# ESTIMATING UNOBSERVED INDIVIDUAL HETEROGENEITY THROUGH PAIRWISE COMPARISONS

ELENA KRASNOKUTSKAYA, KYUNGCHUL SONG, AND XUN TANG

Abstract. This paper proposes a new method to study environments with unobserved agent heterogeneity. We focus on settings where the heterogeneous factor takes values from an unknown finite set, and the economic model yields testable implications in the form of pairwise inequalities. The method produces a consistent classification of economic agents according to their unobserved types. The paper verifies that the method performs quite well in Monte Carlo simulations. We demonstrate empirical usefulness of this method by estimating a model of first-price auction characterized by both agent and auction level unobserved heterogeneity using data from the California highway procurement market.

Key words. Unobserved Agent Heterogeneity; Discrete Unobserved Heterogeneity; Pairwise Comparisons; Asymmetric Bidders; Auction Models; Nonparametric Classifications

JEL Classification: C12, C21, C31

## 1. Introduction

Empirical analysis of many economic settings requires accounting for unobserved agent heterogeneity. The latter arises because some of the agent-specific factors influence the decisions of economic agents and yet are not or could not be recorded in the data available to a researcher. Failing to account for unobserved heterogeneity may lead to biased estimates and may affect the quality of counterfactual predictions.

This paper focuses on the environments where economic agents are characterized by a finite number of unobserved types.[1] That is, we assume that heterogeneity is captured by a discrete (nonstochastic) parameter taking values from a finite ordered set. We discuss identification and propose a method to recover the implied unobserved group

*Date*: December 23, 2014.

[1]While discreteness can be a restriction in certain applications, in many cases the model with discrete types is capable of closely approximating more general model with continuum of types, especially when the number of types can be chosen on the basis of the data. Unobserved heterogeneity in a discrete form has been extensively used in econometrics literature. See below the literature review.

structure. Our approach relies on pairwise comparisons derived from the model that are related to agents' unobserved types.

One of the contributions of this paper is to demonstrate that such pairwise inequalities arise in the context of many models with strategic agent interdependence. The paper discusses a number of examples falling into this framework. For the sake of specificity, we focus on auction environments associated with unobserved bidder heterogeneity. The examples include unobserved bidder attributes used in the allocation decisions under multi-attribute auctions, unobserved cost asymmetry of bidders in standard auctions, and collusion. In these examples, the unobserved group structure reflects the competitive environment faced by agents and thus may be a source of omitted variable bias if it is ignored in estimation.

We investigate identification of the ordered group structure from pairwise inequality restrictions. Specifically, we show that if some pairs of agents cannot be compared in the data, then the group structure may not be identified even if all agents are connected through the directly comparable pairs. We thus study a subset of data structures that permit all pairs of agents to be directly compared.

Even when identification is established, the estimation of the group structure remains quite challenging. One approach would be to establish the ordering of the types by testing the inequality restrictions for each pair separately. However, this may not deliver a coherent estimate of the group structure since transitivity of ordering across pairs may not be preserved in finite samples. Furthermore, it may be computationally infeasible to consider every possible group structure while verifying that restrictions hold for the group structure as a whole. Indeed, even if the types are known to take values from a two-point set and the number of agents is equal to $N$, the total number of the candidate group structures is $2^N$ which can easily become a large number with only a moderate number of agents.

The main idea of the approach proposed in this paper is to recover the whole group structure by sequentially subdividing the set of agents on the basis of information contained in the $p$-values of the tests of pairwise inequality restrictions until the desirable number of groups is reached. Further, we propose an appropriate goodness-of-fit measure and a penalization scheme for overfitting, that together allow us to consistently select the correct number of groups. We formally show that this method delivers a consistent estimator of the group structure under mild regularity conditions on the pairwise restrictions.

Thus, the proposed method has the following three properties. First, it does not require any prior knowledge of the group structure such as the number of groups or group sizes. Second, the estimator for the group structure is consistent under mild regularity conditions. Third, the method is computationally feasible in most applications.

We believe that the method proposed in this paper is most useful as a pre-estimation step in the structural analysis of economic settings with strategic interdependence such as dynamic industry models or auction models with asymmetric bidders.[2] The brute force approach of allowing the types to be arbitrarily idiosyncratic across agents and estimating the types jointly with the structural parameters is often computationally infeasible, because one needs to solve such games for each configuration of agents types observed in the data during the estimation step while maintaining that the number of types is equal to the number of agents. In contrast, our approach of using the nonparametric classification as a first step tends to yield a coarser level of heterogeneity that makes it feasible to estimate the complex model while properly accounting for unobserved agent heterogeneity.[3]

Our approach may also be useful in non-parametric estimation. Brute force estimation is often infeasible in this setting as well since the researcher has to account for the heterogeneity of the units of observation (auctions or markets) in estimation. In contrast, pairwise restrictions and the classification procedure may often aggregate over such heterogeneity. The classification is thus feasible even in cases where brute force nonparametric estimation is not. Further, once the group structure is recovered, the data are organized at a higher level of coarseness which may permit nonparametric estimation of other primitives.

We investigate the performance of the classification method in a Monte Carlo study. The study is based on data generated by the model of first-price auction with asymmetric bidders where the means of the distributions of private values differ across bidders in an unobserved way. We report the outcome for various numbers of bidders and group structures. The estimation works quite well in general. The performance is better when the number of bidders and groups are smaller and the differences between groups are larger.

[2]The first step estimation of discrete unobserved heterogeneity does not affect the second step estimation due to its discreteness in terms of pointwise asymptotics. However, establishing uniform asymptotics remains an open question. This problem is analogous to the one of post-model selection inference that arises from using consistent model selection in the first step estimation. For discussion on the issues, see Pötcher (1991), Leeb and Pötcher (2005), and Andrews and Guggenberger (2009) and references therein. Unlike the problem of variable selection, uniform asymptotics in our set-up is very complex, because we need to consider every possible direction in which a true group structure may be locally perturbed. We believe that a full theoretical investigation of that issue in our context merits a separate paper.

[3]See Krasnokutskaya, Song, and Tang (2014) for an application of this approach in the analysis of online service markets.

We demonstrate methodological usefulness of the classification method through an empirical analysis of the California highway procurement market. Previous studies document cost asymmetries associated with observable characteristics of bidders. In this paper, we allow for potentially unobservable differences in the means of cost distributions. To eliminate other sources of cost heterogeneity we control for the bidder's distance to the project site as well as account for the possible endogeneity of the competitive auction structure. We first use the classification method to recover the underlying group structure associated with permanent unobserved cost asymmetries. In the second step, we recover the group-specific distributions of costs using Generalized Method of Moments estimation. The classification step makes estimation of such a model computationally feasible by reducing the number of auction games which have to be solved in estimation. We identify three unobserved groups of bidders that (as our estimates indicate) are characterized by important costs differences. Our findings thus confirm the importance of accounting for unobserved costs asymmetries in this market.

Researchers often turn to a finite mixture approach to model unobserved agent heterogeneity. Our pairwise-restrictions approach is distinguished from the finite mixture approach in two important aspects. First, our approach allows unobserved heterogeneity to be nonstochastic. Therefore, the unobserved heterogeneity is allowed to be arbitrarily correlated with other random components in the model, as long as it is consistent with the empirical model that produces the pairwise inequality restrictions.

Second, applying the finite mixture method is often non-trivial in environments with strategic interdependence for the following two reasons. First, even though identification of models with nonstochastic unobserved types directly follows from inequality restrictions in many settings, the identification analysis of finite mixture model is often much more involved.[4] Second, the cardinality of the support for the vector of individual types in most applications is so large that the implementation of finite mixture method becomes computationally infeasible.

Modeling part of the unobserved component as a nonstochastic parameter taking values from a finite set has precedents in the literature on panel models. Sun (2005) and Phillips and Sul (2007) study an unobserved group structure of cross-sectional units in the context of growth models based on large panel data. Song (2005) explores consistency

[4]Kasahara and Shimotsu (2009) show how a finite mixture model of individual dynamic decisions with unobserved types can be point-identified using the variation in the covariates reported in the data and its impact on the conditional choice probabilities across different types. Henry, Kitamura, and Salanie (2014) study the partial identification of finite mixture model when there is exogenous source of variation in mixture weights that leave the mixture component distribution invariant. In environments with strategic interdependence finite mixture are mostly used to model unobserved heterogeneity at the market rather than individual level (see, for example, Hu, McAdams, and Shum (2013)). In contrast, the method proposed in our paper targets an environment with agent-specific unobserved heterogeneity.

of the heterogeneous parameter estimators and proposes a consistent estimation method of unobserved group structure in large panel models. Lin and Ng (2012) also study estimation of panel models with unknown group structure. More recently, Bonhomme and Manresa (2014) consider a panel model where the unknown group structure is time-varying.

This paper is organized as follows. Section 2 introduces the basic environment where pairwise inequality restrictions are defined. This section also discusses three examples that fall into such a framework. Section 3 discusses identification and proposes a consistent estimator of the unobserved group structure. The exposition for the estimator is organized in three steps. First, we outline our algorithm in the simple case of two groups. Next, we extend it to the case of multiple groups while maintaining that the number of the groups is known. Finally, we consider the general case with an unknown number of groups. Section 4 presents and discusses Monte Carlo results. Section 5 presents the empirical application. Section 6 concludes, discussing the findings of the paper. Technical proofs and derivations are provided in the appendix.

## 2. Unobserved Heterogeneity and Pairwise Comparisons

2.1. **Identification of the ordered group structure.** We consider a setting with $N$ agents who are characterized by some economically relevant factor $q$ that takes values from an ordered discrete set $Q_{K_0} = \{\bar{q}_1, ..., \bar{q}_{K_0}\}$, with $\bar{q}_1 < \cdots < \bar{q}_{K_0}$. We think of $q$ as a factor impacting agent's payoff such as a parameter characterizing agent's preferences, costs, or budget constraint. Each agent $i$ is thus associated with a type $\tau(i) \in \{1, ..., K_0\}$ such that $q_i = \bar{q}_{\tau(i)}$.

The factor $q$ induces a partition of the set of agents into a *group structure* $T$ which is an ordered collection of disjoint subsets $(\mathbb{N}_1, \mathbb{N}_2, ..., \mathbb{N}_{K_0})$ such that for each $k = 1, ..., K_0$,

$$\mathbb{N}_k = \{i \in \mathbb{N} : \tau(i) = k\}.$$

We are interested in settings where the econometrician does not observe $T$.

As we demonstrate below with several examples, economic models often yield testable implications for the equilibrium actions and outcomes that are formulated in the form of pairwise inequalities associated with agents of distinct types.

More formally, let $\mathcal{P}$ denote the collection of distributions of observed random vectors that reflect the data generating process associated with a specific economic model. We say that a model generates testable pairwise inequality restrictions if for each pair $(i, j)$ of agents, data generated by $P \in \mathcal{P}$ identify three *comparison indices* $\delta^+_{ij}$, $\delta^0_{ij}$ and $\delta^-_{ij}$

that satisfy the following restrictions

$$
\begin{aligned}
\delta_{ij}^{+} &> 0 \text{ if and only if } \tau(i) > \tau(j); \\
\delta_{ij}^{0} &= 0 \text{ if and only if } \tau(i) = \tau(j); \\
\delta_{ij}^{-} &> 0 \text{ if and only if } \tau(i) < \tau(j).
\end{aligned} \tag{2.1}
$$

Our objective is to propose a method to recover the unobserved group structure $T$ from the available data on equilibrium actions and outcomes and on the basis of the pairwise restrictions implied by the data.

2.2. **Example 1: Unobserved Quality in Multi-Attribute Procurement Auctions.** Industry procurement is often implemented through the multi-attribute auction mechanism. Recently, multi-attribute auction became the mechanism of choice for many on-line auction procurement markets. This example is based on Krasnokutskaya, Song, and Tang (2014) which analyzes one of such on-line procurement markets. The abbreviated details of this model are summarized below.

The supply side of the market consists of sellers who are described by a scalar characteristic, $q$, which admits a finite number of values, $\bar{q}_1 < \cdots < \bar{q}_{K_0}$. Extensive communication between buyers and sellers as well as the record of sellers' performance measures provided by the platform allow buyers to be well-informed about sellers' attributes. We allow, however, that a researcher may not observe realizations of the vertical characteristic $q$. Sellers from group $k$ are characterized by private costs, $C_{i,l}$, for completing the project indexed by $l$ drawn from a distribution $F_C^k(\cdot)$.[5]

Buyers allocate projects through multi-attribute auctions that allow them to deviate from allocation based solely on price (as in standard auctions) in favor of service providers who are associated with higher buyer-specific value. We use $A_l$ to denote the set of sellers who submit a bid for project $l$ and refer to such sellers as *active* bidders. These sellers form the buyer's choice set.

Buyer $l$ associates a private value, $\Delta_{i,l}$, with an active seller $i \in A_l$ and awards his project to the active seller with the highest level of $\Delta_{i,l} - b_{i,l}$ if this level exceeds buyer's outside option $U_{0,l}$; otherwise, he leaves the project unassigned. The buyer's value depends on a weighted seller's attribute with a buyer-specific weight $\alpha_l$ and intercept $\epsilon_{i,l}$ (the residual value assigned by buyer $l$ to a specific seller), i.e.,

$$\Delta_{i,l} = \alpha_l q_i + \epsilon_{i,l}. \tag{2.2}$$

In keeping with the definition of a multi-attribute auction, sellers do not observe the weights or the outside option of a specific buyer, and consider them to be a random draw

[5]Krasnokutskaya, Song, and Tang (2014) also allow for observed characteristics to be present.

from some joint distribution of weights and outside options characterizing the population of buyers.

In a simple model, the sellers from a group $k$ become active with an exogenous probability $\rho_k$, although results of the section could be extended to the setting with endogenous participation. An active bidder observes his cost of completing the project but not the participation status of other potential sellers and is thus unaware of the composition of the set of active bidders. He then submits a price $B_{i,l}$ based on his information set. We assume that in quoting prices, the sellers adopt pure strategies prescribed in a type-symmetric Bayesian Nash equilibrium.

For each pair of sellers $i$ and $j$, and each $p$ from the intersection of supports of $B_{i,l}$ and $B_{j,l}$, define

$$r_{i,j}(p) \equiv P\left\{\text{Procurer chooses seller } i | B_i = p,\ i \in \mathbb{A},\ j \notin \mathbb{A}\right\}.$$

Notice that in this index we do not condition on the identities of active participants. Rather, we require that that "the set of entrants contains $i$ but not $j$".

Krasnokutskaya, Song, and Tang (2014) (Proposition 1) show that when sellers' private costs are independent from the procurer's taste,

$$sign(r_{i,j}(p) - r_{j,i}(p)) = sign(q_i - q_j), \tag{2.3}$$

for any $p$ on the intersection of bid supports.

