hasn't decreased. Since bid rigging either lowers the revenue collected or increases the cost of procurement if the government tries to raise funds to meet the deficiencies through distortionary taxes, this creates further inefficiencies. Thus, the increased revenue spent in procurements because of collusion is not simply a wealth transfer.

This paper considers a procurement auction with independent cost and exogenous entry under two environments: competitive and collusive. In view of the data we consider three asymmetric bidders: fringe, regular bidders who might collude and might not collude.[2] Then, assuming that the bidding ring can control the bids of the member and suppress all ring

[1] See also Comanor and Schankerman [1976]; Feinstein, Block, and Nold [1985]; Lang and Rosenthal [1991] and for a summary of the literature on cartels see Harrington [2008].

[2] Asymmetry amongst bidders can be attributed to the location, carrying capacity, informational differences and hence any realistic model of procurement auction should allow asymmetry, (Bajari [2001] and B&Y).

competition and common knowledge of the bidding ring, we characterize the equilibrium bidding strategies for each type, and recover and estimate two different sets of cost distributions as in Guerre, Perrigne, and Vuong [2000]. The only difference between the two sets is in cost that corresponds to the bidders who could collude. For example, if there are five bidders in a ring and five bidders not in the ring then it is equivalent to having only six bidders in total with one from the bidding ring. Therefore, the cost that rationalizes the observed bids must be higher under collusion than under competition. This implies that if there is collusion, then the estimated cost distribution with bidding ring would dominate that without one. To test this stochastic domination we propose a rank based test and a Kolmogorov-Smirnov test, (Lehmann [2006]). But before implementing the test we must determine the potential collusive bidders and determine the asymptotic distribution of the tests because we do not observe the cost but only its estimate, for each bidder, and hence the standard asymptotic distribution of the tests won't work. We use the reduced form test from B&Y for independence and symmetry on each pair of bidders to determine those who fail tests. These bidders will be considered as the regular bidders who can collude. To overcome the second problem we bootstrap the asymptotic distribution of the test statistic.

In this paper we identify two different sets of bidders as potential ring members in the Highway Procurement data in California but find no evidence of collusion even though the reduced form test and 'visual method' on estimated cost distributions supports collusion.[3]

This paper is organized as follows: Section (2) outlines the theoretical models of competition and collusion; Section (3) proposes the two tests; Section (4) discusses the data from CalTrans and how to determine the set of colluding bidders; Section (5) contains the results and Section (6) concludes. Appendix (A-1) explains our estimation procedures; Appendix (A-2) shows how to extend the tests to allow for unobserved auction specific heterogeneity. All tables and figures are collected in Appendix (A-3).

[3] In visual method, we eyeball the estimated cost densities for the suspected bidders and see if the densities are sensitive to the number of bidders. The idea there is that under exogenous entry assumption there should be little or no variation if competition is the true model, which is not what we find, see Section (3) for more.

2. Models, Identification and Estimation

2.1. **Competitive Model (Model A).** A single and indivisible project is procured to $N \geq 2$ risk neutral bidders using sealed bids. We assume that there are three types ($k = 0, 1, 2$) of bidders with n_k type k bidders such that $\sum_k n_k = N$ and the cost $C_i \sim$ IID $F_k(\cdot)$ with absolutely continuous and nowhere vanishing density $f_k(C) > 0, \forall C \in [\underline{c}, \overline{c}]$, for all $i \in n_k$.[4] We also assume that the number of bidders is exogenously given for each auction. Each bidder $i \in n_k$ submits a bid, b_{ik}, to solve

$$\max_{\tilde{b}_{ik}} \left\{ \begin{aligned} \pi_{ik} &= (\tilde{b}_{ik} - c_{ik}) \Pr\left(\tilde{b}_{ik} < \min_{j \neq i} B_{jk}\right) \Pr\left(\tilde{b}_{ik} < \min_{j=1,\dots,n_1} B_{j1},\right) \Pr\left(b_{ik} < \min_{j=1,\dots,n_{(2-k)}} B_{j(2-k)}\right) \\ &= (\tilde{b}_{ik} - c_{ik}) \left(1 - F_k[s_k^{-1}(\tilde{b}_{ik})]\right)^{n_k - 1} \left(1 - F_1[s_1^{-1}(\tilde{b}_{ik})]\right)^{n_1} \left(1 - F_{(2-k)}[s_{(2-k)}^{-1}(\tilde{b}_{ik})]\right)^{n_{(2-k)}} \end{aligned} \right\},$$

for $k = 0, 2$ and

$$\max_{\tilde{b}_{i1}} \left\{ \begin{aligned} \pi_{i1} &= (\tilde{b}_{i1} - c_{i1}) \Pr\left(\tilde{b}_{i1} < \min_{j=1,\dots,n_0} B_{j0}\right) \Pr\left(\tilde{b}_{i1} < \min_{j \neq i} B_{j1}\right) \Pr\left(\tilde{b}_{i1} < \min_{j=1,\dots,n_2} B_{j2}\right) \\ &= (\tilde{b}_{i1} - c_{i1}) \left(1 - F_0[s_0^{-1}(\tilde{b}_{i1})]\right)^{n_0} \left(1 - F_1[s_1^{-1}(\tilde{b}_{i1})]\right)^{n_1 - 1} \left(1 - F_2[s_2^{-1}(\tilde{b}_{i1})]\right)^{n_2} \end{aligned} \right\}$$

for $k = 1$ where where $s_k(\cdot)$ denotes type $k's$ equilibrium strategy. As shown in Lebrun [1996, 1999] and Maskin and Riley [2000a,b, 2003], type specific bidding strategy $s_k(\cdot), k = 0, 1, 2$ exists and is unique and are characterized as a solution to the simultaneous differential equations

$$\begin{aligned} s_k'(c_k, b_k, n) &= (b_k - c_k) \Bigg[(n_k - 1) \frac{f_k(c_k)}{1 - F_k(c_k)} + n_1 \frac{f_1(s_1^{-1}(b_k))}{1 - F_1(s_1^{-1}(b_k))} \frac{s_0'(c_0)}{s_1'(s_1^{-1}(b_k))} \\ &\quad + n_{(2-k)} \frac{f_{(2-k)}(s_{(2-k)}^{-1}(b_k))}{1 - F_{(2-k)}(s_{(2-k)}^{-1}(b_k))} \frac{s_k'(c_k)}{s_{(2-k)}'(s_{(2-k)}^{-1}(b_0))} \Bigg] \quad \text{for } k = 0, 2 \end{aligned}$$

[4] We abuse the notation to use n_k as both the number and set of type k bidders. We allow for the asymmetry of this kind to be consistent with the data: type 1 characterizes large firms that bid simultaneously (on a pairwise basis) more often than others; type 2 bidders are the remaining large firms and type 0 bidders are the other (small/fringe) bidders. Type 1 bidders are the large bidders who are candidates for collusion; we detail how we chose the types of each bidder in Section (4).

and

$$
\begin{aligned}
s_1'(c_1,b_1,n) &= (b_1-c_1)\Bigg[n_0\frac{f_0(s_0^{-1}(b_1))}{1-F_0(s_0^{-1}(b_1))}\frac{s_1'(c_1)}{s_0'(s_0^{-1}(b_1))}+(n_1-1)\frac{f_1(c_1)}{1-F_1(c_1)} \\
&\quad +n_2\frac{f_2(s_2^{-1}(b_1))}{1-F_2(s_2^{-1}(b_1))}\frac{s_1'(c_1)}{s_2'(s_2^{-1}(b_1))}\Bigg], \qquad (1)
\end{aligned}
$$

such that : $s_0(\underline{c}) = s_1(\underline{c}) = s_2(\underline{c}) = \underline{c}$, and $s_0(\bar{c}) = s_1(\bar{c}) = s_2(\bar{c}) = \bar{c}$.

2.2. **Collusive Model (Model B).** We assume that the bidding ring can control the bids of the members and can eliminate all ring competition. Then, both competitive bidders and cartel bidders participate in an auction. As before, there are 3 types of bidders. We label cartel bidders as type 1 bidders. Large competitive bidders are named as type 2 bidders and small competitive (fringe bidders) will be type 0 bidders. From the perspective of a type 1 bidder, there is only one such bidder participating (seriously) in an auction. Hence, $n_1 = 1$ for this group of bidders. This model is the most favorable for collusion and for our purpose we do not have to spell out the exact rules of sharing the surplus.[5] As before, there are n_0 and n_2 bidders of type 0 and type 2, respectively. We maintain the assumption that bidders of type k draw their private costs independently from a distribution $F_k(\cdot)$, $k = 0, 1, 2$. Each bidder i of type k with c_{ik} choses b_{ik} that solves

$$
\begin{aligned}
\max_{\tilde{b}_{ik}}\Bigg\{\pi_{ik} &= (\tilde{b}_{ik}-c_{ik})\Pr\left(\tilde{b}_{ik}<\min_{j\neq i}B_{jk}\right)\Pr\left(\tilde{b}_{ik}<\min_{j=1,\dots,n_1}B_{j1},\right)\Pr\left(\tilde{b}_{ik}<\min_{j=1,\dots,n_{(2-k)}}B_{j(2-k)}\right) \\
&= (\tilde{b}_{ik}-c_{ik})\left(1-F_k[s_k^{-1}(\tilde{b}_{ik})]\right)^{n_k-1}\left(1-F_1[s_1^{-1}(\tilde{b}_{ik})]\right)^{n_1}\left(1-F_{(2-k)}[s_{(2-k)}^{-1}(\tilde{b}_{ik})]\right)^{n_{(2-k)}}\Bigg\}; \\
\max_{\tilde{b}_{i1}}\Bigg\{\pi_{i1} &= (\tilde{b}_{i1}-c_{i1})\Pr\left(\tilde{b}_{i1}<\min_{k=1,\dots,n_0}B_{j0}\right)\Pr\left(\tilde{b}_{i1}<\min_{j=1,\dots,n_2}B_{k2}\right) \qquad (2) \\
&= (\tilde{b}_{i1}-c_{i1})\left(1-F_0[s_0^{-1}(\tilde{b}_{i1})]\right)^{n_0}\left(1-F_2[s_2^{-1}(\tilde{b}_{i1})]\right)^{n_2}\Bigg\},
\end{aligned}
$$

respectively for $k = 0, 2$ and $k = 1$.

