International Journal of Industrial Organization 31 (2013) 26–35

Contents lists available at SciVerse ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

Testing for collusion in asymmetric first-price auctions[☆]

Gaurab Aryal [a], Maria F. Gabrielli [b,c,*]

[a] The Australian National University, Canberra, Australia
[b] CONICET, Buenos Aires, Argentina
[c] Universidad Nacional de Cuyo, Mendoza, Argentina

ARTICLE INFO

Article history:
Received 25 February 2011
Received in revised form 2 October 2012
Accepted 10 October 2012
Available online 17 October 2012

JEL classification:
C1
C4
C7
D4
L4

Keywords:
Asymmetric auctions
Collusion
Nonparametric testing

ABSTRACT

This paper proposes a two step procedure to detect collusion in asymmetric first-price procurement (auctions). First, we use a reduced form test to short-list bidders whose bidding behavior is at-odds with competitive bidding. Second, we estimate the (latent) cost for these bidders under both competition and collusion setups. Since for the same bid the recovered cost must be smaller under collusion—as collusion increases the mark-up—than under competition, detecting collusion boils down to testing for first-order stochastic dominance, for which we use the classic Kolmogorov–Smirnov and Wilcoxon–Mann–Whitney tests. Our bootstrap based Monte Carlo experiments for asymmetric bidders confirm that the procedure has good power to detect collusion when there is collusion. 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. This highlights potential pitfalls of inferring collusion based only on reduced form tests.

1. Introduction

Auction is the most widely used selling mechanism for both private and public goods. For example, in the U.S. the federal government is the biggest auctioneer as it is used to sell the offshore oil leases, timbers from national forests, rights to construct highways and is also used to liquidate the assets of bankrupt businesses. However, auctions are susceptible to bid rigging where bidders collude to dwarf the competition, thereby hurting the taxpayers. Bid rigging is pervasive in various markets, such as public construction, school milk supply, stamps; see Asker (2008), Bajari (2001), Comanor and Schankerman (1976), Feinstein et al. (1985), Harrington (2008), Lang and Rosenthal (1991), Pesendorfer (2000), Porter and Zona (1993), Porter and Zona (1999) and municipal bonds among others. Although in the recent years, criminal enforcement of the antitrust laws has deterred price–fixing in some markets it has not deterred bidder collusion (Marshall and Meurer, 2001), and hence, the social value of any method to detect collusion has not decreased. Since bid rigging either lowers the revenue collected or increases the cost of procurement and if this shortfall were met through some distortionary taxes then it creates further inefficiencies. Thus, the increased revenue spent on procurements due to collusion is not simply a wealth transfer from taxpayer to the colluders.[1] It is important to detect and stop collusion as soon as possible.

Previous empirical work on collusion either rely on data from civil lawsuits to estimate the welfare cost of collusion or use reduced form estimation that ignores potential strategic interactions amongst colluders leading to misspecification errors.[2] It is not an exaggeration to say that such data from lawsuits are very hard to come by and even if they do, in most cases, it is already too late. Therefore, it is desirable to have a method that relies on publicly available bid data to determine if bidders could be colluding. The main objective of this paper is to contribute in this direction by proposing a simple and usable two step procedure to test whether or not the bids data are consistent with collusion, thereby aiding in a potential civil lawsuit.

The basic idea of the procedure is the following. Suppose we have a set of potential members of a collusive ring. If we use the same bid data to recover the underlying cost (by using the Bayesian Nash Equilibrium conditions for asymmetric auction to map bids to the costs, à

☆ The paper supersedes the paper "Detecting Collusion on Highway Procurement," by Gabrielli. Gabrielli thanks Joris Pinske for insights and comments on the earlier draft. We are grateful to the referees and the co-editor for their insightful comments and suggestions. We also thank the audiences in ANU micro econometrics seminar. All remaining errors are ours.

* Corresponding author. Tel.: +54 261 4247624.
E-mail addresses: gaurab.aryal@anu.edu.au (G. Aryal), florgabrielli@gmail.com (M.F. Gabrielli).

[1] For example in the case of United States of America v. Carollo, Goldberg and Grimm, the accused bidders are charged for rigging bids in many municipal bonds auctions cost state and local governments billions of dollars; see (Taibbi, 2012).

[2] For an exception see (Asker, 2008).

0167-7187/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijindorg.2012.10.002

la (Guerre et al., 2000)) twice, once under competition (Model A) and once under collusion (Model B), then the cost under Model A must be stochastically greater than cost under Model B because collusion increases the mark-up (i.e. increases the difference between bids and costs) by dwarfing the competition. And this suggests that if indeed those bidders colluded then the empirical CDF of cost under Model A should first order stochastically dominate (f.o.s.d) empirical CDF under collusion. We assume that the ring is efficient and can control bids of all its members, and hence, there is only one serious bid for the ring, the bid that pertains to the member with the lowest cost.[3] Once we recover the two sets of costs, we can test for f.o.s.d using Wilcoxon–Mann–Whitney (W-MW) and Kolmogorov–Smirnov (KS) tests.[4] Since costs are nonparametrically estimated, to account for the error in estimation we bootstrap both the tests to determine the correct p-values.

