SUPPLEMENTAL NOTE FOR "ESTIMATING UNOBSERVED AGENT HETEROGENEITY USING PAIRWISE COMPARISONS"

ELENA KRASNOKUTSKAYA, KYUNGCHUL SONG, AND XUN TANG

1. Introduction

This note is a supplemental note to Krasnokutskaya, Song, and Tang (2016a). It consists of five parts. Section 2 gives details about how to derive pairwise comparison index in examples of first-price and English auctions where bidders have asymmetric private values, or collusive behavior. Section 3 discusses a bootstrap method to construct a confidence set for the group structure. Section 4 presents further simulation results regarding the performance of our classification algorithm proposed in Krasnokutskaya, Song, and Tang (2016a), and Section 5 reports summary statistics for the data used in the empirical application in the paper.

2. Bidders with Asymmetric Values or Collusive Behavior

2.1. First-Price Auctions with Asymmetric Bidders

We derive the pairwise comparison inequalities in an example of asymmetric first-price auctions with independent private values.

Let the bidders in the population be partitioned into K_0 groups defined by distinct private value distributions F_k for $k = 1, 2, ..., K_0$. Assume that the value distributions are stochastically ordered with the same support. Without loss of generality, let them be stochastically increasing in the subscript k. That is, $F_{k'}$ first-order stochastically dominates F_k whenever $k' > k$. Also assume that the ordering of the distributions is strict ($F_1(v) > F_2(v) > ... > F_{K_0}(v)$) at least for v within some non-degenerate interval on the support. Let $N_{(k)}$ denote the set of all agents in group k.

For simplicity, suppose that a bidder from group k becomes active with a fixed probability that is exogenously given. Let A denote the set of entrants in a given auction and $\lambda(A)$ denote the structure, or the profile, of entrants. That is, $\lambda(A)$ is a K_0-vector of integers $(|A_{(1)}|, ..., |A_{(K_0)}|)$, with $A_{(k)}$ being the set of entrants from group k. An entrant i submits bid B_i according to his private value v_i, taking into account the competitive structure of an auction $\lambda(A)$ which he observes at the time of bidding. Across auctions in the data, A and $\{C_i\}_{i \in A}$ are independent draws from the same population distribution.

Let $G_k(\cdot; \lambda)$ be the distribution of B_i when $i \in N_{(k)}$. The private values are independent of λ_A under exogenous entry. Part (i) of Corollary 3 in Lebrun (1999) showed that, given any realization of $\lambda(A)$, the supremum of the support of bids is the same for all bidder types. That

Date: January 11, 2017.

is, for any λ, $\beta_1(\overline{v}|\lambda) = \beta_2(\overline{v}|\lambda) = ... = \beta_K(\overline{v}|\lambda) \equiv \eta(\lambda) < \infty$ for some $\eta(\lambda) \in (\underline{v}, \overline{v})$, where β_k denotes the equilibrium bidding strategy for a bidder from group k. Furthermore, the corollary also showed that for any $\lambda(A)$,

$$F_{k'}(\beta_{k'}^{-1}(b|\lambda(A))) \leq F_k(\beta_k^{-1}(b|\lambda(A)))$$

for all $b \in [\underline{v}, \eta(\lambda(A))]$ and $k < k'$, and the inequality holds strictly at least over some interval on $[\underline{v}, \eta(\lambda(A))]$. Consider $i \in N_{(k')}$ and $j \in N_{(k)}$ with $k' > k$. It then follows that

$$\begin{aligned} P\{B_i \leq b|i,j \in A\} &= \sum\nolimits_{\lambda(A)} F_{k'}(\beta_{k'}^{-1}(b|\lambda(A)))P\{\lambda(A)|i,j \in A\} \\ &\leq \sum\nolimits_{\lambda(A)} F_k(\beta_k^{-1}(b|\lambda(A)))P\{\lambda(A)|i,j \in A\} = P\{B_j \leq b|i,j \in A\}, \end{aligned}$$

with the inequality holding strictly over some non-degenerate interval in the shared bid support. The inequality does not condition on the identities of the entrants other than i and j.

