ESTIMATING REVENUE UNDER COLLUSION-PROOF AUCTIONS

Gaurab Aryal† and Maria F. Gabrielli‡

† The Australian National University, (corresponding author) gaurab.aryal@anu.edu.au
‡ CONICET and Universidad Nacional de Cuyo: florgabrielli@gmail.com

The first author acknowledges the financial support provided by the CBE research grant.
Estimating Revenue Under Collusion-Proof Auctions

GAURAB ARYAL† AND MARIA F. GABRIELLI‡

ABSTRACT. We propose a method to nonparametrically estimate the revenue under a auction that is efficient and resilient to collusion [Chen and Micali, 2012]. Efficiency is achieved on account of a lower revenue and we propose a method to quantify this efficiency-revenue trade-off, i.e. the extra cost for which efficient allocation can be guaranteed even when bidders collude. We implement a local polynomial estimation method on sample of California highway procurements data and find that to achieve efficiency the cost of procurement must increase by at least 10.8% and can go up to 48.8% depending on the size of bidding-ring.

Keywords: Public Procurement, Collusion-Proof Auction; Local Polynomial, Efficiency-Revenue Trade-off

JEL: C14, C4, C7, D44, L4.

1. INTRODUCTION

In this paper we propose and implement a method to estimate the efficiency-revenue trade-off of adopting a new auction [Chen and Micali, 2012] that guarantees efficient allocation even when bidders collude. Auction is one of the most widely used mechanisms to allocate goods and services. Its performance (efficiency and revenue), however, depends whether or not bidders collude and as shown by [Comanor and Schankerman, 1976; Feinstein, Block, and Nold, 1985; Lang and Rosenthal, 1991; Porter and Zona, 1993; Bajari, 2001; Porter and Zona, 1999; Pesendorfer, 2000; Harrington, 2008], collusion is prevalent in many markets. For instance, [Ausubel and Milgrom, 2006] showed that presence of even two collusive bidders can lead to an inefficient allocation for both first and second-price auctions. Moreover, since it is easier to sustain collusion in the latter than in the former, [Marshall and

† Date: December 10, 2012.

‡
an auctioneer should choose the former if the auctioneer suspects bidder collude, even though the two formats are revenue equivalent without collusion.\footnote{We interchange between “auction” and “procurements” and likewise between “revenue” and “cost”, without loss of generality because procurement is just the reverse auction.}

A perennial question for auctioneer is what rules to adopt that deters collusion. For example, an auctioneer could ban resale, whenever possible, or write a contract to that effect, if he/she believes that colluders use resale market to sustain collusion. In other words, auctioneers could learn how collusion works and design specific rules to deter that. This method, however, is infeasible as it requires an auctioneer to know the exhaustive list of ways to sustain. Some of the means could be quite complex as exemplified by \cite{Asker, 2008}. A feasible way, instead, could be to design statistical tests that can detect collusion. Even then, testing and detecting collusion is very difficult and precarious; see \cite{Bajari and Ye, 2003; Aryal and Gabrielli, 2012}.\footnote{\cite{Aryal and Gabrielli, 2012} implemented their tests and found no evidence of collusion even though they focused only on bidders who failed \cite{Bajari and Ye, 2003}'s tests.} In view of these difficulties, we should consider mechanisms and solution concepts that are resilient to collusion, that guarantee efficiency and perform well in terms of revenue/cost.

\cite{Chen and Micali, 2012} proposed a collusive dominant strategy truthfulness auction (henceforth, CM-auction) that is efficient. They achieve this by: (a) allowing bidders to bid and report their affiliation to any bidding ring;\footnote{CM-auction requires a punishing stage off the equilibrium path, where bidders pay a fine if their affiliation reports are inconsistent. For example bidder 1 says he colludes with bidder 2 but bidder 2 does not reciprocate. For more on coalition incentive compatibility see \cite{Green and Laffont, 1979}.} and (b) by adopting a variant of Vickrey auction wherein the winner bids lowest but gets a price that is equal to highest bid by the bidder who do not belong to the winner’s ring as reported in (a).\footnote{This idea of making the winning price determined outside of the coalition was also independently used by \cite{Aryal and Gabrielli, 2012} to design the test for collusion.} Efficiency is guaranteed under the auction because the auction provides incentives for truthful bidding – as the price is determined outside of the coalition – but at the same time it increases cost of procurements. But, by how much, the answer to which would help auctioneer decide whether or not to use CM-auction.
We view this as primarily an empirical question and propose a method to estimate this cost and hence the efficiency-revenue trade-off. Knowing the revenue/cost effect of implementing CM-auction is important for auctioneers when they choose an auction format. We implement the procedure in California highway procurements data. For the estimation of cost under CM-auction, we need both the cost distribution and the set of coalitions. We use asymmetric procurement model to map the observed bids to the unobserved cost following [Guerre, Perrigne, and Vuong, 2000]. Then we implement the tests proposed by [Bajari and Ye, 2003] to determine the set of colluding bidders. Out of twenty-five regular (type 1) bidders, fifteen of them bid on a pairwise basis more than handful times. Therefore in order to have enough observations for potential cartel members we treat them all as members of the same cartel. Our estimates show that the CM-auction increases the cost of total procurements by 48.8%. Since we treated all fifteen to be in one coalition, this cost should be interpreted as maximum possible increment: if there are more but smaller coalitions, competition will lower the extra cost. For this case, however, we use all the bids to recover the costs, therefore we ignore the fact that there is already collusion (according to the test) in the data and therefore the model of competition is potentially misspecified.

To correct this and to identify the cost, we also use another model that allows some bidders to collude. In particular, under the assumption that coalition is rational, i.e. all bidders maximize sum of total payoff, we only use the minimum bid from coalition members to estimate the bid distribution. This is because, the remaining bids are only “cover bids” that aim to lower competition and hence these bids need not directly correspond to the true cost, as is required for the identification. Furthermore, while we incorporate collusion in bid data we also exploit other features in the data, like the frequency of simultaneous bids amongst these fifteen bidders and conclude that only four out of fifteen bidders seem to be the most likely candidate colluders. So, in the second case we consider coalition of only
four bidders. Once we implement the CM-auction with this smaller coalition we find that the percentage increase in procurement falls to 10.8%.\footnote{The range of the trade-off is wide because in the data, and unlike in the theory, we do not know the true size and membership of coalition(s).} In other words, when we take into account the collusion in the data when we estimate the cost, and when the coalition is smaller, then the auctioneer can guarantee efficiency at the cost of 10.8%.

On a practical level we never know the exact nature of collusive rings, so we have settle with indirect approach, like the tests we used and the other features in the data. Therefore we report two estimates of costs (10.8% and 48.8%) in our attempt to be objective and show that the result does depends on how we account for collusive rings, i.e. how many coalitions, how big and so on, which are all empirical questions and should be explored case by case, for which one could follow the method outlined in this paper.