However, the first step of the procedure involves determining the aforementioned set of potential colluders. Since we do not know who they are, we have to exploit the bidding pattern in the data. To do that we separate regular bidders from fringe bidders, thereby introducing asymmetry, and focus on high-valued (worth at least $1 million) projects where the regular bidders participate more often, and then implement a test proposed in (Bajari and Ye, 2003) to determine if the bidding pattern is consistent with competition and independence or not. We then consider the regular bidders who fail this test as potential colluders (we call them type 1).[5]

It is important to note that although we are interested in determining if the data are inconsistent with competition, our procedure is incapable of deciding in favor of Model B (collusion) or Model A. In other words our procedure exploits only one behavioral feature of collusion, i.e. difference in mark-ups, and using classic hypothesis testing can only answer if the implied cost under collusion is significantly lower than competition to have been generated under competition. Hence, our procedure is not a model selection criterion but is in fact much weaker (in terms of what can be said about the true DGP) than a model selection such as (Vuong, 1989) for non-nested models with finite parameters. Having said this, we have to make sure that if indeed the alternative hypothesis (i.e. collusion) is true then the procedure should pick that up and reject the null of competition. To test the power of the procedure we consider a Monte Carlo exercise. For two asymmetric bidders where type 1 bidders collude, we use the numerical algorithm proposed in Hubbard and Paarsch (2009) based on Mathematical Programming with Equilibrium Constraints (MPEC) method of Su and Judd (2012) to determine the equilibrium bidding strategies. Then using these strategies we generate bids and treat those bids as our data and perform the test on them, while using bootstrapped p-values. We consider many widely used parametric densities (under different parameters) namely Normal, Log Normal, Weibull, Exponential and Uniform and find that the tests detect collusion.[6] Then we apply this procedure on publicly available data on Highway Procurement in the state of California. We find that there is no evidence for stochastic dominance, i.e. collusion, even though the test by Bajari and Ye (2003) detects inconsistencies in bidding pattern.

This paper seeks to contribute to our attempt to understand implications of collusion. However we do not claim our procedure can and should replace wiretapping and thorough criminal investigations. If anything, the procedure should be taken only as a first step in assessing the likelihood of bid rigging. On a technical ground, the paper also seeks to contribute to the literature on empirical auction pioneered by Guerre et al. (2000) by expanding their testable implications of first-price auction models by including a model of collusion; see also Flambard and Perrigne (2006).[7]

This paper is organized as follows: Section (2) outlines the theoretical models of competition and collusion; Section (3) proposes the two tests; Section (4) shows the result of the Monte Carlo exercise; Section (5) discusses the data; Subsection (5.1) collects the results and Section (6) concludes. Appendix (A.1) explains our estimation procedure; 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).

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. In view of the data, we assume that there are three types ($k=0,1,2$) of bidders with $n_k < \infty$ bidders of type k.[8] We abuse the notation to use n_k as both the number and set of type k bidders. The cost $C_i \sim iid\ F_k(\cdot)$ with absolutely continuous and nowhere vanishing density $f_k(C) > 0, \forall C \in [\underline{c}, \bar{c}]$, for all $i \in n_k$. 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

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

for $k=0,1,2$ where $s_k(\cdot)$ denotes type $k's$ equilibrium strategy and B_{ik} is the (random) bid by bidder $i \in n_k$. As shown in Lebrun (1996), Lebrun (1999); Maskin and Riley (2000a,b, 2003), type specific bidding strategy $s_k(\cdot)$ exists and is unique. For each type $k=0,1,2$ they are characterized as a solution to the following system of three ordinary differential equations

$$1 = (b_{ki} - c_{ki})\left[\frac{(n_k - 1) f_k\left(s_k^{-1}(b_{ki})\right)}{\left[1 - F_k\left(s_k^{-1}(b_{ki})\right)\right] s_k\left(s_k^{-1}(b_{ki})\right)} + \sum_{l \neq k, l \in \{0,1,2\}} \frac{n_l f_1\left(s_l^{-1}(b_{ki})\right)}{\left[1 - F_l\left(s_l^{-1}(b_{ki})\right)\right] s_l\left(s_l^{-1}(b_{ki})\right)}\right], \tag{1}$$

such that $\exists! b : s_0(\underline{c}) = s_1(\underline{c}) = s_2(\underline{c}) = b$, 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 and hence there is only one serious bidder, the most efficient bidder, i.e. $n_1 = 1$. As mentioned earlier, this is the most favorable environment for collusion and for our purpose we do not have to spell out the exact rules of sharing

[3] This is the most favorable condition for collusion and the failure to detect collusion in this scenario means that it is unlikely to detect collusion when there is no centralized ring-mechanism. For our purpose, we do not need to know the side payments that are necessary to sustain collusion; see Marshall and Marx (2007) for more on this.

[4] The cost of other types (fringe and regular but non-colluding bidders) are the same under both models and hence do not play any role in the testing procedures.

[5] Asymmetry amongst bidders can be attributed to their locations, carrying capacity, informational differences and hence any realistic model of procurement auction should allow asymmetry, (Bajari, 2001; Bajari and Ye, 2003). Typically, only construction companies who participate mostly on highly valued project are called the regular bidders. It is important to note that using this test to narrow the set of potential colluders is just one possible way. For instance, Conley and Decarolis (2011) exploit some special features in Italian procurement data to identify the bidding rings.

[6] To account for potential numerical errors we run each procedures 1000 times and report the fraction of correct decisions.

[7] We thank a referee for pointing this connection.

[8] Type 1 represents large/regular bidders who bid simultaneously (on a pairwise basis) more often than others and are the candidates for collusion, type 2 bidders are the regular but non-colluding bidders and type 0 are the small/fringe bidders, all of which are explained in detail in Section (5).

the surplus.[9] We assume that the other bidders are unaware of the existence of the bidding ring. Hence, everything else is the same as in Model A except now only type 1 bidder with cost c_{1i} solves

$$\begin{aligned}\max_{\tilde{b}_{1i}} \pi_{i1} &= \left(\tilde{b}_{1i}-c_{i1}\right)Pr\left(\tilde{b}_{i1}<\min_{k=1,\dots,n_0} B_{j0}\right)\Pr\left(\tilde{b}_{1i}<\min_{j=1,\dots,n_2} B_{k2}\right)\\ &= \left(\tilde{b}_{1i}-c_{1i}\right)\left(1-F_0\left[s_0^{-1}\left(\tilde{b}_{1i}\right)\right]\right)^{n_0}\left(1-F_2\left[s_2^{-1}\left(\tilde{b}_{1i}\right)\right]\right)^{n_2}.\end{aligned}$$

The corresponding first-order condition for type 1 bidder becomes

$$1=(b_{1i}-c_{1i})\left[\sum_{l=0,2}\frac{n_l f_1\left(\left(s_l^{-1}(b_{1i})\right)\right)}{\left[1-F_l\left(s_l^{-1}(b_{1i})\right)\right]}s_l'\left(s_l^{-1}(b_{1i})\right)\right], \tag{2}$$

which together with the first-order conditions in (1) determine a new system of differential equations, and together with the same boundary conditions, characterize unique bidding strategies.