Finally, note that by a symmetric argument, a similar inequality holds in first-price procurement auctions with $P\{B_i \geq b|i,j \in A\} \leq P\{B_j \geq b|i,j \in A\}$ (with inequality being strict over some non-degenerate interval in the shared bid support), whenever the private cost distribution for i is stochastically lower than that of j.

2.2. English Auctions with Asymmetric Bidders

Consider the setting in Section 2.1, except that the auction format is English (ascending). The data report the identities of entrants in A and the transaction price W in each auction. In a dominant strategy equilibrium, the price in an auction equals the second highest private value among all entrants.

With independent private values, we show below that

$$P\{W \leq w|i \in A, j \notin A\} \leq P\{W \leq w|j \in A, i \notin A\} \tag{2.1}$$

for all w over the intersection of support, whenever $\tau(i) > \tau(j)$. Furthermore, the inequality holds strictly for some w over a set of positive measure in common support. This implies

$$\mathbf{E}[W|i \in A, j \notin A] > \mathbf{E}[W|j \in A, i \notin A]. \tag{2.2}$$

The intuition behind (2.1) is as follows. Given any structure of entrants who compete with i or j (but not both), the distribution of the transaction price is stochastically higher when i is present but j is not than when j is present but i is not. Loosely speaking, when j is replaced by the stronger type i in the set of entrants, the overall profile of value distributions becomes "stochastically higher". Then the law of iterated expectations implies (2.1) and (2.2).

To infer the group structure, define the following indexes:

$$\begin{aligned} \delta_{i,j}^+ &= \mathbf{E}[W|i \in A, j \notin A] - \mathbf{E}[W|j \in A, i \notin A], \\ \delta_{i,j}^0 &= |\mathbf{E}[W|i \in A, j \notin A] - \mathbf{E}[W|j \in A, i \notin A]|\,, \text{ and} \\ \delta_{i,j}^- &= \mathbf{E}[W|j \in A, i \notin A] - \mathbf{E}[W|i \in A, j \notin A], \end{aligned}$$

One can then use our procedure proposed in the text to classify the bidders based on pairwise comparison.

We now derive (2.1) formally. Let V_i denote the private value for bidder i. Consider the case where $i \in N_{(k')}$ and $j \in N_{(k)}$ where $k' > k$. Let $\lambda(A)$ denote the K_0-vector of integers that summarizes the group structure of the set of entrants A. Let 1_k denote the unit vector with the k-th component being 1. Then define:

$$\begin{aligned} H_{j,i}(w;\lambda^*) &\equiv P\{W \leq w | j \in A, i \notin A, \lambda(A\backslash\{j\}) = \lambda^*\} \\ &= P\left\{\max_{s\in A} V_s \leq w \;\; \lambda(A) = \lambda^* + 1_k\right\} + P\left\{\max_{s\in A} V_s > w, W \leq w \;\; \lambda(A) = \lambda^* + 1_k\right\}, \end{aligned}$$

where the first term on the right-hand side equals $F_k(w)\left(\prod_{l=1}^{K} F_l(w)^{\lambda_l^*}\right)$; and the second on the right-hand side is:

$$\begin{aligned} & P\left\{\max_{s\in A\backslash\{j\}} V_s \leq w, V_j > w \;\; \lambda(A\backslash\{j\}) = \lambda^*\right\} + P\left\{V_j \leq w, \max_{s\in A\backslash\{j\}} V_s > w \;\; \lambda(A\backslash\{j\}) = \lambda^*\right\} \\ = \;\; & [1 - F_k(w)]\left(\prod_{l=1}^{K} F_l(w)^{\lambda_l^*}\right) + F_k(w)\varphi(w;\lambda^*), \end{aligned}$$

where $\varphi(w;\lambda^*)$ denotes the probability that the maximum value in $A\backslash\{j\}$ is strictly greater than w while the second highest value in $A\backslash\{j\}$ is less than or equal to w conditional on the classification $\lambda(A\backslash\{j\}) = \lambda^*$. Therefore

$$H_{j,i}(w;\lambda^*) = \left(\prod_{l=1}^{K} F_l(w)^{\lambda_l^*}\right) + F_k(w)\varphi(w;\lambda^*).$$

