ESTIMATING UNOBSERVED AGENT HETEROGENEITY USING 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 well in Monte Carlo simulations. We demonstrate empirical usefulness of this method by estimating a model of a 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, Nonparametric Classification
JEL CLASSIFICATION: C12, C21, C31

1. Introduction

The 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 recorded in the data. Failing to account for unobserved heterogeneity may lead to biased estimates and may affect the quality of counterfactual predictions.

We focus on the environments where economic agents' unobserved heterogeneity is captured by a discrete (non-stochastic) parameter taking values from a finite set. In such cases, the underlying population is organized as a collection of groups consisting of same-type agents. Thus unobserved heterogeneity defines a latent group structure on the set of agents.

This paper proposes a method to recover the unobserved group structure from data. Our approach relies on pairwise comparisons derived from the model which are related to agents' unobserved types. We establish the necessary and sufficient conditions which ensure that a given set of pairwise comparisons identifies the whole unobserved group structure. However, the estimation of the group structure requires additional insight even when identification is established. One possible 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

---

Date: January 10, 2017.
We are grateful to the Associate Editor and three referees for valuable comments. All errors are ours.

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.[1] The method we propose resolves these issues.

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 a given number of groups is reached. We recover the whole group structure for each given number of the groups, and select the number of the groups (and the associated group structure) by using an appropriate goodness-of-fit measure and a penalization scheme for over-fitting. We show that under mild regularity conditions for p-values, this method allows us to consistently select the correct number of groups and delivers a consistent estimator of the group structure.

As this paper shows, pairwise restrictions arise from many economic models in empirical research. The examples we consider include unobserved product/provider attributes, unobserved heterogeneity in firms' costs, and assortative matching in labor markets. In these examples, the unobserved agent's type as well as the profile of the competitor types (in multi-agent settings) serve as important determinants of the agent's choices and of agent-specific outcomes and thus may be a source of omitted-variable bias if they are ignored in estimation.

In many settings, recovering the unobserved group structure can be of independent interest. For example, it can be used to identify the colluding groups of agents, the cost asymmetries or product quality differences. Moreover, the method proposed in this paper is useful as a pre-estimation step in the structural analysis of complex economic settings such as dynamic industry models or auction models with asymmetric bidders.[2] Specifically, uncovering the unobserved group structure may aid in identification of other primitives. In fact, in some settings it is necessary to know agent-specific unobserved heterogeneity in order to recover other primitives of the model.[3] Further, classification step may help to reduce computational time in estimation of games where no explicit solution for equilibrium exists. Indeed, for structural estimation, one needs to relate the primitives of the model (including agents' unobserved heterogeneity) to the observed outcomes. The brute-force approach of modeling unobserved heterogeneity as agent-specific fixed effects and estimating them jointly with the structural parameters may be computationally infeasible in such settings since equilibrium has

[1]Even if the types are known to take values from a two-point set, the total number of the candidate group structures for n agents is 2^n which is large even with a moderate size n.

[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 that of post-model selection inference which 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]Such situation arises, for example, in in English auctions with asymmetric bidders when only transaction price and the identity of the winner are observed. Athey and Haile (2007) show that in such settings the underlying distributions of private values can only be identified if the groups to which winning players belong are known.

to be numerically computed for each configuration of agents' types in each iteration within the estimation routine. Recovering agents' unobserved groups in the first-step reduces the number of different agent types present in the population relative to the brute approach which may in turn reduce computational burden.[4]

Additionally, our method is advantageous in settings where the number of agents in the population is moderately large but each market (observation) in the data involves only a small number of participants.[5] In this case, despite the large number of markets observed, the researcher may have only a small number of markets which contain the same set of participants. We call this issue *the problem of the sparsely common players*. In such settings, the researcher cannot build inference on the conditional moments *given the set of participants* in a market, as is typically done in the structural empirical literature. Hence in the interest of the accuracy of inference, he needs to "aggregate" the markets or the agents in order to be able to work with consistently estimable objects. Pairwise restrictions can be testable with accuracy even when the data exhibit sparsely common sets of players, since the number of markets where a given pair of agents is present tends to be large even if the number of markets with the same set of participants may be small. Thus, pairwise restrictions and the classification procedure offer a natural way to aggregate agents into groups which permits estimation of other primitives.

In summary, our proposed method of recovering the unobserved group structure has the following advantages. First, it is nonparametric. It does not require parameterization of other model primitives or prior knowledge of the number of groups. Second, the estimator for the group structure is consistent under mild regularity conditions. Third, the method is flexible, and applicable to a wide range of complex structural models because the method is built to depend only on the p-values from the pairwise inequality restrictions. Fourth, the method provides a computationally feasible way to deal with unobserved heterogeneity in many applications.

We investigate the finite sample performance of the classification method in a Monte Carlo study. The study is based on data generated by the model of a first-price procurement auction with asymmetric bidders where the means of the distributions of private costs differ across bidders in an unobserved way. We report the outcomes for various numbers of bidders and group structures. The estimation overall works quite well. The performance is better when the number of bidders and groups are smaller relative to the number of the markets and the differences between groups are larger. We also investigate the finite sample impact of the first step classification error on structural parameter estimates. We find that in our setting the second-step estimator's quality is not substantially affected by the first-step classification error.

[4]See Krasnokutskaya, Song, and Tang (2016) for an application of this approach in the analysis of online service markets.

[5]For example, the total number of agents in population may be several hundreds but each market may attract only a few participants.

We analyze the California highway procurement market using our classification method. Empirical studies of auction markets often emphasized asymmetries in private values associated with observable bidder characteristics.[6] In this paper, we additionally allow for unobservable differences in the means of cost distributions. To account for 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 use the classification method to recover the underlying group structure associated with unobserved cost asymmetries. Next, we recover the group-specific distributions of costs using Generalized Method of Moments estimation. Our estimates indicate that several (unobserved) groups with important cost differences are present. We further determine that ignoring unobserved bidder heterogeneity leads to biased estimates of the impact of several factors on bidders' costs.

This paper is organized as follows. Section 2 discusses related literature. Section 3 introduces the basic environment where pairwise inequality restrictions are defined. This section also discusses examples that motivate our classification method in various contexts. Section 4 discusses identification of the underlying group structure from pairwise restrictions. Section 5 proposes a consistent estimator of the unobserved group structure. Section 6 presents and discusses Monte Carlo results. Section 7 presents the empirical application. Section 8 concludes. Technical proofs and derivations are provided in the Appendix.

2. Related Literature

Researchers often turn to a finite mixture approach when dealing with unobserved agent heterogeneity. This method represents individual unobserved heterogeneity as a latent discrete random variable, whose distribution needs to be recovered from the data. In contrast, our classification method treats individual's unobserved heterogeneity as a non-stochastic parameter of the environment, and aims to recover *the exact type for each individual in the population*.

The data requirements for the two methods are different. Specifically, to implement our classification method the researcher needs to observe each agent participating in moderately many observations (markets) so that pairwise comparison between individuals is feasible based on some consistently estimable quantities. In comparison, the finite mixture approach does not require each individual to show up in many observations. This is because the objective of estimation is to recover the distribution of unobserved heterogeneity in the population (conditional on covariates) rather than each agent's type.

Our classification method recovers the group structure from the pairwise comparisons, which naturally arise in the context of many economic models, and does not rely on parametrization of model primitives. Further, we allow for unobserved individual heterogeneity to be arbitrarily correlated with any other observed or unobserved stochastic element of the environment.

[6] Athey, Levin, and Seira (2011), Roberts and Sweeting (2013), Aradillas-Lopez, Gandhi, and Quint (2013) account for the bidder heterogeneity associated with size in timber market ('mills' vs 'loggers'); Krasnokutskaya and Seim (2011), Jofre-Bonet and Pesendorfer (2003), and Gentry, Komarova, and Shiraldi (2016) incorporate bidder participation differences in highway procurement market ('regular' vs 'fringe' bidders); Conley and Decarolis (2016), Asker (2010) and Pesendorfer (2000) account for possible bidder heterogeneity arising due to collusion.

This feature is particularly useful in the settings with strategic interdependence. The finite mixture approach usually requires some independence between individual unobserved heterogeneity and (at least) a sub-vector of observed covariates.

Both finite mixture and our approach could be applied in the estimation of the full structural model. The identification of model primitives under a finite mixture approach may be nontrivial and requires insights into exclusion restrictions.[7] In contrast, the classification step usually simplifies the identification of other model primitives. Additionally, in some empirical settings, the classification of agents is necessary for the subsequent estimation of other structural parameters.

Our method delivers classification at a low computational cost, due to the sequential split algorithm we developed. In particular, the estimation of the number of groups does not present a computational burden. In a finite mixture approach the assumption about the number of groups has to be imposed at the beginning of analysis while considering alternative values of this parameter can be computationally prohibitive.[8]

In the sense that we model the unobserved heterogeneity as a non-stochastic discrete parameter, our approach is similar in spirit to the literature on panel data models where some of the parameters differ across unknown groups of cross-sectional units.[9] Our approach differs from this literature in terms of the assumed empirical setting. The panel data literature assumes that the dependent variable of each cross-sectional unit is determined explicitly by the observed and unobserved variables within the same cross-sectional unit. In contrast, we focus on settings where the dependent variables for each cross-sectional unit may depend *in an implicit way* on observed and unobserved variables associated with all other cross-sectional units due to equilibrium constraints. Furthermore, the details of such dependence are model-specific. To accommodate this structure, we develop a method which uses only p-values from model-specific pairwise tests.

---

[7]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)).

[8]This is true for most finite mixture methods with the exception of the approach developed in Hu and Schennach (2008), Hu and Shum (2012) and Hu and Shiu (2013) which recovers the number of groups in estimation, using some full-rank (completeness) condition on the component distribution in the finite mixture. However, this method have been developed in the contexts where a single unobserved factor per observation is present such as single agent settings or a game-specific rather than agent-specific unobserved heterogeneity is present. Additionally, the completeness condition may be difficult to interpret in many settings.

[9]See Sun (2005) and Phillips and Sul (2007) who study an unobserved group structure of cross-sectional units in the context of growth models based on large panel data; Song (2005) who explores consistency of the heterogeneous parameter estimators and proposes a consistent estimation method of unobserved group structure in large panel models; Lin and Ng (2012) also studies estimation of panel models with unknown group structure. More recently, Bonhomme and Manresa (2014) make use of a k-means clustering algorithm in the first step to recover the group structure before proceeding to estimate deep parameters characterizing data-generating process. Su, Shi, and Phillips (2016) develop a new Lasso method to recover the unknown group-specific parameters.

Our classification method is somewhat related to the computer science literature on clustering.[10] However, this literature pursues objectives which are fundamentally different from those considered in this paper. More specifically, the clustering methods aim to group objects/agents that are similar in their observable attributes. In comparison, our objective is to classify individuals on the basis of an unobserved attribute exploiting the relationship between the outcome (endogenous) variables and agents' type implied by the economic model.

Our approach is also designed to be applied to data structures which are quite different from the data typically used in clustering analysis. Specifically, clustering analysis is performed on data sets which include many cross-sectional units and where each unit is observed once or a small number of times. The performance of the procedure is evaluated by the fraction of misclassified units. In contrast, our method applies in settings where many observations per cross-sectional unit are available. We aim to classify each individual unit consistently. This is because we would like to allow for the possibility that classification is not an end objective of the analysis. Instead, it may be used as a step enabling estimation of other structural primitives. In this context, it is important that the classification error should have an asymptotically negligible impact on the second-step estimation of other structural primitives. Our (pointwise) consistency result formulates conditions under which our procedure delivers this property.

3. The Model and Motivating Examples

3.1. The Basic Set-Up

The population consists of a set of players denoted by N such that $|N| = n$. The agents are engaged in a payoff-generating activity. The researcher has access to the data which consist of L observations summarizing individuals' decision-making. The decision-making environment may involve a single agent or multiple agents (when it reflects market competition, for example). For the sake of convenience we will refer to a single observation as a game even if only a single player is involved. Further, we will use S_l to denote the set of players involved in l-th game.

Importantly, each agent i is characterized by a non-stochastic factor q_i, which is not observed to the researcher, such that $q_i \in Q_0 = \{\bar{q}_1, ..., \bar{q}_{K_0}\}$, with $\bar{q}_1 < \cdots < \bar{q}_{K_0}$. This factor induces a partition of the set of agents into a *group structure* which is an ordered collection of disjoint subsets $(N_1, N_2, ..., N_{K_0})$ such that $N = \cup_k N_k$. The membership function $\tau : N \to \{1, ..., K_0\}$ links the identity of a player to his type so that $q_i = \bar{q}_{\tau(i)}$ and for each $k = 1, ..., K_0$,

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

The data available to the researcher contain for every observation $l = 1, ..., L$: a vector of observable characteristics for all the players involved, $\{X_{j,l}\}_{j \in S_l}$, as well as at least one but possibly multiple vectors of outcome variables, $Y_l = \{Y_{j,l}\}_{j \in S_l}$. The outcome variables may

[10]The same applies to the literature on statistical classification which is primarily concerned with adaptive learning and prediction. See Hastie, Tibshirani, and Friedman (2009) for a comprehensive review of both literatures.

reflect the actions chosen by players in the course of the game or the various aspects of the payoff realized for individual players such as, for example, whether the player won the game or how many utils or dollars he collected at the end of the game. The relationship between the outcome variables in observation l and players' characteristics can be summarized by a reduced-form expression:

$$Y_{i,l} = \varphi_l(X_{i,l}, q_i, \Omega_l; \theta_0) \tag{3.1}$$

where Ω_l summarizes a decision environment associated with observation l which includes characteristics of other participating players, $\{X_{j,l}, q_j\}$, and other environment-specific variables some of which may be unobserved. Outcome variables also depend on the structural parameters of the model, θ_0. These parameters may include parameters of the utility function, parameters of the cost function, etc. These parameters may be the ultimate objects of interest. Unlike the literature on panel data models, the outcome $Y_{i,l}$ for cross-sectional unit i is allowed to depend on other cross-sectional units j through a potentially nonlinear function φ_l. Often the function φ_l takes a complex form whose existence is ensured but its explicit form is not available to the researcher.

Our method applies in the settings where for each given pair of players (i, j) a pairwise monotone relationship between some functions of outcome variables and these players' unobserved factors can be established which allows us to construct three *comparison indexes* δ_{ij}^+, δ_{ij}^0 and δ_{ij}^- that satisfy the following properties:

$$\begin{aligned} \delta_{ij}^+ &> 0 \text{ if and only if } \tau(i) > \tau(j); \\ \delta_{ij}^0 &= 0 \text{ if and only if } \tau(i) = \tau(j); \\ \delta_{ij}^- &> 0 \text{ if and only if } \tau(i) < \tau(j). \end{aligned} \tag{3.2}$$

These pairwise comparison indexes δ_{ij}^+, δ_{ij}^0 and δ_{ij}^- need to be consistently estimable using data generated in the settings described above. We illustrate these concepts below using several settings broadly studied in the literature.