2.3. Nonparametric identification

The model primitives are $\{F_k(\cdot|X,N)\}$ for $k=0,1,2$, which are type specific conditional cost distributions given the auction specific characteristics X and the set of bidders N (see Assumption (1) below). The data provide information on the characteristics of the 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

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

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 $n_\ell\in\{\underline{n},\bar{n}\}$ 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 $\ell=1,2,\dots L$.[10]
3. For each ℓ and each $k\in\{0,1,2\}$ the variables $C_{ki\ell}, i\in n_{k\ell}\sim iid\ F_k(\cdot|\cdot,\cdot)$ with density $f_k(\cdot|\cdot,\cdot)$ conditional on (X_ℓ,N_ℓ).
4. The type k bid $B_k\sim iid\ G_k(\cdot|\cdot)$ with density $g_k(\cdot|\cdot)$ for $k=0,1,2$.
5. (Exogenous participation): The cost distributions do not depend on the number of bidders, i.e. $F_k(\cdot|X,N)=F_k(\cdot|X)$ for all $k=0,1,2$ and X.

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 which is still consistent with the exogenous entry assumption. Identification follows from Guerre et al. (2000): (1) Using $n_{k\ell}$ type k bids estimate $G_k(\cdot|X_\ell,N_\ell)$ and $g_k(\cdot|X_\ell,N_\ell)$ nonparametrically (Kernel density estimation method; see Appendix A.1); (2) then using the first-order condition for optimal bids and the estimates from the first step we can recover the cost for each bidder as

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

Under competition (Model A), for every ℓ (suppressing the dependence on X_ℓ and N_ℓ) for all $i\in n_{1\ell}$ we have

$$\hat{c}_{1i}=\xi_{1i}(\cdot)=b_{1i}-\frac{1}{n_{0\ell}\frac{\hat{g}_0(b_{1i}|\cdot)}{1-\hat{G}_0(b_{1i}|\cdot)}+(n_{1\ell}-1)\frac{\hat{g}_1(b_{1i}|\cdot)}{1-\hat{G}_1(b_{1i}|\cdot)}+n_{2\ell}\frac{\hat{g}_2(b_{1i}|\cdot)}{1-\hat{G}_2(b_{1i}|\cdot)}}. \tag{4}$$

and under collusion (Model B) $n_1=1$ and hence

$$\hat{c}_{1i}=\xi_{1i}(\cdot)=b_{1i}-\frac{1}{n_{0\ell}\frac{\hat{g}_0(b_{1i}|\cdot)}{1-\hat{G}_0(b_{1i}|\cdot)}+n_{2\ell}\frac{\hat{g}_2(b_{1i}|\cdot)}{1-\hat{G}_2(b_{1i}|\cdot)}}, \tag{5}$$

where we abuse the notation and use b_{1i} to mean the lowest bid amongst type 1 bidders because under collusion the remaining bids are just cover bids and hence arbitrary. Henceforth, in estimating and testing Model B, we shall always only use the lowest bid.[11]

3. Detecting collusion

In this section we present our formal testing procedure to discriminate between two models, that is a model of competition from a model in which a subgroup of bidders collude. These tests do not require the researcher to have any previous knowledge about the existence of rings in that data.

3.1. Collusion as stochastic dominance

As mentioned before, we pose the problem of "collusion" as a problem of testing for stochastic dominance, 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

$$\begin{aligned}M_A&:=\left\{X_\ell,n_{1\ell},\left\{\hat{c}_{1i}^A\right\}_{i=1}^{n_{1\ell}}\right\},\ell=1,2,\dots L,\\ M_B&:=\left\{X_\ell,n_{1\ell},\left\{\hat{c}_1^B\right\}\right\},\ell=1,2,\dots L,\end{aligned}$$

where $\hat{c}_{1i}^j$ and $\hat{c}_1^j$ are the recovered type 1 cost from (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 structural elements of the model 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 (Theorem 1, (Guerre et al., 2000)) are: (i) the observed bids are *iid* conditional on (X_ℓ,N_ℓ); and (ii) given N_ℓ the distribution $G(\cdot|\cdot,\cdot)$ of observed bids can be rationalized by $F(\cdot|\cdot)$ only if $\xi(\cdot)$, the inverse bidding strategy, in (3) is strictly increasing. These restrictions, however, are weak and are insufficient for model selection. Since the only difference between the two models is the recovered type 1 cost, it is natural to see that the criteria to select appropriate model should rely solely on these two sets of costs. Under our assumption of exogenous entry and the assumptions (A1), from (4) and (5) it follows that for the same bid b the implied cost $c_1^A(b)\geq c_1^B(b)$. In other words, because collusion lowers competition, if we observe the same bid under both competition and collusion, then it must be the case that the cost under collusion is smaller than under competition. This is consistent with the fact that under collusion, the effective mark-up, defined as the difference between the lowest cost among the type 1 bidders and the bids must be higher than under competition.[12] A necessary implication of this is that the recovered cost density under M_A would f.o.s.d. the cost density under M_B.[13] To explain the proposed test we simplify the

[9] (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.

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

[11] Since for the test we do not have to estimate the cost density, so strictly speaking we only use the first step of Guerre et al. (2000).

[12] This inequality holds only for one of type 1 bidders, the one with lowest bid.