2.3. **Nonparametric Identification.** The model primitives are $\{F_k(\cdot|X_\ell, N_\ell)\}$ for $k = 0, 1, 2$, which are type specific conditional cost distributions given the auction specific characteristics X_ℓ and the set of bidders N_ℓ (see Assumption (1) below) . The data provides information on the characteristics of the

[5] Marshall and Marx [2007] show that only in the first-price auction, if the ring cannot control the bids then the equilibrium entails multiple bids and the model need not be identified.

project that is being procured, the number of bidders in each auction and their bids. Using the previous notation, the set of observables W are

$$W := \left\{X_\ell, n_{0\ell}, n_{1\ell}, n_{2\ell}, \{b_{0i}\}_{i=1}^{n_{0\ell}}, \{b_{1i}\}_{i=1}^{n_{1\ell}}, \{b_{2i}\}_{i=1}^{n_{2\ell}}\right\}, \ell = 1, 2, \ldots L.$$

where b_{ki} is the bid of type $k \in \{0, 1, 2\}$ bidder $i \in n_{k\ell}$ in the auction ℓ. We make the following assumptions:

Assumption 1. *(A1)*

(1) *An auction ℓ has $n_\ell \in \{\underline{n}, \overline{n}\}$ risk-neutral bidders with $\underline{n} \geq 2$.*
(2) *The $(d+3)$- dimensional vector $(X_\ell, (n_{k\ell}; k = 0, 1, 2)) \sim$ IID $Q_m(\cdot, \cdot)$ with density $q_m(\cdot, \cdot)$ for all $\ell = 1, 2, \ldots L$.*[6]
(3) *For each ℓ and each $k \in \{0, 1, 2\}$ the variables $C_{ki\ell}, i \in n_{k\ell} \sim$ IID $F_k(\cdot|\cdot, \cdot)$ and density $f_k(\cdot|\cdot, \cdot)$ conditional on (X_ℓ, N_ℓ).*
(4) *The observed type k bids $b_k \sim$ IID $G(\cdot)$ with density $g(\cdot)$ for $k = 0, 1, 2$.*
(5) *(Exogenous Participation) For each $N < \overline{n}$, and all $(C_1, \ldots, C_N)$:*

$$F_N(C_1, \ldots, C_N) = F_{\overline{n}}(C_1, \ldots, C_N, \underbrace{\infty, \ldots, \infty}_{\overline{n}-N}).$$

All the assumptions are standard in the literature with exogenous entry and note that this assumption does not require (X_ℓ, N_ℓ) to be independent but still be consistent with the exogenous entry assumption. Identification follows from the two-steps procedure in Guerre, Perrigne, and Vuong [2000]: (1) Using $n_{k\ell}$ type-k bids estimate $G_k(\cdot|X_\ell, N_\ell)$ and $g_k(\cdot|X_\ell, N_\ell)$ non parametrically (we use Kernel density estimator, see the estimation section in Appendix); (2) then using the first order condition for optimal bids and the estimates form first step we can recover the cost for each bidder as

$$\hat{c}_{ki\ell} \equiv \xi_{ki}(b_{ki}, \{\hat{G}_k(\cdot|\cdot), \hat{g}_k(\cdot|\cdot), n_{k\ell}; k = 0, 1, 2\}). \tag{3}$$

With competition (Model A), for every ℓ (suppressing the dependence on X_ℓ) we have

[6] We abuse the notation to use $n_{k\ell}$ to represent both the random variable and its realization.

$$\begin{aligned}
\xi_{ki}(\cdot) &= b_{ki} - \frac{1}{(n_{k\ell}-1)\frac{\hat{g}_k(b_{ki}|\cdot)}{1-\hat{G}_k(b_{ki}|\cdot)} + n_{1\ell}\frac{\hat{g}_1(b_{ki}|\cdot)}{1-G_1(b_{ki}|\cdot)} + n_{(2-k)\ell}\frac{\hat{g}_{(2-k)}(b_{ki}|\cdot)}{1-\hat{G}_{(2-k)}(b_{ki}|\cdot)}}, \\
\xi_{1i}(\cdot) &= b_{1i} - \frac{1}{n_{0\ell}\frac{\hat{g}_0(b_{1i}|\cdot)}{1-\hat{G}_0(b_{1i}|\cdot)} + (n_{1\ell}-1)\frac{\hat{g}_1(b_{1i}|\cdot)}{1-\hat{G}_1(b_{1i}|\cdot)} + n_{2\ell}\frac{\hat{g}_2(b_{1i}|\cdot)}{1-\hat{G}_2(b_{1i}|\cdot)}},
\end{aligned} \tag{4}$$

where the first equation holds for $k = 0, 2$ and $i \in n_{k\ell}$ and the second equation holds for $k = 1$ and $i \in n_{1\ell}$. Similarly, with collusion (Model B) because $n_{1\ell} = 1$ we have

$$\begin{aligned}
\xi_{ki}(\cdot) &= b_{ki} - \frac{1}{(n_{k\ell}-1)\frac{\hat{g}_k(b_{ki}|\cdot)}{1-\hat{G}_k(b_{ki}|\cdot)} + n_{1\ell}\frac{\hat{g}_1(b_{ki}|\cdot)}{1-G_1(b_{ki}|\cdot)} + n_{(2-k)\ell}\frac{\hat{g}_{(2-k)}(b_{ki}|\cdot)}{1-\hat{G}_{(2-k)}(b_{ki}|\cdot)}}, \\
\xi_{1i}(\cdot) &= b_{1i} - \frac{1}{n_{0\ell}\frac{\hat{g}_0(b_{1i}|\cdot)}{1-\hat{G}_0(b_{1i}|\cdot)} + n_{2\ell}\frac{\hat{g}_2(b_{1i}|\cdot)}{1-\hat{G}_2(b_{1i}|\cdot)}}.
\end{aligned} \tag{5}$$

3. DETECTING COLLUSION

3.1. **Visual Method.** In this section, we present a heuristic method that could be used as a preliminary first step in assessing the possibility of collusion in the data. The criteria to decide whether or not there is collusion is, for the lack of better name, termed as "intuitive" method. The logic of the method is very simple and straightforward and relies heavily on the exogenous entry assumption. Suppose the true data generating process (DGP) is competition (Model A), then the conditional density of the recovered cost of bidders will be independent of the number of opponents, in other words, the recovered density should remain the same even when the number of actual bidders in each category changes. However, the estimated density under the misspecified model of collusion will be very sensitive to the number of bidders in each auction. This property is a direct consequence of exogenous entry assumption and is also symmetric because if the true DGP was collusion (Model B) then it would lower the competition faced by bidders from other type 1 bidders but the recovered cost distribution is still independent of the number of other bidders. But, under competition we expect the density to vary with the number of bidders.

We show that using this intuitive method a collusive model does rationalize the observed bids suggesting that the bidders (in our empirical application) might be colluding. We want to emphasize that this method should only be used as a supplement but not a substitute to any formal method to detect collusion (see the sections below). The result of this method will be explained using the recovered conditional densities under various scenarios and the conclusion about the true DGP will be reached by way of "eyeballing" the figures (collected in the Appendix). Although the conclusion of this method is sensitive to the way bidders' type are determined, because it affects the effective competition by affecting size of the collusive ring, the method can be used with all forms of auction data.