3.2. Examples

In this section we provide examples of the circumstances where the pairwise restrictions exist. We also explain the reasons for using a classification method.

3.2.1. **Unobserved Attributes**

In many choice settings an economic agent chooses among multiple alternatives summarized by vectors of attributes where one of the attributes may not be observed by the econometrician. Such a model has been extensively used to analyze markets with differentiated products. One approach to modeling unobserved product heterogeneity in such a setting is to associate a fixed effect with each product. The methodology which pursues this approach relies on the property that a one-to-one correspondence exists between the unobserved attributes and the probabilities of choosing these alternatives conditional on the full choice set of the buyer (see, for example, Berry, Levinsohn, and Pakes (1995)). This property allows researchers to recover unobserved attributes, as well as control for their impact in estimation. In the settings where the choice sets vary a lot across buyers, however, such an approach may be impractical since the choice probabilities conditional on the choice set may not be consistently estimable from the data. In such settings the aggregation of products (or sellers) into groups and/or aggregation over the choice sets may be required. The method proposed in this paper can be applied to infer the group structure which may serve as a basis for aggregation in such a context. It may be useful in the analysis of markets with local competition such as service markets. Specifically, Krasnokutskaya, Song, and Tang (2016) adopt this approach to study an online market for programming services where transactions are implemented in the form of multi-attribute auctions (i.e., buyers base their allocation decisions on sellers' attributes in addition to prices, similar to the differentiated product markets).

To fix ideas, consider a simplified version of the model in Krasnokutskaya, Song, and Tang (2016) that abstracts away from observed auction and seller's heterogeneity. Denote by N the population of the sellers in this industry and by S_l the set of sellers who submitted bids for the project l. As in our basic set up, each seller is characterized by unobservable quality from a finite ordered support: $q_i \in \{\bar{q}_1, \bar{q}_2, ..., \bar{q}_{K_0}\}$, with $\bar{q}_k < \bar{q}_{k'}$ whenever $k < k'$. Sellers' qualities are known to buyers but are not reported in the data.

The buyer for project l selects a seller among those who submitted bids in his auction or chooses an outside option so as to maximize his payoff. The payoff to the buyer from engaging services of seller $i \in S_l$ is given by $U_{i,l} = \alpha_l q_i + \epsilon_{i,l} - B_{i,l}$ whereas the payoff from an outside option is $U_{0,l}$. Here α_l is a non-negative weight the buyer gives to the seller's quality relative to the seller's bid, whereas $\epsilon_{i,l}$ reflects a buyer-seller match-specific stochastic component.

Let us suppress the auction subscript l and define for any two sellers i, j,

$$r_{ij}(b) = \Pr\left(i \, \text{wins} \,|\, B_{i,l} = b, \, i \in S, \, j \notin S \right)$$

for all b on the intersection of the supports of $B_{i,l}$ and $B_{j,l}$.

If $\alpha_l, \, S_l, \{C_{i,l}, \epsilon_{i,l}\}_{i \in S_l}$ are mutually independent[11] then

$$sign(r_{ij}(b) - r_{ji}(b)) = sign(q_i - q_j), \tag{3.3}$$

for any b on the intersection of bid supports. Intuitively, suppose 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 a higher unobserved quality has a higher chance of winning. Note that the identity of the winner is not deterministic due to uncertainty about the buyer's tastes in α_l and $\epsilon_l = \{\epsilon_{i,l}\}_{i \in S_l}$. The ranking of winning probabilities above 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 excludes both i and j. This reasoning is formally developed in Proposition 1 of Krasnokutskaya, Song, and Tang (2016).

[11]This holds, for example, when the participating sellers are not informed of the weights or the outside option of the buyer and are not informed about identities of other sellers they are competing with (i.e., about sellers in S_l).

On the basis of this property the comparison indexes can be constructed as follows: $\delta_{ij}^{+} \equiv \int \max\{r_{ij}(b) - r_{ji}(b), 0\} db$, $\delta_{ij}^{0} \equiv \int |r_{ij}(b) - r_{ji}(b)| \, db$ and δ_{ij}^{-} is defined by swapping i and j in δ_{ij}^{+}. We integrate over observed values of bids, b, to aggregate all the information available for the comparison of i and j in order to maximize the power of this comparison. Furthermore, note that the comparison indexes allow the researcher to use observations from the auctions with different profiles of competitors' qualities. They also do not depend on specific parametric assumptions for the distribution of buyers' tastes which allow the researcher to recover the unobserved group structure outside of the formal structural estimation.

3.2.2. **Bidder Asymmetry in First-Price Auctions**

In the markets which rely on auctions as the allocation mechanism the participants often have private information about their valuations for the object which is being sold. In many settings the participants differ not only in their realizations of the private information but also in the distribution from which their private values are drawn.These are settings with asymmetric bidders.[12]

Inferring such bidder asymmetries may be important from a purely positive point of view, since they could be indicative of collusive arrangements, quality differences, or differences in information. Additionally, auction theory suggests that in order to answer mechanism design questions or to assess the potential outcomes of policy interventions in auction markets the researcher needs to know the distribution of private information at the level at which it is observable by the market participants. The existing empirical studies, however, limit their attention to the asymmetries associated with observable bidder groups due to data limitations or because of the computational burden associated with more sophisticated modeling. Our method can help to overcome some of these difficulties. We summarize the insight enabling the recovery of unobserved groups in the case of the first price auction while relegating a similar analysis for the English auction and for the asymmetries arising due to the collusive behavior to the Online Appendix.

Formally, similar to our basic setting, let the population of bidders, N, be partitioned into K_0 groups, such that each group is characterized by a distinct distribution of private values, $F_k(.)$. For simplicity, we assume that the distributions of values associated with various groups differ only in their means and that $\bar{q}_1 < \bar{q}_2 < ... < \bar{q}_{K_0}$, where $\bar{q}_k$ refers to the mean of the distribution F_k. Hence for a bidder i with $\tau(i) = k$ the mean of the cost distribution is given by $q_i = \bar{q}_k$.[13] Let S_l denote the set of participants in auction l and $B_l = \{B_{i,l}\}$ a vector of bids submitted by these participants where $B_{i,l} = \beta_k(V_{i,l})$ with $\beta_k(.)$ reflecting type-symmetric equilibrium bidding function and $V_{i,l}$ are distributed according to $F_k(.)$ if $\tau(i) = k$.

[12]Some references to the auction studies of the environments with asymmetric bidders are provided in introduction to the paper.

[13]The result here holds more generally when F_k are stochastically ordered. By strict stochastic dominance we mean $F_{k'} < F_k$ whenever $k' < k$ over a non-degenerate interval on the shared support.

Define $G_{ij}(b) = \Pr(B_{i,l} \leq b|\, i,j \in S_l)$.[14] Then $G_{ij}(b) \leq G_{ji}(b)$ for all b in the common support of $B_{i,l}$ and $B_{j,l}$ whenever $\tau(i) \leq \tau(j)$. 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 participating bidders S (see Corollary 3 of Lebrun (1999)). Such an inequality also holds unconditionally when aggregated over the identities of the competitors and auction characteristics. The latter is possible because the index for agents i and j is based on the bids submitted in the auctions where both bidders participate. A similar property holds in the settings where allocations are implemented through first-price procurement auctions.[15] The only difference is that in these settings $G_{ij}(b)$ should be defined as $G_{ij}(b) = \Pr(B_{i,l} \geq b|\, i,j \in S_l)$.

Returning to the case of a standard first-price auction, the property above serves as a basis for pairwise comparison. Specifically, comparison indexes can be constructed as follows:

$$\delta_{ij}^{+} \equiv \int \max\left\{G_{ji}(b) - G_{ij}(b), 0\right\} db. \tag{3.4}$$

Likewise, define δ_{ij}^{-} by swapping the role of i, j in δ_{ij}^{+}; and define δ_{ij}^{0} by taking the absolute value of the difference in the integrand in δ_{ij}^{+}.

Notice that the indexes do not condition on the set of bidders participating in the auction l. This feature is particularly attractive in the settings where the set of participants varies across auctions due to exogenous reasons or due to bidders actively making participation decisions since it allows to utilize observations from a large number of auctions when constructing a comparison index for i and j.

3.2.3. **Firms' Cost Efficiency and Pricing Decisions**

Pairwise comparisons based on chosen equilibrium actions also arise in the settings without private information. Continuing with the differentiated products example, assume for simplicity that each firm produces a single product and N denotes the population of firms comprising a given industry. The data are organized by markets ($l = 1, ..., L$) and S_l denotes the set of products available in market l. Assume that the researcher has been able to estimate the demand system and thus to recover the relationship between the firms' market shares ($\{\sigma_{j,l}(.)\}$) and the products' attributes ($\mathbf{x}_l = \{x_{j,l}\}_{j \in S_l}$), and prices ($\mathbf{p}_l = \{p_{j,l}\}_{j \in S_l}$) for a given set of competing products S_l and given other market factors included in Ω_l: $\sigma_{j,l}(\mathbf{x}_l,\, \mathbf{p}_l, \Omega_l)$.

Further, the marginal cost for firm i on market l is $c_{i,l} = \varphi(w_{i,l}, q_i, \eta_{i,l})$, where $w_{i,l}$ are cost shifters that may overlap with $x_{i,l}$, q_i is a brand-specific unobserved heterogeneity fixed across markets, and $\eta_{i,l}$'s are i.i.d. idiosyncratic noises independent from cost shifters and unobserved heterogeneity. We may interpret q_i as a measure of the firm i's cost efficiency. Firms have

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

[15]In contrast to standard first-price auctions where the bidder who submitted the highest bid wins, under procurement first-price auction the object is allocated to the bidder who submitted the lowest bid.

complete information about each others' cost efficiencies.[16] As above the population of firms is partitioned into groups corresponding to different levels of q_i: $N = \cup_k N_k$ where $i \in N_k$ if $\tau(i) = k$.

The profit for firm i in market l is: $\pi_{i,l} = (p_{i,l} - c_{i,l})\sigma_{i,l}(\mathbf{x}_l, \mathbf{p}_l)M_l$, where M_l is a measure of potential consumers in market l. In any pricing equilibrium with an interior solution, the first-order condition implies

$$c_{i,l} = p_{i,l} + \frac{\sigma_{i,l}}{\partial \sigma_{i,l}/\partial p_{i,l}}. \tag{3.5}$$

Notice that if $\varphi(w_{i,l}, q_i, \eta_{i,l})$ is monotone in q_i then so is the right-hand side of the expression in (3.5) which can be constructed from the the estimates of the demand-side primitives. Hence, for any pair $i, j \in N$, we have $q_i \geq q_j$ if and only if $\mathbf{E}[z_{i,l}|w_{i,l} = w_0] \geq \mathbf{E}[z_{j,l}|w_{j,l} = w_0]$, for all w_0, where $z_{i,l} = p_{i,l} + \frac{\sigma_{i,l}}{\partial \sigma_{i,l}/\partial p_{i,l}}$, and the statement is true when both inequalities are strict.

This insight is exploited in the literature analyzing productivity of firms where the researchers use information available in firms' output and input choices while controlling for the influence of the market conditions through the demand-side estimates (see, for example, DeLoecker and Warzynski (2012)). It also serves as a basis for constructing pairwise indexes. Specifically, define the pairwise comparison index

$$\delta^+_{ij} \equiv \int \max\{\mathbf{E}[z_{i,l}|w_{i,l} = w_0] - \mathbf{E}[z_{j,l}|w_{j,l} = w_0], 0\} dF(w_0). \tag{3.6}$$

Our model equilibrium implies that $\delta^+_{ij} > 0$ if and only if $q_i > q_j$. Likewise, define δ^-_{ij} by swapping the roles of i and j in δ^+_{ij}, and define δ^0_{ij} by replacing $\max\{\cdot, 0\}$ in the integral with the absolute value.

Notice that pairwise comparison above does not condition on the set of the firm's competitors in a specific market (beyond the dependency between the products' market shares on this set) or require that the firms that are being compared are present in the same market. Using an estimation approach which includes the classification step may be useful in the setting where the population of firms is large and their participation varies across markets and is potentially non-random. In such environments, recovering the model primitives from the observed outcomes while allowing for the firm-specific unobserved heterogeneity may prove challenging either computationally or inferentially. The approach proposed in this paper allows the researcher to substantially reduce these burdens.

3.2.4. **Assortative Matching in Labor Market**

A growing recent literature in labor economics studies sorting of heterogeneous employees across heterogeneous firms.[17] In the settings considered in this literature firms are heterogeneous so that the productivity of a given worker varies across firms when all other things are held equal. Similarly, workers differ in their ability which is analogous to q_i introduced

[16] This assumption is plausible in certain industries where production efficiency is mostly determined by firms' technology or equipment that is publicly observable.

[17] See, for example, Lentz and Mortensen (2010), Abowd, Kramarz, and Margolis (1999), and Lise, Meghir, and Robin (2011).

above. Under some further restrictions (see Eeckhout and Kircher (2011) and Hagedorn, Law, and Manovskii (2016)) it can be shown that, everything else equal, among the two workers employed by the same firm the worker with higher ability would earn a higher wage in equilibrium. This forms a basis for pairwise comparisons. Specifically, let $w_{i,f,t} = W(q_i, X_{i,t}, \Omega_{f,t})$ denote the wage individual i earns at time t while employed by firm f, where W is a non-stochastic function. Here $\Omega_{f,t}$ captures all the relevant firm-specific factors while $X_{i,t}$ reflects individual $i's$ characteristics other than q_i. Using $N_{f,t}$ to denote the set of workers employed by firm f at time t, define the comparison index as: for each pair $i, j \in N_{f,t}$,

$$\delta_{ij}^{+} = \int \max\{m_{i,f,t}(x_0) - m_{j,f,t}(x_0), 0\} dF(x_0),$$

where $m_{i,f,t}(x_0) = \mathbf{E}[w_{i,f,t}|\Omega_{f,t}, X_{i,t} = x_0]$. Then $\delta_{ij}^{+} > 0$ if and only if $q_i > q_j$. Likewise before, define δ_{ij}^{-} by swapping the roles of i and j in δ_{ij}^{+}, and define δ_{ij}^{0} by replacing the max operator with its absolute value. Notice that in this setting comparison of workers is complicated by the (unobserved) firm heterogeneity and sorting of workers across firms. Pairwise comparisons allow researcher to circumvent these issues by focusing workers' wages earned while they are employed by the same firm. It is also worthwhile emphasizing that this environment is not characterized by strategic interdependence.