The intuition for (2.3) is as follows. If $i$ and $j$ participate in two separate but ex-ante identical auctions (in terms of the realized set of competitors) and submit the same price then the seller with the higher value of $q$ has the higher chance of winning. Note that the winner is not deterministic in the presence of uncertainty about buyers' weights. The ranking of winning probabilities is preserved when aggregated over different sets of competitors as long as the probability of encountering a given set of competitors is the same for both sellers. This condition holds if, for example, the pool from which competitors are drawn does not include either $i$ or $j$.

Thus if we define

$$\begin{aligned}
\delta_{ij}^{+} &= \int \max\{r_{i,j}(p) - r_{j,i}(p), 0\} dp; \\
\delta_{ij}^{0} &= \int \left| r_{i,j}(p) - r_{j,i}(p) \right| dp; \\
\delta_{ij}^{-} &= \int \max\{r_{j,i}(p) - r_{i,j}(p), 0\} dp.
\end{aligned}$$

The restrictions in (2.3) yield the pairwise comparisons as in (2.1).

2.3. **Example 2: Standard Auctions with Unobserved Cost Asymmetries.** Auction pricing and allocation outcomes are generally quite sensitive to the degree of cost asymmetries (or bidders' heterogeneity with respect to the distribution of their costs). Most existing empirical studies, however, focus on the cost asymmetries associated with observed characteristics (see Athey, Levin, and Seira (2011), Krasnokutskaya and Seim (2011), Campo, Perrigne, and Vuong (2003)).

We now demonstrate how pairwise comparisons could be derived in the case of standard auctions with asymmetric bidders.

2.3.1. *First-Price Procurement Auctions.* Let the population of potential bidders be partitioned into several groups with bidders from group $k$ drawing private costs independently from $F_k$. For simplicity, we assume that the distributions of costs associated with various groups differ only in their means so that $\bar{q}_1 < \bar{q}_2 < ...\bar{q}_K$, where $\bar{q}_k$ refers to the mean of the distribution $F_k$. Hence each bidder $i$ from group $k$ has type $\tau(i) = \bar{q}_k$. The result we refer to in this section holds more generally when $F_k$ are stochastically ordered.

As before, let $\mathbb{A}$ denote the set of entrants in a given auction and suppose again for simplicity that bidder from group $k$ becomes active with a fixed probability $\rho_k$. The result also holds under strategic participation. The set of entrants is characterized by a group structure implied by the group affiliations of its members. We denote this group structure $\lambda_A$.

Entrant $i$ submits bid $B_i$ according to his private costs and taking into account the competitive structure of an auction as summarized by $\lambda_A$ which he observes at the time of bidding. We assume that bidders' actions are consistent with type-symmetric Bayesian Nash equilibrium.

Define $G_{i,j}(b) = P\{B_i \leq b | i, j \in \mathbb{A}\}$ to be the distribution of bids submitted by bidder $i$ when both $i$ and $j$ participate in the same auction.[6] Then, under mild regularity conditions,

$$G_{i,j}(b) \leq G_{j,i}(b), \tag{2.4}$$

whenever bidder $i$ is of a stronger type than bidder $j$ (i.e., $\tau(i) < \tau(j)$) for all $b$ in the common support of $B_i$ and $B_j$. The inequality holds strictly at least over some interval with positive Lebesgue measure.

This regularity has been previously established in the literature for a given configuration of the set of active bidders (see Corollary 3 of Lebrun (1999).) We show in

[6]Here we need to maintain the assumption that the bid data is rationalized by a single BNE, which is standard in the literature on empirical auctions.

the Appendix that the inequality also holds unconditionally when the set of competitors is integrated out. Such an unconditional property is useful since it lowers data requirements for the procedure discussed later in the paper.

To see how the results in this section translate into our general framework define

$$\begin{aligned}\delta^+_{i,j} &= \int \max\{G_{j,i}(b) - G_{i,j}(b), 0\}\, db, \\ \delta^0_{i,j} &= \int |G_{i,j}(b) - G_{j,i}(b)|\, db, \text{ and} \\ \delta^-_{i,j} &= \int \max\{G_{i,j}(b) - G_{j,i}(b), 0\}\, db.\end{aligned} \tag{2.5}$$

Let the inequality $\tau(i) < \tau(j)$ indicate that bidder $j$ is of a stronger type than bidder $i$. Then we arrive at the testable implications as in (2.1).

2.3.2. *English Auctions.* This example is based on the analysis in Athey and Haile (2002). Consider the setting in Section 2.3.1 except that the price is determined in an English (open ascending) auction. The data report the identity of the winner and the final transaction price in each auction. In a dominant strategy equilibrium, the price in an auction equals the second-highest private value among its entrants.

With independent private values, it can be shown that

$$P\{W \le w | i \in \mathbb{A}, j \notin \mathbb{A}\} \le P\{W \le w | j \in \mathbb{A}, i \notin \mathbb{A}\} \tag{2.6}$$

whenever $i$ is from the stronger type. Furthermore, the inequality holds strictly for some $w$ over a set of positive measure in the intersection of supports. This implies

$$\mathbf{E}[W | i \in \mathbb{A}, j \notin \mathbb{A}] > \mathbf{E}[W | j \in \mathbb{A}, i \notin \mathbb{A}]. \tag{2.7}$$

The intuition behind (2.6) is as follows. For any given group structure of entrants that $i$ or $j$ complete with, the distribution of the transaction price is stochastically higher when $i$ is present rather than $j$. 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". The implications (2.6) and (2.7) then follow as a consequence of the law of iterated conditional expectations. We provide details about the regularity conditions in the appendix.

To translate this result into the classification framework, we define

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

and let the inequality $\tau(i) < \tau(j)$ indicate that bidder $j$ is of a stronger type than bidder $i$ as before. Then the pairwise comparisons as in (2.1) follows.

2.4. **Example 3: Detecting the Identities of Bidding Cartel Members.** Pesendorfer (2000) analyzes first-price procurement auctions in which a bidding cartel competes with competitive non-colluding bidders. Our method can be used in this setting to detect the identities of cartel members from the bidding data.

Specifically, let $\mathbb{N}$ denote the population of companies in the data, which is partitioned into a set of colluders $\mathbb{N}_c$ and non-colluders $\mathbb{N}_{nc}$. Consequently, in each auction, the set of potential bidders (i.e. those who are interested in bidding for the contract) $\mathbb{A}$ is partitioned into $\mathbb{A}_c$ and $\mathbb{A}_{nc}$. The cardinality of $\mathbb{A}_c$ is common knowledge among the bidders. The potential bidders in $\mathbb{A}_c$ collude by refraining from participation except for one bidder $i^*$ who is chosen among them to submit a bid.[7] In an efficient truth-revealing mechanism such as that considered in Pesendorfer (2000), the cartel member that has the lowest costs is selected to be the sole bidder from the cartel. That is, $i^*(\mathbb{A}_c) = \arg\min_{j \in \mathbb{A}_c} C_j$ 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 $\mathbb{A}^* \equiv \{i^*(\mathbb{A}_c)\} \cup \mathbb{A}_{nc}$. (The set of colluding potential bidders is unknown to the researcher.)

Entrants know the composition of the set of competitors. Specifically, they know that a representative of the cartel is participating in bidding. They follow Bayesian Nash equilibrium bidding strategies. Bidders are ex ante symmetric in the sense that each interested bidder's private cost is drawn independently from the same distribution.

An empirical question is how to detect potential colluders $\mathbb{N}_c$ from observations on bidding and participation. Let $\mathbb{N}'_c \subseteq \mathbb{N}$ denote the set of bidders such that no two bidders in $\mathbb{N}'_c$ are ever observed to compete with each other in the bidding stage. By construction, $\mathbb{N}_c \subseteq \mathbb{N}'_c$ so the latter should be interpreted as a set of suspects for collusion. However, the set $\mathbb{N}'_c$ 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 $\mathbb{N}'_c$, that is, to separate $\mathbb{N}_c$ from $\mathbb{N}'_c \backslash \mathbb{N}_c \equiv \mathbb{N}_{nc} \cap \mathbb{N}'_c$.

Pesendorfer (2000) (Remark 3) shows that in any given auction with participants $\mathbb{A}_c \cup \mathbb{A}_{nc}$, 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 $i^*$.[8]

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

[8]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 $\mathbb{N}$ into $\mathbb{N}_c$ and $\mathbb{N}_{nc}$ is common knowledge among all bidders.

Specifically, for any $i^*$ and $j$ defined as above,

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

for all $b$ on the common support of the two distributions.[9] This result is similar to the one obtained in section 2.3.1 since in Pesendorfer (2000) the sole bidder representing a cartel has a higher hazard rate than a non-colluding bidder. That is, relative to a competitive bidder from the cartel representative has a higher probability of having a low cost conditional on the costs being above any fixed threshold.

Further, the ex ante symmetry assumption implies:

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

whenever $i, j \in \mathbb{N}_c$ or $i, j \in \mathbb{N}_{nc} \cap \mathbb{N}_c'$.

Now the pairwise indices $\delta^+_{i,j}$, $\delta^0_{i,j}$, and $\delta^-_{i,j}$ can be constructed in a way similar to (2.3.1). Note that on the superficial level, this example is close to the asymmetric first-price auction example in Section 2.2.1. Nevertheless in the current example, no two colluders ever enter the same auction. This necessitates the use of a different index.

## 3. Identification and Estimation of the Ordered Group Structure

3.1. **Identification of the Ordered Group Structure.** As we saw previously in various examples, economic theory often yields pairwise inequality restrictions on actions and outcomes associated with agents of distinct types. Here we explain how such restrictions could be used to establish identification of the group structure of the set of agents.

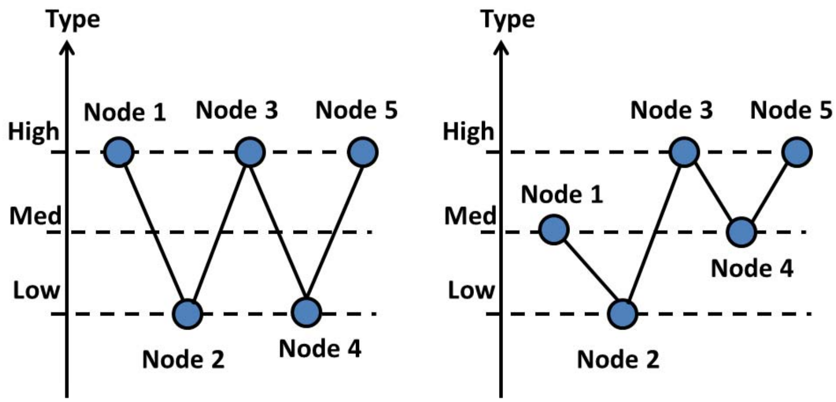
More formally, let $\mathcal{P}$ denote a collection of distributions of observed random vectors generated by the model. Further, let $\mathcal{E}$ be a collection of pairs $(i, j)$, with $i, j \in \mathbb{N}$ and $i \neq j$ such that there exist three *pairwise comparison indices* $\delta^+_{ij}$, $\delta^0_{ij}$ and $\delta^-_{ij}$ that are identified by $P \in \mathcal{P}$, and satisfy (2.1).[10] We call the pair $G = (\mathbb{N}, \mathcal{E})$ the *comparability network* in this paper.

We say that the ordered group structure $T$ is *identified*, if each $P \in \mathcal{P}$, and thus compatibility network $G = (\mathbb{N}, \mathcal{E})$, uniquely determines $T$. Throughout this paper, we assume that the econometrician knows the comparability network $G = (\mathbb{N}, \mathcal{E})$. This assumption is weak and satisfied in many empirical applications. For example, the assumption is satisfied when for each given pair $(i, j)$ in $\mathcal{E}$, sufficiently many observations

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

[10]Note that some agents may not be directly comparable on the basis of available data. That is why $\mathcal{E}$ does not necessarily include all possible pairs $(i, j)$.

Figure 1: Underidentification of the Group Structure




Note: This figure illustrates the case with 5 nodes and comparison edges given by $\mathcal{E} = \{(1,2),(2,3),(3,4),(4,5)\}$. We can determine the order between each pair $(i,j)$ in $\mathcal{E}$. However, we cannot identify the group structure from these pairwise orderings. The two panels depict two different group structures that are compatible with the same pairwise ordering.

are available for comparison of agents $i$ and $j$ and therefore the indices $\delta_{ij}^{+}$, $\delta_{ij}^{0}$, and $\delta_{ij}^{-}$ can be consistently estimated from the data.

It is easy to see that the identifiability of $T$ depends on the "size" of comparability network $G$. Certainly, if $\mathcal{E}$ contains only a small subset of possible pairs, we may not be able to identify the group structure. This is true even if $G$ is a connected graph,[11] as we illustrate below.

**Counterexample:** Suppose that we have five nodes in the network and $\mathcal{E}$ is given by

$$\mathcal{E} = \{(1,2),(2,3),(3,4),(4,5)\}.$$

Then for each pair $(i,j)$ in $\mathcal{E}$, we can determine whether $i$ has a higher, lower, or the same type, relative to node $j$. However, we cannot identify the whole group structure from this comparability network. For example, the following two group structures are entirely compatible with the pairwise comparisons along the edges in $\mathcal{E}$:

Group Structure 1 : $1, 3$ and 5 are high type, and 2 and 4 are low type.

Group Structure 2 : 3 and 5 are high type, 1 and 4 middle type, and 2 low type.

[11]We call a network $G = (\mathbb{N}, \mathcal{E})$ *connected*, if any two agents $i$ and $j$ could be linked by some intermediate pairs of agents $(i_1, j_1), ..., (i_m, j_m)$ such that $(i, i_1), (i_1, j_1), (j_1, i_2), ..., (i_m, j_m)$ and $(j_m, j)$ are in $\mathcal{E}$.

Both group structures yield pairwise comparisons such that 2 is of lower type than 1, 3 of higher type than 2, 4 of lower type than 3, and 5 of higher type than 4. Therefore, it shows that the identification of $T$ may fail depending on the configuration of $G$. ■

In this paper we focus on the case where the comparability network $G$ is complete. In other words, $\mathcal{E}$ contains all the pairs $(i,j)$ such that $i \neq j$. It is easy to see that in this case the group structure $T$ is identified through the following algorithm. In the first step, start with the full set $N$: select a group of agents $N_1$ such that $\delta_{ij}^{-} > 0$ for all $i \in N_1$ and $j \notin N_1$ whereas $\delta_{ik}^{0} = 0$ for all $i,\ k\ \in N_1$. Then, in the next step consider the set complementary to $N_1$ in $N$: select a group of agents $N_2$ such that $\delta_{ij}^{-} > 0$ for all $i \in N_2$ and $j \notin N_1 \cup N_1$ whereas $\delta_{ik}^{0} = 0$ for all $i,\ k\ \in N_2$. Continue like this until all agents are allocated into equi-type groups.