3.2. **Collusion as Stochastic Dominance.** Following up on the intuitive method, we pose the problem of "collusion" as a problem of testing for independence, for which we use the rank test. Then, for each model (competition and collusion) we can derive the underlying cost associated with each bid for each bidder, therefore we have two sets of random variables

$$M_A := \left\{X_\ell, n_{0\ell}, n_{1\ell}, n_{2\ell}, \{\hat{c}^A_{0i}\}_{i=1}^{n_{0\ell}}, \{\hat{c}^A_{1i}\}_{i=1}^{n_{1\ell}}, \{\hat{c}^A_{2i}\}_{i=1}^{N_{2\ell}}\right\}, \ell = 1, 2, \ldots L,$$
$$M_B := \left\{X_\ell, n_{0\ell}, n_{1\ell}, n_{2\ell}, \{\hat{c}^B_{0i}\}_{i=1}^{n_{0\ell}}, \{\hat{c}^B_{1i}\}_{i=1}^{n_{1\ell}}, \{\hat{c}^B_{2i}\}_{i=1}^{n_{2\ell}}\right\}, \ell = 1, 2, \ldots L,$$

where the only difference between M_A (competition) and M_B (collusion) is the recovered cost parameters $\hat{c}^j_{ki}$ for each type $k \in \{0, 1, 2\}$ bidder $i \in n_{k\ell}$ for each model as given in (4) and (5), respectively. A benefit of estimating a structural model is the possibility that we could not only have transparent identification from which we could also infer the exact channel that links the data to the parameters but also provide conditions on the data that are necessary to rationalize the model. Choosing between two models would then be the same as testing which of the two conditions hold in the data. Although very intuitive, in first price auction, the only testable conditions (see Theorem 1 in Guerre, Perrigne, and Vuong [2000]) are: (i) the observed bids are IID conditional on (X_ℓ, N_ℓ); and (ii) given N_ℓ the distribution $G(\cdot)$

of observed bids can be rationalized by $F(\cdot)$ only if $\xi(\cdot)$ in (3) is strictly increasing. So, we cannot differentiate the two models based on these conditions because the first condition is the same for both models and the second condition is redundant because we assume that bidders in both models use strictly increasing bidding strategies. Therefore, we need to look for other variations in the data to differentiate the two models and the only testable difference is the latest cost parameter of the type 1 bidders. Under our assumption of exogenous entry and the assumptions (A1), the recovered costs parameter for each bidder must be independent across all bidders type for both models. However, under Model B, the recovered cost parameters for type 1 must be larger than the cost for the same type under Model A. Under collusion, a necessary implication would be that the recovered cost density under M_A would be stochastically dominated by the cost density under M_B. To explain the proposed test we simplify the notation and say that the random variable c_1^A, c_1^B are the cost parameters under M_A and M_B, respectively, with $F_{1A}(\cdot)$ and $F_{1B}(\cdot)$ as corresponding distributions. Therefore, as a necessary condition for collusion we wish to test the hypothesis that $F_{1A} = F_{1B}$ against the alternative that F_{1B} first order stochastically dominates F_{1A}.

3.2.1. *Rank Based Test.* To test the dominance, we use rank sum test that relies on "U" statistic. To define the relative ranking of the random variables $c^A = \{c_{11}^A, c_{12}^A, \ldots, c_{1n_1^\ell}^A\}$ and $c^B = \{c_{11}^B, c_{12}^B, \ldots, c_{1n_1^\ell}^B\}$ for every auction ℓ we begin with a combined sample in ascending order $(c_{1:2n_1^\ell}, c_{2:2n_1^\ell}, \ldots, c_{2n_1^\ell:2n_1^\ell})$ where the subscript $c_{t:2n_1^\ell}$ is for the t^{th} small private–cost amongst the total of $2n_1^\ell$ such variables. Then we define $R_i^\ell = 1$ if the i^{th} observation of the combined and ordered sample is from M_B and zero otherwise. Then the test statistic is

$$R = \frac{1}{L \times 2n_1^\ell} \sum_{\ell=1}^{L} \sum_{i=1}^{2n_1^\ell} R_i^\ell. \tag{6}$$

Intuitively, we use the entire sample to create a new sample of zeros and ones, such that 1 is chosen only when the private–cost for a bidder $i \in n_1^\ell$ (to rationalize the observed bid) is more under collusion than under competition. Then, averaging across all $L \times n_1^\ell$ observations, we are looking at the empirical measure of the probability of private–cost from M_B being higher than M_A.[7] If $\Pr(R \leq r^*) = \alpha$ under the null hypothesis, the test will be considered significant at the significance level α if $R \leq r^*$ and the hypothesis of identical distributions of c_1^A and c_1^B is rejected in fair of stochastic dominance and hence collusion. A subtle but important issue in implementing the test with our data is that we never observe the private–cost directly but only recover them nonparametrically. This could change the asymptotic variance if not the asymptotic distribution.

3.2.2. *Kolomgorov-Smirnov (KS) Test.* We formulate KS test for whether (under the null hypothesis) the type 1 cost distribution is the same for M_A and M_B against the alternative that the the distribution for M_B stochastically dominates that for M_A:

$$\begin{aligned} H_0 &: \forall c \in [\underline{c}, \bar{c}] \qquad F_{1A}(c) = F_{1B}(c); \\ H_1 &: \exists c \in [\underline{c}, \bar{c}] \qquad F_{1A}(c) \geq F_{1B}(c). \end{aligned}$$

The test statistic is

$$KS_L = \sqrt{\frac{\sum_{\ell=1}^{L} n_{1\ell}}{2}} \sup_{c \in [\underline{c}, \bar{c}]} \Big| F_{1A}(c) - F_{2A}(c) \Big|,$$

which can be shown to be consistent. However, because we do not observe the cost but only the pseudo costs, using the analytical asymptotic distribution only could be misleading. To circumvent that, we Bootstrap the density of the test under the null, to compute the critical point t^*_α at α% significance level.

4. Application to The Procurement Data

In this section, we describe the California Highway procurement market where the rights to maintain and construct highways and roads are granted through sealed low bid auction (procurement) by the California Department

[7] Observe that this intuition is straightforward once we note that the average of R_i^ℓ across all type 1 bidders in auction ℓ is $\frac{1}{2n_1^\ell} \sum_{i=1}^{2n_1^\ell} (R_i^\ell) = \Pr_\ell(c_{2i}^\ell > c_{1i}^\ell)$.

of Transportation (Caltrans), between January 2002 and January 2008.[8] The data include important characteristics about the project that was let, the name of the actual bidders and the set of potential bidders i.e. those who showed interest in the project, their bids and the identity of the winning bidder.

This process of selling the rights is conducted in three steps: First, during the *advertising period*, which lasts between three to ten weeks depending on the size of the project, the Caltrans Headquarters Office Engineer announces a project that is going to be let and solicits bids from bidders/companies. Potential bidders express their interest by buying the project catalogue. Second, sealed bids are revived only from among the potential bidders. Third, on the letting day, the received bids are ranked and the project is awarded to the lowest bidder, provided that the bidder fulfills certain responsibility criteria determined by federal and state law. We ignore any such rules for this paper and treat the lowest bid as winning bid.[9] After each letting, the information about all bids and their ranking is made public. When a company submits bid, it is also required to submit detail information about subcontractors, their fees and obligation(s) of each subcontractor. There is a significant overlap of subcontractors across bidders of similar sizes and bidders tend to have different operational sizes, suggesting that bidders are asymmetric. We divide bidders into two broad types: the main bidders and the fringe bidders, and further allow some of the main bidders to collude. Therefore, we assume there to be three asymmetric types of bidders: the fringe bidders (type 0), the main bidders who can collude (type 1) and finally the main bidders who do not collude (type 2), each with a different cost distribution.[10]

Our data consist of 2,152 contracts that were awarded by Caltrans for a total of $7,645 millions but only 1,907 projects had at least two bidders, with

[8] The data is available from Caltrans web site: `http://www.dot.ca.gov/hq/esc/oe/awards/bidsum/`.

[9] For an example of effect of one such "bid preference" policy, see Krasnokutskaya and Seim [2011].

[10] The parameter of cost is a reduced form for the real production function. So by allowing each type to have unique distribution function we intend to capture the differences in the technology of each bidder.

a total of 823 bidders who bid on at least one project. One of the first challenge for us is to identify the type of each bidder. Determining main and fringe bidders is relatively easy (see Jofre-Bonet and Pesendorfer [2003]) but to determine the bidding ring is not straightforward. In the remaining of this section by way of explaining the data we also explain how we determine the bidding ring. To identify the ring members, we consider large projects that are worth between $1 million and $20 million because smaller project typically do not have margin for profit and hence might not be worth the risk and within that subsample use the reduced methods that includes the test in B&Y to determine bidders who could collude.[11] There are 724 such projects worth $2,408 millions (31% of the total) with 413 bidders out of which 202 win at least once. Further, we consider only 25 bidders who have a nontrivial revenue share (at least 1% revenue share) in the market as the bidders who participate in many auctions and might find it profitable to collude. Although we are agnostic about the exact nature of collusion and how it is sustained, we think having subcontractors facilitates collusion as main bidders compete for the same subcontractors. And this effect is more pronounced for the bidders who participate in multiple auctions and have some non-trivial market share, hence the 1% cutoff. Table A-1 summarizes the bidding activity of these 25 (type 1 and type 2) bidders. All of the remaining bidders will be treated as fringe/small bidders type 0. [12]

The first column in Table (A-1) gives the number of bids of each main bidder and this represents 34% of all bids in the sample. To access the market power of each bidder we define "expected win" (see below) and compare it with the actual numbers of win: bidders with consistently higher actual win than the expected win will be termed as those who have higher market

[11]The test checks if the observed bids are dependent or independent. Competition requires the observed bids be uncorrelated and symmetric across bidders, which are testing using Pearson correlation test and test for exchangeability (see below for the implementation).