4. Identification of the Ordered Group Structure

4.1. Identification Analysis with a Growing Number of Agents

Our pairwise comparison method is attractive in a setting where a simultaneous ordering of all players is infeasible in practice. For example, a global index that would lead to a simultaneous ordering of all players can be difficult to derive theoretically, or cannot be reliably estimated due to data limitation. Such data limitation is illustrated in Figure 1, where each column symbolizes a "market" and each row an individual agent. The ellipses in each column represent agents participating in a market. The second panel shows an example of a data set where only very few markets share exactly the same set of participants. Nevertheless, lots of markets share the same pattern of participation decisions between the first and the third agent. The third panel shows that these two players jointly participate in many markets in the data; the fourth panel highlights many markets in which the first agent participates while the third does not. In such a data set, accurate inference based on any quantity that conditions on the whole set of participants is not possible due to the small number of markets available. On the other hand, pairwise indexes which only condition on the participation pattern between a fixed pair of agents can be accurately estimated.

To accommodate this data feature formally in our asymptotic theory, let us define for any $S \subset N$,

$$\mathcal{L}(S) = \{1 \leq l \leq L : S_l = S\}.$$

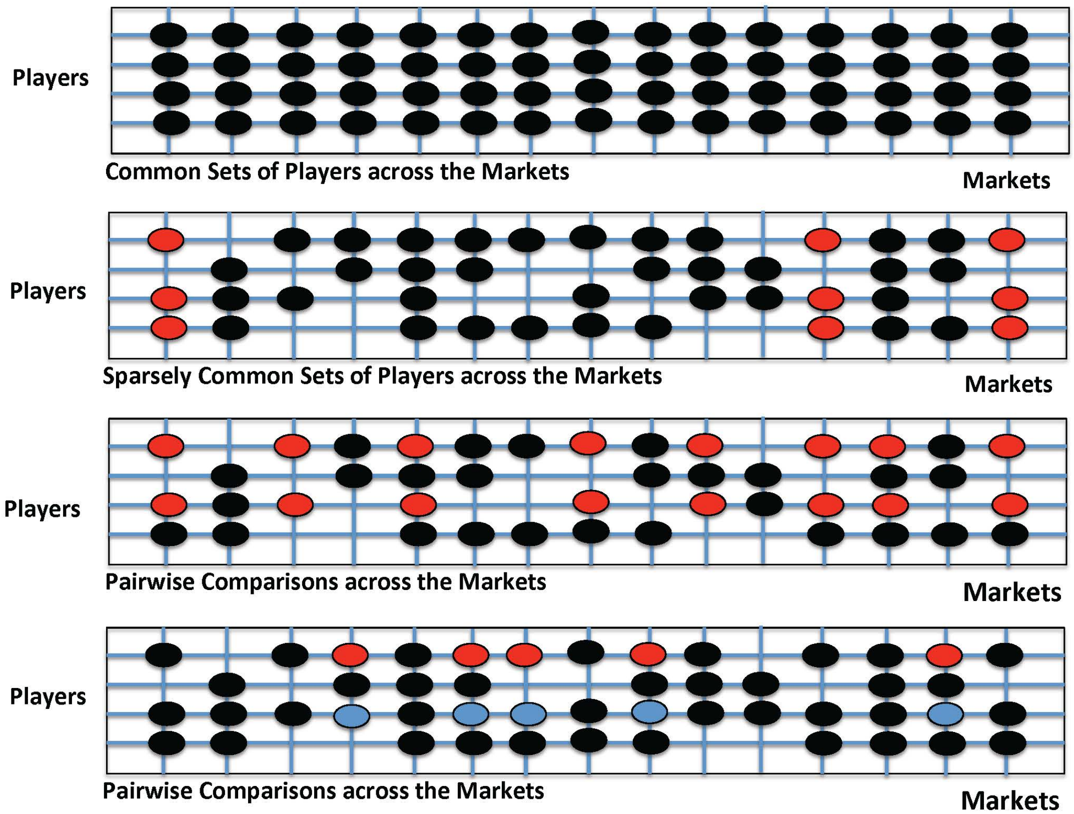

FIGURE 1. The first panel shows an example of a data set where the set of participants is the same across all markets. The second panel shows another example of data where only few markets share the same set of participants. The third panel shows that there are lots of markets in the second data set where both the first and the third player participate. Thus a researcher can estimate population quantities that condition on the joint participation of these two players with better accuracy than quantities that condition on the whole set of participants. The fourth panel highlights many markets in which only one out of these two players participates.

Thus $\mathcal{L}(S)$ represents the set of markets where the set of participants in a market l is precisely S. In this paper, we say that the sets of players are *sparsely common across the markets*, if

$$\max_{S \subset N} |\mathcal{L}(S)|/L \to 0,$$

as $L \to \infty, n \to \infty$. Thus this asymptotic theory requires us to allow both the number of agents in the population n and the number of markets L to go to infinity, though L goes to infinity faster than n.

The identification of unobserved heterogeneity when both n and L increase to infinity is non-standard. A standard identification analysis assumes the data consists of i.i.d. observations drawn from a representative probability. Such an analysis investigates whether the parameter of interest can be uniquely determined once this representative probability is known. However, such a standard identification analysis is cumbersome when both the number of agents and the number of markets grow to infinity. In such a framework, there is no representative probability on which we can base our identification analysis.

In this paper, our identification method is based on quantities that are consistently estimable as $n \to \infty$ and $L \to \infty$. We say that our parameter of interest is *identified*, if it is uniquely pinned down by such quantities. This identification concept encompasses as a special case the standard identification analysis based on random sampling.[18]

We say that agents (i, j) are *comparable* if there exist consistently estimable pairwise indexes δ_{ij}^{+}, δ_{ij}^{0}, and δ_{ij}^{-} such that (3.2) holds. We maintain that a researcher knows whether each pair of agents is comparable through some pairwise comparison index or not. The determination of such comparability can be done in practice by checking whether the data contains sufficiently many markets which allow for reliable estimation of the pairwise indexes.

Let $\mathcal{E}$ be the collection of pairs (i, j) that are comparable. We refer to comparable agents as adjacent (or linked), so that the set $\mathcal{E}$ forms the set of edges in a graph on the set of agents N. We call this graph (denoted by $G = (N, \mathcal{E})$) the *comparability graph*.[19] We say the ordered group structure τ is *identified* if it is uniquely determined once the comparability graph G and the vectors of pairwise indexes $(\delta_{ij}^{+}, \delta_{ij}^{0}, \delta_{ij}^{-})_{ij \in \mathcal{E}}$ are known.

4.2. Identification of the Ordered Group Structure

Let us explore the identification of τ given the comparability graph G and the vector of pairwise indexes. It is easy to see that if $\mathcal{E}$ contains only a small subset of possible pairs, we may not be able to identify the group structure. The identification of the ordered group structure τ is not guaranteed even when many pairs of agents are comparable. For example, even if G is a connected graph (where any two agents are connected at least indirectly), the ordered group structure τ may not be identified. This is illustrated in a counterexample in Figure 2. Certainly, when every pair of agents are adjacent in the graph G, i.e., G is a complete graph, the ordered group structure τ is identified.[20]

Below we establish a necessary and sufficient condition for the group structure to be identified from an incomplete graph G and the pairwise comparison indexes. Let us introduce some definitions.

Definition 4.1. (i) A graph G_τ is the *τ-collapsed graph* of G if (a) any two adjacent vertices i and j in G with $\tau(i) = \tau(j)$ collapse to a single vertex (denoted by (ij)) in G_τ, (b) any edge

---

[18] As this notion suggests, identification analysis based on the unique recoverability of parameters from some "population" quantities would not be useful for finite sample inference, if such quantities are not consistently estimable in large samples.

[19] In a graph (or network) $G = (N, \mathcal{E})$ the set N represents the set of vertices (or nodes) and $\mathcal{E}$ consists of some pairs ij, with $i, j \in N$, where each pair ij is called an edge (or link). If $(i, j) \in \mathcal{E}$, we say that i and j are *adjacent*. A path is a set of vertices $\{i_1, i_2, ..., i_M\}$ such that $i_1 i_2, i_2 i_3, ... i_{M-1} i_M \in \mathcal{E}$. Two vertices are called *connected* if there is a path having i and j as end vertices. A graph is called *connected* if all pairs of vertices are connected in the graph.

[20] If all pairs of agents are comparable, we can split the set of agents into one group with the lowest type and the other group with the remaining agents. Then we split these remaining agents into one group with the lowest type within these agents and the remaining agents. By continuing this process, we can identify the whole group structure.

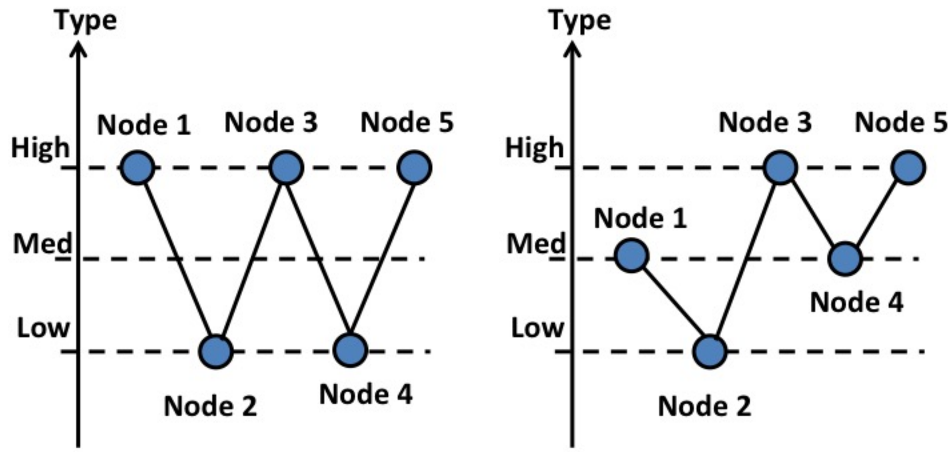

FIGURE 2. This figure illustrates an example where the comparability graph does not identify the group structure even when all the vertices (or nodes) are connected. Here the comparability graph is $G = (N, \mathcal{E})$, where $N = \{1, 2, ..., 5\}$ and $\mathcal{E} = \{12, 23, 34, 45\}$. We cannot identify the group structure from these pairwise orderings. The two panels depict two different group structures that are compatible with the same pairwise ordering.

in G joining a vertex k to either i or j joins vertex k to (ij) in G_τ and (c) all the remaining vertices and edges in G_τ consist of the remaining vertices and edges in G.
(ii) A path in G_τ is *monotone* if $\tau(i)$ is monotone as i runs along the path.
(iii) A vertex i is said to be *identified* if its type $\tau(i)$ is identified.

The τ-collapsed graph of G is constructed by reducing any comparable pair of agents in G who have the same type to a single "agent", and retaining edges as in the original graph of G. Certainly, a τ-collapsed graph G_τ is uniquely determined by δ_{ij}^0's and G. Any pair of adjacent agents in the τ-collapsed graph must have different types, and hence the types of agents on a monotone path are strictly monotone. This means that every vertex on a monotone path in G_τ of lengh $K_0 - 1$ is identified. Also by similar logic, every vertex on a monotone path with end vertices i_H and i_L is identified if the path has length $\tau(i_H) - \tau(i_L)$ and the end vertices i_H and i_L are identified. Using these two facts, we can recover the set of vertices that are identified as follows.

First, set $N_{[1]} \subset N$ to be the set of vertices such that each vertex in the τ-collapsed graph G_τ is on a monotone path in G_τ of length $K_0 - 1$. For $j \geq 1$ generally, let $N_{[j+1]}$ be the set of vertices each of which belongs to a monotone path, say, P, such that its end vertices i_H and i_L

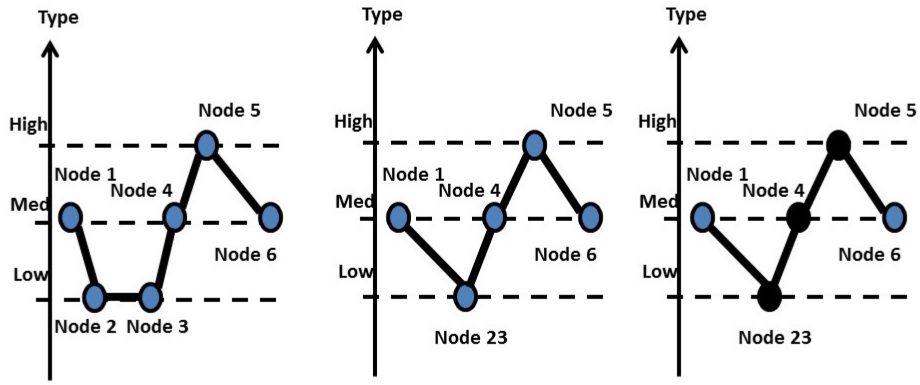

FIGURE 3. This figure shows an example where the condition $N = N^*$ in Theorem 3.1 is violated. The first panel depicts the comparability graph as a line graph connecting 6 vertices (or nodes). The second panel shows the τ-collapsed graph where the two comparable nodes 2 and 3 that have the same type are collapsed into one node named 23. In the last panel, it is shown that Nodes 23, 4, and 5 (expressed as solid black nodes) are identified, because they are on a monotone path of length $K_0 - 1 = 2$. In this example, Nodes 1 and 6 are not identified and thus the comparable graph does not lead to the identification of the group structure.

are from $N_{[j]}$ and $\tau(i_H) - \tau(i_L)$ is equal to the length of the monotone path P. Then define

$$N^* \equiv \bigcup_j N_{[j]}.$$

Given G_τ, N^* is uniquely determined as a subset of N. It is not hard to see that if $N = N^*$ and K_0 is identified, the type structure τ is identified. The following theorem shows that this condition is in fact necessary for the identification of τ as well.

Theorem 4.1. *Let G be a given comparability graph and G_τ be its τ-collapsed graph. The type structure τ is identified if and only if there exists a monotone path in G_τ whose length is equal to $K_0 - 1$ and*

$$N = N^*.$$

Any monotone path in G_τ cannot have length greater than $K_0 - 1$. Note that there exists a monotone path in G_τ whose length is equal to $K_0 - 1$ if and only if K_0 is identified. The conditions in the theorem are obviously satisfied if G contains a monotone Hamiltonian path, i,e., a path that is monotone and covers all the vetrices. The latter condition is trivially satisfied when G is a complete graph. Figure 3 gives a counterexample where the condition that there exists a monotone path in G_τ whose length is equal to $K_0 - 1$ is satisfied, but $N \neq N^*$ so that the comparability graph does not lead to the identification of the group structure.