To estimate the bid distribution, instead of the widely used kernel-smoothed density estimators that are inconsistent at the endpoints of the support, we use the local polynomial estimation (henceforth, LPE) method of [Fan and Gijbels, 1996]. Using LPE automatically takes care of this boundary issue without trimming the data.\footnote{As far as we know, the only other paper to use LPE in empirical auction is [Gabrielli and Vuong, 2010], who use it to propose a $\sqrt{n}$– consistent semiparametric estimation method.}

This paper is organized as follows: Section (2) outlines the models, identification and estimation; Section (3) proposes the collusion-proof mechanism; Section (4) explains the data and the ways to determine the colluding bidders; Section (5) presents the empirical findings; Section (6) concludes. Appendix (A-1) explains the details about LPE and bandwidth selections. We also summarize the exact steps we used for estimation in Appendix (A-2). All tables and figures are collected in Appendix (A-3).

2. Model, Identification and Estimation

In this section we consider a procurement auction, i.e. a low-price sealed bid auction with asymmetric bidders: regular and fringe bidders. The section is divided into two subsections. The first subsection considers the
model with and without collusion and covers nonparametric identification. The second subsection deals with estimation.

2.1. Model and Identification. For every auction $\ell = 1, 2, \ldots, L$, a single and indivisible project is procured to $N_\ell \geq 2$ risk neutral bidders using first price sealed bids mechanism. The essential characteristic of the project for each auction is summarized by a random variable $X_\ell \in \mathbb{R}_{++}$, which for us will be the engineer’s estimate of the project. We assume that there are two types ($k = 0, 1$) of bidders with $n_{k\ell}$ type $k$ bidders, for auction $\ell$, such that $N_\ell = n_{0\ell} + n_{1\ell}$.\footnote{We abuse the notation to use $n_k$ as both the number and set of type $k$ bidders.} In every auction $\ell$ a type $k$ bidder draws his cost, i.i.d. across all other bidders, from $F_k(\cdot|X_\ell, N_\ell)$. Further, we assume that the costs are independent across auctions. Now, we consider two cases: competition and collusion.

2.1.1. Competition. The set of observables $W$ is

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

where $b_{ki}$ is the bid by type $k \in \{0,1\}$ bidder $i \in n_{k\ell}$. Then we make the following assumptions:

**Assumption 1. (A1)**

1. (Exogenous Participation) $F_k(\cdot|X, N) = F_k(\cdot|X)$ for $k = 0, 1$.
2. For each $\ell$ and each $k \in \{0, 1\}$ the variables $C_{k\ell i}, i \in n_{k\ell} \sim iid F_k(\cdot|\cdot)$ with density $f_k(\cdot|\cdot)$ conditional on $X_\ell$.
3. An auction $\ell$ has $N_\ell \in \{n, \bar{n}\}$ risk-neutral bidders with $n \geq 2$.
4. The three-dimensional vector $(X_\ell, n_{0\ell}, n_{1\ell}) \sim iid Q_m(\cdot, \cdot, \cdot)$ with density $q_m(\cdot, \cdot, \cdot)$ for all $\ell = 1, 2, \ldots L$.\footnote{We abuse the notation to use $n_{k\ell}$ to represent both the random variable and its realization and $Q(\cdot)$ is a product of absolutely continuous measure and a counting measure.}
5. The observed type $k \in \{0,1\}$ bids $B_k \sim iid G_k(\cdot|X_\ell, N_\ell)$ with density $g_k(\cdot|X_\ell, N_\ell)$. 
A strategy for bidder $i$ of type $k$ is a strictly increasing, type symmetric bidding strategy $s_k : [c, \bar{c}] \rightarrow [c, \bar{c}]$. Type $k$ bidder $i$ solves

$$\max_{b_i} \Pi_k(b_i, c_i, X_i, N_i) = \max_{b_i} (b_i - c_i) \prod_{j \in n_k \setminus \{i\}} (1 - F_k(s_k^{-1}(b_i)|X_i))^j \prod_{j \in n'_k} (1 - F_{k'}(s_{k'}^{-1}(b_i)|X_i))^j$$

$$= \max_{b_i} (b_i - c_i) \prod_{j \in n_k \setminus \{i\}} (1 - G_k(b_i|X_i, N_i))^j \prod_{j \in n'_k} (1 - G_{k'}(b_i|X_i, N_i))^j,$$

where $k \neq k' \in \{0, 1\}$ and $G_k(b|X_i, N_i) = F_k(s_k^{-1}(b)|X_i)$ is the probability that bidder $j \in n_k \setminus \{i\}$ will bid less than $b_i$ and likewise for $k'$. The first order condition for $i \in n_{k\ell}$ is

$$\frac{b_{ki} - c_{ki}}{(n_{k\ell} - 1) \frac{g_k(b_{ki}|X_i, N_i)}{1 - C_k(b_{ki}|X_i, N_i)} + n_{k'} \frac{g_{k'}(b_{ki}|X_i, N_i)}{1 - C_{k'}(b_{ki}|X_i, N_i)}}.$$  \hspace{1cm} (1)

This first order condition with the boundary conditions $s_k(\bar{c}) = \bar{c}, k = 0, 1$ uniquely characterizes optimal bidding strategy for all bidders. The model structure is the type specific conditional distribution of cost $\{F_k(\cdot|X_\ell)\}$ for $k = 0, 1$ given $X$. But since the data provide information on the characteristics of the project that is being procured, $X_\ell$ in the $\ell^{th}$ project, we can consider only the type specific conditional cost distribution $F_k(\cdot|X_\ell), k = 0, 1$ as the structural parameter. The question of identification is to ask if there are two pairs of cost distributions $\{F_0(\cdot|X_\ell), F_1(\cdot|X_\ell)\}$ and $\{F'_0(\cdot|X_\ell), F'_1(\cdot|X_\ell)\}$ that are observationally equivalent. Evaluating (1) at the estimated bid distribution and densities, we see that for each auction $\ell$, bid $b_{ki}$ uniquely determines the cost

$$\hat{c}_{ki} = b_{ki}^\ell - \frac{1}{(n_{k\ell} - 1) \frac{\hat{g}_k(b_{ki}^\ell|X_i, N_i)}{1 - \hat{C}_k(b_{ki}^\ell|X_i, N_i)} + n_{k'} \frac{\hat{g}_{k'}(b_{ki}^\ell|X_i, N_i)}{1 - \hat{C}_{k'}(b_{ki}^\ell|X_i, N_i)}},$$  \hspace{1cm} (2)

thereby identifying $\{F_0(\cdot|X_\ell), F_1(\cdot|X_\ell)\}$ that is consistent with the data.