By the same argument,

$$\begin{aligned} H_{i,j}(w;\lambda^*) &\equiv P\{W \leq w | i \in A, j \notin A, \lambda(A\backslash\{i\}) = \lambda^*\} \\ &= \left(\prod_{l=1}^{K} F_l(w)^{\lambda_l^*}\right) + F_{k'}(w)\varphi(w;\lambda^*). \end{aligned}$$

It is then straightforward to show that for any λ^*, that $F_{k'} \succeq_{F.S.D.} F_k$ implies $H_{i,j}(w;\lambda^*) \leq H_{j,i}(w;\lambda^*)$ over the union of the K_0 supports of $\{F_l : 1 \leq l \leq K_0\}$, and the inequality holds strictly at least for some w in an interval on the intersection of the K_0 supports of $\{F_l : 1 \leq l \leq K_0\}$. Under exogenous entry, we get

$$P\{W \leq w | i \in A, j \notin A\} \leq P\{W \leq w | j \in A, i \notin A\}$$

after integrating out λ^*. The inequality holds strictly for some w over common support.

Further Discussions

One may wonder whether we can recover the classification of bidders in the English auction example through a "global" approach when the identity of the winner is reported in the data. That is, by comparing the distribution of transaction prices when i is the winner versus that when j is the winner, as opposed to the pairwise comparison approach proposed above. Let us explain why this is not feasible.

For any $i \in N_{(k')}$ and $j \in N_{(k)}$ and $F_{k'} \succeq_{F.S.D.} F_k$, let $A\backslash\{i,j\}$ denote the set of entrants out of $N\backslash\{i,j\}$ and let $M(A\backslash\{i,j\}) \equiv \max\{V_s : s \in A\backslash\{i,j\}\}$. Let $\phi(w;\lambda^*)$ denote the distribution of $M(A\backslash\{i,j\})$ conditional on $\lambda(A\backslash\{i,j\}) = \lambda^*$. Let D denote the identity of the winner in the auction; and S_k denote the survival function for the private value of a type-k bidder. Then,

$$\begin{aligned} & P\left\{W \leq w, D = i | i \in A\right\} \\ = \; & p_j P\{W \leq w, D = i | i,j \in A\} + (1-p_j) P\left\{W \leq w, D = i | i \in A, j \notin A\right\}, \end{aligned}$$

where p_j is shorthand for j's entry probability. Also note that, by construction, once conditioned on the realized set of entrants from $A\backslash\{i,j\}$, we have

$$\begin{aligned} & P\left\{W \leq w, D = i | i,j \in A, \lambda(A\backslash\{i,j\}) = \lambda^*\right\} \\ = \; & \int_{-\infty}^{w} F_k(t)\phi(t;\lambda^*)dF_{k'}(t) + S_{k'}(w)F_k(w)\phi(w;\lambda^*), \end{aligned}$$

and

$$\begin{aligned} & P\left\{W \leq w, D = i | i \in A, j \notin A, \lambda(A\backslash\{i,j\}) = \lambda^*\right\} \\ = \; & \int_{-\infty}^{w} \phi(t;\lambda^*)dF_{k'}(t) + S_{k'}(w)\phi(w;\lambda^*). \end{aligned}$$

Likewise $P\{W \leq w, D = j | j \in A\}$ can be written by swapping the roles of i and j and swapping the roles of k and k' respectively. Then it can be shown that

$$P\{W \leq w, D = i | i \in A, j \in A\} > P\{W \leq w, D = j | i \in A, j \in A\}. \tag{2.3}$$ [1]

To see why the inequality in (2.3) holds, note for any λ^*,

$$\begin{aligned} P\{W & \leq w, D = i | i,j \in A, \lambda(A\backslash\{i,j\}) = \lambda^*\} \\ -P\{W & \leq w, D = j | i,j \in A, \lambda(A\backslash\{i,j\}) = \lambda^*\}, \end{aligned}$$

where the difference is written as

$$\begin{aligned} & \left[\int_{-\infty}^{w} F_k(t)\phi(t;\lambda^*)dF_{k'}(t) - \int_{-\infty}^{w} F_{k'}(t)\phi(t;\lambda^*)dF_k(t)\right] \\ & + \phi(w;\lambda^*)\left[S_{k'}(w)F_k(w) - S_k(w)F_{k'}(w)\right]. \end{aligned}$$