5. Consistent Estimation of the Ordered Group Structure

5.1. Pairwise Hypothesis Testing Problems

In this section, we develop a method to estimate the group structure consistently for the case where the comparability graph is complete, so that we take $\mathcal{E}$ to be all ij with $i, j \in N, i \neq j$. The main challenge 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 from data, but these pairwise comparisons do not necessarily generate an ordered group structure, because transitivity of the pairwise orderings may be violated in finite samples. Thus we need to develop an estimation method that imposes transitivity in finite samples.

Our classification method is based on individual p-values from the hypotheses on pairwise inequalities in (3.2). We first formulate three pairwise hypothesis testing problems for each comparable pair $ij \in \mathcal{E}$:

$$
\begin{aligned}
H_{0,ij}^{+} &: \delta_{ij}^{+} \leq 0 \text{ against } H_{1,ij}^{+} : \delta_{ij}^{+} > 0, \\
H_{0,ij}^{0} &: \delta_{ij}^{0} = 0 \text{ against } H_{1,ij}^{0} : \delta_{ij}^{0} \neq 0 \text{ and} \\
H_{0,ij}^{-} &: \delta_{ij}^{-} \leq 0 \text{ against } H_{1,ij}^{-} : \delta_{ij}^{-} > 0.
\end{aligned}
\tag{5.1}
$$

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 (i.e., the number of the markets) that is used to construct these p-values. We will explain conditions for the p-values later and explain details for construction of p-values using bootstrap later in Section 5.3.

5.2. Classification Method

5.2.1. **Estimation for 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 $i \in N$, $q_i \in \{\bar{q}_h, \bar{q}_l\}$ for some two unknown numbers $\bar{q}_h$ and $\bar{q}_l$ such that $\bar{q}_h > \bar{q}_l$. In this case, there could potentially be several different ways of partitioning N into two groups using the p-values. Here we develop a method that permits a natural extension to a more general case of $K_0 > 2$.

First, let

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

so that the group structure is given by

$$T = (N_l, N_h).$$

Even with this case of two groups, the total number of potential partitions of N is 2^n, $n = |N|$. Instead of checking all the incidences of the potential partitions, we propose a split algorithm that estimates T in two steps. First, for each $i \in N$, we consider a pair of partitions based on pairwise comparisons of the other players with i. Then we choose one that is most likely to be the right partition.

Step 1: For each $i \in N$, we consider two partitions of N:

$$\begin{aligned}\hat{T}_1(i) &= (\hat{N}_1(i), N \setminus \hat{N}_1(i)), \text{ and} \\ \hat{T}_2(i) &= (N \setminus \hat{N}_2(i), \hat{N}_2(i)),\end{aligned}$$

where for each $i \in N$,

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

The group $\hat{N}_1(i)$ is the estimated set of agents with types lower than the agent i and the group $\hat{N}_2(i)$ is the estimated set of agents with types higher than the agent i. Thus the first partition $\hat{T}_1(i)$ regards i as high type and the second partition $\hat{T}_2(i)$ regards i as low type. It remains to choose among the partitions.

Step 2: For each $i \in N$, we define[21]

$$s(i) = \min\{s_1(i), s_2(i)\},$$

where for $k \in \{1, 2\}$,

$$s_k(i) = \frac{1}{|\hat{N}_k(i)|} \sum_{j \in \hat{N}_k(i)} \log \hat{p}_{ij}^{0}.$$

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. The quantity $s(i)$ measures how unlikely $\hat{T}_1(i)$ or $\hat{T}_2(i)$ is the right partition. Then we take

$$\hat{T} = \begin{cases} \hat{T}_1(i^*), & \text{if } s(i^*) = s_1(i^*) \\ \hat{T}_2(i^*), & \text{if } s(i^*) = s_2(i^*), \end{cases}$$

where $i^* = \operatorname{argmin}_{i \in N} s(i)$.

5.2.2. **Estimation for a Known Number of Groups using a Sequential Split Algorithm**

We generalize the procedure to the case where the econometrician knows the number of distinct types K_0 that is allowed to be greater than two. The main idea is that we split the groups sequentially using the previous algorithm.

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

[21] 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.

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

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

We choose r^* such that

$$\hat{p}_{r^*} = \min_{r=1,2} \hat{p}_r \tag{5.2}$$

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

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

$$\hat{p}_r = \min_{i,j \in \hat{N}_r : i \neq j} \hat{p}^0_{ij}. \tag{5.3}$$

We choose r^* such that

$$\hat{p}_{r^*} = \min_{r=1,...,k} \hat{p}_r$$

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

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

We continue until the total number of groups obtained becomes K_0.

As mentioned after Theorem 5.1 below, the probability that a split divides an equi-type group into two is negligible. When we are given classifications $\hat{N}_1, \hat{N}_2, ..., \hat{N}_k$, we define $\hat{p}_r$ as in (5.3) which is used as a group homogeneity index. Intuitively, when $\hat{p}_r$ is low, the group $\hat{N}_r$ is likely to be heterogeneous. As long as the current number of groups is smaller than K_0, we select a group with the lowest homogeneity index and split the group.

5.2.3. **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 misspecification bias and a penalty term for overfitting.

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

$$\hat{T}_K = (\hat{N}_1, \hat{N}_2, ..., \hat{N}_K). \tag{5.4}$$

We define

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

We define our criterion function as follows:

$$\hat{Q}(K) \equiv \hat{V}(K) + K g(L),$$

where $g(L)$ is slowly increasing in L. The precise condition is given in Theorem 5.2 below.[22] The component $\hat{V}(K)$ measures the goodness-of-fit of the classification, and the second component $Kg(L)$ plays the role of a penalty term that prevents overfitting. We select $\hat{K}$ as follows:

$$\hat{K} = \text{argmin}_{K=1,\dots,n}\hat{Q}(K).$$

In a later section, we show that the estimated number of groups $\hat{K}$ is equal to K_0 with probability approaching one.

5.2.4. **Discussion about the Sequential Split Algorithm**

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

First, the split algorithm into two groups is designed so that its consistency can be extended to a general case of $K_0 \geq 2$. The intuition is as follows. Suppose that we have K_0 groups that are ordered, so that we have $T = (N_1, ..., N_{K_0})$. The design of Step 1 ensures that whenever i is of type k, $\hat{N}_1(i)$ coincides with the union of j's that have lower type than i with probability approaching one, and $\hat{N}_2(i)$ 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 $(N\backslash\hat{N}_2(i), \hat{N}_2(i))$ and $(\hat{N}_1(i), N\backslash\hat{N}_1(i))$ 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, 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. We have determined our sequential algorithm after checking its finite sample performance through various Monte Carlo experiments.

[22]The choice of $g(L) = \log\log L$ appears to work very well from our numerous Monte Carlo simulation experiments.

5.2.5. **Construction of p-Values Using Bootstrap**

In most applications, we can use bootstrap to construct p-values for testing the inequality restrictions of (5.1).[23] We explain this procedure using the environment considered in Monte Carlo experiments and in an empirical application of our method. In these sections we consider data from first-price auction markets.

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 $r_{ij}(\cdot)$.

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 indexes:

$$\begin{aligned} \hat{\delta}_{ij}^{+} &= \int \max\{\hat{r}_{ij}(b), 0\}\, db, \\ \hat{\delta}_{ij}^{-} &= \int \max\{\hat{r}_{ji}(b), 0\}\, db, \text{ and} \\ \hat{\delta}_{ij}^{0} &= \int |\hat{r}_{ij}(b)|\, db. \end{aligned} \tag{5.5}$$

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

For p-values, we re-sample $\{Z_l^*\}_{l=1}^{L}$ (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:

$$\begin{aligned} \hat{\delta}_{ij}^{+*} &= \int \max\big\{\hat{r}_{ij}^*(b) - \hat{r}_{ij}(b), 0 \quad db, \\ \hat{\delta}_{ij}^{-*} &= \int \max\big\{\hat{r}_{ji}^*(b) - \hat{r}_{ji}(b), 0 \quad db \text{ and} \\ \hat{\delta}_{ij}^{0*} &= \int \ \hat{r}_{ij}^*(b) - \hat{r}_{ij}(b) \ \, db. \end{aligned} \tag{5.6}$$

[23]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.

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 $\hat{\delta}_{ij}^{+*}$, $\hat{\delta}_{ij}^{-*}$, and $\hat{\delta}_{ij}^{0*}$ respectively.[24]

5.3. Consistency of Classification

Let us explore the sense in which $\hat{T}$ is a "reliable" estimator of T. Let $\mathcal{T}_K$ be the collection of all the ordered K-partitions of N. Here we formally present conditions under which $\hat{T} = T$ with probability approaching one as $n, L \to \infty$. ((In this paper, we consider asymptotics where n and L increase jointly. As for the players, we adopt asymptotics where we have $N_n \subset N_{n+1} \subset N_{n+2} \subset ....$ [25])) Since we are dealing with an increasing set of agents, we need to clarify the meaning of a sequence of probabilities in the asymptotic theory. Let $\mathcal{P}_n$ be the collection of the distributions P of the whole vector of the observations in the data. For each $\varepsilon > 0$, $ij \in \mathcal{E}$ and $s \in \{+, 0, -\}$, we define

$$\begin{aligned} \mathcal{P}_{0,ij}^{s} &= \{P \in \mathcal{P}_n : \delta_{ij}^{s}(P) \le 0\}, \text{ and} \\ \mathcal{P}_{\varepsilon,ij}^{s} &= \{P \in \mathcal{P}_n : \delta_{ij}^{s}(P) > \varepsilon\}, \end{aligned}$$

where we write the pairwise indexes δ_{ij}^{s} as $\delta_{ij}^{s}(P)$ to reflect that the pairwise indexes depend on P. Thus $\mathcal{P}_{0,ij}^{s}$ is the collection of probabilities under the pairwise null hypothesis $H_{0,ij}^{s}$ and $\mathcal{P}_{\varepsilon,ij}^{s}$ that of probabilities under the pairwise alternative hypotheses $H_{1,ij}^{s}$ such that $\delta_{ij}^{s}(P)$ is away from zero at least by ε. Then for each subset $\mathcal{E}_s' \subset \mathcal{E}_s$ and subset $\mathcal{E}_s \subset \mathcal{E}$, let

$$\mathcal{P}_{0,n}^{s}(\mathcal{E}_s', \mathcal{E}_s) = \bigcap_{ij \in \mathcal{E}_s'} \mathcal{P}_{0,ij}^{s}, \text{ and } \mathcal{P}_{\varepsilon,n}^{s}(\mathcal{E}_s', \mathcal{E}_s) = \bigcap_{ij \in \mathcal{E}_s \setminus \mathcal{E}_s'} \mathcal{P}_{\varepsilon,ij}^{s}.$$

Hence under any probability in $\mathcal{P}_{0,n}^{s}(\mathcal{E}_s', \mathcal{E}_s)$, the pairwise null hypotheses for all the adjacent pairs in $\mathcal{E}_s'$ hold, and under any probability in $\mathcal{P}_{\varepsilon,n}^{s}(\mathcal{E}_s', \mathcal{E}_s)$, the pairwise index of all the pairs in $\mathcal{E}_s \setminus \mathcal{E}_s'$ is away from zero by more than ε.

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

Assumption 5.1. For each $s \in \{+, -, 0\}$, $i \in N$, $\mathcal{E}_s' \subset \mathcal{E}_s \subset \mathcal{E}$, and $\varepsilon > 0$,
(i) $P_n \left\{ \min_{j \in N : ij \in \mathcal{E}_s'} \hat{p}_{ij}^{s} \le 0 \right. = o(1/n)$, along $P_n \in \mathcal{P}_{0,n}^{s}(\mathcal{E}_s', \mathcal{E}_s)$, as $n, L \to \infty$, and
(ii) $P_n \left\{ \max_{j \in N : ij \in \mathcal{E}_s \setminus \mathcal{E}_s'} \hat{p}_{ij}^{s} \le 0 \right. = 1 - o(1/n)$, along $P_n \in \mathcal{P}_{\varepsilon,n}^{s}(\mathcal{E}_s', \mathcal{E}_s)$, as $n, L \to \infty$.

Assumption 5.1 is satisfied when L increases sufficiently faster than n and pairwise individual hypotheses allow for consistent testing. (See Appendix B for further details.)

[24] When $a_L\{\hat{r}_{ij}(\cdot) - r_{ij}(\cdot)\}$ for an appropriate normalizing sequence $a_L \to \infty$ 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 $b_L\{\hat{r}_{ij}(\cdot) - r_{ij}(\cdot)\}$ for an appropriate normalizing sequence $b_L \to \infty$ 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.

[25] This asymptotics simplifies the development by removing the need to keep track of players entering and exit the set N and the links in $\mathcal{E}_n$ forming and disappearing as n increases. The asymptotics is a mathematical tool to obtain approximation of inference when a given sample size is "large" but yet finite.

Define for each $\varepsilon > 0$,

$$\mathcal{P}_{0,\varepsilon,n} = \bigcup_{\mathcal{E}_+,\mathcal{E}_0,\mathcal{E}_-} \bigcap_{s\in\{+,0,-\}} \bigcup_{\mathcal{E}_s'\subset\mathcal{E}_s} \left(\mathcal{P}_{0,n}^s(\mathcal{E}_s',\mathcal{E}_s) \cup \mathcal{P}_{\varepsilon,n}^s(\mathcal{E}_s',\mathcal{E}_s)\right),$$

where the first union is union over any disjoint subsets $\mathcal{E}_+, \mathcal{E}_0, \mathcal{E}_-$ of $\mathcal{E}$. The collection $\mathcal{P}_{0,\varepsilon,n}$ represents that of probabilities under which for each $ij \in \mathcal{E}$, the null hypothesis $H_{0,ij}^s$ holds or the alternative hypothesis $H_{1,ij}^s$ holds with the pairwise index greater than ε. In the following, we show that the classification with a correctly specified number of groups is consistent.

Theorem 5.1. *Suppose that Assumption 5.1 holds and that the number of the groups K_0 is known to the econometrician. Assume furthermore that $\hat{T}_{K_0}$ is obtained at Step $K_0 - 1$. Then, for each $\varepsilon > 0$ and along the sequence of probabilities $P_n \in \mathcal{P}_{0,\varepsilon,n}$, as $n, L \to \infty$,*

$$P_n\{T = \hat{T}_{K_0}\} \to 1.$$

Theorem 5.1 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, when the number of groups K_0 is correctly specified.

Let us consider the case where the number of groups K_0 is not known. We make the following further assumptions.