3.2. **Pairwise Hypothesis Testing Problems.** In this section, we develop a method to estimate the group structure consistently. The main challenge in this attempt lies in the fact that we are given only pairwise comparisons. One may be able to determine fairly accurately the type ordering between each pair of agents $i$ and $j$, but these pairwise comparisons do not necessarily generate a coherent estimate of the ordered support $Q_{K_0}$. In other words, while the elements of the population set $Q_{K_0}$ satisfy transitivity, the estimated ordering between two agents may not satisfy transitivity in finite samples. Thus this paper proposes estimating the whole group structure simultaneously with transitivity imposed on the structure.

We first formulate the following three pairwise hypothesis testing problems for each $(i,j) \in \mathbb{N}_2$, where $\mathbb{N}_2 = \{(i,j) \in \mathbb{N}^2 : i \neq j\}$:

$$
\begin{aligned}
H_{0,ij}^{+} &: \ \delta_{ij}^{+} \leq 0 \text{ against } H_{1,ij}^{+} : \delta_{ij}^{+} > 0, \\
H_{0,ij}^{0} &: \ \delta_{ij}^{0} = 0 \text{ against } H_{1,ij}^{0} : \delta_{ij}^{0} \neq 0 \text{ and} \\
H_{0,ij}^{-} &: \ \delta_{ij}^{-} \leq 0 \text{ against } H_{1,ij}^{-} : \delta_{ij}^{-} > 0.
\end{aligned} \tag{3.1}
$$

In most examples, we have various tests available. Instead of committing ourselves to a particular method of hypothesis testing, let us assume generally that we are given $p$-values $\hat{p}_{ij}^{+}$, $\hat{p}_{ij}^{0}$ and $\hat{p}_{ij}^{-}$ from the testing of $H_{0,ij}^{+}$, $H_{0,ij}^{0}$ and $H_{0,ij}^{-}$, against $H_{1,ij}^{+}$, $H_{1,ij}^{0}$ and $H_{1,ij}^{-}$ respectively. Let $L$ be the size of the sample that is used to construct these $p$-values. We explain details for construction of $p$-values using bootstrap later in Section 3.3.4.

As for these pairwise $p$-values, we make the following assumptions.

**Assumption 1:** For each pair $(i,j) \in \mathbb{N}_2$, and each $s \in \{+,-,0\}$, the following is satisfied.[12]

(a) Under $H^s_{0,ij}$, $\log \hat{p}^s_{ij} = O_P(1)$, as $L \to \infty$.

(b) There exists a sequence $r_L \to \infty$ as $L \to \infty$ such that under $H^s_{1,ij}$,

$$\begin{aligned} r_L^{-1} \log \hat{p}^s_{ij} &= O_P(1) \text{ and} \\ r_L'^{-1} \log \hat{p}^s_{ij} &\to {}_P -\infty, \text{ as } L \to \infty, \end{aligned}$$

for all $r'_L \to \infty$ such that $r'_L/r_L \to 0$,

Assumption 1(a) tells us that the $p$-values are from consistent tests of pairwise hypotheses in (3.1). This means that we can distinguish between the null and the alternative hypotheses in each hypothesis testing problem with probability approaching one for large samples. Assumption 1 is mostly satisfied when we have consistent estimators of $\delta^+_{ij}$, $\delta^0_{ij}$ and $\delta^-_{ij}$ for each pair $(i,j)$. Then we can construct consistent tests based on the consistent estimators.

Assumption 1(b) requires a rate of convergence for the logarithm of the $p$-values under the alternative hypothesis. For example, Assumption 1(b) excludes the case where for all $(i,j) \in \mathbb{N}_2$, $\hat{p}^0_{ij}$ converges in probability to 1. When $i$ and $j$ are in the same group, $\hat{p}^0_{ij}$ becomes close to a uniform random variable in a typical situation, and when $i$ and $j$ are in different groups, $\hat{p}^0_{ij}$ converges in probability to zero. Therefore, Assumption 1(b) is often satisfied. Usually, the rate of divergence $r_L$ is equal to $\sqrt{L}$ when the rate of the test is $\sqrt{L}$. Of course our procedure also admits other rates.

We explain our estimation method in three steps. First, we focus on the case where it is known that there are two groups. Second, we turn to the case where there are multiple groups with a known number of the groups. Third, we extend the method to the case where we do not know the number of the distinct types.

## 3.3. **Consistent Estimation.**

### 3.3.1. *Estimation of the Group Structure with Two Groups using a Split Algorithm.*

Suppose that the econometrician knows that there are two distinct types, i.e., $K_0 = 2$, so that for each $i \in \mathbb{N}$, $q_i \in \{\bar{q}_h, \bar{q}_l\}$ for some two unknown numbers $\bar{q}_h$ and $\bar{q}_l$ such that $\bar{q}_h > \bar{q}_l$. In this case, there are several different ways of partitioning $\mathbb{N}$ into two groups using the $p$-values satisfying Assumption 1. Here we choose a method that permits a natural extension to a more general case of $K_0 > 2$.

[12] For a given sequence of random variables $\{X_L\}_{L=1}^\infty$, the divergence in probability ($X_L \to_P -\infty$, as $L \to \infty$) is defined as follows: for any $M > 0$, $P\{X_L < -M\} \to 1$, as $L \to \infty$.

First, let

$$\begin{aligned}\mathbb{N}_h &\equiv \{i \in \mathbb{N} : q_i = \bar{q}_h\} \text{ and} \\ \mathbb{N}_l &\equiv \{i \in \mathbb{N} : q_i = \bar{q}_l\},\end{aligned}$$

so that the group structure is given by

$$T = (\mathbb{N}_l, \mathbb{N}_h).$$

Even with this case of two groups, the total number of potential partitions of $\mathbb{N}$ is $2^N$, $N = |\mathbb{N}|$. Instead of checking all the incidences of the potential partitions, we propose a split algorithm that estimates $T$ in three steps as follows.

**Step 1:** Define for each $i \in \mathbb{N}$,

$$\begin{aligned}\hat{\mathbb{N}}_1(i) &= \{j \in \mathbb{N}\backslash\{i\} : \log \hat{p}_{ij}^+ < \log \hat{p}_{ij}^- - \varepsilon r_L'\} \text{ and} \\ \hat{\mathbb{N}}_2(i) &= \{j \in \mathbb{N}\backslash\{i\} : \log \hat{p}_{ij}^+ > \log \hat{p}_{ij}^- + \varepsilon r_L'\},\end{aligned}$$

where $r_L'$ is as in Assumption 1(b) and $\varepsilon > 0$ is a small number.[13] The group $\hat{\mathbb{N}}_1(i)$ is the estimated set of agents whose type is construed to be lower than the agent $i$ and the group $\hat{\mathbb{N}}_2(i)$ is the estimated set of agents whose type is construed to be higher than the agent $i$. We call the agent $i$ in this step a *pivotal agent*. Next, we need to decide whether to add the pivotal agent to $\hat{\mathbb{N}}_1(i)$ or $\hat{\mathbb{N}}_2(i)$ to form an ordered partition of $\mathbb{N}$.

**Step 2:** For each $i \in \mathbb{N}$, we define[14]

$$\begin{aligned}s_1(i) &= \frac{1}{|\hat{\mathbb{N}}_1(i)|} \sum_{j \in \hat{\mathbb{N}}_1(i)} \log \hat{p}_{ij}^0 \text{ and} \\ s_2(i) &= \frac{1}{|\hat{\mathbb{N}}_2(i)|} \sum_{j \in \hat{\mathbb{N}}_2(i)} \log \hat{p}_{ij}^0,\end{aligned}$$

and let

$$s(i) = \begin{cases} s_1(i), \text{ if } i \in \mathbb{N}\backslash\hat{\mathbb{N}}_1(i) \\ s_2(i), \text{ otherwise} \end{cases},$$

where $|\hat{\mathbb{N}}_2(i)|$ and $|\hat{\mathbb{N}}_1(i)|$ denote the cardinalities of the sets $\hat{\mathbb{N}}_1(i)$ and $\hat{\mathbb{N}}_2(i)$. For each index $i$, we have two classifications, $(\hat{\mathbb{N}}_1(i), \mathbb{N}\backslash\hat{\mathbb{N}}_1(i))$ or $(\mathbb{N}\backslash\hat{\mathbb{N}}_2(i), \hat{\mathbb{N}}_2(i))$. The first classification regards $i$ as high type and the second classification regards $i$ as low type.

[13]The truncation with a small number $\varepsilon$ here is made for theoretical convenience. In our simulation studies, we have ignored this truncation and found the results satisfactory.

[14]Alternatively, one could use $\hat{p}_{ij}^+$ in the definition of $s_1(i)$ and $\hat{p}_{ij}^-$ in the definition of $s_2(i)$. The consistency results of this paper are not affected by this modification.

The index $s(i)$ measures the degree of misclassification caused by each of the two cases. When most agents are correctly classified, $s(i)$ becomes severely negative.

**Step 3:** We choose $i^*$ that minimizes $s(i)$ over $i \in \mathbb{N}$, i.e.,

$$i^* = \text{argmin}_{i \in \mathbb{N}} s(i). \tag{3.2}$$

Now we take

$$(\hat{\mathbb{N}}_l, \hat{\mathbb{N}}_h) = \begin{cases} (\mathbb{N} \backslash \hat{\mathbb{N}}_2(i^*), \hat{\mathbb{N}}_2(i^*)), & \text{if } s_1(i^*) \geq s_2(i^*) \\ (\hat{\mathbb{N}}_1(i^*), \mathbb{N} \backslash \hat{\mathbb{N}}_1(i^*)), & \text{if } s_1(i^*) < s_2(i^*). \end{cases}$$

We take the estimator of $T$ as

$$\hat{T} = (\hat{\mathbb{N}}_l, \hat{\mathbb{N}}_h).$$

Let us explore the sense in which $\hat{T}$ is a "reliable" estimator of $T$. This requires comparing two different ordered partitions $\hat{T}$ and $T$ of the same set. We introduce a metric of classification discrepancy as follows. Let $\mathcal{T}_2$ be the collection of all the ordered 2-partitions of $\mathbb{N}$. Then, we define a map $\delta : \mathcal{T}_2 \times \mathcal{T}_2 \to \{0\} \cup \mathbb{N}$ as

$$\delta(T_1, T_2) = \max \left\{ \left| \mathbb{N}_h^1 \triangle \mathbb{N}_h^2 \right|, \ \left| \mathbb{N}_l^1 \triangle \mathbb{N}_l^2 \right| \right\},$$

for any two different ordered partitions, $T_1 = (\mathbb{N}_h^1, \mathbb{N}_l^1)$ and $T_2 = (\mathbb{N}_h^2, \mathbb{N}_l^2)$, of $\mathbb{N}$, where for any two sets $A$ and $B$, $A \triangle B$ denotes the set difference between $A$ and $B$ (that is, the difference between their intersection and their union).

The theorem below establishes that $\hat{T}$ is consistent for $T$ in the sense that $\delta(\hat{T}, T) = 0$ with probability approaching one as $L \to \infty$.

**Theorem 1:** *Suppose that Assumption 1 holds. Then as $L \to \infty$,*

$$P\left\{ \delta(\hat{T}, T) = 0 \right\} \to 1.$$

Theorem 1 shows that using consistent pairwise tests of $H_{0,ij}^+$, $H_{0,ij}^0$, and $H_{0,ij}^-$, we can determine the classification of each agent with the probability of misclassification vanishing with the growing sample size $L$.

While there may be other alternative methods to obtain the ordered classification, the three-step split method is deliberately designed to satisfy the following properties.

First, the split algorithm is designed so that it can be used in the extension to a general case of $K > 2$. The intuition is as follows. Suppose that we have $K$ groups that are ordered, so that we have $T = (\mathbb{N}_1, ..., \mathbb{N}_K)$. The design of Step 1 ensures that whenever $i$ is of type $k$, $\hat{\mathbb{N}}_1(i)$ coincides with the union of $j$'s that have lower type than $i$ with probability approaching one, and $\hat{\mathbb{N}}_2(i)$ coincides with the union of $j$'s that have higher type than $i$ with probability approaching one. In other words, however $i$ may be chosen, the probability that the two splits $(\mathbb{N}\backslash\hat{\mathbb{N}}_2(i), \hat{\mathbb{N}}_2(i))$ and $(\hat{\mathbb{N}}_1(i), \mathbb{N}\backslash\hat{\mathbb{N}}_1(i))$ splitting an equi-type group into two different groups is negligible when the sample size $L$ is large. It only remains to find estimated groups that are likely to be of heterogeneous types and continue to split such groups.

Second, the algorithm is computationally feasible in many practical set-ups. The algorithm does not require $2^N$ comparisons of candidate group structures as a brute-force approach would. As this split algorithm forms a basic tool for the general case of unknown groups later, it is crucial that the algorithm do not incur heavy computational cost at this simple set-up of two groups.

Third, in this paper avoids comparing directly the $p$-values with a level of the test. Our use of test is not for its own sake but a tool for the consistent estimation of the group structure. Therefore, it is not clear what level one should use in practice. Furthermore, we need to carefully design an algorithm so that it treats the two cases of $\tau(i) < \tau(j)$ and $\tau(j) > \tau(i)$ symmetrically, despite the fact that the individual hypothesis testing problem treats the null hypothesis of $\tau(i) \leq \tau(j)$ and the alternative hypothesis of $\tau(j) > \tau(i)$ asymmetrically. Our algorithm compares $p$-values with $p$-values to minimize the use of tuning parameters left to choose in practice, and treats the inequalities symmetrically.

Fourth, designing a consistent classification method does not always ensure good finite sample properties. Note that there can be numerous variations to the method that do not affect the consistency of the estimated groups. However, these variations typically affect the finite sample performance of the estimator. The three-step algorithm is developed through various Monte Carlo experiments so that it works well in finite sample environments.

3.3.2. *Estimation of a Classification with a Known Number of Groups.* We generalize the procedure to the case where the econometrician knows the number of distinct types $K$ that can be greater than 2. The main idea here is that we split the groups sequentially using the previous algorithm. For any $1 \leq k \leq K$, let $\mathbb{J}_k \equiv \{1, 2, ..., k\}$.

**Step 1**: Split $\mathbb{N}$ into $\hat{\mathbb{N}}_+$ and $\hat{\mathbb{N}}_-$ using the split algorithm in the previous section.

**Step 2**: We relabel $\hat{\mathbb{N}}_1 = \hat{\mathbb{N}}_-$ and $\hat{\mathbb{N}}_2 = \hat{\mathbb{N}}_+$, and compute

$$\hat{p}_1 = \min_{i,j \in \hat{\mathbb{N}}_1 : i \neq j} \hat{p}^0_{ij} \text{ and } \hat{p}_2 = \min_{i,j \in \hat{\mathbb{N}}_2 : i \neq j} \hat{p}^0_{ij}.$$

We choose $r^*$ such that

$$\hat{p}_{r^*} = \min_{r \in \mathbb{J}_2} \hat{p}_r \tag{3.3}$$

and use the algorithm in the previous section to split $\hat{\mathbb{N}}_{r^*}$ into $\hat{\mathbb{N}}_{r^*,h}$ and $\hat{\mathbb{N}}_{r^*,l}$ to obtain a classification of $\mathbb{N}$ into 3 groups.