The first square bracket in the display above is positive because

$$\int_{-\infty}^{w} F_k(t)\phi(t;\lambda^*)dF_{k'}(t) > \int_{-\infty}^{w} F_{k'}(t)\phi(t;\lambda^*)dF_{k'}(t) > \int_{-\infty}^{w} F_{k'}(t)\phi(t;\lambda^*)dF_k(t).$$

Furthermore, the second square bracket in the display is also positive because "$F_{k'} \succeq_{F.S.D.} F_k$" implies

$$S_{k'}(w) \geq S_k(w) \text{ and } F_k(w) \geq F_{k'}(w) \text{ for all } w$$

and these inequalities hold strictly for some set of w with positive measure. Integrating out λ^* on both sides of the inequality

$$\begin{aligned} & P\{W \leq w, D = i | i,j \in A, \lambda(A\backslash\{i,j\}) = \lambda^*\} \\ > \; & P\{W \leq w, D = j | i,j \in A, \lambda(A\backslash\{i,j\}) = \lambda^*\}, \end{aligned}$$

yields the first inequality in (2.3).

Similarly, the difference between $P\{W \leq w, D = i | i \in A, j \notin A, \lambda(A \backslash \{i,j\}) = \lambda^*\}$ and $P\{W \leq w, D = j | j \in A, i \notin A, \lambda(A \backslash \{i,j\}) = \lambda^*\}$ equals

$$\left[\int_{-\infty}^{w} \phi(t;\lambda^*) dF_{k'}(t) - \int_{-\infty}^{w} \phi(t;\lambda^*) dF_k(t)\right] + \phi(w;\lambda^*)[S_{k'}(w) - S_k(w)]$$

which must be positive because the two terms in the square brackets are positive.

Now we write

$$\begin{aligned} (2.4) \qquad & P\{W \leq w, D = i | i \in A\} \\ = \; & p_j P\{W \leq w, D = i | i, j \in A\} + (1 - p_j) P\{W \leq w, D = i | i \in A, j \notin A\} \end{aligned}$$

where $p_j \equiv P(j \in A)$. A similar expression exists for $P\{W \leq w, D = j | j \in A\}$ by swapping the roles of i and j in (2.4). The difference between the two positive differences

$$P\{W \leq w, D = i | i, j \in A\} - P\{W \leq w, D = j | i, j \in A\}$$

and

$$P\{W \leq w, D = i | i \in A, j \notin A\} - P\{W \leq w, D = j | j \in A, i \notin A\}$$

is indeterminate in the absence of knowledge about p_i, p_j. Therefore the difference between $P\{W \leq w, D = i | i \in A\}$ and $P\{W \leq w, D = j | j \in A\}$ is also indeterminate.

2.3. Bidding Cartel in First-Price Procurement Auctions

Our method can be used to detect the identities of cartel members in a model of first-price procurement auctions in which a bidding cartel competes with competitive non-colluding bidders (Pesendorfer (2000)). Let the population of companies N be partitioned into a set of colluding firms $N_{(c)}$ and non-colluding firms $N_{(nc)}$. In each auction, the set of potential bidders (who are interested in bidding for the contract) A is partitioned into $A_{(c)}$ and $A_{(nc)}$. The cardinality of $A_{(c)}$ is common knowledge among the bidders. The potential bidders in $A_{(c)}$ collude by refraining from participation except for one bidder i^* who is chosen among them to submit a bid.[2]

In an efficient truth-revealing mechanism considered in Pesendorfer (2000), the cartel member that has the lowest cost is selected to be the sole bidder from the cartel. That is, $i^*(A_{(c)}) = \arg\min_{j \in A_{(c)}} C_j$ where C_j is the private cost of bidder j. Thus, the set of final entrants who are observed to submit bids in the data is $A^* \equiv \{i^*(A_{(c)})\} \cup A_{(nc)}$. (The set of colluding potential bidders is not reported in the data available to the researcher.)