Assumption 5.2. There exist sequences $r_{1,n}, r_{2,n} \to 0$ such that $r_{1,n}/r_{2,n} \to 0$ as $n \to \infty$ and the following holds for each $\varepsilon > 0$ and $\mathcal{E}_0' \subset \mathcal{E}_0 \subset \mathcal{E}$, as $n, L \to \infty$:
(i) $P_n\left\{\min_{ij\in\tilde{N}(P_n)} \log \hat{p}_{ij}^0 \leq -r_{1,n}\right\} = o(1)$ along $P_n \in \mathcal{P}_{0,n}^0(\mathcal{E}_0', \mathcal{E}_0)$,
(ii) $P_n\left\{\max_{ij\notin\tilde{N}(P_n)} \log \hat{p}_{ij}^0 \leq -r_{2,n}\right\} = 1 - o(1)$ along $P_n \in \mathcal{P}_{\varepsilon,n}^0(\mathcal{E}_0', \mathcal{E}_0)$,
where $\tilde{N}(P_n)$ is the set of ij's such that i and j are in the same group under P_n.

Assumption 5.2 essentially requires that the test of equal type between i and j should be a consistent test such that we can distinguish the null from the alternative hypothesis with probability approaching one as $n, L \to \infty$. The condition is often satisfied when L diverges to infinity sufficiently larger than n. For example, following the arguments after Assumption 5.1, one can show that such $r_{1,n}$ and $r_{2,n}$ exist if L^{-h_1} and L^{-h_2} decrease faster than n^{-1}.

Theorem 5.2. *Suppose that Assumptions 5.1 - 5.2 hold and that $r_{1,n}/g(L) + g(L)/r_{2,n} \to 0$ as $n, L \to \infty$. Let $\varepsilon > 0$ and let P_n be a sequence of probabilities from $\mathcal{P}_{0,\varepsilon,n}$. Then, we have as $n, L \to \infty$,*

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

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

$$P_n\{T = \hat{T}_{\hat{K}}\} \to 1.$$

The main part of Theorem 5.2 is to show that $\hat{K}$ is consistent for K_0. When $K < K_0$, the component $\hat{V}(K)$ is $o_P(r_{1,n})$, while the penalty term increases faster than $r_{1,n}$ as $L \to \infty$. When $K > K_0$, the component $\hat{V}(K)$ diverges at a rate faster than $g(L)$. From this, we obtain that $\hat{K}$ is consistent for K_0.

Table 1: Group Structure in Experiments

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	D_μ	μ_1	μ_2	μ_3	μ_4
P1	0.6	2.0	2.6	3.2	3.8
P2	0.4	2.0	2.4	2.8	3.2
P3	0.2	2.0	2.2	2.4	2.6

Note: For S1 and S3 (with two groups), we use μ_1 and μ_2 only. In S2 and S4 (with four groups), we use μ_1, μ_2, μ_3, and μ_4. Here D_μ denotes the constant increment between μ_k and μ_{k-1}.

6. Monte Carlo Simulations

Our Monte Carlo study proceeds in two parts. The first part focuses on the performance of our classification method. The second part investigates the role of the first step classification in the second stage estimation. The details concerning computational time associated with various steps are reported in the Online Appendix.

6.1. Finite Sample Performance of the Classification

6.1.1. **Simulation Design**

The Monte Carlo simulation study is based on an example of a first-price procurement auction with asymmetric, independent private costs. 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.[26]

Table 1 summarizes the group structures we used in our design of Monte Carlo experiments. The first two structures involve a total of 12 bidders and the last two 40 bidders. Also, the first and third are designed to be coarser group structures than the second and fourth respectively. Table 2 summarizes the specification of group means of unobserved heterogeneity μ_k. The increments between group means is the largest in P1 and smallest in P3. Intuitively, the task to classify bidders into groups is harder to perform when these increments are smaller. The variance σ^2 is taken to be 0.25.

[26]Specifically, L represents the number of auctions in which any given pair of bidders participates.

Table 3: Performance of the Classification Estimator with One Group
($K_0 = 1$ and unknown)

n	L	$\hat{K}_0$	EAD	DH(10)	DH(25)	DH(50)	DH(75)	DH(90)
12	400	1.020	0.024	0.001	0	0	0	0
12	200	1.020	0.024	0.008	0	0	0	0
12	100	1.020	0.028	0.004	0.002	0	0	0
40	400	1.032	0.102	0.002	0.004	0	0	0
40	200	1.064	0.114	0.018	0.010	0	0	0
40	100	1.092	0.136	0.014	0.010	0	0	0

Note: n is the number of bidders in population; and L the number of the markets. $\hat{K}_0$ is the average number of estimated groups in 500 simulation samples. EAD indicates the average number of mismatched bidders across true groups and simulated samples. DH(100λ) measures the distribution of mismatched bidders. For example, DH(10) = 0.002 means that except for 1 out of 500 simulation samples, the average number of mismatched bidders was not more than 10 percent of the total number of the bidders.

We construct p-values following the procedure described in Section 5.2.5 and obtain group classification from 500 simulated samples. For each estimate, we used 200 bootstrap iterations while calculating p-values.

To measure the performance of our classification method, we define a measure of discrepancy between two ordered partitions T_1 and T_2 as follows:

$$\delta\left(T_1, T_2\right) \quad = \quad \frac{1}{K_1} \sum_{k=1}^{K_1} \min_{1 \leq j \leq K_2} |N_k^1 \triangle N_j^2|, \tag{6.1}$$

for any two different ordered partitions, $T_1 = (N_1^1, ..., N_{K_1}^1)$ and $T_2 = (N_1^2, ..., N_{K_2}^2)$, of N, where $\triangle$ denotes set-difference. Then we report the finite sample performance of our classification method based on the following two measures:

$$\begin{aligned} \text{Expected Average Discrepancy (EAD)} \quad &: \quad \mathbf{E}\left[\delta(T, \hat{T}_{\hat{K}})\right] \\ \text{Hazard Rate of EAD at } \lambda \text{ (DH}(100p)) \quad &: \quad \Pr(\delta(T, \hat{T}_{\hat{K}}) > \lambda n), \end{aligned}$$

where $\delta(T, \hat{T}_{\hat{K}})$ is as defined in (6.1). Here we chose $\lambda \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. We use $g(L) = \log\log L$ for the penalty scheme. (We also report results from the case when the number of groups is known in the Online Appendix.)

6.1.2. **Results**

Table 3 reports estimates when there is no unobserved heterogeneity among bidders ($K_0 = 1$). In this case, the estimates for the number of groups $\hat{K}$ are mostly one. This suggests our procedure detects the absence of unobserved heterogeneity effectively . For a given number of bidders, there is a moderate increase in the accuracy of our classification results as the number of markets increases (in terms of both EAD and DH(100λ)).

Table 4: Performance of the Classification Estimator with Two Groups: ($K_0 = 2$ and unknown)

n	L	D_μ	$\hat{K}_0$	EAD	DH(10)	DH(25)	DH(50)	DH(75)	DH(90)
12	400	0.6	2.000	0.012	0.012	0	0	0	0
12	400	0.4	2.000	0.014	0.014	0	0	0	0
12	400	0.2	2.000	0.004	0.004	0	0	0	0
12	200	0.6	2.000	0.004	0.004	0	0	0	0
12	200	0.4	2.000	0.006	0.006	0	0	0	0
12	200	0.2	2.100	1.376	0.496	0.224	0.158	0	0
12	100	0.6	2.000	0.006	0.006	0	0	0	0
12	100	0.4	2.004	0.080	0.078	0.002	0	0	0
12	100	0.2	2.050	1.721	0.654	0.480	0.154	0.038	0
40	400	0.6	2.004	0.025	0.002	0.002	0	0	0
40	400	0.4	2.000	0.011	0	0	0	0	0
40	400	0.2	2.002	1.085	0.094	0	0	0	0
40	200	0.6	2.000	0.018	0	0	0	0	0
40	200	0.4	2.000	0.020	0	0	0	0	0
40	200	0.2	2.132	3.352	0.700	0.376	0.200	0.122	0.012
40	100	0.6	2.004	0.025	0	0	0	0	0
40	100	0.4	2.022	0.433	0.008	0	0	0	0
40	100	0.2	2.046	4.574	0.888	0.578	0.265	0.092	0.015

Note: $\hat{K}_0$, EAD and DH(100λ) are defined as in Table 3. D_μ is the difference between group means μ_1 and μ_2. Conditional on the number of markets (L) and the number of bidders in population (n), the classification task is harder when the difference between group means D_μ is smaller.

Table 4 reports results for $K_0 = 2$. The estimates for the number of groups are mostly 2, except for a few incidences of overestimation. The simulation results suggest estimation accuracy is lower when the difference between group means is smaller. Nevertheless, even with $D_\mu = 0.2$ and $n = 40$ in moderate-sized samples with L = 200 or 100, the average estimates for the number of groups is 2.132 or 2.046.

The performance in terms of EAD and DH(100λ) shows a similar pattern. Except for the hardest cases with $D_\mu = 0.2$, EAD is small and ranges from 0.006 to 0.080. For example, when $D_\mu = 0.4$, $n = 40$ and $L = 200$, the bidders are mostly accurately classified into two groups. However, when $D_\mu = 0.2$, the performance deteriorates substantially in this case as expected. This pattern is also reflected in DH(100λ). For example, with $n = 40$, $L = 200$ and $D_\mu = 0.4$, we observed that DH(10) is equal to zero, which means that on average less than 10 percent of the bidders (i.e., fewer than 4 bidders) were mis-allocated in 500 simulated samples. However, this percentage increases to 70 percent when $D_\mu = 0.2$.

Table 5: Performance of the Classification Estimator with Four Groups: ($K_0 = 4$ and unknown)

n	L	D_μ	$\hat{K}_0$	EAD	DH(10)	DH(25)	DH(50)	DH(75)	DH(90)
12	400	0.6	3.932	0.051	0.034	0.034	0	0	0
12	400	0.4	3.892	0.081	0.054	0.054	0	0	0
12	400	0.2	3.860	0.105	0.070	0.070	0	0	0
12	200	0.6	3.908	0.069	0.046	0.046	0	0	0
12	200	0.4	3.936	0.048	0.032	0.032	0	0	0
12	200	0.2	3.298	1.045	0.850	0.200	0.048	0	0
12	100	0.6	3.952	0.036	0.024	0.024	0	0	0
12	100	0.4	3.880	0.090	0.060	0.060	0	0	0
12	100	0.2	3.021	1.598	0.870	0.326	0.134	0	0
40	400	0.6	3.882	0.303	0.062	0.060	0	0	0
40	400	0.4	3.864	0.340	0.068	0.068	0	0	0
40	400	0.2	3.258	2.672	0.739	0.314	0	0	0
40	200	0.6	3.868	0.330	0.066	0.066	0	0	0
40	200	0.4	3.910	0.231	0.046	0.046	0	0	0
40	200	0.2	3.024	2.975	0.892	0.548	0.084	0.014	0
40	100	0.6	3.846	0.391	0.078	0.078	0	0	0
40	100	0.4	3.790	1.048	0.308	0.208	0.010	0	0
40	100	0.2	3.006	2.929	0.998	0.764	0.340	0.170	0.014

Note: $n, L, \hat{K}_0$, D_μ, EAD and DH(100λ) and defined as in Table 3.

Table 5 reports the case with four groups. Estimates for the number of groups are mostly equal to 4, except when $D_\mu = 0.2$. In such cases, the average estimates for K_0 can be as low as 3.021. However, this underestimation is alleviated when we increase the sample size L. For example, when the number of bidders is 12 and the number of markets is 100 and $D_\mu = 0.2$, the average estimated number of groups $\hat{K}_0$ is 3.021. However, as the number of markets increases to 400, the average estimated number of groups $\hat{K}_0$ becomes 3.952.

Overall, misclassification arises less often when the number of true groups is smaller. Intuitively, this is because when we have fewer groups given the same number of bidders, we can use more testable implications to classify the bidders more accurately. This is confirmed by the better performance of results in Table 4 than in Table 5.

6.2. Two-Step Estimation in a Structural Model

6.2.1. **The Simulation Design**

In this simulation study, we estimate a simple model of a procurement market. Our goal is to investigate the impact of classification errors in the first step on the estimation of structural parameters in a second step. The population is represented by the set N of providers (bidders). Depending on the distribution of private costs, a provider belongs to one of K_0 types. Let N_k denote the set of providers in N with type $k \in \{1, 2, ..., K_0\}$, and let $|N_k| \equiv n_k$ denote its cardinality. For a provider i with type $\tau(i) \in \{1, 2, ..., K_0\}$, his costs are given by

$$c_{i,l} = \mu_{\tau(i)} + \epsilon_{i,l},$$

where $\epsilon_{i,l}$ follows $N(0, \sigma)$ with the support $[\underline{c}, \bar{c}]$.[27]

An auction game is structured as follows. First, an auction l is announced. The set of auction participants is determined in two steps: (a) two out of K_0 groups, $\tau_{l,1}$ and $\tau_{l,2}$, are chosen at random; then (b) n_1 and n_2 providers are randomly drawn from the corresponding groups $N_{\tau_{l,1}}$ and $N_{\tau_{l,2}}$. Then participants draw costs realizations from their corresponding distributions, and construct their bids. The participant with the lowest bid wins. The identities of participants, their bids and the identity of the winner are reported in the data.

For the simulations, we consider two specifications, (S_1) with $K_0 = 4$, $|N_k| = 4$ for all $k = 1, ..., K_0$, and (S_2) with $K_0 = 4$, $|N_k| = 10$ for all $k = 1, ..., K_0$. For both specifications, we set $\mu = (2,\ 2.4,\ 2.8,\ 3.2)$ and $\sigma = 0.5$. We run the following four experiments with different specification and sample sizes: (A) S_1, $L = 200$; (B) S_2, $L = 200$; (C) S_1, $L = 400$; and (D) S_2, $L = 400$. We additionally set $n_1 = 1$ and $n_2 = 1$.

Structural parameters estimated from the simulated data are: $K_0,\ \tau(.),\ \theta = \{(\mu_k)_{k=1}^{K_0},\ \sigma\}$. The estimation consists of two steps. In the first step our classification procedure is implemented. Here, the estimates of the grouping ($\hat{\tau}(.)$ and $\hat{K}$) are recovered. In the second step, a GMM procedure is used to recover the remaining parameters. The standard errors are computed using the analytic expression for the variance-covariance matrix of the estimator's asymptotic distribution.

We use the following moments are used: for $k = 1, ..., K_0$,

1. The within-group mean of bids: $\sum_{i=1}^{n} \mathbf{E}[B_{i,l} - \mu_{B,k}(\theta;\ I)]1\{\tau(i) = k\} = 0$.
2. The within-group second moment of bids: $\sum_{i=1}^{n} \mathbf{E}[B_{i,l}^2 - (\mu_{B,k}(\theta;\ I)^2 + \sigma_{B,k}(\theta;\ I)^2)]1\{\tau(i) = k\} = 0$,

where $\mu_{B,k}(\theta;\ I)$ and $\sigma_{B,k}(\theta;\ I)$ denote the mean and standard deviation of the equilibrium bid distribution for the bidders from group k, the parameter vector θ and the set of auction participants summarized by I.