[12]Hence, we only look at those bidders who are supposed to be colluding according to B&Y but one can use any other method to choose the bidders and our method would still work. As mentioned earlier, it is very difficult to sustain collusion in first-price auction so assuming that the bidding ring can implement any bidding strategy in the auction is enough for us, see Marshall and Marx [2007].

power. To define expected win, consider bidder A, who bids on a total of 50 projects against a varying number of bidders, n_ℓ for $\ell = 1, \ldots, 50$ then his expected win is defined to be $\sum_{\ell=1}^{50} 1/n_\ell$. By comparing column 2 and 3, we see that with the exception of five bidders, all bidders win more contracts than expected. The fourth column reports the average bid of each main bidder in the sample and the fifth column the revenue share computed as the total value of the bidder's winning bid as a fraction of the total value of winning bids for all contracts. The last column is the participation rate (i.e. the bid frequency rate), and bidder D is the one that stands out at 44%. Table (A-2) provides summary statistics with the following conclusions: (i) on an average there are slightly more than four bidders; (ii) average winning bid is \$3.33 millions, which is less than the average engineers' estimate of \$3.77 millions while the average bid is \$3.79 millions;[13] (iii) money on the table– defined as the difference between the highest and the second highest bid– is on average \$300,000 suggesting informational asymmetry among bidders. We also find that distance between the bidder's office and the site of project has no bearing on the bids, which could be because of the subcontracting and each bidder having mobile units. And in general higher valued projects (between \$1 million and \$20 millions) attract relatively smaller bidders, suggesting that it is the main bidders who can gain the most by colluding and moreover, larger projects are more profitable, ceteris paribus, see Figure (1).

In the remaining part of this section we present a method of finding bidders who could be colluding from the twenty five bidders listed in table A-1. To determine potential colluders, we look at patterns that might facilitate collusion or support the presence of collusion. First, from the theoretical literature on collusion we know that members of a bidding ring participate in the same auctions. For the twenty five bidders we consider all combinations of subgroups and select those bidders that have at least fifteen simultaneous bids, see Table (A-3). The identity of the bidder is in first column while the number of simultaneous bids is in the second. Comparing the "expected win" with the actual win for these pairs, we do see that at least one member of the pair wins often which is in line with previous findings, see Table

[13]Even though there bids are highly correlated (corr. coef. 0.95) with the engineer's estimate, the estimates are not binding as 30% of winning bids are above the estimates.

(A-1). When we compare Table (A-1) and Table (A-3) we see:(i) Firm A exclusively bids against firm D; (ii) Firm E bids remarkably frequently with both firm A and firm D; (iii) the pairs (D,P) and (A,D) have the highest simultaneous bids. All of these suggests that the triplet (A,D,E) and the pair (D,P) could be considered as potential candidates for collusive rings. Now, we use the procedure in B&Y to test the criteria of competition developed by those authors. That is (i) conditional on observables, bids are independently distributed; and (ii) bid distributions should satisfy exchangeability. This set of conditions are necessary for competitive bidding but rejection does not imply that bidding is collusive.

First, we test independence using a regression–based (reduced form)approach and consider the fifteen pairs of bidders bidding frequently described above.[14] The model used is the following

$$\frac{BID_{i\ell}}{EE_{\ell}} = \beta_0 + \beta_1 LDIST_{i\ell} + \beta_2 CAP_{i\ell} + \beta_3 UTIL_{i\ell} + \beta_4 LMDIST_{i\ell} + u_{i\ell} \quad (7)$$

$$\frac{BID_{i\ell}}{EE_{\ell}} = \alpha_0 + \alpha_1 LDIST_{i\ell} + \alpha_2 CAP_{i\ell} + \alpha_3 UTIL_{i\ell} + \alpha_4 LMDIST_{i\ell} + \varsigma_{i\ell}, \quad (8)$$

where $\mathrm{LDIST}_{i\ell}$ refers to the logarithm of distance and $\mathrm{LMDIST}_{i\ell}$ refers to the logarithm of the minimum of distances of all bidders on project ℓ, excluding i and $UTIL_{i\ell}$ is the utilization rate.[15] For the bidders listed in Table (A-3) we use (7) with bidder–varying coefficients and for the rest we use (8) and use the pooled data to estimate both models with a project fixed effect. Let ρ_{ij} be the correlation between the residual to bidder i's bid function ($\hat{u}_{i\ell}$) and bidder j's bid function ($\hat{u}_{j\ell}$), then when we use Pearson's correlation test for independence and find that for all but one pair, we reject the null hypothesis of independence at 5% level. To test exchangeability we follow B&Y and construct two kinds of tests: exchangeability at the market level by pooling the fifteen bidders in one group and exchangeability on a pairwise basis. The null hypothesis of the test is: $H_0 : \beta_{ik} = \beta_{jk}$ for all $i, j, i \neq j$ and for all

[14] The main reason for conducting pairwise tests is basically driven by the amount of data because there are relatively few observations for the triplet (A,D,E) in the sample.

[15] We define the rate as $\mathrm{Util}_{\mathrm{it}} = \mathrm{Backlog}_{\mathrm{it}}/\mathrm{Cap}_{\mathrm{i}}$ (if Cap=0, then Util=0 for all t) and as an explanatory variable because it could be important in explaining bids (see Jofre-Bonet and Pesendorfer [2003]). Approximately 60% of bids in the data are explained by capacity but it varies a lot across bidders.

$k = 1, \ldots, 4$. Let $T = 3,347$ be the number of observations, m the number of regressors and r the number of constraint implied by H_0 then under the null hypothesis we have

$$F = \frac{(SSR_C - SSR_U)/r}{SSR_U/(T-m)} \longrightarrow^d F(r, T-m).$$

At the market level, the restricted model imposes that the effect of the four explanatory variables is the same for potential ring members and the remaining bidders (i.e this is the exchangeability hypothesis). The null hypothesis of exchangeability is rejected when comparing the group of potential cartel members against the remaining bidders. Next, we conduct pairwise tests by pooling bidders accordingly and find that the hypothesis of exchangeability is rejected at conventional levels for 13 out of 15 pairs including the pair (D,P), (A,D) and (D,E). Based on the previous analysis all pairs of bidders considered do not pass at least one of the tests for competitive bidding. However, as mentioned above, taking into account the number of simultaneous bids, bidders D and P bid simultaneously more than a handful of times. Also, the triplet (A,D,E) is chosen as a potential cartel candidate. Therefore for the subsequent analysis we concentrate on two groups of candidates, namely the pair (D,P) and the triplet (A,D,E) as type 1 bidders. Firms D and P bid, on average, in projects of smaller size than the remaining thirteen large bidders (i.e type 2 bidders in the model) and roughly of the same size as the small bidders (type 0 bidders). At least one of the bidders participates in 325 projects winning 113 out of 724 contracts with and average winning bid of \$3.67 million. On average the engineers' estimate in these projects is above the winning bid. The average number of bidders participating in the 325 contracts is 4.65. Generally speaking, the data suggest that this pair tends to participate more often in small size projects with less competition. The other main bidders tend to bid on larger projects and participate in 312 projects. Type 0 bidders participate in almost all auctions (666 out of 724). Table A-4 below contains summary statistics per type when type 1 bidders are the pair (D,P).[16]

[16] We want to emphasize that this is just one of potentially many ways to "identify" bidding ring(s) and it depends on the nature of the data. For example, Conley and Decarolis [2011] uses some special features in Italian procurement data to identify the bidding ring.

The triplet (A,D,E) also tends to bid in smaller size projects relative to type 2 bidders. At least one of the bidders participate in 329 projects winning 117 times. The average winning bid for this group is \$3.70 million which is below the average of the engineers' estimate. There are about five bidders participating in the projects where the triplet bids see Table (A-5) for some summary statistics. Hence, when we implement the test, we consider two cases: one when type 1 bidders are A, D and E (the triplet) and second, when they are D and P (the pair).

5. Implementation of the Tests

5.1. **Visual Method.** In this section we look at the estimated cost densities, for both triplet and pair. Equations (4) and (5) summarize the effect of the collusive ring on the implied cost as under collusion the effective/actual number of type 1 bidders is reduced to only one and one should expect to find the greatest differences when both n_0 and n_2 are small and n_1 is large. To exemplify the process we only present the results for two values of the (log) engineers' estimate, namely 6.1 and 6.5 but the result is same for all other values. The first corresponds to fairly small projects (around \$1.3 million) while the second one corresponds to the average estimate in the sample. To begin with, we first start by changing the number of bidders in group zero and two i.e. n_0 and n_2 to see the effect of this on the estimated type 0 and type 2 densities. Figure (2) contain the estimated densities of private values for type 0 bidders in the triplet–case and the pair–case, respectively. The distribution of private costs for type 0 bidders exhibits some variation with respect to n_0 for both the triplet–case and the pair–case, nevertheless, type 0 bidders are fringe bidders which hardly ever win a contract. Similar exercise for type 2 bidders does not show great variation for different values of n_2 as shown in Figs. (3) and (4) for the triplet and the pair cases, respectively. Unlike the case of type 0 bidders, these results are more in line with what we would expect if bidders are symmetric (within types) and entry is exogenous, as we assume.

In order to make the pictures as informative as possible, by controlling other sources of variation, we look at how the distributions for type 1 bidders change as n_2 changes for various values of n_1. For the triplet–case, Fig.