We maintain that across the auctions in the data bidders' private costs are independent draws from the same distribution. Bidders are ex ante symmetric in that each bidder's private cost is drawn independently from the same distribution. Entrants know that a representative of the cartel is participating in bidding, and all follow Bayesian Nash equilibrium bidding strategies.

We are interested in detecting the identities of the set of colluding firms in $N_{(c)}$ from the reported bidding and participation decisions. Let $N'_{(c)} \subset N$ denote the set of bidders such

[2] The cartel is sustained through side payments among its members.

that no two bidders in $N'_{(c)}$ are ever observed to compete with each other in the bidding stage. By construction, $N_{(c)} \subseteq N'_{(c)}$ so the latter should be interpreted as a set of suspects for collusion. However, the set $N'_{(c)}$ could also contain innocent non-colluding bidders who are never observed to compete with each other in the data because of finite sample limitation. Our goal is to use bidding data to identify the group structure of $N'_{(c)}$, that is, to separate $N_{(c)}$ from $N'_{(c)} \backslash N_{(c)} \equiv N_{(nc)} \cap N'_{(c)}$.

Pesendorfer (2000) (Remark 3) shows that in any given auction with participants $A_{(c)} \cup A_{(nc)}$, the distribution of bids from a non-colluding bidder j first-order stochastically dominates the distribution of the bids from the sole bidder representing the cartel i^*.[3] Specifically, for any such i^* and j,

$$P\{B_{i^*} \leq b | i^* \in A^*, |A^*|\} > P\{B_j \leq b | j \in A^*, |A^*|\} \tag{2.5}$$

for all b on the common support of the two distributions.[4]

The intuition, as is noted in Pesendorfer (2000), is that the sole bidder representing a cartel has a higher hazard rate than a non-colluding bidder. That is, relative to a competitive bidder, the cartel representative has a higher probability of having a low cost conditional on the costs being above any fixed threshold. Besides, ex ante symmetry among bidders implies that

$$P\{B_i \leq b | i \in A^*, |A^*|\} = P\{B_j \leq b | j \in A^*, |A^*|\}$$

whenever $i, j \in N_{(c)}$ or $i, j \in N_{(nc)} \cap N'_{(c)}$.

We can then construct pairwise comparison indexes $\delta^+_{i,j}$, $\delta^0_{i,j}$, and $\delta^-_{i,j}$ by replacing $G_{i,j}$ and $G_{j,i}$ in equation (3.4) of Krasnokutskaya, Song, and Tang (2016a) with the left- and right-hand side of (2.5).

3. Confidence Sets for the Group Structure

The web appendix of Krasnokutskaya, Song, and Tang (2016b) proposes a method to construct a confidence set for each group of agents having the same type. Here for the sake of readers' convenience, we reproduce the procedure here using the notation of this paper. Let us consider a set-up where we have K_0 groups and the set N of agents. Let $\hat{K}$ be the consistent estimator of K_0 as proposed in Krasnokutskaya, Song, and Tang (2016a). As for confidence sets, we construct a confidence set for each group of agents who have the same type. First, we fix $k = 1, ..., \hat{K}$ and construct a confidence set for the k-th type group N_k. In other words, we construct a random set $\hat{\mathcal{C}}_k \subset N$ such that

$$\text{liminf}_{n,L\rightarrow\infty} P\{N_k \subset \hat{\mathcal{C}}_k\} \geq 1 - \alpha,$$

[3]Pesendorfer (2000) proved this result using the implicit assumption that the distribution of costs for non-colluding bidders and that for the sole cartel is common knowledge among all participants in an auction. (See proof of Remark 3 in Pesendorfer (2000).) This assumption is consistent with the informational environment that the partition of N into $N_{(c)}$ and $N_{(nc)}$ is common knowledge among all bidders.

[4]Note that the statement is conditional since the bidding strategies depend on the cardinality of the final set of bidders $|A^*|$.