**Step** $k$: In general, suppose that we have classifications $\hat{\mathbb{N}}_1, \hat{\mathbb{N}}_2, ..., \hat{\mathbb{N}}_k$ (after relabeling the groups). For each $r = 1, ..., k$, we compute

$$\hat{p}_r = \min_{i,j \in \hat{\mathbb{N}}_r : i \neq j} \hat{p}^0_{ij}. \tag{3.4}$$

We choose $r^*$ such that

$$\hat{p}_{r^*} = \min_{r \in \mathbb{J}_k} \hat{p}_r$$

and use the algorithm in the previous section to split $\hat{\mathbb{N}}_{r^*}$ into $\hat{\mathbb{N}}_{r^*,h}$ and $\hat{\mathbb{N}}_{r^*,l}$ to obtain

$$\hat{\mathbb{T}}_{k+1} = (\hat{\mathbb{N}}_1, \hat{\mathbb{N}}_2, ..., \hat{\mathbb{N}}_{r^*-1}, \hat{\mathbb{N}}_{r^*,l}, \hat{\mathbb{N}}_{r^*,h}, \hat{\mathbb{N}}_{r^*+1}, ..., \hat{\mathbb{N}}_k).$$

We continue until the total number of groups obtained becomes $K$.

As mentioned after Theorem 1, the probability that a split divides an equi-type group into two is negligible. When we are given classifications $\hat{\mathbb{N}}_1, \hat{\mathbb{N}}_2, ..., \hat{\mathbb{N}}_k$, we define $\hat{p}_r$ as in (3.4) which is used as a group homogeneity index. Intuitively, when $\hat{p}_r$ is low, the group $\hat{\mathbb{N}}_r$ is likely to be heterogeneous. We select a group with the lowest homogeneity index and split the group.

Let $\mathcal{T}_K$ be the collection of all the ordered $K$-partitions of $\mathbb{N}$. Then, we define a map $\delta : \mathcal{T}_K \times \mathcal{T}_K \rightarrow \{0\} \cup \mathbb{N}$ as

$$\delta(T_1, T_2) = \max_{1 \leq k \leq K} \left\{ \left| \mathbb{N}^1_k \triangle \mathbb{N}^2_k \right| \right\},$$

for any two different ordered partitions, $T_1 = (\mathbb{N}^1_1, ..., \mathbb{N}^1_K)$ and $T_2 = (\mathbb{N}^2_1, ..., \mathbb{N}^2_K)$, of $\mathbb{N}$. In the following, we show that this generalized version of classification is consistent.

**Theorem 2:** *Suppose that Assumption 1 holds. Assume furthermore that $\hat{T}_K$ is obtained at Step $K-1$. Then, we have as $L \rightarrow \infty$,*

$$P\left\{ \delta(\hat{T}_K, T) = 0 \right\} \rightarrow 1.$$

Theorem 2 shows that the estimated group structure $\hat{T}_K$ is consistent for $T$, as $L \rightarrow \infty$, when the number of groups $K$ is known.

3.3.3. *Consistent Selection of the Number of Groups.* Let us extend the method to the case where the number of groups is not known. Our proposal selects the number of groups that minimizes a criterion function which balances a measure of goodness-of-fit that captures a misspecification bias and a penalty term for over-fitting.

We assume that there is a known upper bound for the number of groups.

**Assumption 2:** There exists a known upper bound integer $\bar{K}$ such that $K_0 \leq \bar{K} \leq N$.

Assumption 2 is not very restrictive in practice, because one can choose a plausibly large enough $\bar{K}$.

Suppose that we assume $K$ groups and follow the sequential process in the previous subsection, and obtain the group structure:

$$\hat{T}_K = (\hat{\mathbb{N}}_1, \hat{\mathbb{N}}_2, ..., \hat{\mathbb{N}}_K). \tag{3.5}$$

We define

$$V(K) = \frac{1}{K}\sum_{k=1}^{K}\left|\min_{i,j\in\hat{\mathbb{N}}_k} \log \hat{p}_{ij}^0\right|.$$

Let $r_L$ be the sequence in Assumption 1(b) and let $g(L) \to \infty$ be such that $g(L)/r_L \to 0$ as $L \to \infty$.[15] We define our criterion function as follows:

$$\hat{Q}(K) \equiv V(K) + \frac{V(\bar{K})Kg(L)}{\bar{K}},$$

where the factor $V(\bar{K})/\bar{K}$ is made as scale normalization. The component $V(K)$ measures the goodness-of-fit of the classification, and the second component $V(\bar{K})Kg(L)/K$ plays the role of a penalty term that prevents overfitting. We select $\hat{K}$ as follows:

$$\hat{K} = \text{argmin}_{1\leq K\leq \bar{K}}\hat{Q}(K).$$

Now we consider the definition of the discrepancy measure between $\hat{T}_{\hat{K}}$ and $T$ by extending the previous measure to the case where $\hat{K}$ is not equal to $K_0$ in finite samples:

$$\delta(\hat{T}_{\hat{K}}, T) = \max_{1\leq k\leq \max\{K_0,\hat{K}\}}\left|\hat{\mathbb{N}}_k \triangle \mathbb{N}_k\right|, \tag{3.6}$$

where we set

$$\hat{\mathbb{N}}_k = \varnothing \text{ if } \hat{K} < k \leq K_0 \text{ and } \mathbb{N}_k = \varnothing \text{ if } K_0 < k \leq \hat{K}.$$

The following theorem shows that this selection procedure is a consistent one.

**Theorem 3:** *Suppose that Assumptions 1-2 hold. Then, we have as* $L \to \infty$,

$$P\{\hat{K} = K_0\} \to 1,$$

[15]There are multiple choices possible for $g(L)$. In our simulation study, we used $g(L) = \log\log L$.

*and hence the estimated group structure $\hat{T}_{\hat{K}}$ with selected $\hat{K}$ satisfies that as $L \to \infty$,*

$$P\left\{\delta(\hat{T}_{\hat{K}}, T) = 0\right\} \to 1.$$

The main part of Theorem 3 is to show that $\hat{K}$ is consistent for $K_0$. When $K < K_0$, the component $V(K)$ cannot shown to be $O_P(1)$, while the penalty term increases as $L \to \infty$. When $K > K_0$, the component $V(K)$ diverges at a rate faster than $g(L)$ by Assumption 1(b). From this, we obtain that $\hat{K}$ is consistent for $K_0$.

3.3.4. *Construction of p-Values Using Bootstrap.* In most applications, we can use bootstrap to construct $p$-values for testing the inequality restrictions of (3.1).[16] We explain this procedure using the environment considered in Monte Carlo experiments and in an empirical application of our method as an illustration. In these sections we consider data that are generated by the first-price auction market.

Formally, suppose that we are given observations $\{Z_l\}_{l=1}^L$, where $Z_l = (Z_{i,l})_{i=1}^n$ denotes the observations pertaining to auction $l$ and $Z_{i,l}$ denotes the vector of observations for bidder $i$. The random vector $Z_{i,l}$ includes the bid submitted by bidder $i$ and other observed bidder or auction characteristics at auction $l$. Suppose that for each pair of bidders $i$ and $j$, there exists a nonparametric function, say, $r_{ij}(b)$ such that

$$\begin{aligned}
\tau(i) &> \tau(j) \text{ if and only if } r_{ij}(b) > 0 \text{ for all } b \text{ in } B, \\
\tau(i) &= \tau(j) \text{ if and only if } r_{ij}(b) = 0 \text{ for all } b \text{ in } B, \text{ and} \\
\tau(i) &< \tau(j) \text{ if and only if } r_{ji}(b) > 0 \text{ for all } b \text{ in } B,
\end{aligned}$$

where $B$ is the domain of the function $r_{ij}(b)$.

To construct a test statistic, we first estimate $r_{ij}(b)$ using the sample $\{Z_l\}_{l=1}^L$ to obtain $\hat{r}_{ij}(b)$. Then we construct the following test statistics:

$$\begin{aligned}
T_{ij}^+ &= \int \max\{\hat{r}_{ij}(b), 0\}\, db, \\
T_{ij}^- &= \int \max\{\hat{r}_{ji}(b), 0\}\, db, \text{ and} \\
T_{ij}^0 &= \int |\hat{r}_{ij}(b)|\, db.
\end{aligned} \tag{3.7}$$

For concreteness, we use the integration to form a test statistic, but one may choose to use other functionals such as supremum over $b \in B$.

[16] For a more recent strand of research, see Bugni (2010), Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Lee, Song, and Whang (2013), and Lee, Song, and Whang (2014), among many others, and references therein.

For $p$-values, we resample $\{Z_l^*\}_{l=1}^L$ (with replacement) from the empirical distribution of $\{Z_l\}_{l=1}^L$ and construct a nonparametric estimator $\hat{r}_{ij}^*(b)$ for each pair $(i,j)$ in the same way as we did using the original sample. Using these bootstrap estimators, we construct the following bootstrap test statistics:

$$
\begin{aligned}
T_{ij}^{+*} &= \int \max\left\{\hat{r}_{ij}^*(b) - \hat{r}_{ij}(b), 0\right\} db, \\
T_{ij}^{-*} &= \int \max\left\{\hat{r}_{ji}^*(b) - \hat{r}_{ji}(b), 0\right\} db \text{ and} \\
T_{ij}^{0*} &= \int \left|\hat{r}_{ij}^*(b) - \hat{r}_{ij}(b)\right| db.
\end{aligned}
\tag{3.8}
$$

Note that the bootstrap test statistic involves recentering to impose the null hypothesis. Now, the $p$-values, $\hat{p}_{ij}^+$, $\hat{p}_{ij}^-$, and $\hat{p}_{ij}^0$ can be constructed from the bootstrap distributions of $T_{ij}^{+*}$, $T_{ij}^{-*}$, and $T_{ij}^{0*}$ respectively.[17]

## 4. Monte Carlo Simulations

The Monte Carlo simulation study is based on an example of independent private value (IPV) auction with bidder asymmetry. We let the bidders be classified into $K_0$ groups. We abstract away from details in the formation of equilibrium strategies, and draw bids from a normal distribution $N(\mu_k, \sigma^2)$ directly, so that we have $\{B_{i1}, ..., B_{iL}\}$ for each bidder $i$ whenever the bidder belongs to the group $k$. The number of observations here is $L$, which represents the number of auctions observed in the data.[18]

In the simulation study we consider the configurations of group structures as shown in Table 1. Structures S1 and S2 involve a total of 12 bidders and S3 and S4 involve a total of 40 bidders. Also, S1 and S3 use a coarser group structure than S2 and S4. The number of observed auctions ($L$) and specifications of parameters $\mu_k$ and $\sigma^2$ used in our simulation study are given in Table 2.

The specification P1 uses $\mu_1, ..., \mu_4$ that are most set apart from each other among the three groups of specifications. On the other hand, P3 uses means that are closer to each other. In our simulation below, we also report performance of our classification estimator in small samples (with 100 observations) when $N$ is as large as 40 and the difference in the unobserved heterogeneity is as small as in P3.

[17]When $p_L\{\hat{r}_{ij}(\cdot) - r_{ij}(\cdot)\}$ converges weakly to a Gaussian process, the distribution of the statistic and the bootstrap test statistic is derived as a functional of that process through the continuous mapping theorem. When $p_L\{\hat{r}_{ij}(\cdot) - r_{ij}(\cdot)\}$ does not weakly converge, as in the case of kernel regression/density estimators or local polynomial estimators, the test statistic has a limiting normal distribution after appropriate scale-location normalization. See Lee, Song, and Whang (2014) for details in this latter case.

[18]More specifically, $L$ is equal to the number of observations available for each pair of bidders.

Table 1: Group Structures

| Structure | $N$ | $K_0$ | $N_k$ |
|---|---|---|---|
| S1 | 12 | 2 | 6 |
| S2 | 12 | 4 | 3 |
| S3 | 40 | 2 | 20 |
| S4 | 40 | 4 | 10 |

Note: $N$ denotes the total number of the bidders; $K_0$ denotes the number of the groups; $N_k$ denotes the number of actual bidders from group $k$. For each structure in the simulation design, groups all have the same number of bidders.

Table 2: Parameter Specifications

| Specification | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ |
|---|---|---|---|---|
| P1 | 2.0 | 2.6 | 3.2 | 3.8 |
| P2 | 2.0 | 2.4 | 2.8 | 3.2 |
| P3 | 2.0 | 2.2 | 2.4 | 2.6 |

Note: For S1 and S3 (with two groups), we use $\mu_1$ and $\mu_2$ only. In S2 and S4 (with four groups), we use $\mu_1$, $\mu_2$, $\mu_3$, and $\mu_4$.

We construct $p$-values following the procedure described in (3.3.4) and obtain group classifications from 500 simulated samples. For each estimate we used 100 bootstrap iterations.

To measure the performance of our classification method, we consider three criteria:

$$\text{(Expected Maximum Discrepancies (EMD)):} \quad \mathbf{E}\left[\delta(\hat{T}_{\hat{K}}, T)\right]$$

$$\text{(Expected Sum of Discrepancies (ESD))} \quad : \quad \mathbf{E}\left[\Sigma_{k=1}^{\max\{K,\hat{K}\}} |\hat{\mathbb{N}}_k \triangle \mathbb{N}_k|\right] \text{ and}$$

$$\text{(Discrepancy Hazard at Proportion } p \text{ (DH(100}p\text{)))} \quad : \quad P\{\delta(\hat{T}_{\hat{K}}, T) > pN\},$$

where $\delta(\hat{T}_{\hat{K}}, T)$ is as defined in (3.6). The EMD is the expected value of the maximum discrepancy between the estimated group structure and the true group structure. The ESD is the expected sum of the discrepancies across groups. Finally the discrepancy hazard at proportion $p$ (abbreviated to be DH($100p$)) represents the probability that the maximum discrepancy is larger than the $p$ proportion of the total bidders. We report the results for $p \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. Also, for the choice of the penalty scheme, we used $g(L) = \log \log L$.

4.1. **The Case of Known Number of Groups.** The discussion of the results is divided into two cases: the case of a known number of groups and the case of an unknown number of groups. Considering the first case, we investigate the performance of the classification algorithm separately from the issue of data-dependent selection of the number of groups. The results are shown in Tables 3-4.

In Table 3, we see that the performance in the case of S1 is better than that of S2. Therefore, given the same number of bidders, it is easier to estimate the group structure when the number of groups is smaller. As expected, the performance is better when the difference between $\mu_1, ..., \mu_4$ is larger (as in P1). Also, the large number of samples helps procuring a better estimate of the group structure.