[27]For the upper and lower bounds of the cost, we set

$$\underline{c} = \frac{1}{K_0} \sum_k (\mu_k - 1.96 \times \sigma),\ \text{and}\ \bar{c} = \frac{1}{K_0} \sum_k (\mu_k + 1.96 \times \sigma).$$

We have set the true parameters so that $\underline{c}$ is above zero.

Table 6: Simulation Results from Specifications A and B

(Specification A: $K_0 = 4$, $n_k = 4$, and $L = 200$)

	μ_1	μ_2	μ_3	μ_4	σ
Using True Groups					
Rej. Prob.	0.0150	0.0515	0.0523	0.0546	0.0149
Bias	-0.0189	-0.0252	-0.0610	-0.0511	0.0242
MSE	0.0005	0.0008	0.0035	0.0039	0.0039
Using Estimated Groups					
Rej. Prob.	0.0148	0.0542	0.0510	0.0485	0.0151
Bias	0.0059	0.0329	-0.0241	-0.0225	-0.0549
MSE	0.0041	0.0083	0.0035	0.0027	0.0383

(Specification B: $K_0 = 4$, $n_k = 10$, and $L = 200$)

	μ_1	μ_2	μ_3	μ_4	σ
Using True Groups					
Rej. Prob.	0.0120	0.0515	0.0512	0.0514	0.0111
Bias	-0.0211	-0.0233	-0.0621	-0.0622	0.0236
MSE	0.0005	0.0007	0.0039	0.0039	0.0034
Using Estimated Groups					
Rej. Prob.	0.0131	0.0550	0.0540	0.0530	0.0160
Bias	-0.0213	-0.0218	-0.0763	-0.0765	0.0211
MSE	0.0004	0.0015	0.0411	0.0441	0.0023

Note: n_k is the number of the bidders in group k, and L is the number of the markets. The rejection probabilities are from t-tests for the individual parameters. The nominal rejection probability is set to be 0.05. Among other things, note that the rejection probabilities using the true groups and those using the estimated groups are very similar. This shows that the first step estimation error of the classification does not play a major role in determining the finite sample performance of the second step estimator.

To compute $\mu_{B,k}(\theta_0;\, I)$ and $\sigma_{B,k}(\theta_0;\, I)$ for a given vector of trial parameter values (θ_0) and given a profile of the types of auction participants, $I = (\tau_{l,1},\ \tau_{l,2},\ n_1,\ n_2)$ we simulate the equilibrium bidding functions. Specifically, we use the analytical solution for the bidding functions when all auction participants belong to the same group and a modified version of the numerical method in Marshall, Meurer, Richard, and Stromquist (1994) to solve the system of differential equations which define the bidding strategies when multiple groups are present.[28] The bidding functions are then combined with the cost distributions implied by a vector of trial parameters, θ_0, to obtain the distribution of bids: $F_{B,k}(b|\,\theta_0,\, I) = F_{C,k}(\beta_k^{-1}(b)|\,\theta_0)$. We then compute the mean and the standard deviation of thus computed distribution of bids.

6.2.2. **Results**

Tables 6 and 7 report the bias and mean squared errors (MSEs) of two estimators for the structural parameters $(\mu_k)_{k=1}^{K_0}, \sigma$. The first is an "infeasible" estimator that uses the knowledge

[28]In estimation, we impose the sample version of the support constraint $\underline{c} = \max\{0, \frac{1}{K}\sum_k(\mu_k - 1.96 \times \hat{\sigma}),$ $\bar{c} = \frac{1}{K}\sum_k(\mu_k + 1.96 \times \hat{\sigma})$.

Table 7: Simulation Results from Specifications C and D

(Specification C: $K_0 = 4$, $n_k = 4$, and $L = 400$)

	μ_1	μ_2	μ_3	μ_4	σ
Using True Groups					
Rej. Prob.	0.0149	0.0500	0.0513	0.0520	0.0149
Bias	-0.0214	-0.0222	-0.0615	-0.0713	0.0237
MSE	0.0005	0.0007	0.0039	0.0039	0.0006
Using Estimated Groups					
Rej. Prob.	0.0151	0.0485	0.0506	0.0505	0.0151
Bias	0.0131	0.0412	-0.0615	-0.0556	-0.0211
MSE	0.0060	0.0008	0.0053	0.0031	0.0054

(Specification D: $K_0 = 4$, $n_k = 10$, and $L = 400$)

	μ_1	μ_2	μ_3	μ_4	σ
Using True Groups					
Rej. Prob.	0.0133	0.0480	0.0510	0.0520	0.0108
Bias	-0.0227	-0.0219	-0.0611	-0.0231	0.0236
MSE	0.0005	0.0007	0.0039	0.0039	0.0006
Using Estimated Groups					
Rej. Prob.	0.0126	0.0468	0.0520	0.0520	0.0128
Bias	-0.0229	-0.0123	-0.0761	-0.0361	0.0098
MSE	0.0093	0.0056	0.0068	0.0061	0.0001

Note: As in Table 6, n_k is the number of the bidders in group k and L the number of the markets and the nominal rejection probability is set to be 0.05. As compared with Table 6, the results in Table 7 show that the role of the estimation error of the estimated group structure is very small. This confirms that with a larger number of the markets, our classification method performs better given the same number of within-group bidders.

of the true group structure. The second is the two-step estimator we propose, which requires bidder classification in the first step. These two tables also report the rejection probabilities from t-tests of individual parameters.

Table 6 contains the results for a smaller sample size $L = 200$. It shows that the rejection probabilities are close to the nominal rejection rate 0.05, except for parameters μ_1 and σ. For both parameters, inference turns out to be conservative. However, in terms of MSE and Bias, the performance of estimators for these two parameters is not substantially different from that for the other parameters. We conclude that the asymptotic inference of the model works reasonably well in finite samples.

Next, we compare the performance of the estimator using the true group structure and the two-step version using the estimated group structure. Table 6 suggests the rejection probabilities are mostly similar between the two estimators. There is some minor difference in the MSE of some group means. The discrepancy seems more prominent when the size of each group is increased from $n_k = 4$ to $n_k = 10$. Nevertheless the impact of first-stage classification errors appears to be reasonably small. It is also remarkable that both the infeasible and the two-step version of the estimators of μ_1 and σ are conservative in terms of finite sample rejection probabilities. This is true regardless of whether n_k is equal to 4 or 10. Overall, Table 6 shows that the classification errors in the first-step do not have any major impact on the finite sample performance of the two-step estimator.

Table 7 reports results for larger samples with $L = 400$. The performance of the estimators improves slightly relative to the case with $L = 200$. For example, the rejection probabilities in the t-test for μ_2, μ_3, and μ_4 are closer to the nominal level 0.05. The rejection probabilities for the two sets of estimators are also quite similar. This confirms our conclusion that the classification errors in the first step do not aggravate the performance of the two-step estimators in experiments.

In summary, it appears that our classification method seems to work quite well in the environment of two-step estimators, especially when the size of each group is not too large relative to the sample size.

7. Empirical Application: California Market for Highway Procurement

We apply our methodology to data on highway procurement auctions conducted by the California Department of Transportation (CalTrans). Our goal is to demonstrate the performance of our method in the empirical setting, and to highlight the consequences of ignoring agent unobserved heterogeneity in estimation.

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.

As in other auction markets, the costs of contractors vary across firms and across projects. The firms' costs are private information and are summarized by the distribution from which the costs of contractors are drawn. The distribution of costs for a given project may differ across contractors on the basis of their observable characteristics (for example, the distance from contractor's location to the project site). However, some of the pertinent characteristics may be unobservable to the researcher.

The most straightforward way to account for possible (unobserved) 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 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, Perrigne, 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 afore-mentioned 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.

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 procedure to allocate regular firms into groups characterized by the same cost distribution. This step reduces the number of possible types (relative to the case when each firm is a separate type) 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.[29] 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 bidding decisions, thus, generating endogeneity of the competitive structure of a given auction. We take this feature into account in our analysis.

7.1. Model

We denote the population of regular firms operating in California procurement market by N. Each project l auctioned in this market is summarized by a set of characteristics X_l which are observable to a researcher and an unobservable factor U_l. The project is associated with a set of potential bidders, S_l. We assume that S_l 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 σ_U.

A contractor 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, are independent of U_l, and are distributed according to the exponential distribution with a rate parameter $\lambda_{\kappa,i,l}$. The costs of completing the work are drawn from a Lognormal distribution with mean $\mu_{C,i,l}$ and standard deviation σ_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

[29] 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.

(invariant across projects) 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. As in our basic set up this cost factor induces partitioning of the population of firms into the groups: $N = \cup_{k=1,\ldots,K_0} N_k$ with $N_k = \{i :\ q_i = \bar{q}_k\}$ and so that $\tau(i) = k$ if and only if $i \in N_k$. We also 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 $\bar{q}_0$.

Reflecting these features, the mean of the costs distribution and the parameter of the entry distribution are given by

$$\begin{aligned} \mu_{C,i,l} &= X_l\alpha_1 + D_{i,l}\alpha_2 + \sum_{k=1}^{K_0} \bar{q}_k 1\{\tau(i) = k\} + U_l \text{ and} \\ \lambda_{\kappa,i,l} &= X_l\gamma_1 + \sum_{k=1}^{k=K} \tilde{q}_k 1\{\tau(i) = k\}. \end{aligned}$$

Note that the groups capture differences in the contractors' cost efficiencies related to the project work. While entry costs may also vary across groups, there is no reason for the group differences in project costs to coincide with the group differences in entry costs. That is why we explicitly distinguish between the parameters capturing the former ($\bar{q}_k$) and the latter ($\tilde{q}_k$) effects.

A potential bidder decides whether to enter the auction on the basis of information about the set of potential bidders, S_l, and the realization of $\kappa_{i,l}$. 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, A_l. An active bidder chooses a bid $B_{i,l}$ on the basis of information about the set $A_{i,l}$ and his private cost $C_{i,l}$ for completing the project which he observes upon entry.

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. $i, j \in S_l$ such that $q_i = q_j$, and $D_{i,l} = D_{j,l}$) adopt the same strategies. Specifically, we define bidders' type in auction l as $(d,\, k)$ if $D_{i,l} = d$ and $q_i = \bar{q}_k$, where d is a discretized measure of the distance between the contractor and the location of the project.

For briefness we suppress the auction subscript l in notation from now till the end of the section. Thus, an equilibrium of an auction game for project l is characterized by a set of equilibrium entry and bidding strategies: $\{\sigma_i^E(\,.\,;\,(d,\,k)),\, \sigma_i^B(\,.\,;\,(d,\,k))\}_{(d,\,k)}$.For a given type $(d,\, k)$, realized cost c and composition of the set of entrants A, the bidding strategy of an entrant maximizes expected profit from bidding. That is,

$$\begin{aligned}\sigma^B(c,\, I_A;\, (d,\, k)) &= \arg\max \pi_{(d,\, k)}(b,\, c,\, I_A;\, \sigma^B_{-i}), \text{ where} \\ \pi_{(d,\, k)}(b,\, c,\, I_A;\, \sigma^B_{-i}) &= (b-c)P\left\{i \textit{ wins} \mid b,\, I_A,\, \text{“}i \textit{ is type-}(d,\, k)\text{”};\, \sigma^B_{-i}\right.,\end{aligned}$$

and σ^B_{-i} denotes the profile of the other bidders' strategies that they would use should they become active in a given project and I_A summarizes information about the composition of the set of entrants; similarly I_S below summarizes information about the set of potential bidders.[30]

A set of participation thresholds $\{\bar{K}_{(d,k)}(I_S)\}_{(d,k)}$ with $\bar{K}_{(d,k)}(I_S) = \mathbf{E}[\pi_{(d,k)} \mid I_S]$ characterizes the equilibrium entry strategies. In the last expression, the expectation is taken over the distribution of C_i and I_A conditional on a potential bidder's information set prior to the entry decisions I_S (and of course on X and U, which are suppressed in notation). This implies that the equilibrium probability of participation of the bidders of type $(d,\, k)$ is given by

$$p_{(d,k)} = \Pr(\kappa_{i,l} \le \mathbf{E}[\pi_{(d,k)} \mid I_S]).$$

7.2. Estimation Details

We begin by recovering the group structure of the population consisting of the regular participants of procurement market (the number of groups, K_0, and the membership function $\tau(.)$). After that we use a GMM procedure to estimate the remaining structural parameters of the model,

$$\theta_1 = (\alpha_1, \alpha_2, \gamma_1, \gamma_2, \sigma_C, \sigma_U, \{\bar{q}_k\}_{k=1,..,K_0}, \{\tilde{q}_k\}_{k=1,..,K_0}).$$

In the first step we use the pairwise comparison indexes derived in the second example of Section 3.2 to recover the unobserved group structure. Specifically, in accordance with the notation used in the paper, we define $\delta^+_{ij} \equiv \int \max\{r_{ij}(b), 0\}\, db$; $\delta^-_{ij} \equiv \int \max\{r_{ji}(b), 0\}\, db$ and $\delta^0_{ij} \equiv \int |r_{ij}(b)| db$ with $r_{ij}(b) = G_{ji}(b|\, d) - G_{ij}(b|d)$ and $G_{ij}(b|d) = \Pr(B_{i,l} \ge b|\, D_{i,l} = d,\, D_{j,l} = d,\, i \in A_l,\, j \in A_l)$.[31]

We implement classification using the bootstrap testing procedure described previously. We recover group structure on the basis of the indexes which aggregate over the values 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.

In the second step, we estimate the parameters of the model by a GMM procedure while imposing the recovered group structure in estimation (the number of groups, $\hat{K}_0$, and the membership function $\hat{\tau}(.)$). Since the participation stage of our game may potentially generate multiple equilibria, we do not explicitly solve this part of the game. Instead, we discretize the

[30]Specifically, $I_{S,l}$, and $I_{A,l}$ contain information on the number of potential and active bidders from each (d,k)-group.

[31]We obtain empirical counterparts of these indexes by replacing $G_{ij}(b|d)$ with $\hat{G}_{ij}(b|d)$ obtained as follows

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

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

support of auction characteristics (X_l, U_l) and treat the probabilities of participation for bidders of various $(k,d)-$types corresponding to these grid values, $p_{(k,d)}(x_m, u_m, \bar{I}_{S,p})$, as auxiliary parameters. We follow the spirit of Dube, Fox, and Su (2012) by maximizing a moment-based objective function subject to the constraints that the optimality of the participation strategies is satisfied on the grid of the project characteristics' values.[32]

We consider the following moments: (a) the first and the second moment of bid distribution for a given level of d and for a given group of bidders; (b) the covariance between bids and the observable project characteristics; (c) the covariance between any two bids submitted in the same auction; (d) the expected number of participants in any given auction for every $(d,\, q)$-group; (e) the covariance between the number of participants and the 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 below.