For this, we need to devise a way to approximate the finite sample probabilities like $P\{N_k \subset \hat{\mathcal{C}}_k\}$. Since we do not know the cross-sectional dependence structure among the agents, we use a bootstrap procedure that preserves the dependence structure from the original sample. The remaining issue is to determine the space in which the random set $\hat{\mathcal{C}}_k \subset N$ can take values in. It is computationally infeasible to consider all possible such sets. Instead, we proceed as follows. First we estimate $\hat{N}_k$ as prescribed in the paper and also obtain $\hat{\delta}^0_{ij}$, the test statistic defined in the main text. Given the estimate $\hat{N}_k$, we construct a sequence of sets as follows:

Step 1: Find $i_1 \in N \backslash \hat{N}_k$ that minimizes $\min_{j \in \hat{N}_k} \hat{\delta}^0_{i_1,j}$, and construct $\hat{\mathcal{C}}_k(1) = \hat{N}_k \cup \{i_1\}$.
Step 2: Find $i_2 \in N \backslash \hat{\mathcal{C}}_k(1)$ that minimizes $\min_{j \in \hat{C}_k(1)} \hat{\delta}^0_{i_2,j}$, and construct $\hat{\mathcal{C}}_k(2) = \hat{C}_k(1) \cup \{i_2\}$.
Step m**:** Find $i_m \in N \backslash \hat{\mathcal{C}}_k(m-1)$ that minimizes $\min_{j \in \hat{C}_k(m-1)} \hat{\delta}^0_{i_m,j}$ and construct $\hat{\mathcal{C}}_k(m) = \hat{\mathcal{C}}_k(m-1) \cup \{i_m\}$.

Repeat Step m up to $n = |N|$.

Now, for each bootstrap iteration $s = 1, ..., B$, we construct the sets $\hat{N}^*_{k,s}$ and $\{\hat{\mathcal{C}}^*_{k,s}(m)\}$ following the steps described above but using the bootstrap sample. (Note that this bootstrap sample should be drawn independently of the bootstrap sample used to construct bootstrap p-values $\hat{p}_{ij}$ in the classification.)

Then, we compute the following:

$$\hat{\pi}^k(m) \equiv \frac{1}{B} \sum_{s=1}^{B} 1 \left\{ \hat{N}_k \subset \hat{\mathcal{C}}^*_{k,s}(m) \right\}.$$

Note that the sequence of sets $\hat{\mathcal{C}}^*_{k,s}(m)$ increases in m. Hence the number $\hat{\pi}^k(m)$ should also increase in m. An $(1-\alpha)$%-level confidence set is given by $\hat{\mathcal{C}}^*_k(m)$ with $1 \leq m \leq n$ such that

$$\hat{\pi}^k(m-1) < 1 - \alpha \leq \hat{\pi}^k(m).$$

Note that such m always exists, because $\hat{\mathcal{C}}^*_{k,s}(n) = N$.

4. Further Simulation Results

Tables 4.1 and 4.2 summarize the distribution of estimation errors in our group classification algorithm from 500 simulated data sets, when the number of groups is $K_0 = 2$ and assumed known to the econometrician. The column D_μ shows the difference between the group means chosen in the simulation.

When $K_0 = 2$, the results show that the estimation error, as measured by the expected average discrepancy (EAD), decreases with the distance between group means. Such a reduction in EAD is most substantial when the number of players is larger ($n = 40$) and the size of the data is small ($L = 100$). Given group difference, EAD decreases as sample size increases moderately from $L = 100$ to 400. This pattern is most obvious when $D_\mu = 0.2$.

The other measure of estimation errors, DH(100p), also shows encouraging results. DH(100p) is zero for most of the cells in both panels (a) and (b), which shows that the empirical distribution of proportion of mis-classified bidders is reasonably skewed to the right. Besides, the reduction in DH(100p) as the sample size increases is most pronounced with closer group means, regardless of the number of bidders in the population.

When $K_0 = 4$, the results demonstrate very similar patterns. Most remarkably, both measures of mis-classification errors only increase very marginally relative to the case with $K_0 = 2$.