The results in Table 4 show that our estimator performs better when the number of the bidders ($N$) and the number of groups ($K_0$) are smaller. For example, it is not surprising that the performance is less satisfactory in S4 when we try to estimate the classification of 40 bidders into four groups using bids from only 100 auctions.

4.2. **The Case of Unknown Number of Groups.** We now turn to the case where the number of groups is unknown to the researcher. In this case the number of groups needs to be selected through our consistent selection method. The results are reported in Tables 5-6.

Like Table 3, Table 5 contains the results for S1 and S2. The results are mostly similar to the case with a known number of groups. Also when we compare Table 6 with Table 4, the difference does not seem substantial either. This demonstrates that the consistent selection of the number of the groups works well in the sense that the estimation error in $\hat{K}$ has little impact on the subsequent estimation of bidder classifications.

In summary, the estimated group structure does not have more than two bidders misclassified on average, in most cases of 400 or 200 auctions with twelve bidders with two or four groups or with forty bidders with two groups. In other cases, the estimated group structure can be unstable.

## 5. Empirical Application: California Market for Highway Procurement

We apply our methodology to the data on highway procurement auctions conducted by the California Department of Transportation (CalTrans).

CalTrans is responsible for the construction and maintenance of roads and highways in California. The services for the related projects are procured by means of first-price sealed-bid auctions. The projects are formulated as lists of tasks such as resurfacing, replacing the base or filling in cracks. The supply side of this market consists of 20-30 regular participants (referred to as "firms" or "contractors") who have been in the industry for many years and participated in many auctions. Auctions also attract fringe bidders that enter only very few auctions throughout the data.

Table 3: Performance of the Classification Estimator in S1 and S2
($K_0$ assumed known)

| | | $L$ | EMD | ESD | DH(10) | DH(25) | DH(50) | DH(75) | DH(90) |
|---|---|---|---|---|---|---|---|---|---|
| | P1 | 400 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P2 | 400 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P3 | 400 | 0.46 | 0.46 | 0.34 | 0.07 | 0.03 | 0.01 | 0 |
| | P1 | 200 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| S1 | P2 | 200 | 0.02 | 0.02 | 0.02 | 0 | 0 | 0 | 0 |
| | P3 | 200 | 1.54 | 1.54 | 0.46 | 0.24 | 0.16 | 0.03 | 0 |
| | P1 | 100 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P2 | 100 | 0.66 | 0.66 | 0.45 | 0.14 | 0.06 | 0.02 | 0 |
| | P3 | 100 | 1.64 | 1.58 | 0.66 | 0.40 | 0.18 | 0.06 | 0 |
| | P1 | 400 | 0.04 | 0.02 | 0.02 | 0.01 | 0 | 0 | 0 |
| | P2 | 400 | 0.12 | 0.06 | 0.04 | 0.03 | 0.01 | 0.01 | 0 |
| | P3 | 400 | 1.29 | 0.66 | 0.54 | 0.18 | 0.08 | 0.03 | 0 |
| | P1 | 200 | 0.07 | 0.04 | 0.03 | 0.02 | 0 | 0 | 0 |
| S2 | P2 | 200 | 0.12 | 0.06 | 0.06 | 0.03 | 0 | 0 | 0 |
| | P3 | 200 | 3.52 | 1.80 | 0.98 | 0.95 | 0.5 | 0.28 | 0.08 |
| | P1 | 100 | 0.10 | 0.05 | 0.04 | 0.03 | 0 | 0 | 0 |
| | P2 | 100 | 1.36 | 0.69 | 0.27 | 0.14 | 0 | 0 | 0 |
| | P3 | 100 | 3.87 | 1.96 | 0.98 | 0.96 | 0.65 | 0.35 | 0.14 |

Note: The number of bidders in S1 and S2 are both 12. EMD indicates the expected maximum number of bidders (across groups) that are misclassified. ESD indicates the expected total number of bidders that are misclassified. DH(10) indicates the probability that the maximum number of bidders that are misclassified is more than 10% of the 12 bidders.

As in other auction markets, the costs of participating firms vary across firms and across projects. The firms' costs are private information and are not directly observed by the government or by competitors. Auction theory (e.g. Myerson (1981) and Milgrom and Weber (1982)) suggests that in order to answer mechanism design questions or to assess the potential outcomes of policy interventions in this market, the researcher needs to know the distribution of private information at the level at which it is observable by the market participants. It is thus important to account in estimation for the firms' cost asymmetries (heterogeneity in cost distributions) known to all market participants. These concerns are especially important for long-term participants since their business practices are likely to become known in the market over time.

The most straightforward way to account for possible cost asymmetries is to estimate firm-specific cost distributions. This approach, however, is not feasible in most auction studies. This is because the primitives of an auction game (cost distributions) are linked to the observed auction outcomes (bid distributions) through a set of non-linear bidding

Table 4: Performance of the Classification Estimator in S3 and S4
($K_0$ assumed known)

| | | $L$ | EMD | ESD | DH(10) | DH(25) | DH(50) | DH(75) | DH(90) |
|---|---|---|---|---|---|---|---|---|---|
| S3 | P1 | 400 | 0.08 | 0.05 | 0 | 0 | 0 | 0 | 0 |
| | P2 | 400 | 1.35 | 1.35 | 0.08 | 0.06 | 0.04 | 0.01 | 0 |
| | P3 | 400 | 3.49 | 3.46 | 0.69 | 0.56 | 0.26 | 0.06 | 0 |
| | P1 | 200 | 0.12 | 0.12 | 0 | 0 | 0 | 0 | 0 |
| | P2 | 200 | 1.65 | 1.65 | 0.13 | 0.12 | 0.11 | 0.09 | 0.05 |
| | P3 | 200 | 3.66 | 3.66 | 0.79 | 0.77 | 0.39 | 0.07 | 0 |
| | P1 | 100 | 0.38 | 0.22 | 0.05 | 0.02 | 0 | 0 | 0 |
| | P2 | 100 | 1.84 | 1.76 | 0.24 | 0.21 | 0.13 | 0.09 | 0.05 |
| | P3 | 100 | 3.96 | 3.89 | 0.94 | 0.81 | 0.46 | 0.17 | 0.03 |
| S4 | P1 | 400 | 3.03 | 1.43 | 0.13 | 0.22 | 0.04 | 0.02 | 0 |
| | P2 | 400 | 3.06 | 1.95 | 0.25 | 0.23 | 0.07 | 0.03 | 0.03 |
| | P3 | 400 | 5.64 | 3.83 | 0.56 | 0.50 | 0.32 | 0.30 | 0.08 |
| | P1 | 200 | 3.07 | 1.63 | 0.15 | 0.28 | 0.08 | 0.02 | 0.02 |
| | P2 | 200 | 3.06 | 1.53 | 0.27 | 0.27 | 0.06 | 0.02 | 0.02 |
| | P3 | 200 | 9.53 | 7.27 | 0.78 | 0.58 | 0.44 | 0.18 | 0.10 |
| | P1 | 100 | 3.20 | 1.60 | 0.25 | 0.25 | 0.04 | 0.02 | 0.02 |
| | P2 | 100 | 7.33 | 5.17 | 0.53 | 0.45 | 0.29 | 0.09 | 0.06 |
| | P3 | 100 | 10.94 | 9.48 | 0.92 | 0.77 | 0.49 | 0.24 | 0.18 |

Note: The number of bidders in S3 and S4 are both 40.

strategies which have to be obtained by solving a system of differential equations that has a degeneracy on the boundary. If the cost asymmetries are defined at the level of an individual firm, the estimation would involve solving many different auction games for every parameter vector that is evaluated in estimation.

Such concerns do not arise in non-parametric studies since the bidding strategy and the underlying cost distribution could be recovered from the first-order conditions by applying them to appropriate bid distributions (see Guerre, Isabelle, and Vuong (2000)). However, this means that estimation has to be implemented conditional on the composition of the set of participants which summarizes the competitive structure of an auction known to all market participants and is reflected in bidding strategies. Thus, the aforementioned procedure is likely to be infeasible due to data limitations if the asymmetries are defined at the level of an individual firm.

Recent empirical studies tend to resolve these issues by estimating a parametrized distribution of bids and then using optimal first-order conditions to solve for the distribution of costs. However, such an approach is likely to lead to estimation bias since it is difficult to capture the impact of the competitive structure on bids through parameterization in the case of asymmetric bidders.

Table 5: Performance of the Classification Estimator in S1 and S2
($K_0$ unknown)

| | | $L$ | $\hat{K}$ | EMD | ESD | DH(10) | DH(25) | DH(50) | DH(75) | DH(90) |
|---|---|---|---|---|---|---|---|---|---|---|
| S1 | P1 | 400 | 2.00 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P2 | 400 | 2.00 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P3 | 400 | 2.01 | 0.46 | 0.45 | 0.34 | 0.07 | 0.03 | 0.01 | 0 |
| | P1 | 200 | 2.00 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P2 | 200 | 2.00 | 0.02 | 0.02 | 0.02 | 0 | 0 | 0 | 0 |
| | P3 | 200 | 2.07 | 1.50 | 1.43 | 0.56 | 0.40 | 0.15 | 0.08 | 0 |
| | P1 | 100 | 2.00 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 |
| | P2 | 100 | 2.00 | 0.66 | 0.66 | 0.45 | 0.13 | 0.05 | 0.02 | 0 |
| | P3 | 100 | 1.89 | 1.95 | 1.32 | 0.68 | 0.52 | 0.27 | 0.14 | 0 |
| S2 | P1 | 400 | 3.91 | 0.45 | 0.12 | 0.07 | 0.07 | 0.03 | 0 | 0 |
| | P2 | 400 | 3.87 | 0.43 | 0.11 | 0.07 | 0.07 | 0.03 | 0 | 0 |
| | P3 | 400 | 3.06 | 3.06 | 0.83 | 0.50 | 0.47 | 0.21 | 0.01 | 0 |
| | P1 | 200 | 3.84 | 0.52 | 0.14 | 0.08 | 0.08 | 0.04 | 0 | 0 |
| | P2 | 200 | 3.85 | 0.48 | 0.13 | 0.07 | 0.07 | 0.04 | 0 | 0 |
| | P3 | 200 | 2.58 | 4.14 | 1.07 | 0.92 | 0.50 | 0.38 | 0.19 | 0 |
| | P1 | 100 | 3.88 | 0.39 | 0.11 | 0.06 | 0.06 | 0.03 | 0 | 0 |
| | P2 | 100 | 3.43 | 3.29 | 3.29 | 0.56 | 0.50 | 0.21 | 0.01 | 0 |
| | P3 | 100 | 2.48 | 4.41 | 4.41 | 0.96 | 0.52 | 0.39 | 0.26 | 0 |

Note: The column of $\hat{K}$ indicates the average of the estimated number of groups across simulations. In S1, the true number of groups is 2, and in S2, the true number of groups is 4. As in Table 3, both specifications involve 12 bidders.

For this application, we rely on an alternative approach which is made possible by the classification methodology proposed in this paper. Specifically, we structure our estimation in two steps. In the first step we apply the classification routine to allocate regular firms into groups characterized by the same cost distribution. This step reduces the number of possible types and hence the computational burden of the estimation methodology. In the second step we recover the group-specific distributions of costs through a Generalized Method of Moments procedure.

Recent empirical studies of highway procurement emphasize the importance of taking bidders' participation decisions into account.[19] Specifically, they recognize that some of the project-specific factors that influence bidders' costs may not be observed in the data (unobserved auction heterogeneity). These factors may drive bidders' participation and

[19]For example, Hong and Shum (2002) study the evidence of the presence of winner's curse, Krasnokutskaya and Seim (2011) evaluate participation behavior and the impact of disadvantage business enterprise, Li and Zheng (2009) evaluate applicability of various models of participation, etc.

Table 6: Performance of the Classification Estimator in S3 and S4 (with $K_0$ unknown)

| | | $L$ | $\hat{K}$ | EMD | ESD | DH(10) | DH(25) | DH(50) | DH(75) | DH(90) |
|---|---|---|---|---|---|---|---|---|---|---|
| S3 | P1 | 400 | 2.00 | 0.12 | 0.12 | 0 | 0 | 0 | 0 | 0 |
| | P2 | 400 | 2.01 | 1.12 | 1.10 | 0.06 | 0.04 | 0.02 | 0 | 0 |
| | P3 | 400 | 2.04 | 3.43 | 3.23 | 0.69 | 0.55 | 0.28 | 0.06 | 0.04 |
| | P1 | 200 | 2.00 | 0.24 | 0.22 | 0 | 0 | 0 | 0 | 0 |
| | P2 | 200 | 2.02 | 1.62 | 1.60 | 0.13 | 0.12 | 0.11 | 0.08 | 0.05 |
| | P3 | 200 | 2.16 | 3.73 | 3.54 | 0.74 | 0.62 | 0.31 | 0.11 | 0.08 |
| | P1 | 100 | 2.00 | 0.49 | 0.42 | 0.18 | 0.12 | 0.08 | 0 | 0 |
| | P2 | 100 | 2.16 | 1.74 | 1.70 | 0.23 | 0.13 | 0.11 | 0.09 | 0.05 |
| | P3 | 100 | 2.56 | 4.34 | 4.24 | 0.83 | 0.72 | 0.39 | 0.15 | 0.08 |
| S4 | P1 | 400 | 3.83 | 3.96 | 2.07 | 0.23 | 0.23 | 0.07 | 0 | 0 |
| | P2 | 400 | 3.51 | 6.91 | 2.13 | 0.27 | 0.27 | 0.12 | 0 | 0 |
| | P3 | 400 | 2.85 | 9.32 | 4.14 | 0.81 | 0.73 | 0.46 | 0 | 0 |
| | P1 | 200 | 3.76 | 3.96 | 2.07 | 0.22 | 0.23 | 0.07 | 0 | 0 |
| | P2 | 200 | 3.31 | 7.29 | 2.15 | 0.31 | 0.28 | 0.14 | 0 | 0 |
| | P3 | 200 | 2.44 | 10.44 | 5.37 | 0.89 | 0.78 | 0.52 | 0.16 | 0.07 |
| | P1 | 100 | 3.22 | 4.04 | 2.19 | 0.28 | 0.22 | 0.14 | 0.05 | 0 |
| | P2 | 100 | 2.34 | 9.25 | 3.15 | 0.36 | 0.31 | 0.28 | 0.07 | 0.01 |
| | P3 | 100 | 2.05 | 15.00 | 7.08 | 0.95 | 0.81 | 0.64 | 0.25 | 0.13 |

Note: In S3, the true number of groups is 2 and in S4, the true number of groups is 4. As in Table 4, both S3 and S4 involve 40 bidders.

bidding decisions, thus, generating endogeneity of the competitive structure of a given auction. We take this feature into account in our analysis.