[13] It is important to note that this dominance criteria is applicable only to type 1 bidders. Furthermore, we are comparing the distribution for the same bidder type and not across different types. Thus, even though in an asymmetric auctions, weaker bidders (type 0) will bid aggressively than stronger bidders (type 1 or 2) (Maskin and Riley, 2000a), the testing criteria is *not* picking that aggressive bidding behavior. In fact the procedure aims to pick the weak bidding (hence high mark-up) within type 1.

notation and say that the random variable c_1^A,c_1^B are the (recovered) costs 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 $F_{1A}=F_{1B}$ (i.e. competition) against the alternative that F_{1A} first order stochastically dominates F_{1B} (i.e. collusion).

3.1.1. Rank based test

To test the dominance, we use rank sum test that relies on "U" statistics. To define the relative ranking of the random variables, let $c^A = \left\{c_{11}^A, c_{12}^A, \dots, c_{1n_1^\ell}^A\right\}$ and $c^B=\{c_1^B\}$ for every auction ℓ. We begin with a combined sample in ascending order. Then we define $R_i^\ell = 1$ if the i^{th} observation of the combined and ordered sample is from M_A and zero otherwise. Then the test statistic is

$$R = \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{(n_1^\ell+1)}\sum_{i=1}^{n_1^\ell+1} R_i^\ell. \tag{6}$$

Intuitively, the test uses the entire sample to create a new sample of zeros and ones, such that it is equal to one when the cost for a bidder $i\in n_1^\ell$ (that rationalizes the observed bid) is less under collusion than under competition, and zero otherwise. Then, averaging across all $L\times(n_1^\ell+1)$ observations, we are looking at the empirical measure of the probability of private cost from M_A being higher than M_B.[14] If $Pr(R\le r^*)=\alpha$ under the null hypothesis, the test will be considered significant at the level α if $R\le r^*$ and the hypothesis of identical distributions of c_1^A and c_1^B is rejected in favor f.o.s.d., i.e., we conclude that we have found sufficient evidence in favor of collusion. Since costs are estimated and not observed, we use the Bootstrap to compute corrected p-value. We use the numerical procedure from Reiczigel et al. (2005) to estimate bootstrap p-value.

3.1.2. Kolmogorov–Smirnov (KS) test

We formulate the 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 distribution for M_A stochastically dominates that for M_B:

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

The test statistic is

$$KS_L = \sup_{c\in[\underline{c},\bar{c}]} |F_{1A}(c)-F_{1B}(c)|,$$

which can be shown to be consistent.[15] However, because the costs are estimated, we have to allow for estimation errors while computing the distribution of the test statistic. So, we bootstrap the density of the test under the null, to compute the critical point t_α^* at α% significance level for a modified test statistic[16]:

$$\widetilde{KS}_L = \sup_{c\in[\underline{c},\bar{c}]}\left|\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{n_{1\ell}}\sum_{i=1}^{n_{1\ell}}1\left\{\hat{c}_{1i}^A\le c\right\}-\frac{1}{L}\sum_{\ell=1}^{L}1\{\hat{c}_{\ell B}\le c\}\right|.$$

Note:Even though by construction $\hat{c}_1^B<\hat{c}_1^A$, it does not mean that the test will always reject the null even when the null is true because this inequality is true only for L pseudo-costs with no restrictions imposed on the remaining $\left(\sum_{\ell=1}^{L} n_{1\ell}-L\right)$ pseudo-costs. Moreover the difference between the two costs $\hat{c}_1^A$ and $\hat{c}_1^B$ is not a constant but (under the null) a mean zero random variable. The intuition behind the test is that under the null, the differences between L observations will not matter (in terms of test outcome) when we add the remaining samples because adding these $\left(\sum_{\ell=1}^{L} n_{1\ell}-L\right)$ i.i.d. observations will only change the scale and both the tests use ranks and rank based tests do not get affected by change in scale. On the other hand, if alternative hypothesis is true by using only the minimum bids (amongst type 1 bidders) the two ECDFs differ. In the Monte Carlo exercise, we verify if the test can detect collusion when the true DGP is collusion. But when we implement the tests in the data we cannot reject the null, refer to Section (5.1) for more. Now, we present the results from the Monte Carlo exercise.

4. Monte Carlo exercise

In this section, we simulate bids data from asymmetric auctions when type 1 bidders collude and see if the test can detect collusion. Since the bidders do collude, we will have confidence in our procedure if we reject the null of equal distribution in favor of the alternative that $F_{1A}(\cdot)$ f.o.s.d. $F_{1B}(\cdot)$. We repeat this test for many widely used parametric families and for each case run the test 1000 times. This multi-step exercise, however, is nontrivial because, unlike in symmetric auctions, there is no closed form solutions for asymmetric auctions- except for uniform distributions. So, we have to rely on numerical approximation of the bidding strategies and solving this has been identified as a major computational problem; see Marshall et al. (1994), Bajari (2001), Gayle and Richard (2008), Hubbard and Paarsch (2009), Fibich and Gavish (2011) among others. After exploring various methods and keeping in mind the stability of the solution, the documentation and ease of use, we decided to follow the numerical procedure proposed by Hubbard and Paarsch (2009) by extending it to our environment.[17] This method, in turn, is based on Bajari (2001) and solves the equilibrium bidding strategies by casting the problem within the MPEC framework proposed by Su and Judd (2012). For simplicity, in this simulation exercise we consider homogenous auctions with only two types of asymmetric bidders with four bidders of each type. Then the steps we choose are as follows: (1) We choose F_1 and F_2, cost distributions, for type 1 and type 2 bidders, respectively[18]; (2) We consider four type 1 and three type 2 bidders and numerically compute competitive and collusive bidding strategies for each parametrization; (3) We draw four random costs for type 1 and three random costs for type 2 from respective distribution; (4) We map costs to bids using the bidding strategy from Step (2); (5) We repeat this for $L=1000$ homogenous auctions and at the end have4,000 type 1 bids, of which only 1,000 corresponds to the real cost (under Model B) and can be used for our purpose[19]; (6) We perform the tests using the simulated bid data from step (5) while Bootstrapping the asymptotic distribution. To account for possible numerical errors, we repeat each test, for each parametrization, 1000 times and report the fraction of times the procedure makes correct decision.[20] The results are collected in Table 1.[21]

[14] 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 $\frac{1}{2n_1^\ell}\sum_{i=1}^{2n_1^\ell}\left(R_i^\ell\right)=Pr_\ell(c_{2i}^\ell<c_{1i}^\ell)$.