Tables 4.3 and 4.4 report results from the full, feasible classification procedure when the number of groups is estimated through the penalization scheme proposed in the text. For most of the specifications used in these two tables, the estimates for the number of groups $\hat{K}_0$ are tightly clustered around the correct K_0. Compared with the results for infeasible classification under known K_0, EAD and DH(100p) increase in most cases. Nonetheless such an increase is quite moderate, suggesting that our feasible classification algorithm performs reasonably well relative to its infeasible counterpart.

In Tables 4.3 - 4.4, we report the analysis of computation time for the classification procedure. In Table 4.3, we give a decomposition of the time that it took for the classification procedure. The table clearly shows that the major computation time spent is when we construct bootstrap p-values. Once the p-values are constructed, the classification algorithm itself runs fairly fast.

In Table 4.4 , the computation time is shown to vary depending on the number of the agents (n), the number of the true groups (K_0), and the number of the markets (L). The results show that the most computation time increase arises when the number of the bidders increases rather than when the number of the markets or the number of the groups increases. Our simulation studies are based on our MatLab code. The program was executed using a computer with the following specifications: Intel(R) Xeon (R) CPU X5690 @3.47 GHz 3.46 GHz.

Table 4.1 : Performance of the Classification Estimator with Two Groups: ($K_0 = 2$ and known)

n	L	D_μ	EAD	DH(10)	DH(25)	DH(50)	DH(75)	DH(90)
12	400	0.6	0.012	0.012	0	0	0	0
12	400	0.4	0.014	0.014	0	0	0	0
12	400	0.2	0.004	0.004	0	0	0	0
12	200	0.6	0.004	0.004	0	0	0	0
12	200	0.4	0.006	0.006	0	0	0	0
12	200	0.2	1.118	0.560	0.252	0.158	0	0
12	100	0.6	0.006	0.006	0	0	0	0
12	100	0.4	0.084	0.078	0.006	0	0	0
12	100	0.2	1.794	0.682	0.478	0.284	0	0
40	400	0.6	0.018	0	0	0	0	0
40	400	0.4	0.022	0	0	0	0	0
40	400	0.2	1.170	0.178	0.014	0	0	0
40	200	0.6	0.018	0	0	0	0	0
40	200	0.4	0.020	0	0	0	0	0
40	200	0.2	2.726	0.404	0.210	0.122	0.021	0.001
40	100	0.6	0.020	0	0	0	0	0
40	100	0.4	0.452	0.010	0	0	0	0
40	100	0.2	3.720	0.902	0.578	0.234	0.132	0.043

Note: This table summarizes the distribution of estimation errors in our classification algorithm from 500 Monte Carlo replications when $K_0 = 4$ and known. Here n represents the number of the individual players, L the number of the observed games in the data, D_μ the distance between population means, EAD the expected average discrepancy, and DH(100p) the hazard rate of EAD at p.

Table 4.2 : Performance of the Classification Estimator with Four Groups: ($K_0 = 4$ and known)

n	L	D_μ	EAD	DH(10)	DH(25)	DH(50)	DH(75)	DH(90)
12	400	0.6	0.011	0.014	0.004	0	0	0
12	400	0.4	0.018	0.016	0.010	0	0	0
12	400	0.2	0.017	0.022	0.006	0	0	0
12	200	0.6	0.013	0.018	0.004	0	0	0
12	200	0.4	0.004	0.008	0	0	0	0
12	200	0.2	1.112	0.764	0.188	0.024	0.008	0
12	100	0.6	0.003	0.006	0	0	0	0
12	100	0.4	0.044	0.040	0.024	0.002	0	0
12	100	0.2	1.504	0.868	0.342	0.106	0.04	0
40	400	0.6	0.115	0.020	0.020	0	0	0
40	400	0.4	0.121	0.020	0.020	0	0	0
40	400	0.2	2.450	0.680	0.368	0.018	0.018	0
40	200	0.6	0.109	0.018	0.018	0	0	0
40	200	0.4	0.140	0.026	0.026	0	0	0
40	200	0.2	3.172	0.810	0.366	0.246	0.026	0
40	100	0.6	0.141	0.024	0.024	0	0	0
40	100	0.4	1.003	0.176	0.176	0.006	0	0
40	100	0.2	4.557	0.904	0.652	0.526	0.202	0.053

Note: This table summarizes the distribution of estimation errors in our classification algorithm from 500 Monte Carlo replications when $K_0 = 4$ and known. Here n represents the number of the individual players, L the number of the observed games in the data, D_μ the distance between population means, EAD the expected average discrepancy, and DH(100p) the hazard rate of EAD at p.