2.1.2. Collusion. Now, we also consider a model where some of the type 1 bidders collude. We maintain all the afore mentioned assumptions. For every auction $\ell$, let $M_\ell \subset n_{1\ell}$ be the set of bidders who collude. We focus on efficient collusion where the colluders have access to a centralized mechanism that can control the bids placed by the members in the real auction. So

\footnote{The second equality follows from Assumption (A1.1)- exogenous participation. This mapping between bids and valuation distribution under equilibrium condition is due to [Guerre, Perrigne, and Vuong, 2000].}
there will be only one serious bid in each auction from the bidders in \( M_\ell \) and the rest will be just cover-bids. We also assume that the bidders outside the ring are unaware about the ring. This means that when we map the observed bids to the underlying cost we consider only the minimum bid \( b_{1\ell}^* \) among the bidders in \( M_\ell \) and the bids by everyone outside \( M_\ell \) to estimate the type 1 bid distribution and density \( G_1^*(\cdot|\cdot,\cdot) \) and \( g_1^*(\cdot|\cdot,\cdot) \). Then, we can use these estimates instead of \( G_1 \) and \( g_1 \) in (2) to recover the pseudo cost for each non-members. But for the the minimum bid we use

\[
\hat{c}_{1\ell} = b_{1\ell}^* - \frac{1}{(n_{1\ell} - (|M_\ell| - 1))} \frac{g_1^*(b_{1\ell}^*|X_\ell,N_{r\ell})}{1 - G_1^*(b_{1\ell}^*|X_\ell,N_{r\ell})} + n_0 \frac{g_0(b_{1\ell}^*|X_\ell,N_{r\ell})}{1 - G_0(b_{1\ell}^*|X_\ell,N_{r\ell})},
\]

where the main difference from the non-colluders is that while the non-colluders think they are competing with \( n_{1\ell} \) type-1 bidders and \( n_{0\ell} \) type 0 bidders, the ring knows that it is competing with only \( n_{1\ell} - (|M_\ell| - 1) \) bidders. For type 0 bidders, just like with the type 1 non-colluders, the only difference is that the appropriate type 1 bid distribution and density are, respectively, \( G_1^*(\cdot|\cdot,\cdot) \) and \( g_1^*(\cdot|\cdot,\cdot) \).

2.2. Estimation. In the first step we estimate the conditional bid distributions \( G_k(\cdot|X,N) \) and the bid densities \( g_k(\cdot|X,N) \) given the engineer’s estimate \( X \) and the set of bidders \( N \), using Local Polynomial Estimation (LPE) method, see [Fan and Gijbels, 1996; Gabrielli and Vuong, 2010].

Consider a bivariate i.i.d. data \( \{X_i, Y_i\}_{i=1}^n \). Our interest is the regression function \( m(x_0) \) and its derivatives \( m'(x_0), m''(x_0) \) and so on till \( m^p(x_0) \). Hence, we regard the model \( E[Y|X] = m(X) \). Under the assumption that \((p+1)^{th}\) derivative of \( m(\cdot) \) at \( x = x_0 \) exists, LPE can approximate \( m(\cdot) \) by a polynomial of order \( p \). Taylor expansion gives

\[
m(x) \approx \sum_{j=0}^{p} \frac{m^j(x_0)}{j!} (x - x_0)^j,
\]

and this polynomial is fitted locally by a weighted least squares regression that minimizes

\[
\sum_{i=1}^{n} \left( Y_i - \sum_{j=0}^{p} \beta_j (x - x_0)^j \right)^2 K_h(X_i - x_0),
\]
where $h$ is the bandwidth, $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$ with $K$ a kernel function. If $\hat{\beta}_j, j = 0, \ldots, p$ is the solution to the weighted least squares then it is clear that $j!\hat{\beta}_j(x_0)$ is the estimator for $m^j(x_0), j = 0, \ldots, p$. For us, $Y$ will be the indicator function and hence $\beta_0(\cdot)$ will be the LPE estimator of the conditional bids distribution and its first derivative $\beta_1(\cdot)$ will be the corresponding density. The exact form used for our estimation is given in Appendix (A-1). We make the following assumptions for estimation.

**Assumption A3:**

(i) The kernels $K_G(\cdot), K_{0g}(\cdot)$ and $K_{1g}(\cdot)$ are symmetric with bounded hyper-cube supports and twice continuous bounded derivatives with respect to their arguments,

(ii) $\int K_G(x)dx = 1, \int K_{0g}(x)dx = 1, \int K_{1g}(b)db = 1$

(iii) $K_G(\cdot), K_{0g}(\cdot)$ and $K_{1g}(\cdot)$ are of order $R - 1$. Thus moments of order strictly smaller than $R - 1$ vanish.

**Assumption A4:** The bandwidths $h_G, h_{1g}$ and $h_{2g}$ satisfy

(i) $h_G \to 0$ and $\frac{Lh_G^d}{\log L} \to \infty$, as $L \to \infty$,

(ii) $h_{0g} \to 0$, $h_{1g} \to 0$ and $\frac{Lh_{0g}^d h_{1g}}{\log L} \to \infty$, as $L \to \infty$.

From this assumption it is clear that it is possible to choose the optimal bandwidths, i.e. the bandwidths proposed in [Stone, 1982]. Unlike GPV we do not need to specify a “boundary bandwidth” since the local polynomial method does not require knowledge of the location of the endpoints of the support. Therefore, it is not necessary to estimate the boundary of the support of the bid distribution. This is necessary when one needs to trim out observations, which we do not given that our estimator is not subject to the so-called boundary effect. We have 3 conditioning variables, one that is continuous and two that are discrete. Thus, we have to adapt the definition of the LPE to the present case. However, the discrete variables do not affect the asymptotic properties of the estimator, so in order to choose the optimal bandwidth the relevant number of covariates to consider is the number of continuous variables. \(^{10}\)

\(^{10}\)Similar observation is made by [Abadie and Imbens, 2006].
We will denote by $p$ the number of continuous variables and by $d$ the total number of conditioning variables. For our application, $d = 3$ and $p = 1$. Let $\hat{\psi} = \frac{\hat{g}(\cdot, \cdot)}{1 - \hat{G}(\cdot, \cdot)}$ be the estimator of $\psi(\cdot, \cdot, \cdot) = \frac{g_k(\cdot, \cdot, \cdot)}{1 - G_k(\cdot, \cdot, \cdot)}$. From Proposition 1 in [Guerre, Perrigne, and Vuong, 2000] we know that $G_k(\cdot, \cdot)$ is $R + 1$ times continuously differentiable on its entire support and therefore $g_k(\cdot, \cdot)$ is $R$ times continuously differentiable on its entire support as well. Given the smoothness of each function we propose to use a LPE ($R$), i.e. a LPE of degree $R$, for $G_k(\cdot, \cdot, \cdot)$ and a LPE ($R - 1$) for $g_k(\cdot, \cdot, \cdot)$. Following [Fan, Gasser, Gijbels, Brockmann, and Engel, 1993] we can show that the bid distribution is consistent and following [Guerre, Perrigne, and Vuong, 2000] it is easy to see that the estimated costs are strongly consistent. The exact econometrics model and the selection of optimal bandwidth are explained in the Appendix A-1.

3. COLLUSION PROOF MECHANISM

We begin with an example, adopted from [Chen and Micali, 2012] and for more formal and through treatment we direct the readers to that paper.

**Example 1.** Consider 4 risk neutral bidders with cost $(1, 1, 100, 100)$ and the first and second bidders collude together. Furthermore, suppose that they have a wrong belief about the types of the other two bidders. In particular they believe that their respective costs are $(0.1, 2)$. Under competitive second price auction one of the two bidders will sell the object at 100. Under collusion they could adopt the following strategy: 1 bids 1 and 2 bids 100. The real bids will then be $(1, 100, 100, 100)$ and hence the coalition gets a surplus of 99 which can be shared easily between the two.