[15] Note that the modified test statistic has unbalanced samples of data, because for collusion we can only use the cost that corresponds to the lowest bid for the type 1 bidders.

[16] For bootstrap version of KS test in the context of policy evaluation see Abadie (2002).

[17] We also tried using the standard numerical method and Matlab's in-built ODE package and also the methods proposed by Fibich and Gavish (2011), but we found the results were sensitive to the parameters and unstable at the boundary. From our experience (Hubbard and Paarsch, 2009) performed the best. The codes are written by Hubbard and is available from his website, for which we are extremely thankful.

[18] For each auction both cost distributions are from the same parametric family.

[19] We know from our model that the only relevant bid is the smallest bid amongst the ring members, while the rest can be arbitrary.

[20] Since the true data generating process is collusion, we should reject the null in favor of the alternative that the cost distribution under competition f.o.s.d the cost distribution under collusion.

[21] To assess the performance of the tests, we repeated the exercise for different number of bidders and different number of auctions (with same number of bidders in each auction). The conclusions from the tests remained the same. Although we do not report the result here, on account of space, they are available upon request from the authors.

5. 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 auctions (procurements) by the California Department of Transportation (Caltrans), between January 2002 and January 2008.[22] 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.

The 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 received 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. After each letting, the information about all bids and their ranking is made public. When a company submits a 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 the possibility for 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.[23]

Our data consist of 2,152 contracts that were awarded by Caltrans for a total of $7,645 million but only 1,907 projects had at least two bidders, with a total of 823 bidders who bid on at least one project. The contracts in our sample include different kinds of projects. The broad categories are asphalt repaving, road paving, bridge reconstruction, among other tasks.[24]

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 we use the reduced methods that includes the test in Bajari and Ye (2003) to determine bidders who could collude.[25] There are 724 such projects that worth $2,408 million (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. This effect is more pronounced for the bidders who participate in multiple auctions and have some nontrivial 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 type 0 fringe/small bidders.[26]

The second 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 number of wins: bidders with consistently higher actual win than the expected win will be termed as those who have higher market 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 columns 3 and 4, we see that with the exception of five bidders, all bidders win more contracts than expected. The fifth column reports the average bid of each main bidder in the sample and the sixth 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 million, which is less than the average engineers' estimate of $3.77 million while the average bid is $3.79 million[27]; (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. In general higher valued projects (between $1 million and $20 million) 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 Fig. 1.

In the remaining part of this section we present a method of finding bidders who could be colluding from the 25 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 25 bidders we consider all combinations of subgroups and select those bidders that have at least 15 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 A.1. When we compare Tables A.1 and 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 suggest that the triplet (A,D,E) and the pair (D,P) could be considered as potential candidates for collusive

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

[23] The private 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.

[24] Some of these tasks include: laying asphalt; installing new sidewalks; striping the highway; constructing, replacing and widening brides; planting, widening, resurfacing and installing irrigation and waste water system in highways; reconstructing interchanges and widen over-crossing; rehabilitating roadways and pavements; repairing and/or remove existing bridge and; storm damage repair, etc. The exhaustive list of the projects is available from the Caltrans webpage.

[25] The test checks if the observed bids are dependent or independent. Competition requires the observed bids be uncorrelated and symmetric across bidders, which we test by means of the Pearson correlation test. We also test for exchangeability (see below for the implementation).

[26] Hence, we only look at those bidders who are supposed to be colluding according to Bajari and Ye (2003) 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).

[27] Even though the 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.

Table 1
Fraction of right decisions when there is collusion.

	Exponential	Exponential	Exponential	Normal	Normal	Normal	Normal
Type 1	1.5	0.8	0.4	(1.5, 1.5)	(1.5, 0.9)	(0.9, 1.2)	(2.9, 0.3)
Type 2	0.5	0.5	0.5	(3.5, 0.8)	(2.5, 0.3)	(2.8, 0.4)	(2.5, 0.3)
W-MW	1	1	1	1	1	1	1
KS	1	1	1	1	1	1	1
	Normal	Log norm	Log norm	Log norm	Weibull	Weibull	Uniform
Type 1	(1.9, 0.3)	(1.2, 1.3)	(2.0, 0.3)	(1.5, 1.0)	(3.0, 2.5)	(2.5, 1.8)	(1.0, 4.0)
Type 2	(2.5, 0.3)	(2.9, 0.7)	(2.8, 0.2)	(2.5, 0.8)	(3.0, 4.0)	(2.5, 3.0)	(1.0, 4.0)
W-MW	1	1	1	1	1	1	1
KS	1	1	1	1	0.97	1	1

The table shows the proportion of tests that rejected the null of equal distribution in favor of f.o.s.d. when there was collusion. Proportion is defined as the number of times the p-value was less than 0.05 and each test was run 1,000 times, for every parameter. The *p*-values were estimated using 10,000 Bootstrap replications.

rings. Now, we use the procedure in Bajari and Ye (2003) 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 is 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 15 pairs of bidders bidding frequently described above.[28] 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 $UTIL_{i\ell}$ is the utilization rate and $\mathrm{LDIST}_{i\ell}$ and $\mathrm{LMDIST}_{i\ell}$ refer to the logarithm of distance and logarithm of the minimum of distance between bidder *i's* registered office and the site of ℓ*th* project, respectively.[29] 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 $\left(\hat{u}_{j\ell}\right)$, then 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 Bajari and Ye (2003) and construct two kinds of tests: exchangeability at the market level by pooling the 15 bidders in one group and exchangeability on a pairwise basis. The null hypothesis of the test is: H_0:$\beta_{is} = \beta_{js}$ for all $i,j,i \neq j$ and for all $s = 1,\ldots,4$. 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

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

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 an 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 suggests that this pair tends to participate more often in small size projects with less competition. The other main bidders (type 2) 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).