(5) shows the effect on the distributions of type 1 bidders in the competitive model (Model A, see the first row) and in the collusive model (Model B, see the second row) when $n_0 = 0$. The results for the case $n_0 > 0$ (not reported) are similar. The distribution of type 1 bidders shows less variation in the collusive setup. That is, under the exogeneity assumption and the assumption of symmetry within types, this piece of evidence suggests that bidders (A,D,E) could be engaged in a collusive agreement. Now, for the pair case (see Fig. (6)), the results are remarkably the same. The distributions of private costs in Model B exhibit less variation than in Model A, thus, providing additional evidence supporting possibility of presence of collusive ring, for both $n_0 = 0, 1$.[17] Overall, the evidence in the sample tends to favor collusive model over the competitive model.

5.2. **Rank Based Test.** In this section we present the result of the rank test. As mentioned above, this is a non-parametric statistical hypothesis test for assessing whether the distributions of two random variables is the same. We implement two versions of this test, namely for matched data (sign rank test) and for unmatched data (rank sum test). We first computed the test for the samples of pseudo costs obtained from the estimation procedure, i.e. assuming that the asymptotic distribution of the tests is not affected by the fact that we use pseudo costs in lieu of true (unobserved) costs. Since is possible that the asymptotic distributions of these tests get affected we also report the bootstrapped version of each test in Table (1).[18]

TABLE 1. Rank–Based Tests

	PAIR CASE			TRIPLET CASE		
TESTS	Statistic	p-value	Bootstrap p-value	Statistic	p-value	Bootstrap p-value
SIGN RANK	1.1310	0.2583	0.2590	1.3010	0.1914	0.2100
RANK SUM	0.2630	0.7926	0.7930	0.5150	0.6067	0.5990

Source: Own calculations. Sign Rank Test is for the matched data and Rank Sum is for unmatched data.

[17] Recall that bidder D is a type 1 bidder in both the triplet–case and the pair–case. Moreover, this bidder participates in 44% of the projects in the sample. Thus, the similarity in the results for the triplet–case and the pair–case could be driven by the fact that bidder D is a type 1 bidder in both cases.

[18] All Bootstrapped results are based on 500 replications.

As can be seen, in all cases the null hypothesis of equal distributions cannot be rejected at conventional levels. Thus, the result of this testing procedure does not support evidence of collusion as we concluded from the visual method. This is not surprising given that the visual method is by no means a robust method of inference. For the asymptotic density of the test see Figure (7).

5.3. **KS Test.** We next show the results from the KS two sample test. As before we first implement this test directly on the two samples of pseudo costs recovered nonparametrically and then we computed the bootstrapped standard error so that we also report the corresponding p-value, see Table (2) and Figure (8). These results are again supporting the hypothesis that both distributions are equal, therefore we conclude that this evidence is in favor of a model of competition for the Caltrans data set used.

TABLE 2. Kolmogorov–Smirnov Test

	PAIR CASE			**TRIPLET CASE**		
TESTS	Statistic	p-value	Bootstrap p-value	Statistic	p-value	Bootstrap p-value
KS two sample test	0.0229	1.000	0.259	0.0212	1.000	0.281

Source:Own calculations.

6. CONCLUSIONS

In this paper we propose nonparametric tests that can be used to detect collusion in first-price asymmetric auctions. The methods is based on structural estimation and does not require any prior knowledge about collusion. The tests exploit the difference between the inverse bidding behavior in auction with and without collusion. The only difference between the two models is that collusion dwarfs competition so the recovered cost from the data on bids must be higher under collusion than under competition for type 1 bidders. This suggests that detecting collusion is equivalent to choose one of the two cost distributions as the true DGP. This, in turn, is equivalent to say that the cost distribution under collusion first order stochastically dominates the cost under competition. We propose a two steps (model selection) procedure to detect collusion: (1) First, we determine those bidders who could be colluding, i.e. whose bids fail independence and symmetry (B&Y);

and (2) For those bidders we apply a rank based test and a Kolmogorov-Smirnov test to test the null hypothesis of no domination against domination. We implement the tests on procurement auction data from California and find no evidence of collusion even though we implemented the tests only on those who failed the test for competition. We also propose a visual method where we look at the effect of bidders on the recovered cost density. Under the assumption of exogenous entry, we find that the estimated densities do vary, which is consistent with collusion.

Several extensions to this paper are possible. We could allow for binding reserve price by implementing the tests on the conditional distribution instead. We can use the truncated bids data to identify the conditional distribution of cost (Guerre, Perrigne, and Vuong [2000]). As long as the bidding ring can control the bids of its members, the implementation would be straightforward.[19] Another important extension would be to derive the asymptotic distribution of the tests that we use on estimated (latent) data and compare it with the Bootstrapped results.[20] We could also look at adapting the test to allow for endogenous entry, when the recovered cost will be correlated with the number of bidders and the bidding ring could potentially use entry decision to facilitate collusion. All of which are very important steps towards understanding the complex nature of collusion.

[19] It could happen that the reserve price is too low for the bidding ring thereby effectively screening them out. If the ring never wins a single auction then we would not be able to detect those rings, which is not that bad because they never win the auction anyway.

[20] This extension would be important for testing any structural models. In Aryal and Gabrielli [2011] we study the problem of non–nested model selection with estimated data i.e. pseudo data.

APPENDIX

A-1. Estimation

We first discuss some practical issues. The skewness of the bid distribution is a typical problem encountered with auction data. In addition, the use of kernel estimators is subject to the so–called boundary effect so that some kind of trimming is often used.[21] As a consequence it is common practice among empirical researchers to use a logarithmic transformation in order to keep a substantial number of observations after trimming (see for example Li and Perrigne [2003]). For notational simplicity we suppress the dependance of the distributions on (X, N). Later, when presenting the estimators we include these variables explicitly. Applying the log transformation to system (4) for M_A and M_B, respectively gives

$$\begin{aligned} c_{kM} &= \xi_k(d_k, n) = 10^{d_k} - \frac{10^{d_k}}{(n_k-1)\frac{g_{kd}(d_k)}{1-G_{kd}(d_k)} + n_1\frac{g_{1d}(d_k)}{1-G_{1d}(d_k)} + n_{(2-k)}\frac{g_{(2-k)d}(d_k)}{1-G_{(2-k)d}(d_k)}}; \\ c_{1A} &= \xi_1(d_1, n) = 10^{d_1} - \frac{10^{d_1}}{n_0\frac{g_{0d}(d_1)}{1-G_{0d}(d_1)} + (n_1-1)\frac{g_{1d}(d_1)}{1-G_{1d}(d_1)} + n_2\frac{g_{2d}(d_1)}{1-G_{2d}(d_1)}}; \\ c_{1B} &= \xi_1(d_1) = 10^{d_1} - \frac{10^{d_1}}{n_0\frac{g_{0d}(d_1)}{1-G_{0d}(d_1)} + n_2\frac{g_{2d}(d_1)}{1-G_{2d}(d_1)}}, \end{aligned} \tag{9}$$

where the first equation is for $k = 0, 2, M = A, B$ and $d_k = \log(b_k)$, $G_{kd}(\cdot), g_{kd}(\cdot)$ are the distribution and density of $\log(b_k), k = 0, 1, 2$. As noted earlier, some kind of trimming is often needed due to the bad behavior of kernel estimators close to the boundaries of the support of bids. Following Guerre, Perrigne, and Vuong [2000] we use

$$\hat{c}_{ki\ell} = \begin{cases} \xi_k(d_{i\ell}) & \text{if } d_{\min} + \varrho \max\{h_g, h_G\}/2 \leq d_{i\ell} \leq d_{\max} - \rho \max\{h_g, h_G\}/2; \\ +\infty & \text{otherwise.} \end{cases}$$

[21] To avoid trimming we could have used LPEs instead of kernels in the first step. However, here it does not matter because we are mainly interested in assessing the center of the distributions of private costs.

for $k = 0,1,2$, $i = 1,\dots,n_k$ and $\ell = 1,\dots,L$, where d_{min} and d_{max} are the minimum and maximum of log bids respectively, h_g, h_G are bandwidths and ϱ is the length of the support of the kernel.[22]

Let $S_{d_k}(d|x,n) = \Pr(D \geq d|x,n)$. Then, the *hazard rate* functions involved in the expressions for private costs given by the system of equations in (9) can be written as

$$\frac{g_{d_k}(d|x,n)}{1 - G_{d_k}(d|x,n)} = \frac{g_{d_k}(d|x,n)}{S_{d_k}(d|x,n)} = \frac{g_{d_k}(d,x,n)}{S_{d_k}(d,x,n)}$$

for $k = 0,1,2$. Let T_k denote the total number of observations for bidders of type k. We consider L auctions in which different types of bidders participate. Thus bidder i, $i = 1,\dots,n_k$ of type k participates in auction $\ell = 1,\dots,L$. Relabeling bidders such that $j = (i,\ell)$, i.e. the ith bidder in auction ℓ, the sample consists of observation (d_j, x_j, n_j).[23,24] Thus, the estimators involved in the first step are

$$\begin{aligned}
\hat{g}_k(d,x,n) &= \frac{1}{T_k h_g^{p+1}} \sum_{j=1}^{T_k} K_g\left(\frac{d - D_j}{h_g}, \frac{x - X_j}{h_g}, \frac{n - n_j}{h_{gn}}\right), \\
\hat{S}_k(d,x,n) &= \frac{1}{T_k h_{G_x}^{p}} \sum_{j=1}^{T_k} \mathbb{I}(d_j \geq d) K_G\left(\frac{x - X_j}{h_G}, \frac{n - n_j}{h_{Gn}}\right).
\end{aligned}$$

With the sample of pseudo private costs $\hat{C}$ in the second step we estimate the cost densities as $\hat{f}_k(c|x,n) = \frac{\hat{f}_k(c,x,n)}{\hat{q}_m(x,n)}$, where

$$\begin{aligned}
\hat{f}_k(c,x,n) &= \frac{1}{T_k h_f^{p+1}} \sum_{j=1}^{T_k} K_f\left(\frac{c - \hat{C}_j}{h_f}, \frac{x - X_j}{h_f}, \frac{n - n_j}{h_{fn}}\right), \\
\hat{q}_m(x,n) &= \frac{1}{T_k h_q^{p}} \sum_{j=1}^{T_k} K_q\left(\frac{x - X_j}{h_q}, \frac{n - n_j}{h_{qn}}\right).
\end{aligned}$$

[22]Without loss of generality we set $d_{\min} = 0$.