Now, suppose bidders report their bids (the price at which they are willing to supply the public good) and their coalition membership, if any. The winner is the bidder who announces the lowest price and the price is equal to the second highest price outside of the winner’s coalition, i.e. the price paid by a coalition is not controlled by them. To see this, consider the following announcement $(1, \{1, 2\})$, $(1, \{1, 2\})$, $(100, \{3\})$ and $(100, \{4\})$. This ensures that strategy proof-ness and the object is sold to coalition for 100.

---

11From Proposition 1 by [Guerre, Perrigne, and Vuong, 2000], we also know that the conditional density $g_0(\cdot, \cdot, \cdot)$ is $R + 1$ times continuously differentiable on a closed subset of the interior of the support and thus the degree of smoothness closed to the boundaries and at the boundaries of the support is not $R + 1$. 
The example shows that the mechanism is a slight variation of classic second price auction, now with respect to coalitions rather than a singleton. Let there be \( N < \infty \) risk-neutral bidders in an independent private value, low bid auction. Each bidder draws i.i.d cost \( C \sim F(\cdot) \).\(^{12}\) Let \( \mathcal{C} \) represent the partition of players such that each element of the partition represents a coalition such that every singleton \( \{i\} \in \mathcal{C} \) is an independent bidder, and \( M \) is a generic element. The set was \( \{\{1,2\},\{3\}\} \) in the example. Let \( \mathcal{M} = \{N, F(\cdot), \mathcal{C}\} \) be the context of the game and is commonly known by all the bidders. Moreover, we assume that for every coalition \( M \in \mathcal{C} \), the \( |M| \)-tuple cost profile \( C_M = \{C_i : i \in M\} \) is common knowledge only amongst the bidders in that coalition. The seller, however, only knows \( \{N, F(\cdot)\} \).

Let \( \{A_i, P_i\}_{i=1}^N \) be an allocation and pricing rule, where \( A_i \in \{0,1\} \) such that \( \sum_{i \in N} A_i = 1 \) and \( P_i \) is the price paid by the bidder \( i \). We assume that all coalitions are efficient and hence when the ex-post utility of a bidder \( i \) is \( (P_i - C_i)A_i \), the utility of the coalition \( M \) is the sum across the members, i.e. \( u_M = \sum_{i \in M}(P_i - C_i)A_i \). Each member \( i \in M \) acts to maximize \( u_M(\cdot) \).

**Definition 1.** An auction, for a context \( \mathcal{M} \), is directly collusive if the set of pure strategies for \( i, s_i(\cdot) \) consist of the set of all mapping from \( C \mapsto (C, M) \).

So, a bidder with cost \( C \) reports his cost and the coalition \( M \). Let \( u_M(s) \) denote the total utility of coalition \( M \) when everyone uses symmetric bidding strategy \( s(\cdot) \). Then we are in a position to define dominant-strategy truthfulness and coalitional rationality.

**Definition 2.** An auction is collusively dominant-strategy truthful if, for all coalition \( M \in \mathcal{C} \) and all strategy profiles \( s_M := \{s_i(\cdot) : i \in M\} \) and \( s_{-M} := \{s_j(\cdot) : j \notin \mathcal{C}\setminus\{M\}\} \),

\[
\forall i \in M : u_M((C_i, M), s_{-M}) \geq u_M((C'_i, M'), s_{-M}).
\]

and is coalitionally rational if \( u_M((C_i, M), s_{-M}) \geq 0 \).

Let \( s(\cdot) = \{(C_1, M_1), \ldots, (C_N, M_N)\} \) be an action profile. Then a disagreement (in \( s(\cdot) \)) is an ordered pair \((i,j)\) such that \( M_i \ni j \) but \( M_j \not\ni i \). In other

\(^{12}\)For notational ease, we treat all bidders to be symmetric, extending it to asymmetric bidders is straightforward.
words, we say that \((i, j)\) is disagreement if \(i\) claims to be a part of collusion ring \(M_i\) that contains \(j\) but \(j\) does not reciprocate. Given a profile \(s(\cdot)\) the outcome \((A, P)\) is computed as follows: First, there is the punishing phase if there is any disagreement, in which case \(A_i = 0\) for all \(i\). Then to determine the price we start with \(P_i = 0\) and for each disagreement \((i, j)\) charge \(P_i = P_i + 2t\) and \(P_j = P_j - t\), while keeping \(t\) with the seller.\(^{13}\) Second, when there is no disagreement we initiate the standard phase where from the reported coalitions the coalition partition \(C\) is constructed. Then lowest bidder wins the auction and determine the winning coalition \(M^*\) and charge

\[
P_i = \begin{cases} 
0 & \text{if } A_{M^*} = \sum_{j \in M^*} A_j = 0 \\
C_{1:(N \setminus M^*)} & \text{otherwise}
\end{cases}
\]

to every bidder \(i\), where \(C_{1:(N \setminus M^*)}\) is lowest bid from the bidder \(j \not\in M\).

**Theorem 2.** [Chen and Micali, 2012] The mechanism outlined above is (a) Collusive dominant-strategy truthful; (b) Coalitional rational and ; (c) Efficient.

Now, in the next section we analyze the data and determine the sets of coalitions that are used for the counterfactual exercise later.

## 4. Data

The aim of this section is to explain the main features of the data and to explain how we determine the colluding rings. The data consist of the Highway procurements in the state of California between January 2002 and January 2008, 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).\(^{14}\) The data include information about the characteristics of the projects that were let, the name of bidders and their bid in each auction. The timing is as follows. First, during the *advertising period* that lasts between three to ten weeks depending on the size of the project, the Caltrans Headquarters Office Engineer announces a

---

\(^{13}\)This punishment phase is off the equilibrium path and does not affect the estimation results.

\(^{14}\)The data is publicly available at [http://www.dot.ca.gov/hq/esc/oe/awards/bidsum/](http://www.dot.ca.gov/hq/esc/oe/awards/bidsum/).
Potential bidders express their interest by buying the project catalogue. Second, sealed bids are received only from among the potential bidders. Third, on the \textit{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.