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).

5.1. Implementation of the tests

Having laid out the model and estimation method in the previous sections, we now move to discuss the results from the tests. To get an idea of what the estimated costs look like we use quantile–quantile plot in Fig. 2 to gauge the difference between the two sets of costs, for both the pair and the triplet. And from the plot we expect both the tests not to reject the null of equal distribution and hence rule out collusion.

5.2. Rank based test

Here we present the result of the rank test. As mentioned above, this is a nonparametric statistical hypothesis test for assessing whether the distributions of two random variables is the same. We first computed the test for the samples of pseudo-costs obtained from the estimation

[28] 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.
[29] We define the rate as $\mathrm{Util}_{it} = \mathrm{Backlog}_{it}/\mathrm{Cap}_i$ (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.

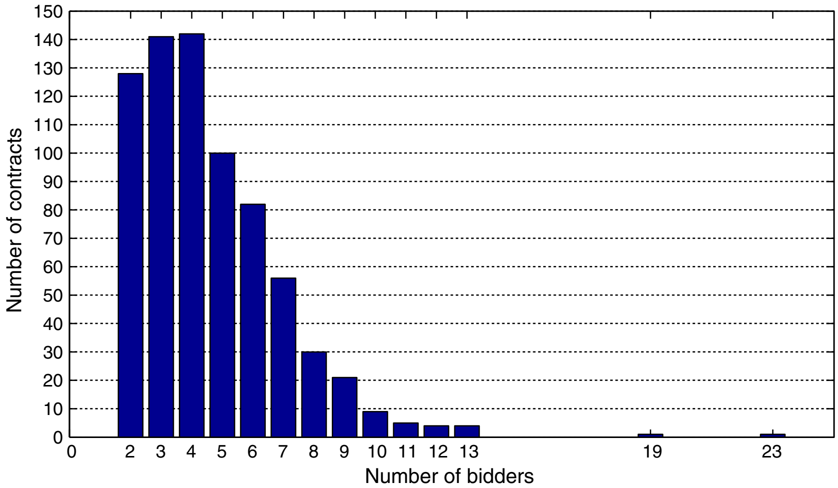

Fig. 1. Bidder concentration.

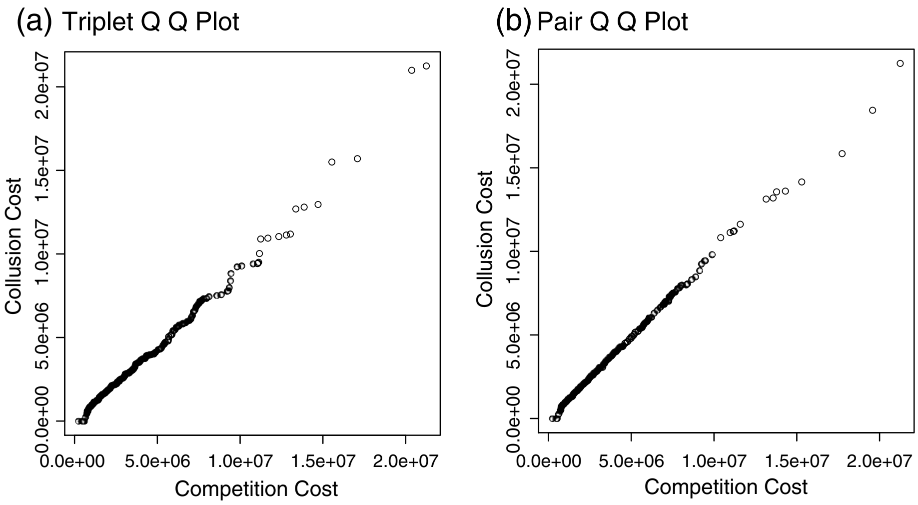

Fig. 2. Quantile–quantile plot of cost.

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 it is possible that the asymptotic distributions of these tests get affected we also report the bootstrapped version of each test in Table 2.[30] As can be seen, in all cases the null hypothesis of equal distributions cannot be rejected at conventional levels.

5.3. KS test

We next present 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 can also report the corresponding *p*-value, see Table 2. 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.

[30] All Bootstrapped results are based on 10,000 replications.

6. Conclusion

In this paper we propose a two step procedure that can be used to detect possible collusion in sealed bid first-price low-bid auctions with asymmetric bidders. The method is based on structural estimation and does not require any prior knowledge about collusion but exploits the difference between the inverse bidding behavior with and without collusion. Since collusion dwarfs competition, when we use the same bid to estimate cost with and without collusion, the estimated cost under collusion should be smaller than that under competition. This

Table 2
Results of the tests.

	Pair case			Triplet case		
	Statistic	*p*-value	Bootstrap *p*-value	Statistic	*p*-value	Bootstrap *p*-value
W-MW	0.503	0.4552	0.481	0.503	0.4434	0.469
KS	0.021	0.9931	0.9985	0.0013	0.995	0.994

difference is consistent with the idea that mark-up increases with collusion, and hence bigger difference from bids to costs. Therefore, checking if the cost distribution under competition f.o.s.d. the cost distribution under collusion might be a valid test to detect collusion. Using the testing procedure of Bajari and Ye (2003) we narrow the pool of bidders to those whose bidding pattern fails to be in line with competitive behavior and independence. For those bidders we map their bids to latent cost and then we test the null hypothesis of equal distribution against first-order stochastic 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 reduced form test for competition and conclude that there is no evidence of collusion.[31]

One caveat of our approach is that we only exploit the minimum bid under collusion. This is because we do not have a theory to interpret the observed "cover bids." This suggests that there is room to consider a more general model of collusion. So, just because our tests reject collusion does not mean that there is no collusion. It just confirms to the common wisdom that detecting collusion can be awfully difficult and many major bid rigging takes place over a long period, sometime even decades. On the flip side, finding collusion by our method need not be admissible in the court of law as a sufficient proof, and cannot substitute more reliable methods such as wiretapping; it should, nonetheless, be used to raise the "red flag" against the bidders.