[23]To keep the notation simple, we just include n_j in the formulas above. However, for the computation of the estimator we have used n_{0k}, n_{1k} and n_{2k} separately.

[24]Recall that X characterizes auction heterogeneity, thus it only varies across auctions. In terms of the notation used this means that $X_j = X_\ell$. In other words, for each auction ℓ the value x is the same for all bidders participating in that auction. A similar argument applies to the number of bidders, N_ℓ.

The functions $K_g(\cdot)$, $K_G(\cdot)$, K_f and $K_q(\cdot)$ are kernels. The bandwidths for the continuous variables are denoted h_G, h_g, h_q and h_f. The bandwidths for the discrete variables are h_{Gn}, h_{gn}, h_{qn} and h_{fn}. Now, we discuss the choices of kernels and bandwidths.

A-1.1. **Choices of Kernels and Bandwidths.** As it is well known in the nonparametric econometric literature, the choice of kernel is not crucial in practice. The estimators in this paper are multivariate kernels which are computed as the product of univariate kernels. That is

$$K_m\left(\frac{a-A_k}{h_g}, \frac{b-B_k}{h_g}, \frac{n-N_k}{h_{gn}}\right) = K_a\left(\frac{a-A_k}{h_g}\right) K_b\left(\frac{b-B_k}{h_g}\right) K_n\left(\frac{n-N_k}{h_{gn}}\right),$$

where $K_m(\cdot)$ refers to the multivariate kernel, $K_a(\cdot)$ and $K_b(\cdot)$ denote the univariate kernels corresponding to the continuous variables A and B, say, and $K_n(\cdot)$ is the kernel for the discrete variables. Recall that $K_n \equiv K_{n_0} K_{n_1} K_{n_2}$.

The econometric procedure follows closely that of Guerre, Perrigne, and Vuong [2000]. The kernels for continuous variables are required to be symmetric with bounded supports (Assumption A3 in Guerre, Perrigne, and Vuong [2000]). Thus, we decide to use the tri-weight kernel function defined as $K(u) = 35/32(1-u^2)^3 \mathbb{I}(|u| \leq 1)$ for these variables, namely d, x and c. The compact support of this function implies that only non–trimmed private costs are used in the second step to obtain the corresponding latent densities.

We use Gaussian Kernel, instead of the tri-weight for the discrete variables because there is relatively small variation in the number of bidders and it is desirable to give more weight to observations farther from the point at which estimation takes place. This is best achieved with a kernel with unbounded support.[25] The smoothness of the distribution of private values is denoted by R, we assume $R = 1$. The bandwidths' choice is critical in nonparametric estimation. To ensure the uniform consistency at the optimal convergence rates of the estimators the bandwidths for the continuous variables are of the following form: $h_g = 1.06 \times 2.978 \times \hat{\sigma} \times (T)^{-1/(2R+4)}$, $h_G = 1.06 \times 2.978 \times \hat{\sigma} \times (T)^{-1/(2R+3)}$, $h_f = 1.06 \times 2.978 \times \hat{\sigma} \times (T_\tau)^{-1/(2R+4)}$

[25] There are no theoretical restrictions to the kernels applied to discrete variables.

$h_q = 1.06 \times 2.978 \times \hat{\sigma} \times (T_\tau)^{-1/(2R+3)}$. The constant term comes from the so–called rule of thumb and the factor 2.978 is the one corresponding to the use of triweight kernels instead of Gaussian kernels (see Hardle [1991]) and T_τ denotes the number of observations kept after trimming. There are 47 bandwidths involved in the whole estimation procedure, with 27 being used in the first step and 20 in the second step. Some bandwidths correspond to the continuous variables, while others to the discrete variables; see Table (A-6) and (A-7).

A-2. Unobserved Heterogeneity

In this section we show how the tests can be implemented when there is unobserved heterogeneity. In particular we consider the unobserved heterogeneity of multiplicative form as in Krasnokutskaya [2011], where the cost of a bidder i in an auction ℓ is given by $\tilde{c}_{i\ell} = y_\ell \times c_{i\ell}$. Krasnokutskaya [2011] shows that (suppressing the index for auction and asymmetry in bidders):

(1) The bids with auction heterogeneity y is just y times the bids without auction heterogeneity;[26]
(2) Under the assumption of independence between y_ℓ and $c_{i\ell}$ the model structure $[F_Y(\cdot), F_C(\cdot)]$ can be nonparametrically identified.

So, in every auction, y_ℓ is common and affects all bid in the same way (bids are multiplied by y), the variation in bids must be through the individual cost, which is independent of y_ℓ. Therefore, whether $y = 1$ or $y \neq 1$, under collusion (M_B) the pseudo–cost recovered must be higher than under competition (M_A) - for type 1 bidders. So, we could estimate the cost distribution using the procedure in section 4 of Krasnokutskaya [2011] and then implement all the tests. But because the estimation procedure is different as it requires using sample characteristic function determining the exact asymptotic distributions is even more difficult and is beyond the scope of this paper and is left for future research.

[26] Because 'no–unobserved heterogeneity' is a special case of unobserved heterogeneity when $y_\ell = 1, \forall \ell = 1, \ldots, L$, if $s_{ik}(\cdot)$ is the bidding strategy when $y = 1$ and $\beta_{ik}(\cdot)$ when $y \neq 1$ then $\beta_{ik}(\tilde{c}_{ik}) = \beta_{ik}(y \times c_{ik}) = y \times s_{ik}(c_{ik})$.

A-3. Tables and Figures

Table A-1. Revenue Shares and Participation of Main Firms

Firm ID	Number of Bids	Number of wins	Exp. Number of wins	Average bid (Mill. $)	Revenue Share	Participation rate
A	50	9	10.34	4.83	0.020	0.07
B	34	13	10.51	3.21	0.012	0.05
C	43	9	10.46	5.32	0.013	0.06
D	319	97	87.32	3.61	0.145	0.44
E	46	11	10.15	4.49	0.015	0.06
F	42	15	10.70	3.63	0.016	0.06
G	25	12	5.84	4.09	0.027	0.03
H	26	6	5.16	5.03	0.011	0.04
I	21	7	4.27	4.54	0.012	0.03
J	20	9	4.69	3.84	0.015	0.03
K	34	4	6.90	8.44	0.019	0.05
L	35	16	7.95	4.32	0.020	0.05
M	29	13	6.94	3.69	0.016	0.04
N	9	3	1.55	6.33	0.012	0.01
O	31	5	6.82	6.37	0.011	0.04
P	50	16	12.95	4.03	0.027	0.07
Q	33	9	6.31	3.35	0.017	0.05
R	28	10	8.10	3.48	0.012	0.04
S	47	12	8.82	4.37	0.021	0.06
T	25	13	5.99	3.75	0.021	0.03
U	68	16	15.22	4.77	0.026	0.09
V	26	7	4.78	5.75	0.025	0.04
W	41	11	7.18	2.92	0.019	0.06
X	41	7	10.27	4.50	0.021	0.06
Y	11	4	1.89	6.04	0.012	0.02
Total	1148	351	282		0.57	

Only bidders with revenue shares $\geq 1\%$ are reported.

Table A-2. Summary Statistics

	No. observations	Mean	SD
No. Bidders	724	4.62	2.37
Winning bid	724	3.33	3.11
Money on the table	724	0.30	0.46
Engineers' Estimate	724	3.77	3.49
All Bids	3347	3.79	3.51
Backlog	3347	4.30	9.76
Distance (miles)	3347	123.98	162.93
Capacity (across bidders)	413	2.30	5.69
Utilization rate	3347	0.20	0.32

All dollar figures are expressed in millions. Utilization rate is the ratio of backlog to capacity.