Formally, let $\mu_{B,k,d}(\theta_1)$, $\sigma_{B,k,d}(\theta_1)$ denote the mean and the standard deviation of the distribution of bids submitted by bidders from group k for the projects which are distance d away when the vector of structural parameters takes value θ_1; $S_{l,k,d}$ denote the subset of potential bidders for project l who are from group k and are at distance d from project l; and $|S_{l,k,d}|$ denote the number of the bidders in the set $S_{l,k,d}$. Then we form moment conditions as follows.

$$(a_1)\ \mathbf{E}[(B_{i,l} - \mu_{B,k,d}(\theta_1)) \times 1\{\tau(i) = k\} \times 1\{d_{i,l} = d\}] = 0$$

$$(a_2)\ \mathbf{E}[(B_{i,l}^2 - (\sigma_{B,k,d}(\theta_1)^2 + \mu_{B,k,d}(\theta_1)^2)) \times 1\{\tau(i) = k\} \times 1\{d_{i,l} = d\}] = 0$$

$$(b)\ \mathbf{E}[B_{i,l}X_{r,l} - \alpha_{1,r}\sigma_X^2] = 0 \text{ where } X_{r,l} \text{ is the r-th component of } X_{r,t}$$

$$(c)\ \mathbf{E}[(B_{i,l} - \mu_{B,\tau(i),d}(\theta_1))(B_{j,l} - \mu_{B,\tau(j),d}(\theta_1)) - \sigma_U^2] = 0$$

$$(d)\ \mathbf{E}[1\{i \in A_l\}1\{\tau(i) = k\}1\{d_{i,l} = d\} - p_{(k,d)}(X_l, U_l, I_{S_l})|N_{l,k,d}|] = 0$$

$$(e)\ \mathbf{E}[(1\{i \in A_l\}1\{\tau(i) = k\}1\{d_{i,l} = d\} - p_{(k,d)}(X_l, U_l, I_{S_l})|N_{l,k,d}|)X_l] = 0.$$

When computing empirical counterparts of these moments in the second step of our estimation procedure we impose the group structure recovered in the first step by replacing $\tau(i)$ with $\hat{\tau}(i)$.

In order to construct these moments we need to map the distributions of the bidders' costs summarized by the parameters vector θ_1 into the corresponding distributions of bids. For this, we solve for equilibrium bidding strategies corresponding to the costs distribution summarized by θ_1 using an extension of the algorithm proposed in Marshall, Meurer, Richard, and Stromquist (1994). We then combine the computed bidding functions with the distribution of costs to compute the moments of the bid distribution. We also use so computed bidding strategies to obtain the values of expected profits we use in the participation constraints.[33]

[32]Specifically, for a grid of (X_l, U_l)-values we impose that

$$p_{(k,d)}(x_{m_1}, u_{m_2}, \bar{I}_{S,p}) = F_\kappa(\mathbf{E}[\pi_{(k,d)}(x_{m_1}, u_{m_2}, \bar{I}_{S,p}; \theta_1)]).$$

This procedure allows us to select participation strategies that are most consistent with the data.

[33]When computing the moments, the values of $p_{(k,d)}(X_l, U_l, I_{S_l})$ between grid points are interpolated using cubic spline.

7.3. 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, the location of the project, category of work, and list of potential bidders. We construct the distance variable to reflect the expected driving distance between the project location and the closest company plant. The projects in our sample are worth $523,000 and last for around three months on average; 38% of these projects are partially supported through federal funds. There are 25 firms that participate regularly in this market. The market also attracts fringe bidders that enter only very few auctions throughout the data. An average auction attracts six regular potential bidders and eight fringe bidders. Since only a fraction of potential bidders submits bids, an entry decision plays an important role in this market. Finally, the distance to the company location varies quite a bit and is around 28 miles on average for regular potential bidders. A table of these statistics can be found in the Online Appendix.

In the first step, we obtain through our classification method the grouping of the bidders into eight groups that consist of 2, 3, 8, 3, 2, 3, 2 and 2 bidders respectively.[34] 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 first two columns present the estimates which are obtained when the unobserved group structure is taken into account in the estimation. The results indicate significant differences in bidders' costs across the groups. Specifically, fringe bidders (the reference group) tend to have the highest costs whereas the difference in costs between the group of fringe bidders and the groups of regular bidders is comparable in impact to the shortening of the distance to the project site by 42.5 (i.e., by 0.051/0.0012), 10.1, 26.67, 48.33, 11.67, 6.67, 7.5, and 41.67 miles respectively. Additionally, the distance increases project costs (additional 8.33 miles result in costs which are 1% higher on average);[35] bridge projects appear to be about 1% cheaper on average, the projects which attract federal aid also tend to have costs which are on average 4% lower than projects which are financed through local funds. This effect probably arises because bidders anticipate a higher degree of scrutiny on these projects and thus make more effort to save costs. Unobserved project heterogeneity while non-negligible is moderate in size: increasing the value of the unobserved factor from its mean (equal to '0') to

[34]We identify two subsets of regular bidders that overlap by two bidders, and two separate non-overlapping subsets. These four subsets satisfy the condition of being fully connected (i.e. every pair of bidders from a given subset participate together in a large number of auctions). We estimate that the underlying group structure of the overlapping subsets consists of three and two groups correspondingly. The bidders that are common to the two subsets are estimated to belong to the same group in both cases. To avoid issues associated with such overlap we treat these two common bidders as a separate group. The non-overlapping subsets are estimated to consist of one group each. This obtains the group structure with eight groups.

[35]Recall that the coefficients reflect the impact on costs in terms of the fraction of the engineer's estimate. The distance resulting in 0.01 increase of average costs can thus be comuted as 0.01/0.012.

Table 8. Parameter Estimates

	Parameter	Std. Error	Parameter	Std. Error
The Distribution of Project Costs				
Constant ($\bar{q}_0$)	0.127***	(0.0129)	0.113***	(0.0119)
Eng. Estimate	-0.0004***	(0.0002)	-0.0005***	(0.0002)
Duration	0.00026*	(0.00036)	0.00022*	(0.00027)
Distance	0.0012***	(0.00022)	0.00086***	(0.00019)
Bridge	-0.0092***	(0.0018)	-0.012***	(0.0011)
Federal Aid	-0.043***	(0.0103)	-0.078***	(0.009)
Regular Bidder			-0.035***	(0.003)
$\bar{q}_1 - \bar{q}_0$	-0.051***	(0.008)		
$\bar{q}_2 - \bar{q}_0$	-0.012***	(0.005)		
$\bar{q}_3 - \bar{q}_0$	-0.032***	(0.009)		
$\bar{q}_4 - \bar{q}_0$	-0.058***	(0.008)		
$\bar{q}_5 - \bar{q}_0$	-0.014***	(0.007)		
$\bar{q}_6 - \bar{q}_0$	-0.008***	(0.006)		
$\bar{q}_7 - \bar{q}_0$	-0.009***	(0.007)		
$\bar{q}_8 - \bar{q}_0$	-0.050***	(0.006)		
σ_C	0.087***	(0.032)	0.112***	(0.022)
σ_U	0.021***	(0.009)	0.0207***	(0.008)
The Distribution of Entry Costs				
Constant ($\tilde{q}_0$)	-0.0114*	(0.0078)	-0.0161*	0.0091
Eng. Estimate	0.0055***	(0.0016)	0.0051***	(0.0012)
Number of Items	0.0018*	(0.0011)	0.0011***	(0.0005)
Regular Bidder			-0.022 ***	(0.004)
$\tilde{q}_1 - \tilde{q}_0$	-0.019***	(0.005)		
$\tilde{q}_2 - \tilde{q}_0$	-0.018***	(0.007)		
$\tilde{q}_3 - \tilde{q}_0$	-0.016***	(0.007)		
$\tilde{q}_4 - \tilde{q}_0$	-0.024***	(0.006)		
$\tilde{q}_5 - \tilde{q}_0$	-0.022***	(0.008)		
$\tilde{q}_6 - \tilde{q}_0$	-0.018***	(0.006)		
$\tilde{q}_7 - \tilde{q}_0$	-0.017***	(0.008)		
$\tilde{q}_8 - \tilde{q}_0$	-0.019***	(0.008)		

Note: In the results above the distance is measured in miles. The fringe bidders are the reference group; the impact of fringe status on costs is summarized by a constant. The first two columns reflects the estimates reflect the specification which allows for the possibility of unobserved bidder heterogeneity. The last two columns reflects the specification which imposes that there no unobserved bidder heterogeneity is present in the data.

a value that corresponds to one standard deviation from the mean is equivalent to increasing the distance to the project site by 30 miles. The entry costs of regular bidders are significantly lower than entry costs of fringe bidders. However, they appear to be quite similar across the groups of regular bidders.

The last two columns of Table 8 show the parameter estimates under the specification when the unobserved group structure of the regular bidders is ignored in the estimation. The parameter estimates are obtained by the GMM estimation procedure using the same set of moments by imposing that only two groups of sellers are present in the data: fringe and regular bidders. Under this specification, the cost reduction due to the federal aid is estimated to be much higher (7.8% rather than 4.3%), the impact of the distance is estimated to be lower (the distance to the project has to be 11.67 miles higher in order to increase the average cost by 1%), and the variance of the cost distribution is estimated to be higher. Additionally, the entry costs are estimated to be lower relative to the baseline specification.

The differences between the results obtained under the two specification are primarily driven by endogenous participation behavior. The model without unobserved bidder heterogeneity treats all the regular bidders as homogeneous. As a result it only rationalizes aggregate participation behavior. The model with unobserved bidder heterogeneity takes into account differences in the participation behavior of a regular bidder. For example, in the data projects receiving federal aid tend to attract regular bidders from groups 1, 4 and 8 more frequently relative to bidders from groups 5, 6 and 7. When heterogeneity is ignored, the model has to rationalize participation behavior by forcing the costs on such projects to be lower than they are in reality to rationalize such observed behavior. Similar reasoning explains the difference in the estimated impact of distance under the two specifications.

Our results thus confirm that regular participants in the highway procurement market are characterized by important unobserved cost differences that persist throughout the data. As we noted above documenting such cost asymmetries may be important from a purely informative point of view. It is also important to correctly measure cost asymmetries from a normative point of view. 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.[36]

8. 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 pairwise 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. Finally, we apply this method to data

[36]For example, the affirmative action programs in government procurement have been largely implemented in the form of discriminating 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 discrimination factor correctly takes into account existing costs asymmetries then such 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).

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.

This method is also likely to be useful in the multi-agent settings characterized by agents' sparse commonality. In such environments only a small subset of agents is present in any given market. As a result, the conditional probability of an agent choosing a specific action given a full set of participants cannot be consistently estimated from the data. Pairwise comparisons could be used to recover the group structure which in turn may serve as a basis for data aggregation which enables accurate inference.

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 mixture approach has lower data requirements but requires stronger assumptions in order to achieve identification of model primitives and can be more computationally costly in settings with strategic interdependence.

References

ABOWD, J., F. KRAMARZ, AND D. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–334.

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.

ARADILLAS-LOPEZ, A., A. GANDHI, AND D. QUINT (2013): "Identification and Inference in Ascending Auctions with Correlated Private Values," *Econometrica*, 81(2), 489–534.

ASKER, J. (2010): "A Study of the Internal Organisation of a Bidding Cartel," *American Economic Review*, 100, 724–762.

ATHEY, S., AND P. HAILE (2007): "Nonparametric Approaches to Auctions," in *Handbook of Econometrics, Vol. 6*, ed. by J. J. Heckman, and E. E. Leamer, chap. 60, pp. 3847–3959. Elsevier, Amsterdam.

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.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

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.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

CONLEY, T., AND F. DECAROLIS (2016): "Detecting Bidder Groups in Collusive Auctions," *American Economic Journal: Microeconomics*, 8(2), 1–38.

DELOECKER, J., AND F. WARZYNSKI (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102(6), 2437–2471.

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.

EECKHOUT, J., AND P. KIRCHER (2011): "Identifying Sorting - In Theory," *Review of Economic Studies*, 78(3), 872–906.

GENTRY, M., T. KOMAROVA, AND P. SHIRALDI (2016): "Preferences and Perfomance in Simulateneous First Price Auctions: Structural Analysis," Discussion paper.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.

HAGEDORN, M., T. LAW, AND I. MANOVSKII (2016): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning Second Edition*. Springer.

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.

HU, Y., AND S. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

HU, Y., AND J. SHIU (2013): "Nonparametric identification of dynamic models with unobserved covariates," *Journal of Econometrics*, 175(2), 116–131.

HU, Y., AND M. SHUM (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171(1), 32–44.

JOFRE-BONET, M., AND M. PESENDORFER (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71(5), 1443–1489.

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 (2016): "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.

LENTZ, R., AND D. MORTENSEN (2010): "Labor Market Model of Worker and Firm Heterogeneity," *Annual Review of Economics*, 2(1), 577–603.

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.

LISE, J., C. MEGHIR, AND J.-M. ROBIN (2011): "Matching, Sortin and Wages," Discussion paper.

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.

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.

ROBERTS, J., AND A. SWEETING (2013): "When Should Sellers Use Auctions," *American Economic Review*.

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

SU, L., Z. SHI, AND P. PHILLIPS (2016): "Identifying Latent Structures in Panel Data," *Econometrica*, 84(6), 2215–2264.

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

Appendix A: Mathemathical Proofs

Proof of Theorem 4.1: Sufficiency is obvious. We focus on necessity. First consider the two facts:

Fact 1: If G_τ does not contain a monotone path of length $K_0 - 1$, τ is not identified.
Fact 2: A vertex i is identified if and only if there is a monotone path P containing i such that its end vertices i_H and i_L are identified and

$$\tau(i_H) - \tau(i_L) = \ell(P),$$

where $\ell(P)$ denotes the length of P.

By Fact 1, the necessity of G_τ containing a monotone path of length $K_0 - 1$ follows, and Fact 2 completes the proof of the necessity part of the theorem.

Now let us prove Fact 1. Suppose that G_τ does not contain a monotone path of length $K_0 - 1$. Let N_{max} be the set of vertices such that for each vertex i in N_{max}, all his G_τ-neighbors have lower type than the vertex i. Then there is no edge in G_τ which joins any two vertices from the set N_{max}. Choose a vertex i^* from N_{max} which is an end vertex of a longest monotone path, say, with length $K - 1 < K_0 - 1$. This identifies a lower bound for K_0 but there is no upper bound for K_0 that we can obtain from G_τ. Take any τ' such that $\tau'(i^*) > \tau(i^*)$ and $\tau'(i) = \tau(i)$ for all $i \in N \setminus \{i^*\}$. Then τ' is compatible with G_τ and the given comparison indexes, proving that τ is not identified from G.