5.1. **Model.** A project $l$ is summarized by a set of characteristics $X_l$ which are observable to a researcher and an unobservable factor $U_l$. It is associated with a set of potential bidders, $\mathcal{N}_l$, consisting of companies that are qualified to perform the required work and are available during the time of an auction. We assume that this set is independent of the unobserved factor $U_l$ and that the factor $U_l$ is distributed according to normal distribution with the mean normalized to zero and a standard deviation $\sigma_U$.

A company $i$ that is a potential bidder for project $l$ is characterized by private entry costs, $\kappa_{i,l}$, and the private cost of completing the project, $C_{i,l}$. We assume that entry costs vary independently across bidders and auctions and are distributed according to the exponential distribution with a rate parameter $\lambda_E = X_l\gamma$. The costs of completing the work are drawn from a lognormal distribution with mean $\mu_C$ and standard deviation $\sigma_C$. The mean of the cost distribution depends on project characteristics, the distance between the project and the bidder's locations, $D_{i,l}$, as well as an unobserved bidder-specific cost factor $q_i$ which takes discrete values in $\{\bar{q}_0,\, \bar{q}_1,\, ...,\, \bar{q}_K\}$. This unobserved cost factor captures the difference in cost efficiencies across firms generated perhaps by the differences in managerial ability or other factors associated with the firm organization. We assume that this factor remains invariant across auctions in the data. To summarize, the mean of the costs distribution of bidder $i$ is given by

$$\mu_{C,i,l} = X_l \alpha + D_{i,l}\beta + q_i + U_l.$$

Finally, we distinguish between the bidders who regularly participate in the procurement market (regular bidders) and those who only appear in a very small number of auctions (fringe bidders). In our model, all fringe bidders are associated with the same fixed level of the unobserved cost factor $q_0$.

A potential bidder decides whether to enter the auction. Upon entry, he decides on the bid to submit. Specifically, during the auction for project $l$ each potential bidder observes his entry costs, $\kappa_{i,l}$, and is aware of the composition of the set of potential bidders, denoted by $I_{\mathcal{N},l}$. It is worth emphasizing that there is a qualitative difference between the composition $I_{\mathcal{N},l}$ and the set of potential bidders $\mathcal{N}_l$ here: $I_{\mathcal{N},l}$ contains information on the number of potential bidders in each $(d, q)$-group,where $d$ is a discretized measure of the distance between the contractor and the location of the project.

Given this information of $I_{\mathcal{N},l}$ and the cost distributions, a potential bidder $i$ decides whether to participate in the auction or not. We denote the entry decision (outcome) by $E_{i,l}$ ($E_{i,l} = 1$ if enters and $E_{i,l} = 0$ otherwise). Potential bidders who decided to participate in the auction form a set of active bidders, $\mathcal{A}_l$. An active bidder observes the composition of the set of active bidders, $I_{\mathcal{A},l}$, and his private cost $C_{i,l}$ for completing the project. He then submits a price $B_{i,l}$ based on this information set. We assume that the information in $I_{\mathcal{N},l}$ and $I_{\mathcal{A},l}$ is not available to the researcher.

In line with the existing empirical auction literature, we assume that the observed outcomes reflect a type-symmetric pure strategy Bayesian Nash equilibrium (psBNE). In such an equilibrium, participants who are *ex ante* identical in an auction $l$ (i.e. $i, j \in \mathcal{N}_l$ such that $q_i = q_j$, and $D_{i,l} = D_{j,l}$) adopt the same strategies. Thus, an equilibrium of an auction game for project $l$ is characterized by a set of equilibrium entry and bidding strategies: $\{\sigma_i^E(\,.\,|\,(d,\, q)),\, \sigma_i^B(\,.\,|\,(d,\, q)\}_{(d,\, q)}$.[20]

For a given type $(d,\, q)$, realized cost $c$ and composition of entrants $I_{\mathcal{A}}$, the bidding strategy of an entrant maximizes expected profit from bidding. That is,

$$\sigma^B(c,\, I_{\mathcal{A}}|\,(d,\, q)) = \arg\max \pi^{(d,\, q)}(b,\, c,\, I_{\mathcal{A}};\, \sigma^B_{-i}),$$

[20] Here we suppress the auction subscript $l$ in notation for expositional simplicity.

where

$$\pi^{(d,\,q)}(b,\, c,\, I_{\mathcal{A}};\, \sigma^B_{-i}) \equiv (b-c)P\left\{i\ wins\mid b,\, I_{\mathcal{A}},\ \text{“}i\ is\ type\text{-}(d,\, q)\text{”};\sigma^B_{-i}\right\},$$

and $\sigma^B_{-i}$ denotes a profile of other bidders' strategies that they would use should they become active in a given project.

The equilibrium entry strategies are characterized by a set of participation thresholds $\{\bar{K}^{(d,q)}(I_{\mathcal{N}})\}_{(d,q)}$ such that

$$\bar{K}^{(d,q)}(I_{\mathcal{N}}) = \mathbf{E}[\pi^{(d,q)} \mid I_{\mathcal{N}}] \text{ for all } (d,q)\text{-type.}$$

In the last expression, the expectation is taken over the distribution of $C_i$ and $I_{\mathcal{A}}$ conditional on a potential bidder's information set prior to the entry decisions $I_{\mathcal{N}}$ (and of course on $X$, which is suppressed in notation). This implies that the equilibrium probability of participation of the bidders of type $(d,\, q)$ is given by[21]

$$p^{(d,q)} = P\{E_{i,l} \leq \mathbf{E}[\pi^{(d,q)} \mid I_{\mathcal{N}}]\}. \tag{5.1}$$

5.2. **Estimation Details.** As in Section 2.2, the regular bidders are partitioned into groups with heterogeneous cost distributions. This implies that conditional on the same value of $d$, the distribution of bids submitted from those with higher $q_i$ first-order stochastically dominates that from bidders with lower $q_i$. We use this implication to recover the unobserved group structure in the first step of our estimation procedure. Specifically, in accordance with the notation used in the paper, we define $\hat{\delta}^+_{ij}(b;d) \equiv \hat{r}_{ij}(b;d) - \hat{r}_{ji}(b;d)$ with

$$\hat{r}_{ij}(b;d) \equiv \frac{\sum_{l=1}^{L} 1\{B_{i,l} \leq b\} 1\{i,\, j\, \in A_l\} K_h(D_{i,l}-d) K_h(D_{j,l}-d)}{\sum_{l=1}^{L} 1\{i,\, j\, \in A_l\} K_h(D_{i,l}-d) K_h(D_{j,l}-d)},$$

where $K_h(v) = K(v/h)/h$ for a univariate kernel function $K$, and $h$ is the bandwidth.

Notice that the index for agents $i$ and $j$ is based on the bids submitted in the auctions where both bidders participate. Thus, the bids of both agents are subject to the same variation in the auction characteristics. This permits aggregation over the auctions with different observable and unobservable characteristics. On the other hand, the distance to the project site may vary across bidders in the same auction. That is why we consider only bids from auctions for the projects from which bidders $i$ and $j$ have the same distance when constructing the index (the index is constructed so that everything else except bidder type is held equal). Since the grouping is invariant to the distance from the project, the index used in estimation could be aggregated over different values

[21] In the full notation that incorporates observed auction heterogeneity $X$, the entry decision would be characterized by thresholds $\{\bar{K}^{(d,q)}(I_{\mathcal{N}}, X)\}_{(d,q)}$ such that $\bar{K}^{(d,q)}(I_{\mathcal{N}}, X) = \mathbb{E}[\pi^{(d,q)} \mid I_{\mathcal{N}}, X]$ for all $(d,q)$-type. Hence the type-specific entry probabilities would be given by $p^{(d,q)}(X) = \Pr\{E_i \leq \mathbb{E}\left(\pi^{(d,q)} \mid I_{\mathcal{N}}, X\right) \mid X\}$.

of the distance $d$. As a robustness check we also compute groupings on the basis of subsets of distances. We find that the results of classification are very similar across these approaches. We implement classification following the bootstrap testing procedure described in (3.3.4).

In the second step, we estimate the parameter of the model by a GMM procedure while imposing the recovered group structure in estimation. Specifically, for a given guess at parameter values and for every configuration of potential bidders observed in our data, we solve the bidding game for every possible configuration of the set of active bidders associated with a given configuration of potential bidders. Among other things we use these solutions to compute a vector of expected profits for a bidder from a given group, for every possible configuration of his active competitors and for a grid of the unobserved auction heterogeneity values. This part of our routine uses an extension of the algorithm proposed in Marshall, Meurer, Richard, and Stromquist (1994). Since the participation stage of our game may generate multiple equilibria, we do not explicitly solve this part of the game. Instead, we maximize a moments-based objective function subject to the constraint that the optimality of the participation strategies is satisfied. We follow the spirit of Dube, Fox, and Su (2012) in treating entry probabilities corresponding to different bidder groups and to auctions with different characteristics as auxiliary parameters. These parameters are anchored in place by a set of constraints in the form of the equations characterized in (5.1).This procedure allows us to select participation strategies that are most consistent with the data.

We simulate the following moments as they are implied by the model at the current value of parameter values: (a) the probability to win the auction for every $q$-group; (b) the expected bid conditional on $(d, q)$; (c) the covariance between bids and the observable project characteristics; (d) the covariance between any two bids submitted in the same auction; (e) the expected number of participants for every $(d, q)$-group; (f) the covariance between the number of participants and observable project characteristics. We search for the set of parameters which minimizes the distance between the empirical and theoretical counterparts of these moments subject to constraints described above.

5.3. **Estimation Results.** We implement the analysis using the data for California Highway Procurement projects auctioned between 2002 and 2012. We use data from 1,054 medium-sized projects that involve paving and bridge work. Available information on project characteristics includes the engineer's estimate, the completion deadline (a measure of duration), the location of the project, category of work, and list of potential bidders. We construct the distance variable to reflect the expected driving time between the project location and the closest company plant. Table 7 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. 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 submit 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 20 miles on average for regular potential bidders.

Table 7. 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 |
| 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 2002 and 2012. It consists of 1,054 projects. The distance variable is computed to reflect the driving time between the project site and the nearest company plant.

In the first step of our estimation procedure we are able to group regular bidders into eight groups that consist of 2, 3, 8, 3, 2, 3, 2 and 2 bidders respectively.[22] The parameter estimates obtained in the second stage of our estimation procedure and their standard errors are summarized in Table 8. We normalize bids by the engineer's estimate in the estimation. Therefore all the parameters measure the effects relative to project size.

The results show that the distance significantly impacts bidders costs. Specifically, an additional increase of 10 miles in the distance raises the project costs by 1.6% of the engineer's estimate. Unobserved project heterogeneity while non-negligible is moderate

[22]We identify two clusters that overlap by two large bidders, and three separate non-overlapping clusters. We estimate that the underlying group structure of the overlapping clusters consists of three groups each. The bidders that are common to the two clusters are estimated to belong to the same group in both clusters. That is why we combine the groups from the two clusters that contain bidders of the same types as the two common bidders. The non-overlapping clusters were estimated to consist of one group each. This obtains the group structure with eight groups.

Table 8. Coefficient Estimates

| | Parameter | Std. Error |
|---|---|---|
| Costs of project work | | |
| Constant | 0.127*** | (0.011) |
| Eng. Estimate | -0.004*** | (0.002) |
| Duration | 0.0011* | (0.0006) |
| Bridge | -0.092*** | (0.018) |
| Distance | 0.026*** | (0.007) |
| Group 1 | -0.051*** | (0.008) |
| Group 2 | -0.058*** | (0.008) |
| Group 3 | -0.032*** | (0.009) |
| Group 4 | -0.012*** | (0.005) |
| Group 5 | -0.014*** | (0.007) |
| Group 6 | -0.008*** | (0.006) |
| Group 7 | -0.009*** | (0.008) |
| Group 8 | -0.010*** | (0.007) |
| $\sigma_C$ | 0.087*** | (0.032) |
| $\sigma_u$ | 0.021*** | (0.009) |
| Entry Costs | | |
| Constant | -0.114* | (0.078) |
| Engineer's Estimate | 0.055*** | (0.016) |
| Regular Bidder | -0.022*** | (0.019) |
| Number of Items | 0.018* | (0.011) |

Note: In the results above the distance is measured in units of ten miles. The fringe bidders are the omitted group; the impact of fringe status on costs is thus summarized by a constant.

in size: increasing the value of the unobserved factor from its mean (equal to '0') to a value that corresponds to one standard deviation from the mean is equivalent to increasing the distance to the project site by 10 miles. Importantly, we estimate statistically significant differences in bidders costs across the groups. Specifically, there appears to be three distinct levels of cost differences among regular bidders. Fringe bidders tend to have the highest costs whereas the difference in costs among groups of regular bidders is comparable in impact to the shortening of the distance to the project site by 20, 13 or 5 miles respectively.

Our results thus confirm that regular participants in the highway procurement market are characterized by important unobserved cost differences that persist throughout the data. Documenting such costs asymmetries is important from a purely informative point of view, since they could be indicative of collusive arrangements, quality differences, or differences in information. Additionally, cost asymmetries play an important role in policy-related decisions. For example, the revenue equivalence of simple auctions breaks down in the presence of costs asymmetries. The exact magnitudes of these asymmetries may influence the government choice of auction format. Affirmative action programs in government procurement have been largely implemented in the form of discriminatory auctions. Such auctions favor disadvantaged businesses, which are likely to have higher costs. They thus have a potential of increasing government procurement costs. It has been shown in the literature (McAfee and McMillan (1989)), however, that if the discriminatory factor correctly takes into account existing costs asymmetries then the discriminatory auction may actually lower government costs. Clearly, to optimally choose the structure of such an auction the exact information about the costs asymmetries is very important. For an extended discussion see Krasnokutskaya and Seim (2011).

## 6. Conclusion

This paper makes a number of contributions to the literature. First, for environments with agent-specific unobserved heterogeneity which takes values from a discrete finite set we show that the underlying group structure associated with the unobserved heterogeneity could be identified from pair-wise inequality restrictions implied by a theoretical model. Second, we demonstrate that such pairwise inequality restrictions exist in a number of settings characterized by strategic interdependence where identification of the primitives of the model with unobserved agent heterogeneity would otherwise be far from obvious. Third, we propose a computationally feasible method which produces consistent estimates of the ordered group structure associated with unobserved heterogeneity component. Finally, we apply this method to data from California highway procurement auctions to show that unobserved bidder heterogeneity plays an important role in this procurement market.

We believe that the classification method proposed in this paper may prove especially useful in settings where the analysis of unobserved agent heterogeneity is complicated by strategic interdependence. To the best of our knowledge this paper offers a novel insight into identification and estimation of such models. Specifically, classification could be used as a pre-estimation step in the structural studies of many environments where analysis would otherwise be infeasible due to the high computational cost.