Table 4.3 : Computational Time for Various Steps of the Procedure
($n = 60$, $K_0 = 2$, unknown, time measured in **seconds**)

Step	Description	L=100	L=200	L=400
1	generating pairwise indexes from the data	0.2987	0.3543	0.4607
2	constructing bootstrap pairwise indexes	81.2178	81.4871	82.0807
3	computing bootstrap p-values	0.0012	0.0014	0.0014
4	division of a group into two	0.0008	0.0008	0.0008
n+4	number of groups selection	0.0002	0.0002	0.0002
	Total Time	81.528	81.852	82.552

Note: The table shows a decomposition of a total time it has taken for the classification procedure. The table shows that the major portion of the time comes from constructing the bootstrap pairwise indexes. Once the bootstrap p-values are constructed, the classification algorithm runs quite fast.

Table 4.4 : Total Computational Time: across n, L, and K_0
(K_0 unknown, time measured in **seconds**)

	L=100 K_0=2	L=200 K_0=2	L=400 K_0=2	L=200 K_0=4	L=200 K_0=6
$n = 12$	3.246	3.224	3.239	3.216	3.219
$n = 24$	13.057	13.177	13.259	13.185	13.189
$n = 48$	51.987	52.272	52.700	52.281	52.291
$n = 60$	81.528	81.852	82.552	81.862	82.874
$n = 72$	116.949	117.213	117.577	116.912	117.328
$n = 96$	209.426	209.971	209.834	209.884	210.058

Note: The table shows the change in the computation time as one changes the number of the groups (K), the number of the markets and the number of the agents (i.e., bidders) ((n)). The major increase in the computation time arises when the number of the bidders increases rather than when the number of the markets or the groups increases.

5. Additional Materials for the Empirical Application

Table 5.1 reports summary statistics for this set of projects. The table indicates that the projects are worth $523,000 and last for around three months on average; 38% of these projects are partially supported through federal funds. There are 25 firms that participate regularly in this market. All other firms in the data are treated as fringe bidders. An average auction attracts six regular potential bidders and eight fringe bidders. Since only a fraction of potential bidders submits bids, an entry decision plays an important role in this market. Finally, the distance to the company location varies quite a bit and is around 28 miles on average for regular potential bidders.

Table 5.1 : Summary Statistics for California Procurement Market

Variable	Mean	Std. Dev
Engineer's estimate (mln)	0.523	0.261
Duration, large projects (months)	3.01	1.56
Federal Aid	0.384	
Number of Potential Bidders:	14.1	8.4
Fringe Bidders	8.2	4.8
Regular Bidders	5.5	3.3
Number of Entrants:	5.4	2.8
Fringe Bidders	3.5	2.7
Regular Bidders	1.9	1.8
Distance (miles):	18.72	6.33
Fringe Bidders	11.21	5.42
Regular Bidders	28.34	11.73

Note: This table reports summary statistics for the set of medium size bridge work and paving projects auctioned in the California highway procurement market between years of 2002 and 2012. It consists of 1,054 projects. The distance variable is measured in miles. It reflects the driving time between the project site and the nearest company plant. The "Federal Aid" variable is equal to one if the project receives federal aid and zero otherwise.

References

KRASNOKUTSKAYA, E., K. SONG, AND X. TANG (2016a): "Estimating Unobserved Agent Heterogeneity Using Pairwise Comparisons," Working paper.

——— (2016b): "The Role of Quality in Internet Service Markets," Working paper.

LEBRUN, B. (1999): "First Price Auctions in the Asymmetric N Bidder Case," *International Economic Review*, 40(1), 125–142.

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