ECONOMETRIC INFERENCE ON LARGE BAYESIAN GAMES WITH HETEROGENEOUS BELIEFS

KYUNGCHUL SONG

Abstract. Most existing econometric models on games assume observation of many replications of a single small-player game. Such a framework is not adequate for a situation where one observes multiple heterogeneous many-player games, as in models of social interactions or heterogeneous complete information games. This paper focuses on a situation where the econometrician observes multiple heterogeneous many-player games. The model does not require a common prior assumption, and allows for the players to form beliefs or higher order beliefs about the other players' types and beliefs differently from each other. The main challenge for the econometrician is to recover the beliefs an individual player forms about other players' actions. By drawing on the main intuition of Kalai (2004), this paper introduces the notion of a hindsight regret which measures each player's ex post value of other players' type information, and establishes its belief-free bound. From this bound, this paper derives a set of moment inequalities associated with the incentive constraints of a Bayesian Nash equilibrium, and develops an asymptotic inference procedure for the structural parameters.

Key words. Large Game; Incomplete Information; Heterogenous Beliefs; Bayesian Nash Equilibria; Ex Post Stability; Hindsight Regrets; Cross-Sectional Dependence; Partial Identification; Moment Inequalities.

JEL Subject Classification. C13, C31.

1. Introduction

Many economic variables arise as a consequence of economic agents' rational decisions in various strategic environments. Education decisions by people, entries into a market by firms, or behavioral choices by students such as smoking, are made directly under the influence of others' choices. For an empirical researcher studying the way economic agents make decisions in such environments, the main challenge lies in the dilemma that the statistical inference based on the law of large numbers and the central limit theorem requires independence

Date: May 21, 2013.
I thank Andres Aradillas-Lopez, Aureo de Paula, Hiro Kasahara, Jinwoo Kim, Sokbae Lee, Wei Li, Vadim Marmer, and Mike Peters for useful conversations and comments, and seminar participants at Seoul National University, UBC, University of Texas at Dallas, and University of Wisconsin-Madison for their comments. I also thank Yoram Halevy, Wei Li, Qingmin Liu, and Mike Peters for their kind answers to my numerous elementary questions on Bayesian games. Last but not least, I thank Bruce Hansen and Joon Park for their encouragements at the early stage of this research. All errors are mine. I acknowledge that this research was supported by Social Sciences and Humanities Research Council of Canada. Corresponding address: Kyungchul Song, Department of Economics, University of British Columbia, Vancouver, BC, Canada. Email address: kysong@mail.ubc.ca.

among observed cross-sectional outcomes, whereas the focus of interest is on the nature of their dependence which potentially arises from their strategic interactions.

A strand of empirical models have been actively developed in response to this challenge by separating statistical independence and strategic interdependence through modeling explicitly the strategic environments. This strand of models have a common feature that the econometrician observes many replications of a single small representative game with i.i.d. draws of certain payoff or information components, so that statistical independence is imposed across the replications, whereas strategic interdependence is maintained within each replication. This separation enables one to deal with both asymmetric players within each game and heterogeneity across the games. (See Bresnahan and Reiss (1991), Tamer (2003), Ciliberto and Tamer (2009), Aradillas-Lopez (2010), Beresteanu, Molchanov, and Molinari (2011), Aradillas-Lopez and Tamer (2008), and de Paula and Tang (2011), among many others, for methodological contributions.[1])

However, there are many strategic interactions where such a nice separation is far from plausible. For example, consider studying peer effects among students. One may view strategic interactions among the students in each school as one game. Each school has a different number of students. In order to view the games as arising from a single representative game and to deal with heterogeneity in the number of the students, one needs to either ascribe it to various observable school characteristics, or assume simply that there exists a potential set of students from which students are randomly assigned to each school, or explicitly model endogeneity in the assignments of the students. Then one needs to specify how the equilibrium is selected across schools or whether the same equilibrium is selected across schools or not. In other words, one is forced to introduce various additional structures to transform the original strategic environment into one that fits the representative game framework. This observation applies to many other examples, such as studies on neighborhood effects on the choice of housing location or on the effect of friendship networks on the students' smoking behavior. Many such examples are found in the literature of social interactions. (See Brock and Durlauf (1995, 2001 and 2007) for their pioneering works on the structural modeling of social interactions. See also surveys by Brock and Durlauf (2001) and a recent monograph by Ioannides (2012) for this literature.)