Our methodology complements the finite mixture approach in the toolbox of an empirical researcher. It offers a straightforward and constructive identification mechanism, combined with computational feasibility at a cost, perhaps, of somewhat higher data requirements. The latter, however, becomes less of a problem as large datasets are made available to modern researchers. In contrast, the finite mixtures approach has lower data requirements but still remains very computationally costly in settings with strategic interdependence.

This paper makes a first step towards using classifications in the structural analysis of models with unobserved agent heterogeneity and strategic interdependence. The directions for future research include exploring the issue of uniform validity of the second step estimator when one uses the classification as a first step estimator, as well as relaxing the assumption that the number of distinct groups is small relative to the sample size. Both of these issues are nontrivial at the current state of the literature and may require an approach that is somewhat different from the one taken in this paper.

# 7. Appendix

## 7.1. Technical Details for Examples in Section 2.

7.1.1. *Equilibrium in Multi-Attribute Auctions.* Assume the outcome in the data are rationalized by type-symmetric pure-strategy Bayesian Nash equilibria (psBNE). In such an equilibrium, participants who belong to the same group $\mathbb{N}_k$ adopt the same strategy. A pure strategy for a seller $i$, denoted $\sigma_i$, maps from private costs to a price quote. It depends on the seller's type $k$, the set of potential bidders $\mathbb{N}$ as well as the exogenously given the type-specific entry probabilities $\rho \equiv (\rho_1, .., \rho_K)$. An entrant $i$'s expected profits from bidding $p$, when others follow strategies $\sigma_{-i} \equiv (\sigma_j : j \in \mathbb{N}\backslash\{i\})$, is given by:[23]

$$\Pi_i(p, c; \sigma_{-i}) \equiv (p - c)P\left\{\text{Procurer buys from } i \mid B_i = p; \sigma_{-i}\right\}. \tag{7.1}$$

Then a psBNE is a profile of strategies $\sigma \equiv (\sigma_i : i \in \mathbb{N})$ such that

$$\sigma_i(c) = \arg\max_p \Pi_i(p, c; \sigma_{-i})$$

for all $c$. Existence of psBNE in this model is discussed in the web supplement of Krasnokutskaya, Song, and Tang (2014).

7.1.2. *First-Price Auctions with Asymmetric Bidders.* Assume 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 the ordering of the distributions are strict $(F_1(v) > F_2(v) > ... > F_K(v))$ at least for $v$ within some non-degenerate interval on the common support.

Let $\lambda(\mathbb{A})$ denote the classification of bidders in the set of entrants $\mathbb{A}$, which is fully characterized by a $K$-vector of integers $(|\mathbb{A}_1|, ..., |\mathbb{A}_K|)$ with $\mathbb{A}_k$ being the subset of $\mathbb{A}$ consisting of bidders from group $k$ only. Let $G_k(\cdot; \lambda)$ be the CDF of $B_i$ when $i \in \mathbb{N}_k$. The marginal distribution of values is independent from $\lambda(\mathbb{A})$. Part (i) of Corollary 3 in Lebrun (1999) showed that, given any realized value of $\lambda(\mathbb{A})$, the supremum of the support of bids is the same for all bidder types. (That is, for any $\lambda$, $\beta_1(\overline{v}|\lambda) = \beta_2(\overline{v}|\lambda) = ... = \beta_K(\overline{v}|\lambda) \equiv \eta(\lambda) < \infty$ for some $\eta(\lambda) \in (\underline{v}, \overline{v})$.) Furthermore, the corollary also showed that for any $\lambda$,

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

[23] To calculate the conditional probability in (7.1), one needs to apply the Law of Total Probability by conditioning on the realizations of the quality classifications of the set of entrants. The latter in turn depends on entry probabilities $\rho$. The dependence of this conditional probability on $\rho$ and $\mathbb{N}$ is suppress to simplify notation here.

for all $b \in [\underline{v}, \eta(\lambda)]$ and $k < k'$ so that $F_{k'} \succeq_{F.S.D.} F_k$, and the inequality holds strictly at least over some interval on $[\underline{v}, \eta(\lambda)]$. Consider $i \in \mathbb{N}_{k'}$ and $j \in \mathbb{N}_k$ with $k' > k$. For any admissible $\lambda \in \mathbf{R}^K$, let $\tau_{i,j}(\lambda) \equiv P\{\lambda(\mathbb{A}) = \lambda | i, j \in \mathbb{A}\}$. It then follows that

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

with the inequality holding strictly over some interval on $[\underline{v}, \eta(\lambda)]$. This allows us to apply the method proposed in this paper to recover the group classifications based on pair-wise comparisons of the marginal distributions of $B_i$ and $B_j$ conditional on $i, j \in \mathbb{A}$. An obvious advantage of this approach is that we do not need to condition on the realization of the identities for the other entrants.

7.1.3. *English Auctions with Asymmetric Bidders.* We adopt the same notation as Section 2.3.1 in the main text. Let $W$ denote the transaction price; $V_s$ denote the private value for bidder $s$. Again, consider the case where $i \in \mathbb{N}_{k'}$ and $j \in \mathbb{N}_k$ where $k' > k$. Let $1_k$ denote the unit vector with the $k$-th component being 1. Then define:

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

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:

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

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

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

By the same argument,

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

It is then straightforward to show that for any $\lambda^*$, 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 $K$ supports of $\{F_l : 1 \leq l \leq K\}$, and the inequality holds strictly at least for some $w$ in an interval on the intersection of the $K$ supports of $\{F_l : 1 \leq l \leq K\}$. After integrating out $\lambda^*$ with respect to the marginal distribution of the set of entrants from $\mathbb{N}_{-i,j}$ (the set of potential bidders in the population excluding $i, j$), we get

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

and the inequality holds strictly for some $w$ over an interval in the intersection of supports. This implies

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

thus suggesting that the method proposed in this paper is applicable.

7.1.4. *Further Discussion about the English Auction Example.* One may wonder whether we can recover the classification of bidders in the English auction example through a "global" approach (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). We now explain why this is not a feasible alternative.

For any $i \in \mathbb{N}_{k'}$ and $j \in \mathbb{N}_k$ and $F_{k'} \succeq_{F.S.D.} F_k$, let $\mathbb{A}\backslash\{i,j\}$ denote the set of entrants out of $\mathbb{N}\backslash\{i,j\}$ and let $M(\mathbb{A}\backslash\{i,j\}) \equiv \max\{V_s : s \in \mathbb{A}\backslash\{i,j\}\}$. Let $\phi(w;\lambda^*)$ denote the CDF of $M(\mathbb{A}\backslash\{i,j\})$ conditional on $\lambda(\mathbb{A}\backslash\{i,j\}) = \lambda^*$. 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:

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

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

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

and

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

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

$$P\{W \le w, D = i | i \in \mathbb{A}, j \in \mathbb{A}\} > P\{W \le w, D = j | i \in \mathbb{A}, j \in \mathbb{A}\}. \tag{7.2}$$[24]

To see why the inequality in (7.2) holds, note for any $\lambda^*$,

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

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

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

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

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

yields the first inequality in (7.2).

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

$$P\{W \le w, D = i | i \in \mathbb{A}\} \tag{7.3}$$
$$= p_j P\{W \le w, D = i | i, j \in \mathbb{A}\} + (1 - p_j) P\{W \le w, D = i | i \in \mathbb{A}, j \notin \mathbb{A}\}$$

where $p_j \equiv P(j \in \mathbb{A})$. A similar expression exists for $P\{W \le w, D = j | j \in \mathbb{A}\}$ by swapping the roles of $i$ and $j$ in (7.3). The difference between the two positive differences

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

and

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

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

7.2. **Proofs of the Main Asymptotic Results.** Throughout the proofs, the asymptotic results are always as $L \to \infty$, unless specified otherwise. Also we write $wp \to 1$ as shorthand for "with probability approaching 1."

**Proof of Theorem 1:** For each $i \in \mathbb{N}$, we let $\mathbb{N}(i) = \mathbb{N} \backslash \{i\}$, and let $\hat{\mathbb{N}}_1(i)$ and $\hat{\mathbb{N}}_2(i)$ be as defined in Step 1 of the three step algorithm. Also, we define

$$\begin{aligned} \mathbb{N}_1^*(i) &\equiv \{j \in \mathbb{N}(i) : \tau(i) > \tau(j)\}, \text{ and} \\ \mathbb{N}_2^*(i) &\equiv \{j \in \mathbb{N}(i) : \tau(i) < \tau(j)\}. \end{aligned}$$

If $i \in \mathbb{N}_h$, we have $\mathbb{N}_l = \mathbb{N}_1^*(i)$ and $\mathbb{N}_h = \mathbb{N} \backslash \mathbb{N}_1^*(i)$. Also, if $i \in \mathbb{N}_l$, we have $\mathbb{N}_l = \mathbb{N} \backslash \mathbb{N}_2^*(i)$ and $\mathbb{N}_h = \mathbb{N}_2^*(i)$. By Assumption 1, for any $i \in \mathbb{N}_h$, if $j \in \mathbb{N}_1^*(i)$, $\log \hat{p}_{ij}^+ < \log \hat{p}_{ij}^- - \varepsilon r_L'$, $wp \to 1$, and if $j \in \mathbb{N} \backslash \mathbb{N}_1^*(i)$, $\log \hat{p}_{ij}^+ \ge \log \hat{p}_{ij}^- - \varepsilon r_L'$, $wp \to 1$. The last inequality follows because $\log \hat{p}_{ij}^+ = O_P(1)$ and $\log \hat{p}_{ij}^- = O_P(1)$, while $r_L' \to \infty$. This implies that whenever $i \in \mathbb{N}_h$, $\mathbb{N}_l \subset \hat{\mathbb{N}}_1(i)$ and $\mathbb{N}_h \subset \mathbb{N} \backslash \hat{\mathbb{N}}_1(i)$, $wp \to 1$. Since $(\hat{\mathbb{N}}_1(i), \mathbb{N} \backslash \hat{\mathbb{N}}_1(i))$ and $(\mathbb{N}_l, \mathbb{N}_h)$ are partitions of $\mathbb{N}$, this also implies that whenever $i \in \mathbb{N}_h$, $\mathbb{N}_l = \hat{\mathbb{N}}_1(i)$ and $\mathbb{N}_h = \mathbb{N} \backslash \hat{\mathbb{N}}_1(i)$, $wp \to 1$. Similarly, whenever $i \in \mathbb{N}_l$, $\mathbb{N}_l = \mathbb{N} \backslash \hat{\mathbb{N}}_2(i)$ and $\mathbb{N}_h = \hat{\mathbb{N}}_2(i)$, $wp \to 1$.

If we let for each $i \in \mathbb{N}$, $\hat{T}_1(i) = (\hat{\mathbb{N}}_1(i), \mathbb{N} \backslash \hat{\mathbb{N}}_1(i))$ and $\hat{T}_2(i) = (\mathbb{N} \backslash \hat{\mathbb{N}}_2(i), \hat{\mathbb{N}}_2(i))$, the two ordered partitions are consistent, i.e., for any $i \in \mathbb{N}$,

$$P\left\{\delta(\hat{T}_1(i), T) \ge 1\right\} \to 0 \text{ and } P\left\{\delta(\hat{T}_2(i), T) \ge 1\right\} \to 0. \tag{7.4}$$

Since $\mathbb{N}$ is a fixed finite set, we have for $j = 1, 2$,

$$P\left\{\delta(\hat{T}_j(i), T) \ge 1 \text{ for some } i \in \mathbb{N}\right\} \le \sum_{i \in \mathbb{N}} P\left\{\delta(\hat{T}_j(i), T) \ge 1\right\}.$$

Since $i^* \in \mathbb{N}$ and $\hat{T} = \hat{T}_1(i^*)$ or $\hat{T} = \hat{T}_2(i^*)$, the probability $P\{\delta(\hat{T}, T) \geq 1\}$ is bounded by the right hand side sum. But this sum converges to zero by (7.4), yielding the desired result. ■

**Proof of Theorem 2:** Take $\mathbb{N}^{[0]} = \mathbb{N}$. For each $i \in \mathbb{N}$, we obtain two ordered partitions $\hat{T}_1^{[1]}(i) \equiv (\mathbb{N}\backslash\hat{\mathbb{N}}_2(i), \hat{\mathbb{N}}_2(i)) \equiv (\hat{\mathbb{N}}_{1,1}^{[0]}(i), \hat{\mathbb{N}}_{2,1}^{[0]}(i))$ and $\hat{T}_2^{[1]}(i) \equiv (\hat{\mathbb{N}}_1(i), \mathbb{N}\backslash\hat{\mathbb{N}}_1(i)) \equiv (\hat{\mathbb{N}}_{1,2}^{[0]}(i), \hat{\mathbb{N}}_{2,2}^{[0]}(i))$. Define $E_{L,1}^{[1]}(i)$ to be the event that $\hat{\mathbb{N}}_{1,2}^{[0]}(i) = \mathbb{N}_{1,2}^{[0]}(i)$ and define $E_{L,2}^{[1]}(i)$ to be event that $\hat{\mathbb{N}}_{2,1}^{[0]}(i) = \mathbb{N}_{2,1}^{[0]}(i)$, where for each $r = 1, ..., K$, if $i \in \mathbb{N}_r$,

$$\begin{aligned} \mathbb{N}_{2,1}^{[0]}(i) &= \mathbb{N}_{r+1} \cup ... \cup \ \mathbb{N}_K \text{ and} \\ \mathbb{N}_{1,2}^{[0]}(i) &= \mathbb{N}_1 \cup ... \cup \ \mathbb{N}_r. \end{aligned}$$

Then by following the same arguments as in the proof of Theorem 1, we find that

$$P\left\{\bigcup_{i\in\mathbb{N}}\left(E_{L,1}^{[1]}(i) \cup E_{L,2}^{[1]}(i)\right)\right\} \to 1. \tag{7.5}$$

Now, generally, suppose that at Step $k-1 \leq K$, we have obtained the estimated ordered partition $(\hat{\mathbb{N}}_r^{[k]})_{r=1}^k$. Let $d_{k-1}^*, = (i_1^*, ..., i_{k-1}^*)$ denote the vector of pivotal agents chosen so far in obtaining the partition and let $\mathcal{D}_{k-1}$ be the set of subvectors of $(1, ..., N)$ with $k-1$ entries. Suppose that for $R_k \equiv \{r_1, ..., r_j\} \subset \{1, ..., k\}$ and $d_{k-1} \in \mathcal{D}_{k-1}$, we define the event $A_k(R_k, d_{k-1})$ to be such that $d_{k-1}^* = d_{k-1}$ and

$$\hat{\mathbb{N}}_r^{[k]} = \mathbb{N}_r \text{ for all } r \in R_k.$$

Now assume that we have at this step $k-1 \leq K$,

$$P\left\{\bigcup_{d_{k-1}\in\mathcal{D}_{k-1}}\bigcup_{R_k\subset\{1,...,k\}} A_k(R_k, d_{k-1})\right\} \to 1. \tag{7.6}$$

We will show that we can extend this convergence to the next step $k$.