We divide bidders into two broad types of asymmetric bidders: the fringe bidders (type 0) who bid in small projects and are infrequent and the main bidders (type 1) who participate frequently and in bigger projects. The private cost is interpreted as a reduced form of real cost of production and depends on many unobservable characteristics of the bidder. The data consist of 2,152 projects awarded by Caltrans for a total of $7,645 millions but we focus only on 1,907 projects that had at least two bidders. Of all bidders, only 823 bidders bid at least once. In the remaining subsection we determine the set of bidders who fail the tests proposed by [Bajari and Ye, 2003]: the colluders. We find that the fifteen type 1 bidders, who bid simultaneously more often on a pairwise basis, fail at least one of the conditions by [Bajari and Ye, 2003] for competitive bidding. Since we do not know the true collusive ring, we further explore some other features in the data in a hope to narrow the set of members in the ring. Looking at the frequency of their bids and their winning patterns we narrow the coalition to only four bidders. The exact process is explained in the reaming of this section. Main difficulty with this exercise is to determine the set of colluders, so the way we determine these two sets should be taken as suggestive and exploratory.\footnote{Some examples of projects include asphalt repaving, road paving, bridge reconstruction, striping the highway, constructing, replacing and widening bridges, storm damage repair, etcetera.}

4.1. \textbf{Bidding Ring}. To identify the ring members, we focus only on the projects that are worth between $1 million and $20 million as smaller projects are unlikely to be worth the risk. There are 724 such projects that worth $2,408 millions (31\% of the total) with 413 bidders out of which 202 win at\footnote{The main point is, we can either use the tests from the literature and be agnostic about the nature of the collusive ring or try and explore some other features besides the test to determine the ring, in a hope that the exploration helps us find the “true” collusive rings.}
least once. Furthermore, following the literature we define regular bidders as bidders who have a nontrivial revenue share (at least 1%) in the market. Twenty-five bidders satisfy this criteria and will henceforth be called the type 1 bidders and the remaining bidders are the fringe (type 0) bidders—see Table A-1. The first column is the index of the bidders while the second column gives the number of bids of each of them. To assess 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. Expected number of wins is defined as follows: consider A, who bids on a total of 50 projects against a varying number of bidders, \( n_\ell \) for \( \ell = 1, \ldots, 50 \). Then his expected win is defined to be \( \sum_{\ell=1}^{50} 1/n_\ell \). By comparing third and fourth column, 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 bidder and the sixth column reports the revenue share—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 contains the summary statistics from which we can conclude: (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; (iii) money on the table—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. In general higher valued projects (between $1 million and $20 millions) attract relatively fewer bidders, suggesting that it is the main bidders who can gain the most by colluding and moreover, larger projects are more profitable, all else equal.

Now, we follow the tests proposed by [Bajari and Ye, 2003] on this subsample of bids for twenty five type 1 bidders. The basic idea behind the tests is to detect those bidders whose bidding pattern systemically violate competitive and independent bidding. To increase the likelihood of picking a coalition we give more emphasis to bidders who participate in the
same auction because as the theoretical literature suggests ring members tend to participate in the same auctions to enforce the bidding agreement. To this end, we consider all combinations of pairs and select those bidders that have at least fifteen simultaneous bids, see Table A-3. There are fifteen bidders who bid frequently together.\footnote{This cutoff is based on the data and is big enough to capture the simultaneous bidding but not too big so that we have enough observations left for the test.}

First, to test independence we consider the fifteen pairs of bidders bidding frequently described above and estimate the following models for fifteen type 1 bidders and the remaining bidders, respectively

\begin{align*}
  \text{BID}_{i\ell}/\text{EE}_{\ell} &= \gamma_0 + \gamma_1 \text{LDIST}_{i\ell} + \gamma_2 \text{CAP}_{i\ell} + \gamma_3 \text{UTIL}_{i\ell} + \gamma_4 \text{LMDIST}_{i\ell} + u_{i\ell} \quad (4) \\
  \text{BID}_{i\ell}/\text{EE}_{\ell} &= \alpha_0 + \alpha_1 \text{LDIST}_{i\ell} + \alpha_2 \text{CAP}_{i\ell} + \alpha_3 \text{UTIL}_{i\ell} + \alpha_4 \text{LMDIST}_{i\ell} + \zeta_{i\ell} \quad (5)
\end{align*}

Here LDIST$_{i\ell}$ is the logarithm of distance, LMDIST$_{i\ell}$ is the logarithm of the minimum of distances of all bidders (except $i$) to the project $\ell$ and UTIL$_{i\ell}$ is the utilization rate of the capacity. We define the utilization rate as Util$_{it} = \text{Backlog}_{it}/\text{Capacity}_i$, where backlog is defined as the past projects that were won but yet to be completed and the capacity is the total capacity of bidder. We find that approximately 60% of bids are explained by capacity, although the effect varies across bidders; for more on the effect of capacity utilization on bidding behavior see \cite{Jofre-Bonet and Pesendorfer, 2003}. For the bidders listed in Table A-3 (those who participate frequently) we estimate (4) with bidder-varying coefficients and for the rest we use (5).

Second, we test exchangeability

\begin{align*}
  H_0 : (\forall i, j, i \neq j), (\forall s \in \{1, 2, 3, 4\}) & \quad \beta_{is} = \beta_{js} \\
  H_A : (\exists i, j, i \neq j), (\exists s \in \{1, 2, 3, 4\}) & \quad \beta_{is} \neq \beta_{js}
\end{align*}

at both market level by pooling the fifteen bidders in one group and on a pairwise basis. Let $T = 3,347$ be the number of observations, $m$ the number of regressors and $k$ the number of constraints implied by $H_0$. Then under the
null the test statistic $F = \frac{(SSR_C - SSR_U)/r}{SSR_U/(T - m)} \Rightarrow^d F(r, T - m)$. At the market level, exchangeability hypothesis imposes that the effect of the four explanatory variables is the same for both potential ring members and the remaining bidders. Since there are fourteen dummies (indexing the bidders) and for each case there are four restrictions (under null), the total number of restrictions imposed under the null is $k = 56$. Here $m = 748$ and $n - m = 2599$ and the estimated $F-$ statistic is 5.934 with the upper tail area equal to 0.0000. Therefore we reject the null of exchangeability when comparing the fifteen bidders (potential cartel members) against the remaining bidders. The assumption thus far is that all fifteen bidders form one single coalition and in our counterfactual exercise of case 1, when we say potential colluders we mean these fifteen bidders.

However, sustaining such a large coalition might be difficult. To see if we can reduce the size of the coalition, 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). See Table A-5 for details. Comparing the “expected win” with the actual win for these pairs, we do see that at least one member of the pair wins often. Comparing Table A-1 and Table A-3 we can conclude that: (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 bidders (A,D,E, P) could be considered as potential collusive ring. 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 often than others. And since the triplet (A,D,E) also fit the collusive behavior, we consider colluding bidders to be (A, D, E, P).\[18]
5. Estimation Results

In this section we present the main findings from estimation of the pseudo-cost and from the counterfactual exercise of implementing CM-auction. As we mentioned in the sections (2) and (4.1), we have two sets of type specific costs and also two sets of colluders. The first case uses all the bids data to recover cost and finds that fifteen out of twenty-five regular bidders could be colluding according to [Bajari and Ye, 2003]. The second case takes into account the fact that some bidders might already be colluding. To determine the set of colluders we use the same tests as above and some other data features, which allows us to identify a set of four bidders. In every auction we discard all but the minimum bid by these four colluders to estimate the cost.\footnote{As mentioned earlier we assume that the coalition is rational and maximizes the total payoff. This means only the bid corresponding to the lowest cost will be serious, the rest will only be “cover-bids.”}

Before we estimate the counterfactual cost, we test if the estimated costs support asymmetry amongst two types. To that end, we use the estimated cost to test if the two empirical CDFs are the same or different:

\[ H_0 : \forall c \in [c_0, c] \quad F_0(c) = F_1(c); \quad H_1 : \exists c \in [c_0, c] \quad F_0(c) \neq F_1(c), \]

using the classic Kolmogorov-Smirnov test. The test statistic when we use all the bids data is

\[ KS_L = \sup_{c \in [c_0, c]} \left| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{n_{0\ell}} \sum_{i=1}^{n_{0\ell}} 1\{\hat{c}_{0i} \leq c\} - \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{n_{1\ell}} \sum_{i=1}^{n_{1\ell}} 1\{\hat{c}_{1i} \leq c\} \right|, \]

and the test statistic when we discard cover bids is similar except that now we have fewer data for type 1. Since we do not observe the costs but only their estimates we use Bootstrap to get the correct \( p \)– value. The \( p \)– values for both the cases, computed using 10,000 replications, are zero, so it confirms that the cost distributions are indeed asymmetric.\footnote{The test was implemented in R; see [Sekhon, 2011].} Similarly we find that the distribution of markup (see Figures 1 and 2), defined as difference between bid and cost, are stochastically different across types.