Let us prove Fact 2. Sufficiency is trivial. Let us focus on necessity. Suppose there is no monotone path with identified end vertices which contains i. Then i is not identified. Therefore it is necessary that there exists a monotone path with identified end vertices which contains i. So it suffices to show that it is necessary that such a monotone path has to have length equal to $\tau(i_H) - \tau(i_L)$. Suppose to the contrary that every monotone path P that contains i and has identified end vertices i_H and i_L also satisfies $\tau(i_H) - \tau(i_L) > \ell(P)$. Then we will show that i is not identified.

First, assume that there exists a monotone path which contains i but not as one of its end vertices. Let i_H^* be a lowest type vertex among all the identified vertices each of which is on a monotone path that contains i and is of higher type than i. Also, let i_L^* be a highest type vertex among all the identified vertices each of which is on a monotone path that contains i and is of lower type than i. Let P be a monotone path between i_H^* and i_L^* that passes through i. Then by construction, the type difference $\tau(i_H^*) - \tau(i_L^*)$ between the two end vertices is smallest among all the monotone paths that go through i. By the condition, we have $\tau(i_H^*) - \tau(i_L^*) > 2$. Therefore, we have multiple different ways to assign $\tau(i_H^*) - 1, \tau(i_H^*) - 2, ..., \tau(i_L^*) + 1$ to the vertex i on the path P. Hence i is not identified.

Second, assume that all the monotone paths that contain i have i as one of their end vertices. Then either all neighbors of i are of higher type than i or all neighbors of i are of lower type than i. Suppose that we are in the former case. (The latter case can be dealt with similarly.) Let i_H^* be a lowest type vertex among all the vertices each of which is on a monotone path that contains i and is of higher type than i. Then by the condition, $\tau(i_H^*) - \tau(i) > 1$. Thus we have multiple different ways to assign $\tau(i_H^*) - 1, \tau(i_H^*) - 2, ..., \tau(i) + 1, \tau(i)$ to the vertex i. Hence i is not identified. ■

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 5.1: Let us first consider the case where $K_0 = 2$. For each $i \in N$, we let $N(i) = N\backslash\{i\}$, and let $\hat{N}_1(i)$ and $\hat{N}_2(i)$ be as defined in Step 1 of the two step algorithm. Also, we define

$$\begin{aligned} N_1^*(i) &\equiv \{j \in N(i) : \tau(i) > \tau(j)\}, \text{ and} \\ N_2^*(i) &\equiv \{j \in N(i) : \tau(i) < \tau(j)\}. \end{aligned}$$

If $i \in N_h$, we have $N_l = N_1^*(i)$ and $N_h = N\backslash N_1^*(i)$. Also, if $i \in N_l$, we have $N_l = N\backslash N_2^*(i)$ and $N_h = N_2^*(i)$. By Assumption 5.1, for any $i \in N_h$, we have

$$P\left\{\min_{j \in N_1^*(i)} \hat{p}_{ij}^- \leq 0\right\} = o(1/n), \text{ and } P\left\{\min_{j \in N\backslash N_1^*(i)} \hat{p}_{ij}^+ \leq 0\right\} = o(1/n),$$

and hence

$$P\left\{\max_{j \in N_1^*(i)} (\hat{p}_{ij}^+ - \hat{p}_{ij}^-) < 0\right\} \geq P\left\{\max_{j \in N_1^*(i)} \hat{p}_{ij}^+ < 0\right\} - o(1/n) = 1 - o(1/n),$$

as $n, L \to \infty$, and

$$P\left\{\max_{j \in N\backslash N_1^*(i)} (\hat{p}_{ij}^- - \hat{p}_{ij}^+) \leq 0\right\} \geq \left\{\max_{j \in N\backslash N_1^*(i)} \hat{p}_{ij}^- \leq 0\right\} - o(1/n) = 1 - o(1/n),$$

as $n, L \to \infty$. Therefore, whenever $i \in N_h$,

$$P\{N_l \subset \hat{N}_1(i)\} = 1 - o(1/n), \text{ and } P\{N_h \subset N\backslash\hat{N}_1(i)\} = 1 - o(1/n).$$

Since $(\hat{N}_1(i), N\backslash\hat{N}_1(i))$ and (N_l, N_h) are partitions of N, this also implies that whenever $i \in N_h$,

$$P\{N_l = \hat{N}_1(i), \text{ and } N_h = N\backslash\hat{N}_1(i)\} = 1 - o(1/n),$$

and similarly, whenever $i \in N_l$,

$$P\{N_l = N\backslash\hat{N}_2(i), \text{ and } N_h = \hat{N}_2(i)\} = 1 - o(1/n).$$

Hence we have for any $i \in N$, $r \in \{1, 2\}$,

$$P\left\{\hat{T}_r(i) \neq T\right\} = o(1/n), \tag{8.1}$$

which implies that

$$P\left\{\hat{T}_r(i) \neq T \text{ for some } i \in N\right\} \leq \sum_{i \in N} P\left\{\hat{T}_r(i) \neq T\right\} \to 0.$$

Since $i^* \in N$ and $\hat{T} = \hat{T}_1(i^*)$ or $\hat{T} = \hat{T}_2(i^*)$, the probability $P\{\delta(\hat{T}, T) > 0\}$ is bounded by the right hand side sum. We call such agent i^* used in the split a *pivotal agent*. But this sum converges to zero, yielding the desired result consistency result when $K_0 = 2$.

Now, suppose that K_0 is allowed to be greater than 2. Take $N^{[0]} = N$. For each $i \in N$, we obtain two ordered partitions $\hat{T}_1^{[1]}(i) \equiv (N \backslash \hat{N}_2(i), \hat{N}_2(i)) \equiv (\hat{N}_{1,1}^{[0]}(i), \hat{N}_{2,1}^{[0]}(i))$ and $\hat{T}_2^{[1]}(i) \equiv (\hat{N}_1(i), N \backslash \hat{N}_1(i)) \equiv (\hat{N}_{1,2}^{[0]}(i), \hat{N}_{2,2}^{[0]}(i))$. Define $E_{n,1}^{[1]}(i)$ to be the event that $\hat{N}_{1,2}^{[0]}(i) = N_{1,2}^{[0]}(i)$ and define $E_{n,2}^{[1]}(i)$ to be event that $\hat{N}_{2,1}^{[0]}(i) = N_{2,1}^{[0]}(i)$, where for each $r = 1, ..., K$, if $i \in N_r$,

$$\begin{aligned} N_{2,1}^{[0]}(i) &= N_{r+1} \cup ... \cup N_K \text{ and} \\ N_{1,2}^{[0]}(i) &= N_1 \cup ... \cup N_r. \end{aligned}$$

Then by following the same arguments as before, we find that

$$P\left\{\bigcup_{i \in N} \left(E_{n,1}^{[1]}(i) \cup E_{n,2}^{[1]}(i)\right)\right\} \to 1. \tag{8.2}$$

Now, generally, suppose that at Step $k-1 \leq K$, we have obtained the estimated ordered partition $(\hat{N}_r^{[k]})_{r=1}^k$. 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

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

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

$$P\left\{\bigcup_{d_{k-1} \in \mathcal{D}_{k-1}} \bigcup_{R_k \subset \{1,...,k\}} A_k(R_k, d_{k-1})\right\} \to 1. \tag{8.3}$$

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 $\hat{N}^{[k-1]} = N \backslash (\cup_{r \in R_k} \hat{N}_r)$ and $N^{[k-1]} = N \backslash (\cup_{r \in R_k} N_r)$. Then in the event $A_k(R_k, d_k)$, we also have $\hat{N}^{[k-1]} = N^{[k-1]}$ by the definition of $A_k(R_k, d_k)$. If $N^{[k-1]}$ is empty or contains i's with the same type, the event $A_k(R_k, d_k)$ is equal to the event $A_k(\{1, ..., k\})$ with $k = K$. 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 (8.3).

Suppose that $N^{[k-1]}$ 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 $\hat{N}^{[k-1]}$ such that $\hat{p}_{ij}^0$ is less than or equal to 0 with probability being $1 - o(1/n)$ by Assumption 5.1, whereas for all (i, j) such that $i, j \in N_r$ for some $r \in R_k$, $\hat{p}_{ij}^0$ is greater than 0 with probability being $1 - o(1/n)$. 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 $\hat{N}^{[k-1]}$ is $o(1/n)$.

We obtain two ordered partitions

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

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

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

Then again by following the same arguments as previously, we find that

$$P\left\{\bigcup_{i\in N}\left(E_{n,1}^{[k]}(i) \cup E_{n,2}^{[k]}(i)\right)\right\} \to 1. \tag{8.4}$$

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

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

and rename $\hat{N}_r^{[k+1]} = \hat{N}_r^{[k]}$ for all $r \in R_k$. Thus we have obtained the augmented partition $(\hat{N}_r^{[k+1]})_{r=1}^{k+1}$. 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 is clear that the convergence in (8.4), combined with (8.3), 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 the conditions of Theorem 5.2 hold..*
(i) *If* $K \geq K_0,$ *then* $\hat{V}(K) = o_P(r_{1,n})$*, as* $n, L \to \infty$*.*
(ii) *If* $K < K_0,$ *then for any* $M > 0$*, as* $n, L \to \infty$*,*

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

Proof: (i) From the proof of Theorem 5.1, for each $k = 1, ..., K$,

$$P\left\{N_k = \hat{N}_k\right\} \to 1.$$

(Here $N_k = \varnothing$ if $K > K_0$.) Therefore, by Assumption 5.2, we have

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

Thus (i) follows.

(ii) Suppose that $K < K_0$. Then for some $k = 1, ..., K$, and for some $i, j \in N_k$, we have $H_{1,ij}^0$ true. Since $g(L)/r_{2,n} \to 0$ as $n, L \to \infty$, we take $r_n \equiv g(L)$ to find that for this pair (i, j),

$\log \hat{p}^0_{ij}/g(L) \to_P -\infty$, as $n, L \to \infty$. Therefore, it follows that $\hat{V}(K)/g(L) \to_P \infty$ as $n, L \to \infty$. ■

Proof of Theorem 5.2: Choose K such that $K_0 < K$ and write

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

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

$$\hat{V}(K_0) - \hat{V}(K) = o_P(r_{1,n}),$$

by Lemma A1(i). Since $g(L)/r_{1,n} \to \infty$, 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 $\hat{V}(K_0) = o_P(r_{1,n})$. Therefore, taking $M > K_0 - K$, 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

$$P\{\hat{T}_{\hat{K}} \neq T\} = P\{\hat{T}_{K_0} \neq T\} + o(1).$$

That the last probability converges to zero follows precisely by the arguments in the proof of Theorem 5.1. ■

Appendix B: Discussion on Assumption 5.1

Suppose that $T^s_{ij,n}$ is the test statistic whose distribution function under the null hypothesis is twice continuously differentiable with bounded derivatives and denoted by $F^s_{ij,n}$. Assume also that the density $f^s_{ij,n}$ of $F^s_{ij,n}$ is bounded away from zero uniformly over i, j. A bootstrap distribution used to compute the critical value is denoted by $\tilde{F}^s_{ij,n}$ and the asymptotic distribution by $F^s_{ij,\infty}$. Suppose further that

$$\max_{ij \in \mathcal{E}} \sup_{t \in \mathbf{R}} |\tilde{F}^s_{ij,n}(t) - F^s_{ij,n}(t)| + \max_{ij \in \mathcal{E}} \sup_{t \in \mathbf{R}} |F^s_{ij,\infty}(t) - F^s_{ij,n}(t)| = O(L^{-h_1}),$$

for some $h_1 > 0$. Now, as for Assumption 5.1(i), note that for $\mathcal{E}' \subset \mathcal{E}$ and for each $i \in N$,

$$P_n\left\{\min_{j \in N: ij \in \mathcal{E}'} \hat{p}^s_{ij} \leq 0\right\} \leq P_n\left\{\max_{j \in N: ij \in \mathcal{E}'} U_{ij} \geq -M_n L^{-h_1} + 1\right\},$$

where $U_{ij} = F^s_{ij,n}(T^s_{ij,n})$ and M_n is any sequence increasing to infinity. Since U_{ij}'s follow the uniform distribution on $[0,1]$, using the Fréchet-Hoeffding lower bound for a copula, we can bound the above probability by

$$1 - \max\left\{1 - |\mathcal{E}'(i)| M_n L^{-h_1}, 0\right. \leq 1 - \max\left\{1 - n M_n L^{-h_1}, 0\right.,$$

where $|\mathcal{E}'(i)|$ denotes the number of agents j such that $ij \in \mathcal{E}'$. Therefore, it suffices that $L^{-h_1} = o(n^{-2})$.[37]

As for Assumption 5.1(ii), observe that for $\mathcal{E}' \subset \mathcal{E}$ and for each $i \in N$,

$$\begin{aligned} P_n\left\{\max_{j\in N: ij\in\mathcal{E}'} \hat{p}_{ij}^s > 0\right\} &\leq n \max_{j\in N: ij\in\mathcal{E}'} P_n\left\{\hat{p}_{ij}^s > 0\right. \\ &\leq n \max_{j\in N: ij\in\mathcal{E}'} P_n\{F_{ij,n}^s(T_{ij,n}^s) > -M_n L^{-h_1} + 1\}. \end{aligned}$$

Suppose that along P_n, the statistic $T_{ij,n}^s$ diverges to infinity in such a way that $T_{ij,n}^s/L^{h_2} \to c > 0$, as $L \to \infty$ for some $h_2 > 0$. Let $T_{ij,n}^{s,\circ}$ be the test statistic (under the null hypothesis) having the distribution as $F_{ij,n}^s$. Then we write the last probability as

$$\begin{aligned} & P_n\{F_{ij,n}^s(T_{ij,n}^s - T_{ij,n}^{s,\circ} + T_{ij,n}^{s,\circ}) - F_{ij,n}^s(T_{ij,n}^{s,\circ}) + F_{ij,n}^s(T_{ij,n}^{s,\circ}) > -M_n L^{-h_1} + 1\} \\ \leq \; & P_n\{f_{ij,\infty}^s(T_{ij,n}^{s,\circ}) c L^{h_2} + o(L^{h_2}) + F_{ij,n}^s(T_{ij,n}^{s,\circ}) > -M_n L^{-h_1} + 1\} \\ \leq \; & P_n\{c_1 L^{h_2} + o(L^{h_2}) + F_{ij,n}^s(T_{ij,n}^{s,\circ}) > -M_n L^{-h_1} + 1\}, \end{aligned}$$

for some constant $c_1 > 0$. The last probability becomes one from some large n, L on, as long as $h_2 > 0$.

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
E-mail address: xun.tang@rice.edu

[37]Note that this rate condition is obtained as conservative because the joint dependence of U_{ij}'s is allowed to be arbitrary. We can weaken this assumption, if U_{ij}'s exhibit a form of local dependence.