We focus on a given event $A_k(R_k, d_{k-1})$. Define $\hat{\mathbb{N}}^{[k-1]} = \mathbb{N}\backslash(\cup_{r\in R_k}\hat{\mathbb{N}}_r)$ and $\mathbb{N}^{[k-1]} = \mathbb{N}\backslash(\cup_{r\in R_k}\mathbb{N}_r)$. Then in the event $A_k(R_k, d_k)$, we also have $\hat{\mathbb{N}}^{[k-1]} = \mathbb{N}^{[k-1]}$. If $\mathbb{N}^{[k-1]}$ is empty or contains $i$'s with the same type, the event $A_k(R_k, d_k)$ is equal to the event that $A_k(\{1, ..., k\})$ and $k = K$. In other words, $A_k(R_k, d_{k-1})$ remains the same for all choices of $R_k$ and $d_{k-1}$ that are consistent with this assumption in this case. Hence the classification is consistent by (7.6).

Suppose that $\mathbb{N}^{[k-1]}$ contains at least $i$ and $j$ with different types. When restricted to the sequence of events $A_k(R_k, d_k)$, there exists some $p$-value $\hat{p}_{ij}^0$ with $i, j$ in $\hat{\mathbb{N}}^{[k-1]}$ such that $\hat{p}_{ij}^0$ converges in probability to zero by Assumption 1, whereas for all $(i, j)$ such that

$i, j \in \mathbb{N}_r$ for some $r \in R_k$, $\hat{p}^0_{ij} = O_P(1)$. Therefore, the probability that the next split in Step $k$ under event $A_k(R_k, d_k)$ is made on a group other than $\hat{\mathbb{N}}^{[k-1]}$ converges to zero.

We obtain two ordered partitions

$$\begin{aligned}\hat{S}_1^{[k]}(i) &\equiv (\mathbb{N}^{[k-1]} \backslash \hat{\mathbb{N}}_2(i), \hat{\mathbb{N}}_2(i)) \equiv (\hat{\mathbb{N}}_{1,1}^{[k-1]}(i), \hat{\mathbb{N}}_{2,1}^{[k-1]}(i)) \text{ and} \\ \hat{S}_2^{[k]}(i) &\equiv (\hat{\mathbb{N}}_1(i), \mathbb{N}^{[k-1]} \backslash \hat{\mathbb{N}}_1(i)) \equiv (\hat{\mathbb{N}}_{1,2}^{[k-1]}(i), \hat{\mathbb{N}}_{2,2}^{[k-1]}(i))\end{aligned}$$

of the set $\mathbb{N}^{[k-1]}$, where $\hat{\mathbb{N}}_1(i)$ and $\hat{\mathbb{N}}_2(i)$ are defined as in Step 1 of the split algorithm in Section 3.3.1 except that we replace $\mathbb{N}$ there by $\mathbb{N}^{[k-1]}$. With $\hat{S}_1^{[k]}(i)$ and $\hat{S}_2^{[k]}(i)$ given, we construct two ordered partitions $\hat{T}_1^{[k]}(i)$ and $\hat{T}_2^{[k]}(i)$ by replacing $\hat{\mathbb{N}}^{[k-1]}$ with $(\hat{\mathbb{N}}_{1,1}^{[k-1]}(i), \hat{\mathbb{N}}_{2,1}^{[k-1]}(i))$ and $(\hat{\mathbb{N}}_{1,2}^{[k-1]}(i), \hat{\mathbb{N}}_{2,2}^{[k-1]}(i))$ respectively. Define $E_{L,1}^{[k]}(i)$ to be the event that $\hat{\mathbb{N}}_{1,2}^{[k-1]}(i) = \mathbb{N}_{1,2}^{[k-1]}(i)$ and define $E_{L,2}^{[k]}(i)$ to be event that $\hat{\mathbb{N}}_{2,1}^{[k-1]}(i) = \mathbb{N}_{2,1}^{[k-1]}(i)$, where for each $r = 1, ..., K$, if $i \in \mathbb{N}_r$,

$$\mathbb{N}_{2,1}^{[k-1]}(i) = \bigcup_{s>r:s\in R_k} \mathbb{N}_s \text{ and } \mathbb{N}_{1,2}^{[k-1]}(i) = \bigcup_{s<r:s\in R_k} \mathbb{N}_s.$$

Then again by following the same arguments as in the proof of Theorem 1, we find that

$$P\left\{\bigcup_{i\in\mathbb{N}} \left(E_{L,1}^{[k]}(i) \cup E_{L,2}^{[k]}(i)\right)\right\} \to 1. \tag{7.7}$$

Let $i_{k,1}$ be the pivotal agent used to split $\hat{\mathbb{N}}^{[k-1]}$ into $(\hat{\mathbb{N}}_{1,1}^{[k-1]}(i_{k,1}), \hat{\mathbb{N}}_{2,1}^{[k-1]}(i_{k,1}))$, and let $i_{k,2}$ be the pivotal agent used to split $\hat{\mathbb{N}}^{[k-1]}$ into $(\hat{\mathbb{N}}_{1,2}^{[k-1]}(i_{k,2}), \hat{\mathbb{N}}_{2,2}^{[k-1]}(i_{k,2}))$, in Step $k$. In the former case, we set $\hat{\mathbb{N}}_{r_{j+1}}^{[k+1]} = \hat{\mathbb{N}}_{2,1}^{[k-1]}(i_{k,1})$ and $d_k = (d_{k-1}, i_{k,1})$. In the latter case, we set $\hat{\mathbb{N}}_{r_{j+1}}^{[k+1]} = \hat{\mathbb{N}}_{1,2}^{[k-1]}(i_{k,2})$ and $d_k = (d_{k-1}, i_{k,2})$. Then we define

$$R_{k+1} = R_k \cup \{r_{j+1}\}$$

and rename $\hat{\mathbb{N}}_r^{[k+1]} = \hat{\mathbb{N}}_r^{[k]}$ for all $r \in R_k$. Thus we have obtained the augmented partition $(\hat{\mathbb{N}}_r^{[k+1]})_{r=1}^{k+1}$. Now we can define $A_{k+1}(R_{k+1}, d_k)$ similarly as we defined $A_k(R_k, d_{k-1})$. Then it clear that the convergence in (7.7), combined with (7.6), implies that

$$P\left\{\bigcup_{d_k\in\mathcal{D}_k} \bigcup_{R_k\subset\{1,...,k+1\}} A_{k+1}(R_{k+1}, d_k)\right\} \to 1.$$

We keep iterating the process until we have $k = K$ at which point the resulting estimated ordered partition is consistent as shown before. ■

**Lemma A1:** *Suppose that Assumptions 1-2 hold and the true type structure is identified and the true number of the groups is $K_0$.*
(i) *If $K \geq K_0$, $\hat{V}(K) = O_P(1)$, as $L \to \infty$.*

(ii) *If* $K < K_0$, *for any* $M > 0$, *as* $L \to \infty$,

$$P\{\hat{V}(K) > g(L)M\} \to 1.$$

**Proof:** (i) From the proof of Theorem 2, we find that for all $i, j \in \mathbb{N}$ such that for some $k$,

$$P\left\{i, j \in \hat{\mathbb{N}}_k\right\} \to 1,$$

we have $i, j$ in the same group $k$. Therefore, by Assumption 1(a), we have

$$\hat{V}(K) = \frac{1}{K}\sum_{k=1}^{K}\left|\min_{i,j\in\hat{\mathbb{N}}_k} \log \hat{p}_{ij}^0\right| = O_P(1).$$

Thus (i) follows.

(ii) Suppose that $K < K_0$. Then for some $k = 1, ..., K$, and for some $i, j \in \mathbb{N}_k$, we have $H_{1,ij}^0$ true. Since $g(L)/r_L \to 0$ as $L \to \infty$, we take $r_L' \equiv g(L)$ in Assumption 1(b) to find that for this pair $(i, j)$,

$$\frac{\log \hat{p}_{ij}^0}{g(L)} \to_P -\infty,$$

as $L \to \infty$. Therefore, it follows that

$$\hat{V}(K)/g(L) \to_P \infty,$$

as $L \to \infty$. ■

**Proof of Theorem 3:** Since $K_0 \leq \bar{K}$, by Lemma A1(i), we have

$$\hat{V}(\bar{K}) = O_P(1).$$

Choose $K$ such that $K_0 < K \leq \bar{K}$ and write

$$\hat{Q}(K_0) - \hat{Q}(K) = \hat{V}(K_0) - \hat{V}(K) + (K_0 - K)\frac{\hat{V}(\bar{K})}{\bar{K}}g(L).$$

As for the leading term on the left hand side, we have

$$\hat{V}(K_0) - \hat{V}(K) = O_P(1),$$

by Lemma A1(i). Since $K_0 < \bar{K} \leq N$, there exists at least one pair $(i, j)$ in the same group. As for the last term, we choose any $c > 0$, and a pair $(i, j)$ in the same group, and write

$$\begin{aligned}(K_0 - K)\frac{\hat{V}(\bar{K})}{\bar{K}}g(L) &\leq (K_0 - K)\frac{\hat{V}(\bar{K})}{\bar{K}}g(L)1\left\{\log \hat{p}_{ij}^0 \leq -c\right\} \\ &\leq (K_0 - K)\frac{cg(L)}{\bar{K}^2}1\left\{\log \hat{p}_{ij}^0 \leq -c\right\}.\end{aligned}$$

Therefore, for any large $M > 0$,

$$P\left\{(K_0 - K)\frac{\hat{V}(\bar{K})}{\bar{K}}g(L) < -M\right\} \geq P\left\{(K_0 - K)\frac{cg(L)}{\bar{K}^2} < -M\right\} - P\left\{\log \hat{p}_{ij}^0 > -c\right\}.$$

The leading probability on the right hand side converges to 1 because $g(L) \to \infty$ and $K > K_0$. Since $i$ and $j$ are in the same group, the last probability converges to zero as we send $c$ to 0 by Assumption 1(b). Therefore, we find that whenever $K > K_0$, we have

$$P\left\{\hat{Q}(K_0) < \hat{Q}(K)\right\} \to 1.$$

And for all $K < K_0$, we have by Lemma A1(ii), for any $M > 0$,

$$P\left\{\hat{V}(K) > g(L)M\right\} \to 1,$$

whereas

$$\hat{V}(K_0) = O_P(1) \text{ and } \hat{V}(\bar{K}) = O_P(1).$$

Therefore, we find again that

$$P\left\{\hat{Q}(K_0) < \hat{Q}(K)\right\} \to 1.$$

We conclude that $P\{\hat{K} = K_0\} \to 1$.

Let us turn to the second statement. Since $P\{\hat{K} = K_0\} \to 1$, we have

$$P\left\{\delta\left(\hat{T}_{\hat{K}}, T\right) \geq 1\right\} = P\left\{\delta\left(\hat{T}_{K_0}, T\right) \geq 1\right\} + o(1).$$

That the last probability converges to zero follows precisely by the arguments in the proof of Theorem 2. ■

## References

Andrews, D. W., and P. Guggenberger (2009): "Incorrect Asymptotic Size of Subsampling Procedures Based on Post-Consistent Model Selection Estimator," *Journal of Econometrics*, 152(6), 19–27.

Andrews, D. W., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Athey, S., and P. A. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107–2140.

Athey, S., J. Levin, and E. Seira (2011): "Comparing Open and Sealed Bid Auctions: Theory and Evidence from Timber Auctions," *Quarterly Journal of Economics*, 126(1), 207–257.

Bonhomme, S., and E. Manresa (2014): "Group Patterns of Heterogeneity in Panel Data," Working paper, Chicago University.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753.

Campo, S., I. Perrigne, and Q. Vuong (2003): "Asymmetry in First-price Auctions With Affiliated Pprivate Values," *Journal of Applied Econometrics*, 18(2), p179 – 207.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Dube, J., J. Fox, and C. Su (2012): "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficient Demand Estimation," *Econometrica*, 80, 2231–2267.

Guerre, E., P. Isabelle, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.

Henry, M., Y. Kitamura, and B. Salanie (2014): "Partial Identification of Finite Mixtures in Econometric Models," *Quantitative Economics*, 5, 123–144.

Hong, H., and M. Shum (2002): "Increasing Competition and the Winner's Curse: Evidence from Procurement," *Review of Economic Studies*, 69(4), 871–898.

Hu, Y., D. McAdams, and M. Shum (2013): "Identification of First-Price Auctions with Non-Separable Unobserved Heterogeneity.," *Journal of Econometrics*, 174, 186–193.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77, 135–175.

Krasnokutskaya, E., and K. Seim (2011): "Bid Preference Programs and Participation in Highway Procurement," *American Economic Review*, 101.

Krasnokutskaya, E., K. Song, and X. Tang (2014): "Quality in On-line Service Markets," Working paper.

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

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing Functional Inequalities," *Journal of Econometrics*, 172, 14–32.

——— (2014): "Testing for a General Class of Functional Inequalities," *arXiv:1311.1595v3[math.ST]*.

Leeb, H., and B. M. Pötcher (2005): "Model Section and Inference: Facts and Fiction," *Econometric Theory*, 11, 537–549.

Li, T., and X. Zheng (2009): "Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions," *Review of Economic Studies*, 76, 1397–1429.

Lin, C., and S. Ng (2012): "Estimation of Panel Data Models with Parameter Heterogeneity When Group Membership is Unknown," *Journal of Econometric Method*, 1, 42–55.

Marshall, R. C., M. J. Meurer, J.-F. Richard, and W. Stromquist (1994): "Numerical Analysis of Asymmetric First Price Auctions," *Games and Economic Behavior*, 7(2), 193–220.

McAfee, R. P., and J. McMillan (1989): "Government procurement and international trade," *Journal of International Economics*, 26(3/4), 291–308.

Milgrom, P. R., and R. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1089–1122.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operation Research*, 6, 58–73.

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

Phillips, P., and Y. Sul (2007): "Transition Modelling and Econometric Convergence Tests," *Econometrica*, 75, 1771–1855.

Pötcher, B. M. (1991): "Effects of Model Selection on Inference," *Econometric Theory*, 7, 163–185.

Song, K. (2005): "Semiparametic Specification Testing in Econometrics and Heterogeneous Panel Modeling," Ph.d. thesis, Yale University.

Sun, Y. (2005): "Estimation and Inference in Panel Structure Models," Unpublished manuscript.

Department of Economics, Johns Hopkins University
*E-mail address*: ekrasnok@jhu.edu

Vancouver School of Economics, University of British Columbia
*E-mail address*: kysong@mail.ubc.ca

Department of Economics, Rice University and University of Pennsylvania
*E-mail address*: xun.tang@rice.edu