Now we are ready to compute the difference in cost from implementing the CM-auction. Let \( M_\ell \) be the set of colluders who are present in auction
and let $i_\ell \in N_\ell$ be the winner and $\tilde{b}_\ell$ the corresponding bid. Let $a_\ell$ be the highest cost amongst all bidders participating in auction $\ell$ who do belong to the coalition $M_\ell$. That is if the winner belongs to the coalition, i.e. $i_\ell \in M_\ell$, then $a_\ell = \max_{j \in n_\ell \setminus M_\ell} \hat{C}_j$ and if the winner is not a member of the coalition then we can set $a_\ell = \tilde{b}_\ell$, the real winning bid. Then the difference between CM-auction and the data is $r_\ell = a_\ell - \tilde{b}_\ell$ if the winner is a cartel member and $r_\ell = 0$ otherwise. Once we compute the change in cost $r_\ell$ for all auctions the total change in cost of procuring is just $\sum_\ell r_\ell$. For the first case the total set of colluders is the fifteen bidders (see section 4.1) and $M_\ell = M \cap n_1$. Likewise for the second case we find that only $M = \{A, B, D, E\}$ bidders are consistent with collusion and hence $M_\ell = n_1 \cap \{A, B, D, E\}$. Figure 3 shows the empirical CDF of $r_\ell$ for these two cases.

We find that, in the first case with large coalition, implementing CM-auction increases the total cost by 48.8%, while for the second case the cost increases by 10.8%. This difference in cost is not surprising given the difference in size of the two rings.

6. Conclusion

In this paper we study the implication of adopting an efficient auction that is resilient to collusion proposed by [Chen and Micali, 2012] on the total cost of procurements. Using the data from California highway procurements, we recover the unobserved costs of bidders and implement the new collusion proof auction in a counterfactual exercise. We find that efficient allocation can be achieved at a cost of either 10.8% or 48.8% increase to current (total) cost of procurement.

We have two values because we do not know the set of colluders and had to use indirect methods. This shows that, as expected, larger ring leads to larger cost, but the decision to model collusion should be specific to the data.

In conclusion, it is our aim to provide a way to construct such measures for many widely available data sets and will help auctioneer and mechanism designers alike, in choosing auction format that is efficient and resilient to collusion.
APPENDIX

A-1. Estimation

In this section we outline the estimation problem and discuss the choice of bandwidths and kernels. To account for the skewness in the bid distribution, a widely observed problem encountered with auction data, we use logarithmic transformation. For notational simplicity we suppress the dependence of the distributions on \((X, N)\), unless otherwise noted. Log transformation of (2) gives

\[
c_{kM} = \xi_k(d_k, n) = e^{d_k} - \frac{e^{d_k}}{(n_k - 1) \frac{g_{kd}(d_k, \cdot)}{1-G_{kd}(d_k, \cdot)} + n_1 \frac{g_{id}(d_k, \cdot)}{1-G_{id}(d_k, \cdot)}}
\]

(6)

where \(d_k = \ln(b_k)\) and \(G_{kd}(\cdot, \cdot), g_{kd}(\cdot, \cdot)\) are the distribution and density of \(\log(b_k)\) for type \(k\). Define \(K_H(u) = |H|^{-1}K(H^{-1}u)\), where \(H\) is a non-singular \(d \times d\) matrix, the bandwidth matrix that usually takes the form \(H = hI_d\) and \(|B|\) denotes its determinant. The observations are given by \(\{(Z_i^T, Y_i) : i = 1, \ldots, n\}\) with \(Z_i = (X_i, N_{0i}, N_{1i})^T\). Let \((x, n_0, n_1)\) be a point in \(\mathbb{R}^3\). The estimators involved are, as mentioned above Local Polynomial Estimators. For our application, \(R = 2\) and therefore we implement a LPE(2) for each cdf involved and a LPE (1) for each pdf involved. Let \(Y^G_{p \ell} = \mathbb{I}(B_{p \ell} \leq b)\). Using a local quadratic approximation to estimate each cdf implies obtaining the solution to the following least squares minimization problem

\[
\sum_{\{\ell: I_{\ell} = i\}} \sum_{p=1}^{L} \left\{ Y^G_{p \ell} - \left[ \beta_0 + \beta_1(X_{p \ell} - x) + \beta_2(N_{1 \ell} - n_1) + \beta_3(N_{0 \ell} - n_0) \\
+ \beta_{11}(X_{p \ell} - x)^2 + \beta_{12}(X_{p \ell} - x)(N_{1 \ell} - n_1) + \beta_{13}(X_{p \ell} - x)(N_{0 \ell} - n_0) \\
+ \beta_{23}(N_{0 \ell} - n_0)(N_{1 \ell} - n_1) + \beta_{22}(N_{1 \ell} - n_1)^2 + \beta_{33}(N_{0 \ell} - n_0)^2 \right] \right\}^2 K_H(z - z)
\]

with respect to \(\beta_G = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{22}, \beta_{23}, \beta_{33})\). In particular we are interested in \(\hat{\beta}_0 = G(b|x, n_0, n_1)\); see [Fan and Gijbels, 1996]. Then, we know from the least squares theory that \(\hat{\beta}_G = (Z_G^T W_G Z_G)^{-1} Z_G^T T_G Y\), where the design matrix \(Z_G\) for the local quadratic case (what we use) is
For the densities involved define $Y_{p\ell}^g = \frac{1}{n_{2g}} K_{2g} \left( B_{p\ell} - b \right)$. We use a local linear estimator, i.e. LPE(1) which, as before, is obtained as the solution to the following least squares problem

$$\sum_{\{i: h = i\}} \sum_{p=1}^L \left\{ Y_{p\ell}^g - \beta_0 + \beta_1 (X_{p\ell} - x) + \beta_2 (N_{1\ell} - n1) + \beta_3 (N_{0\ell} - n0) \right\}^2 K_H (Z - z)$$