We hope this research highlights the importance of considering carefully various alternative environments, such as dynamic bidding, subcontracting and endogenous entry in auction. All of these are important extensions and can help us understand better how to use the bid data to detect and then deter welfare reducing collusive behavior or to use the bid data to help us design a collusion proof mechanism.

Appendix A

A.1. Estimation

We first discuss some practical issues. The skewness of the bid distribution is a typical problem encountered with auction data and as in any kernel estimation, some trimming is usually done to account for the so-called boundary effect. To reduce the number of variables lost under trimming, it is a common practice to use a logarithmic transformation; see Li and Perrigne (2003). We also use this transformation, even though we do not have to, as the tests can be implemented on the ECDF and ranks. 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 (4) gives

$$c_1^A = \xi_1(d_1,n) = 10^{d_1} - \frac{10^{d_1}}{n_0\frac{g_{0d}(d_1)}{1-G_{0d}(d_1)} + (n_1-1)\frac{g_{1d}(d_1)}{1-G_{1d}(d_1)} + n_2\frac{g_{2d}(d_1)}{1-G_{2d}(d_1)}}, \tag{9}$$

where $d_1 = log(b_1)$ and $G_{kd}(\cdot)$,$g_{kd}(\cdot)$ are the distribution and density of $log(b_k)$ for $k = 0,1,2$. Similarly log transformation of (5) for M_B gives

$$c_1^B = \xi_1(d_1) = 10^{d_1} - \frac{10^{d_1}}{n_0\frac{g_{0d}(d_1)}{1-G_{0d}(d_1)} + n_2\frac{g_{2d}(d_1)}{1-G_{2d}(d_1)}}. \tag{10}$$

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 Eqs. in (9) and in (10) can be written as

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

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, $i = 1,\ldots,n_k$ of type k participates in auction $\ell = 1,\ldots,L$. 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).[32,33] Thus, the estimators involved are

$$\hat{g}_{k_d}(d,x,n) = \frac{1}{T_k h_g^{p+1}}\sum_{j=1}^{T_k} K_g\left(\frac{d-D_j}{h_g},\frac{x-X_j}{h_g},\frac{n-n_j}{h_{gn}}\right),$$
$$\hat{S}_{k_d}(d,x,n) = \frac{1}{T_k h_{G_x}^p}\sum_{j=1}^{T_k} 1\mathrm{I}\left(d_j \geq d\right) K_G\left(\frac{x-X_j}{h_G},\frac{n-n_j}{h_{Gn}}\right),$$

where $K_g(\cdot)$ and $K_G(\cdot)$ are the kernels, h_G, and h_g are the bandwidths for the continuous variables and h_{Gn} and h_{gn} are bandwidths for discrete ones. Since, the choice bandwidths are critical to determine the rate of convergence, now we discuss their choice in detail.

A.1.1. Choices of kernels and bandwidths

Since the exact choice of the kernels is not crucial for inference, we use product of univariate kernels to represent the multivariate kernel, i.e.

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

Here, $K_m(\cdot,\cdot,\cdot)$ is the multivariate kernel, $K_a(\cdot)$ and $K_b(\cdot)$ denote the univariate kernels corresponding to the continuous variables A and B, respectively, and $K_n(\cdot)$ is the kernel for the discrete variables such that $K_n := K_{n_0}K_{n_1}K_{n_2}$. The kernels for continuous variables should be symmetric with bounded supports (Guerre et al., 2000); we decided to use the triweight kernel function $K(u) = 35/32\left(1-u^2\right)^3 1\mathrm{I}(|u| \leq 1)$.

For the discrete variables, we use Gaussian Kernel because, as there is less variation in the number of bidders it is desirable to give less weight to observations farther from the point at which estimation takes place and is best achieved with a kernel with unbounded support.[34] We assume the smoothness parameter R for the cost distribution is 1. To ensure the uniform consistency at the optimal rates the bandwidths for the continuous variables are chosen to be $h_g = 1.06 \times 2.978 \times \hat{\sigma} \times (T)^{-1/(2R+4)}$, $h_G = 1.06 \times 2.978 \times \hat{\sigma} \times (T)^{-1/(2R+3)}$. 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). In total, there are 27 bandwidths for both continuous and discrete variables and are reported in Tables 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 $\tilde{c}_{i\ell} = y_\ell \times c_{i\ell}$.

[31] We believe that our method can also be adapted to symmetric multi-unit (share) auction, using (Hortaçsu, 2002) to identify the marginal value, making the procedure important to analyze markets such as municipal bonds and Treasury bills auctions.

[32] 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_{1k} and n_{2k} separately.

[33] 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.

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

(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[35];
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 smaller 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.

A.3. Tables

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 ≥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 million. Utilization rate is the ratio of backlog to capacity.

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

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 (2.36)	325	4.65 (2.46)	312	5.17 (2.77)
Winning bid	488	3.07 (2.93)	113	3.67 (3.08)	123	4.01 (3.65)
Money on the table	488	0.28 (0.46)	113	0.29 (0.34)	123	0.36 (0.53)
Engineers' estimate	666	3.64 (3.38)	325	3.74 (3.27)	312	4.32 (3.72)
All bids	2520	3.69 (3.49)	369	3.66 (3.18)	458	4.41 (3.81)
Backlog	2520	1.37 (3.40)	369	24.60 (16.44)	458	4.05 (6.00)
Distance (miles)	2520	116.98 (168.91)	369	194.29 (98.51)	458	105.85 (157.12)
Capacity (across bidders)	398	1.67 (4.09)	2	39.12 (32.07)	13	15.73 (6.09)
Utilization rate	2520	0.16 (0.32)	369	0.42 (0.26)	458	0.25 (0.32)

All dollar figures are expressed in million.