TABLE A-3. Simultaneous Bids

Firm Pair	Simultaneous Bids	Expected Wins	First Bidder Wins	Second Bidder Wins
(A,D)	44	9.03	9	5
(A,E)	20	4.05	3	6
(B,D)	29	9.51	12	10
(C,D)	17	5.65	5	9
(D,E)	41	8.67	8	9
(D,F)	26	7.46	5	9
(D,H)	19	3.92	7	3
(D,I)	18	3.68	1	7
(D,O)	25	5.16	7	5
(D,P)	44	11.08	13	14
(D,R)	27	7.96	10	10
(D,V)	22	4.20	5	6
(D,W)	19	2.97	2	3
(M,X)	22	4.91	11	2
(W,X)	15	2.81	5	2

TABLE A-4. Summary Statistics per Type

	Type 0		Type 1=(D,P)		Type 2	
	Number of observations	Mean (S.E)	Number of observations	Mean (S.E)	Number of observations	Mean (S.E)
No. Bidders	666	4.81	325	4.65	312	5.17
		(2.36)		(2.46)		(2.77)
Winning bid	488	3.07	113	3.67	123	4.01
		(2.93)		(3.08)		(3.65)
Money on the table	488	0.28	113	0.29	123	0.36
		(0.46)		(0.34)		(0.53)
Engineers' Estimate	666	3.64	325	3.74	312	4.32
		(3.38)		(3.27)		(3.72)
All Bids	2520	3.69	369	3.66	458	4.41
		(3.49)		(3.18)		(3.81)
Backlog	2520	1.37	369	24.60	458	4.05
		(3.40)		(16.44)		(6.00)
Distance (miles)	2520	116.98	369	194.29	458	105.85
		(168.91)		(98.51)		(157.12)
Capacity (across bidders)	398	1.67	2	39.12	13	15.73
		(4.09)		(32.07)		(6.09)
Utilization rate	2520	0.16	369	0.42	458	0.25
		(0.32)		(0.26)		(0.32)

All dollar figures are expressed in millions.

TABLE A-5. Summary Statistics per Type

	Type 0		Type 1=(A,D,E)		Type 2	
	Number of observations	Mean SD	Number of observations	Mean SD	Number of observations	Mean SD
No. Bidders	666	4.81	329	4.66	306	5.08
		2.36		2.45		2.76
Winning bid	488	3.07	117	3.70	119	3.99
		2.93		3.12		3.63
Money on the table	488	0.28	117	0.30	119	0.36
		0.46		0.34		0.54
Engineers' Estimate	666	3.64	329	3.76	306	4.35
		3.38		3.34		3.77
All Bids	2520	3.69	415	3.85	412	4.30
		3.49		3.34		3.75
Backlog	2520	1.37	415	22.75	412	3.62
		3.40		16.64		5.39
Distance (miles)	2520	116.98	415	146.87	412	143.74
		168.91		100.69		172.66
Capacity (across bidders)	398	1.67	3	31.72	12	15.63
		4.09		26.84		5.72
Utilization rate	2520	0.16	415	0.42	412	0.23
		0.32		0.28		0.30

All dollar figures are expressed in millions.

TABLE A-6. Bandwidths used in the triplet–case

First Step					
Continuous Variables		Discrete Variables			
h_{gd_0}	0.276	$h_{g_0n_0}$	0.624	$h_{G_0n_0}$	0.481
h_{gx_0}	0.272	$h_{g_0n_1}$	0.417	$h_{G_0n_1}$	0.321
h_{Gx_0}	0.209	$h_{g_0n_2}$	0.624	$h_{G_0n_2}$	0.481
h_{gd_1}	0.372	$h_{g_1n_0}$	0.826	$h_{G_1n_0}$	0.676
h_{gx_1}	0.382	$h_{g_1n_1}$	0.735	$h_{G_1n_1}$	0.601
h_{Gx_1}	0.313	$h_{g_1n_2}$	0.826	$h_{G_1n_2}$	0.676
h_{gd_2}	0.400	$h_{g_2n_0}$	0.894	$h_{G_2n_0}$	0.732
h_{gx_2}	0.394	$h_{g_2n_1}$	0.734	$h_{G_2n_1}$	0.600
h_{Gx_2}	0.323	$h_{g_2n_2}$	0.894	$h_{G_2n_2}$	0.732
Second Step					
Continuous Variables		Discrete Variables			
h_{f_0c}	0.246	$h_{f_0n_0}$	0.628		
h_{f_0x}	0.224	$h_{f_0n_1}$	0.426		
$h_{f_{1A}c}$	0.334	$h_{f_0n_2}$	0.628		
$h_{f_{1A}x}$	0.326	$h_{f_{1A}n_0}$	0.852		
$h_{f_{1B}c}$	0.334	$h_{f_{1A}n_1}$	0.979		
$h_{f_{1B}x}$	0.316	$h_{f_{1A}n_2}$	0.852		
h_{f_2c}	0.360	$h_{f_{1B}n_0}$	0.854		
h_{f_2x}	0.339	$h_{f_{1B}n_1}$	0.726		
		$h_{f_{1B}n_2}$	0.854		
		$h_{f_2n_0}$	0.932		
		$h_{f_2n_1}$	0.730		
		$h_{f_2n_2}$	0.932		

TABLE A-7. Bandwidths used in the pair–case

First Step					
Continuous Variables		Discrete Variables			
h_{gd_0}	0.276	$h_{g_0 n_0}$	0.624	$h_{G_0 n_0}$	0.481
h_{gx_0}	0.272	$h_{g_0 n_1}$	0.417	$h_{G_0 n_1}$	0.321
h_{Gx_0}	0.209	$h_{g_0 n_2}$	0.624	$h_{G_0 n_2}$	0.481
h_{gd_1}	0.369	$h_{g_1 n_0}$	0.963	$h_{G_1 n_0}$	0.791
h_{gx_1}	0.379	$h_{g_1 n_1}$	0.441	$h_{G_1 n_1}$	0.362
h_{Gx_1}	0.311	$h_{g_1 n_2}$	0.963	$h_{G_1 n_2}$	0.791
h_{gd_2}	0.396	$h_{g_2 n_0}$	1.049	$h_{G_2 n_0}$	0.856
h_{gx_2}	0.392	$h_{g_2 n_1}$	0.539	$h_{G_2 n_1}$	0.439
h_{Gx_2}	0.319	$h_{g_2 n_2}$	1.049	$h_{G_2 n_2}$	0.856
Second Step					
Continuous Variables		Discrete Variables			
$h_{f_0 c}$	0.246	$h_{f_0 n_0}$	0.628		
$h_{f_0 x}$	0.225	$h_{f_0 n_1}$	0.426		
$h_{f_{1A} c}$	0.332	$h_{f_0 n_2}$	0.628		
$h_{f_{1A} x}$	0.324	$h_{f_{1A} n_0}$	0.973		
$h_{f_{1B} c}$	0.342	$h_{f_{1A} n_1}$	0.597		
$h_{f_{1B} x}$	0.316	$h_{f_{1A} n_2}$	0.973		
$h_{f_2 c}$	0.353	$h_{f_{1B} n_0}$	0.978		
$h_{f_2 x}$	0.336	$h_{f_{1B} n_1}$	0.449		
		$h_{f_{1B} n_2}$	0.978		
		$h_{f_2 n_0}$	1.086		
		$h_{f_2 n_1}$	0.548		
		$h_{f_2 n_2}$	1.086		

FIGURE 1. Bidder Concentration

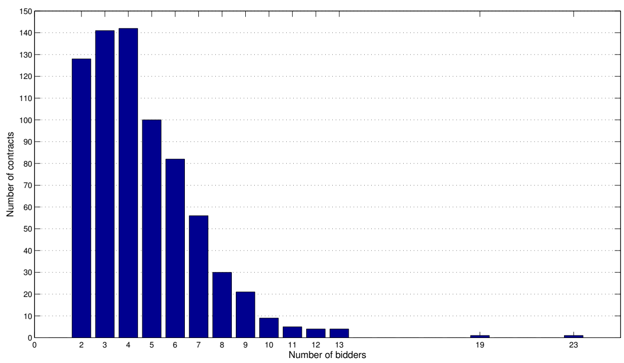
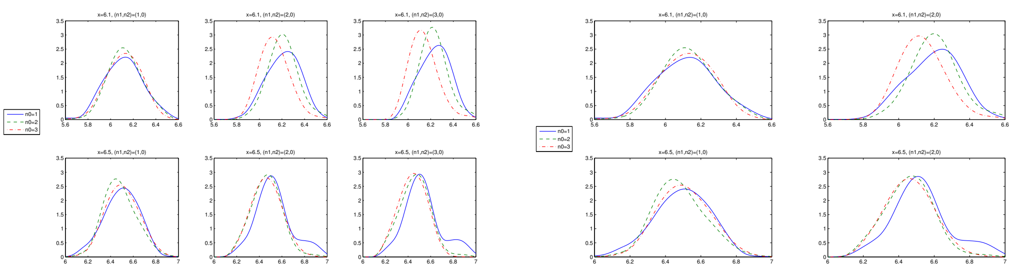
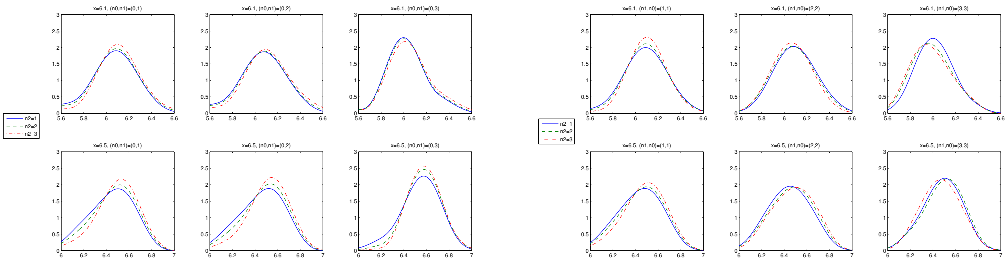