It is well known that $\hat{\beta}_g = (Z^T T_g Z)^{-1} Z^T T_g Y$. The design matrix $Z$ for the local linear case is

$$Z = \begin{pmatrix}
1 & (X_{1,1} - x) & (N_{0,1} - n0) & (N_{1,1} - n1) \\
1 & (X_{1,n1} - x) & (N_{0,n1} - n0) & (N_{1,n1} - n1) \\
\vdots & \vdots & \vdots & \vdots \\
1 & (X_{1,1} - x) & (N_{1,1} - n1) & (N_{0,1} - n0)^2 (N_{1,1} - n1)^2 \\
\vdots & \vdots & \vdots & \vdots \\
(X_{1,1} - x)(N_{1,1} - n1) & (N_{0,1} - n0)(N_{1,1} - n1) & (N_{0,1} - n0)^2 (N_{1,1} - n1)^2 & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
(X_{1,n1} - x)(N_{1,n1} - n1) & (N_{0,n1} - n0)(N_{1,n1} - n1) & (N_{0,n1} - n0)^2 (N_{1,n1} - n1)^2 & \vdots \\
\end{pmatrix}$$

The corresponding weighting matrix for each estimation procedure are $T_g = \text{diag}\{K_H (Z_i - z)\}$ and $T_g = \text{diag}\{K_H (Z_i - z)\}$, respectively. The bandwidths and kernels involved for distributions and densities are different.

### 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.

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

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 (\cdot):= \ldots \ldots $
The kernels for continuous variables should be symmetric with bounded supports, so we decided to use the Epanechnikov Kernel function \( K(u) = \frac{3}{4} (1 - u^2) \mathbb{I}(|u| \leq 1) \), as it is an optimal Kernel in the sense that it minimizes the asymptotic mean squared error over all non-negative functions [Fan, Gasser, Gijbels, Brockmann, and Engel, 1993]. 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.\(^{21}\) We assume the smoothness parameter \( R = 2 \) for the cost distribution. To ensure uniform consistency at the optimal rates, the bandwidths for the continuous variables are chosen to be \( h_g = 1.06 \times 2.214 \times \hat{\sigma} \times (T)^{-1/(2R+1)} \), \( h_G = 1.06 \times 2.214 \times \hat{\sigma} \times (T)^{-1/(2R+3)} \). The constant term comes from the so-called rule of thumb and the factor 2.978 is the one corresponding to the use of Epanechnikov Kernels instead of Gaussian Kernels; see [Hardle, 1991].

A-2. Steps For Estimation

To make the estimation procedure transparent we outline the steps required to estimate the (pseudo) costs:

1. Case 1: Competition case where collusion comes only for counterfactual.
   
   (a) Choose appropriate Kernels and determine the optimal bandwidths, see Section (A-1.1).
   
   (b) Use \( \{ \{ b_0 \}_{i=1}, X_\ell \}_{\ell=1}^L \) to estimate \( \hat{g}_0(b|X_\ell,n_0 \ell) \) and \( \hat{G}_0(b|X_\ell,n_0 \ell) \) on \( [\min b_0, \max b_0] \) such that \( \hat{g}_0(b|\cdot, \cdot) = 0 \) if \( b \notin [\min b_0, \max b_0] \).
   
   (c) Suppressing the conditioning variables, define \( \hat{g}_0(b) = \max(0, \hat{g}_0(b)) \) and \( \hat{G}_0(b) = \min(1, \hat{G}_0(b)) \).
   
   (d) Repeat Steps (3) and (4) for \( \{ \{ b_1 \}_{i=1}, X_\ell \}_{\ell=1}^L \) to estimate \( \hat{g}_1(b|X_\ell,n_1 \ell) \) and \( \hat{G}_1(b|X_\ell,n_1 \ell) \).
   
   (e) Use the estimates \( \{ \hat{G}_0, \hat{g}_0 \} \) evaluated at \( b_0 \) and \( \{ \hat{G}_1, \hat{g}_1 \} \) evaluated at \( b_1 \) to interpolate:

\(^{21}\)There are no theoretical restrictions to the kernels applied to discrete variables.
For every \( b_{0i} \) from the type 0 bids, find the highest lower bound of \( b_{1i} \) from type 1 bids data such that \( \bar{b}_{1i}(b_{0i}) \leq b_{0i} \leq \bar{b}_{1i}(b_{0i}) \).

(i) Determine the weight \( w_{0i} = \frac{b_{0i} - \bar{b}_{1i}(b_{0i})}{\bar{b}_{1i}(b_{0i}) - \bar{b}_{0i}(b_{0i})} \).

(ii) Then define \( \hat{g}_1(b_{0i}) = w_{0i}\hat{g}_1(\bar{b}_{1i}(b_{0i})) + (1 - w_{0i})\hat{g}_1(\bar{b}_{0i}(b_{0i})) \).

(iii) Similarly, determine \( \hat{G}_1(b_{0i}) = w_{0i}\hat{G}_1(\bar{b}_{1i}(b_{0i})) + (1 - w_{0i})\hat{G}_1(\bar{b}_{0i}(b_{0i})) \).

(iv) Repeat (a) - (d) for all bids \( b_{0i} \) not in the domain of observed range of bids \( b_{1i} \).

(v) Repeat (a) - (e) for bids \( b_{1i} \).

(f) The corresponding estimates of the (pseudo) costs are

\[
c_{0i}^\ell = \begin{cases} 
  b_{0i} - \frac{1}{(n_0 - 1)\frac{\hat{g}_0(b_{0i})}{\hat{g}_1(b_{0i})} + n_1\frac{\hat{g}_1(b_{0i})}{\hat{g}_1(b_{0i})}} & \text{if } b_{0i} \in [\min_i b_{1i}, \max_i b_{1i}], \\
  b_{0i} - \frac{1}{(n_0 - 1)\frac{\hat{g}_0(b_{0i})}{\hat{g}_1(b_{0i})} + n_1\frac{\hat{g}_1(b_{0i})}{\hat{g}_1(b_{0i})}} & \text{o/w,}
\end{cases}
\]

\[
b_{1i} - \frac{1}{(n_1 - 1)\frac{\hat{g}_1(b_{1i})}{\hat{g}_1(b_{1i})} + n_0\frac{\hat{g}_0(b_{1i})}{\hat{g}_1(b_{1i})}} & \text{if } b_{1i} \in [\min_i b_{0i}, \max_i b_{0i}], \\
  b_{1i} - \frac{1}{(n_1 - 1)\frac{\hat{g}_1(b_{1i})}{\hat{g}_1(b_{1i})} + n_0\frac{\hat{g}_0(b_{1i})}{\hat{g}_1(b_{1i})}} & \text{o/w,}
\end{cases}
\]

(g) Then we have \( \{c_{0i}^\ell\}_{i=1}^{n_0}, \{c_{1i}^\ell\}_{i=1}^{n_1} \), \( \ell = 1, 2, \ldots, L \) pseudo-cost vectors.

(h) To implement the counterfactual, we do the [Bajari and Ye, 2003] tests on the type 1 bidders. We determine 15 bidders whose bidding is at-odds with independent bidding and competition.

(2) Case 2: Estimating with Collusion:

(a) We begin with \( n_0 \) type-0 and \( n_1 \) type-1 bidders.