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 2.36	329	4.66 2.45	306	5.08 2.76
Winning bid	488	3.07 2.93	117	3.70 3.12	119	3.99 3.63
Money on the table	488	0.28 0.46	117	0.30 0.34	119	0.36 0.54
Engineers' estimate	666	3.64 3.38	329	3.76 3.34	306	4.35 3.77
All bids	2520	3.69 3.49	415	3.85 3.34	412	4.30 3.75
Backlog	2520	1.37 3.40	415	22.75 16.64	412	3.62 5.39
Distance (miles)	2520	116.98 168.91	415	146.87 100.69	412	143.74 172.66
Capacity (across bidders)	398	1.67 4.09	3	31.72 26.84	12	15.63 5.72
Utilization rate	2520	0.16 0.32	415	0.42 0.28	412	0.23 0.30

All dollar figures are expressed in million.

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

Table A.7
Bandwidths used in the pair 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}	0.369	$h_{g_1n_0}$	0.963	$h_{G_1n_0}$	0.791
h_{gx_1}	0.379	$h_{g_1n_1}$	0.441	$h_{G_1n_1}$	0.362
h_{Gx_1}	0.311	$h_{g_1n_2}$	0.963	$h_{G_1n_2}$	0.791
h_{gd_2}	0.396	$h_{g_2n_0}$	1.049	$h_{G_2n_0}$	0.856
h_{gx_2}	0.392	$h_{g_2n_1}$	0.539	$h_{G_2n_1}$	0.439
h_{Gx_2}	0.319	$h_{g_2n_2}$	1.049	$h_{G_2n_2}$	0.856

References

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

Asker, J., 2008. A study of the internal organisation of a bidding cartel. American Economic Reveiw 100 (3), 724–762.

Bajari, P., 2001. Comparing competition and collusion: a numerical approach. Economic Theory 18, 187–205.

Bajari, P., Ye, L., 2003. Deciding between competition and collusion. The Review of Economics and Statistics 85, 971–989.

Comanor, W.S., Schankerman, M.A., 1976. Identical bids and cartel behavior. Bell Journal of Economics 7, 281–286.

Conley, T.G., Decarolis, F., 2011. "Detecting Bidders Groups in Collusive Auctions", Working Paper. University of Wisconsin, Madison.

Feinstein, J.S., Block, M.K., Nold, F.C., 1985. Asymmetric information and collusive behavior in auction markets. The American Economic Review 75 (3), 441–460.

Fibich, G., Gavish, N., 2011. Numerical simulation of asymmetric first-price auctions. Games and Economic Behavior 73, 479–495.

Flambard, V., Perrigne, I., 2006. Asymmetry in procurement auctions: evidence from snow removal contracts. The Economic Journal 116, 1014–1036.

Gayle, W., Richard, J.-F., 2008. Numerical solutions of asymmetric, first price, independent private values auctions. Computational Economics 32, 245–275.

Guerre, E., Perrigne, I., Vuong, Q., 2000. Optimal nonparametric estimation of first-price auctions. Econometrica 68, 525–574.

Hardle, W., 1991. Smoothing Techniques with Implementation in S. Springer-Verlag, New York. (Inc.).

Harrington, J.E., 2008. Detecting Cartels. In: Buccirossi, P. (Ed.), Handbook in Antitrust Economics. MIT Press.

Hortaçsu, A., 2002. Bidding Behavior in Divisible Good Auctions: Theory and Evidence from the Turkish Treasury Auction Market. Working Paper. University of Chicago.

Hubbard, T.P., Paarsch, H.J., 2009. Investigating bid preferences at low-price, sealed-bid auctions with endogenous participation. International Journal of Industrial Organization 17, 1–14.

Jofre-Bonet, M., Pesendorfer, M., 2003. Estimation of a dynamic auction game. Econometrica 71, 1443–1489.

Krasnokutskaya, E., 2011. Identification and estimation of auction models with unobserved heterogeneity. Review of Economic Studies 78, 293–327.

Lang, K., Rosenthal, R.W., 1991. The contractor's game. The RAND Journal of Economics 22, 329–338.

Lebrun, B., 1996. Existence of an equilibrium in first price auctions. Economic Theory 7, 421–443.

Lebrun, B., 1999. First–price auction in the asymmetric N bidder case. International Economic Review 40, 125–142.

Li, T., Perrigne, I., 2003. Timber sale auctions with random reserve prices. The Review of Economics and Statistics 85, 189–200.

Marshall, R.C., Marx, L.M., 2007. Bidder collusion. Journal of Economic Theory 133 (1), 374–402.

Marshall, R., Meurer, M., 2001. The Economics of Auctions and Bidder Collusion. In: Chatterjee, K., Samuelson, W. (Eds.), Game Theory and Business Applications. Kluwer Academic Publishers, Norwell.

Marshall, R., Meurer, M., Richard, J., Stromquist, W., 1994. Numerical analysis of asymmetric first price auctions. Games and Economic Behaviour 7, 193–220.

Maskin, E., Riley, J., 2000a. Asymmetric auctions. Review of Economic Studies 67, 413–438.

Maskin, E., Riley, J., 2000b. Equilibrium in sealed high bid auctions. Review of Economic Studies 67, 439–454.

Maskin, E., Riley, J., 2003. Uniqueness of equilibrium in sealed high-bid auctions. Games and Economic Behavior 45, 395–409.

Pesendorfer, M., 2000. A study of collusion in first-price auctions. Review of Economic Studies 67, 381–411.

Porter, R., Zona, D., 1993. Detection of bid-rigging in procurement auctions. Journal of Political Economy 101, 518–538.

Porter, R., Zona, D., 1999. Ohio School Milk Markets: an analysis of bidding. The RAND Journal of Economics 30, 263–288.

Reiczigel, J., Zakariás, I., Rózsa, L., 2005. A bootstrap test of stochastic equality of two populations. The American Statistician 59 (2), 156–161.

Su, C.-L., Judd, K.L., 2012. Constrained Optimization Approaches to Estimation of Structural Models. Econometrica 80 (5), 2213–2230.

Taibbi, M., 2012. The Scam Wallstreet Learned from the Mafia. The Roling Stones.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.