FIGURE 2. $\hat{f}_0(\cdot)$ with various values of n_0



(1): Triplet (2): Pair

FIGURE 3. $\hat{f}_2(\cdot)$ with varying n_2- triplet–case



(1) : $n_0 = 0$ (2) : $n_0 > 0$

FIGURE 4. $\hat{f}_2(\cdot)$ with varying n_2 - pair–case

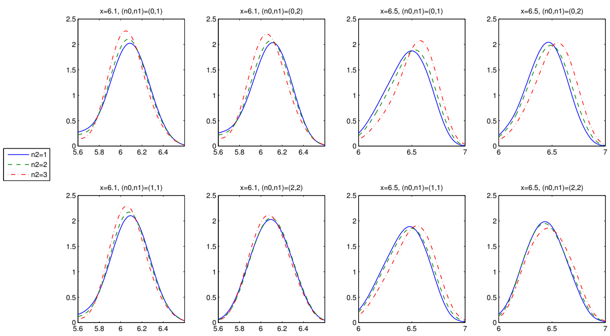


FIGURE 5. $\hat{f}_1(\cdot)$ with varying n_2 - triplet–case

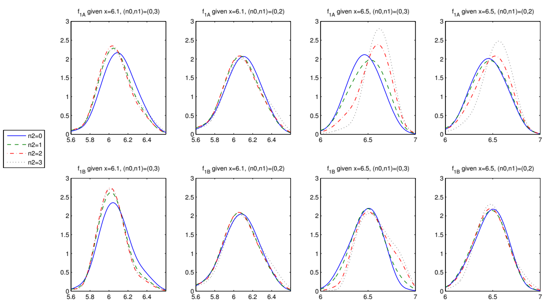
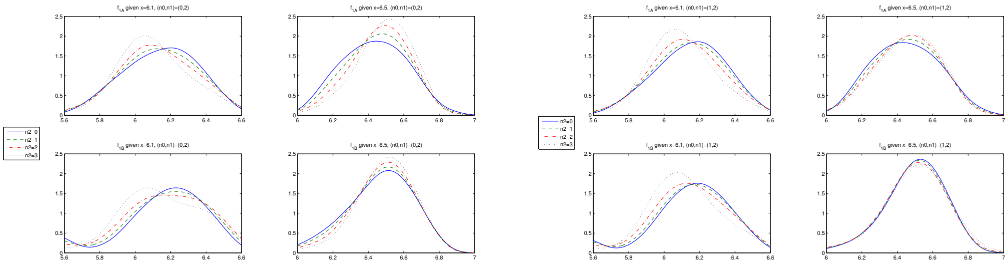
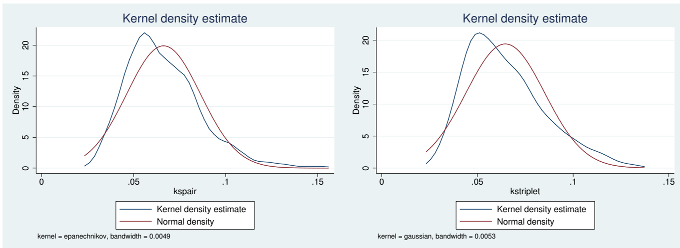


FIGURE 6. $\hat{f}_1(\cdot)$ with varying n_2- pair–case

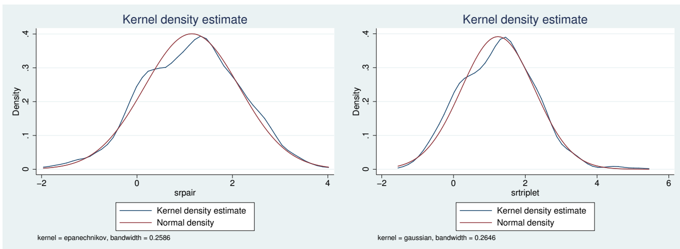


$(1) : n_0 = 0$ $(2) : n_0 = 1$

FIGURE 7. Bootstrapped density for MWW test

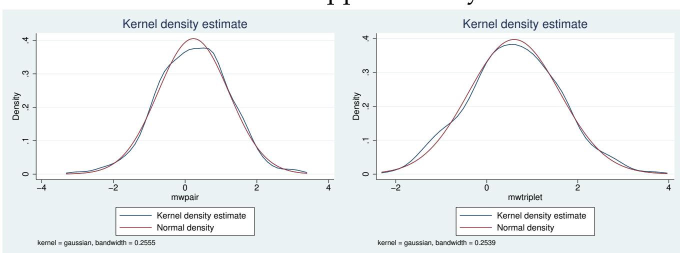


matched: (1): Pair (2): Triplet



unmatched (1): Pair (2): Triplet

FIGURE 8. Bootstrapped density of K-S test.



(1): Pair (2): Triplet

REFERENCES

ARYAL, G., AND M. F. GABRIELLI (2011): "Model Selection Tests with Pseudo Data," *Mimeo, The Australian National University.* 20

ASKER, J. (2008): "A Study of the Internal Organisation of a Bidding Cartel," *American Economic Reveiw*, 100(3), 724–762. 3

BAJARI, P. (2001): "Comparing Competition and Collusion: A Numerical Approach," *Economic Theory*, 18, 187–205. 2, 3

BAJARI, P., AND L. YE (2003): "Deciding Between Competition and Collusion," *Review of Economics and Statistics*, 85, 971–989. 2

COMANOR, W. S., AND M. A. SCHANKERMAN (1976): "Identical Bids and Cartel Behavior," *Bell Journal of Economics*, 7, 281–286. 3

CONLEY, T. G., AND F. DECAROLIS (2011): "Detecting Bidders Groups in Collusive Auctions," *Working Paper, University of Wisconsin, Madison.* 16

FEINSTEIN, J. S., M. K. BLOCK, AND F. C. NOLD (1985): "Asymmetric Information and Collusive Behavior in Auction Markets," *The American Economic Review*, 75(3), 441–460. 3

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. 4, 7, 9, 20, 21, 23

HARDLE, W. (1991): *Smoothing Techniques with Implementation in S*. Springer-Verlag New York, Inc. 24

HARRINGTON, J. E. (2008): "Detecting Cartels," in *Handbook in Antitrust Economics*, ed. by P. Buccirossi. MIT Press. 3

JOFRE-BONET, M., AND M. PESENDORFER (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 1443–1489. 13, 15

KRASNOKUTSKAYA, E. (2011): "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *Review of Economic Studies*, 78, 293–327. 24

KRASNOKUTSKAYA, E., AND K. SEIM (2011): "Bid Preference Programs and Participation in Highway Procurement Auctions," *American Economic Reveiw*, 101, 2653–2686. 12

LANG, K., AND R. W. ROSENTHAL (1991): "The Contractor's Game," *The RAND Journal of Economics*, 22, 329–338. 3

LEBRUN, B. (1996): "Existence of an Equilibrium in First Price Auctions," *Economic Theory*, 7, 421–443. 5

——— (1999): "First–Price Auction in the Asymmetric N Bidder Case," *International Economic Review*, 40, 125–142. 5

LEHMANN, E. L. (2006): *Nonparametrics: Statistical Methods Based on Ranks*. Springer-Verlag New York, Inc., second edn. 4

LI, T., AND I. PERRIGNE (2003): "Timber Sale Auctions with Random Reserve Prices," *The Review of Economics and Statistics*, 85, 189–200. 21

MARSHALL, R., AND M. MEURER (2001): "The Economics of Auctions and Bidder Collusion," in *Game Theory and Business Applications*, ed. by K. Chatterjee, and W. Samuelson. Kluwer Academic Publishers, Norwell. 3

MARSHALL, R. C., AND L. M. MARX (2007): "Bidder Collusion," *Journal of Economic Theory*, 133(1), 374–402. 3, 6, 13

MASKIN, E., AND J. RILEY (2000a): "Asymmetric Auctions," *Review of Economic Studies*, 67, 413–438. 5

——— (2000b): "Equilibrium in Sealed High Bid Auctions," *Review of Economic Studies*, 67, 439–454. 5

——— (2003): "Uniqueness of Equilibrium in Sealed High–Bid Auctions," *Games and Economic Behavior*, 45, 395–409. 5

PESENDORFER, M. (2000): "A Study of Collusion in First–Price Auctions," *Review of Economic Studies*, 67, 381–411. 3

PORTER, R., AND D. ZONA (1993): "Detection of Bid–Rigging in Procurement Auctions," *Journal of Political Economy*, 101, 518–538. 2

——— (1999): "Ohio School Milk Markets: An Analysis of Bidding," *Rand Journal of Economics*, 30, 263–288. 2