(b) We follow [Aryal and Gabrielli, 2012] and identify four bidders \( M = \{A, B, D, E\} \) who are treated as colluders and take this into account while estimating pseudo-costs.

---

22 We have now 4 sets, the cdf/pdf for both types evaluated at their corresponding data and then the remaining 2 sets that are determined from interpolation of the estimated cdf/pdf.
Estimating Revenue Under Collusion-Proof Auctions

23

c) Determine $M_\ell = M \cap n_{1\ell}$, the set of bidders in auction $\ell$, for
every auction.

d) Since type-0 case is unaltered, repeat 1(a)-1(d) to determine $\{\hat{G}_0, \hat{g}_0\}$

e) Determine the collusive bids for every auction, i.e. the set $\{b_{1i} : i \in M_\ell\}, \ell = 1, \ldots, L$, and for every auction determine $\{b^*_1\}_\ell^L$.

Then, the effective type-1 bidder from the point of view of colluding bidders is $n_{1\ell}^* = n_{1\ell} - (|M_\ell| - 1)$ as the coalition $M_\ell$ is effectively treated as a single bidder, while the number of type-1 bidders for those outside $M_\ell$ is still $n_{1\ell}$. So we consider two sub-cases: 23

f) Colluders:

(i) In every auction we discard all but the minimum bid of the ring. Let $\{b^*_{1i}\}_{i=1, 2, \ldots, L}$ be the type 1 bids.

(ii) Using this set repeat 1(e) to estimate the pair $\{\hat{G}^*_1, \hat{g}^*_1\}$.

(iii) Using the estimates from 2(d) and 2(f(ii)), the steps 1(f) and 1(g) to determine the pseudo-cost of the cartel:

$$
c^*_1 = \begin{cases} 
    \frac{1}{(n_{1\ell}^*-1)g_1(b_1|\ell)X_{\ell, n_{1\ell}^*} + n_0g_0(b_1|\ell)X_{\ell, n_{1\ell}^*}}, & \text{if } b^*_1 \in [\min_i b_{0i}, \max_i b_{0i}], \\
    \frac{1}{(n_{1\ell}^*-1)g_1(b_1|\ell)X_{\ell, n_{1\ell}^*} + n_0g_0(b_1|\ell)X_{\ell, n_{1\ell}^*}}, & \text{o/w,}
\end{cases}
$$

(g) Non-colluders: Since they still think they compete with $n_{1\ell}$ bidders, so like Case 1 we get

$$
c^*_1 = \begin{cases} 
    \frac{1}{(n_{1\ell}^*-1)g_1(b_1|\ell)X_{\ell, n_{1\ell}^*} + n_0g_0(b_1|\ell)X_{\ell, n_{1\ell}^*}}, & \text{if } b_{1i} \in [\min_i b_{0i}, \max_i b_{0i}], \\
    \frac{1}{(n_{1\ell}^*-1)g_1(b_1|\ell)X_{\ell, n_{1\ell}^*} + n_0g_0(b_1|\ell)X_{\ell, n_{1\ell}^*}}, & \text{o/w,}
\end{cases}
$$

(h) Then we have $\{c_{0i}\}_{i=1}^{n_{0\ell}}, \{c_{1i}\}_{i=1}^{n_{1\ell}}, \{M_\ell\}, \{c^*_1\}; \ell = 1, 2, \ldots, L$ pseudo-cost vectors. 24

(i) Then we can implement the collusion-proof mechanism.

---

23 It is possible that in an auction, $M_0 = 1$, in which case the minimum is just the bid and everything is the same. This means, we have two sets of $\{G_0, g_0\}$ one for $M_0$ and the other for the rest.

24 To estimate the type 0 cost, we use $\{\hat{G}^*_1, \hat{g}^*_1\}$ and $\{\hat{G}_0, \hat{g}_0\}$.
Table A-1. Revenue Shares and Participation of Main Firms

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

Only bidders with revenue shares ≥ 1% are reported.

Table A-2. Summary Statistics

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

All dollar figures are expressed in millions. Utilization rate is the ratio of backlog to capacity.
**TABLE A-3. Summary of Simultaneous Bids**

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

**TABLE A-4. Conditional Independence Test**

<table>
<thead>
<tr>
<th>Bidder Pair</th>
<th>Test Statistic</th>
<th>p-value</th>
<th>n</th>
<th>deg. freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A,D)</td>
<td>0.7660</td>
<td>0.0000</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>(A,E)</td>
<td>0.7427</td>
<td>0.0002</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>(B,D)</td>
<td>0.7331</td>
<td>0.0000</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>(C,D)</td>
<td>0.9239</td>
<td>0.0000</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>(D,E)</td>
<td>0.6530</td>
<td>0.0000</td>
<td>41</td>
<td>39</td>
</tr>
<tr>
<td>(D,F)</td>
<td>0.7570</td>
<td>0.0000</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>(D,H)</td>
<td>0.4734</td>
<td>0.0406</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>(D,I)</td>
<td>0.7121</td>
<td>0.0009</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>(D,O)</td>
<td>0.7643</td>
<td>0.0000</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>(D,P)</td>
<td>0.8538</td>
<td>0.0000</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>(D,R)</td>
<td>0.8555</td>
<td>0.0000</td>
<td>27</td>
<td>25</td>
</tr>
<tr>
<td>(D,V)</td>
<td>0.6877</td>
<td>0.0004</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>(M,X)</td>
<td>0.6529</td>
<td>0.0010</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>(W,X)</td>
<td>0.6271</td>
<td>0.0123</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>

**TABLE A-5. Exchangeability Test on Pairwise Basis**

<table>
<thead>
<tr>
<th>PAIR</th>
<th>F</th>
<th>UTA</th>
<th>k</th>
<th>m</th>
<th>n-m</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A,D)</td>
<td>5.9560</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(A,E)</td>
<td>2.3540</td>
<td>0.0161</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(B,D)</td>
<td>5.9354</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(C,D)</td>
<td>8.4271</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,E)</td>
<td>7.9549</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,F)</td>
<td>6.8441</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,H)</td>
<td>5.2001</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,I)</td>
<td>4.1670</td>
<td>0.0001</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,O)</td>
<td>5.5088</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,P)</td>
<td>7.1682</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,R)</td>
<td>6.1147</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,V)</td>
<td>3.7384</td>
<td>0.0002</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(D,W)</td>
<td>4.6217</td>
<td>0.0000</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(M,X)</td>
<td>0.7984</td>
<td>0.6040</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
<tr>
<td>(W,X)</td>
<td>0.3509</td>
<td>0.9458</td>
<td>8</td>
<td>796</td>
<td>2551</td>
</tr>
</tbody>
</table>
Figure 1. Histogram of difference between bid and cost (markup) for fringe and regular bidders, respectively, ignoring collusion while estimating cost (competition).

Figure 2. Histogram of difference between bid and cost (markup) for fringe and regular bidders, respectively, incorporating collusion while estimating cost (collusion).

Figure 3. ECDFs of extra cost under competition and collusion, respectively.
REFERENCES