This paper proposes an alternative modeling view in which heterogeneity across the games and heterogeneity across the players are given characteristics of a single large game. For example, in the study of the peer effects mentioned before, this framework views interactions

[1]Also see Chernozhukov, Hong and Tamer (2007), Rosen (2008), Bugni (2010), Andrews and Soares (2010), Andrews and Shi (2011), Chernozhukov, Lee and Rosen (2013), among many others, for general theory of inference for models under moment inequality restrictions. Note that the representative game models have been among the main motivation behind this literature.

within each school as subgames of a large game, where subgames are allowed to be heterogeneous in various aspects such as the number of the players, their payoff specifications, and the way they form beliefs about other players' types or higher order beliefs about other players' beliefs. Since we do not require that there be random components accounting for game heterogeneity that are drawn from a common distribution, we do not need to model the nature of the randomness of those components in relation to other components of the game.

More specifically this paper introduces a static large Bayesian game which is not necessarily reducible to many replications of a small representative game. To accommodate various informational assumptions in games into a unifying framework, this paper defines two layers of group structures on the large number of the players. The first group structure is an *information group structure* where each agent's payoff differential at the change of his action is never affected by the actions of the players outside his group, and players in each group observe commonly a group-specific signal (called a *public signal* here). The second group structure is an *acquaintance group structure* such that each agent has a small group of players whose types he observes. This paper calls this group the player's *acquaintance group*.[2] Throughout this paper, it is assumed that the acquaintance groups form a partition of the players that is finer than that of information groups.

In this game, the beliefs that the players form about the types of other players may be different across individual players, except that the type vectors across disjoint acquaintance groups are conditionally independent given their public signal. Hence we do not make a common prior assumption that ensures a direct link (through Bayes' rule) between the objective probability that the Nature adopts for drawing types at the initial stage of the game and the subjective probability that each player adopts for predicting the other players' types.

The econometrician observes outcomes from a pure strategy Bayesian Nash equilibrium (possibly among multiple equilibria), and attempts to make inference about the structural parameters of the game. While the equilibrium is driven by the subjective beliefs of the players, the validity of the econometrician's inference is measured in terms of the Nature's objective probability.[3]

Examples in this paper's framework are many. First, many small games of complete information similar to those in Tamer (2003) and Ciliberto and Tamer (2009) belongs to this paper's framework as a special case. Each complete information entry game can be viewed as

[2]Kim and Che (1994) analyzed auction models with a group structure similar to the acquaintance group structure. See also Andreoni, Che and Kim (2007).

[3]This paper confines attention to simultaneous-move games with an unordered finite action space. Hence auction models with continuous bids are excluded. Global network games with endogenous network formation or matching games are also excluded because the action space increases as the number of players increases.

an acquaintance group in this paper's game. The main difference is that the games need not have the same number of players or payoff structures. Second, models of social interactions developed by Brock and Durlauf (2001b, 2007) can be accommodated in this framework. One may model each game of social interactions as one information group with agents with private information.

The main challenge for the econometrician in this data environment is that it is hard for the econometrician to recover the heterogeneous beliefs of each player about other players' types from observing only several large games. To address this challenge, this paper develops what this paper calls a *hindsight regret approach*. Using the assumption of conditional independence, this paper draws on the insights of Kalai (2004) and Deb and Kalai (2010), and introduces the notion of *hindsight regret* for each player. Hindsight regret measures the payoff loss due to the players' not being able to observe the other players' actions. This paper establishes a belief-free bound for the regret, using a concentration inequality called McDiarmid's inequality. The bound for the hindsight regret obtained here does not depend on the way each player forms beliefs about other players' types.

Using the belief-free hindsight regret, this paper derives moment inequalities in a spirit similar to Ciliberto and Tamer (2009), and applies their partial identification approach to develop asymptotically valid confidence sets for the structural parameters. When one observes large games, one might think that checking equilibrium constraints jointly for many players in each game raises the problem of dimensionality in practice. However, this paper formulates moment inequalities for each acquaintance group, not for the whole game, and hence such a problem does not arise, as long as acquaintance groups are small.

The paper proposes two kinds of wild bootstrap methods and establishes their asymptotic validity uniformly over the probabilities that the Nature adopts for drawing the players' types. The paper also presents results from a Monte Carlo simulation study based on a social interactions model. First, the results show that the larger the hindsight regrets, the more conservative the inference becomes. This reflects the difficult nature of the problem in which one cannot recover the beliefs while the beliefs about other players' types play an important role in players' strategic decisions. Second, it is also found that even when the coverage probabilities are reported to be 1, the power properties for the test of the parameter values can be reasonably good. Third, the results also show that ignoring the hindsight regrets and proceeding as if the Bayesian Nash equilibrium was from an *ex post* Nash equilibrium may lead to invalid inference for a reasonable range of sample sizes. Hence the hindsight regrets cannot be ignored in general.

Given the general set-up of the games in this paper, one might ask in what kinds of game set-ups the asymptotic approach of the proposed method becomes reliable and produces potentially a nontrivial result. The major condition is that many players in the game should have small hindsight regrets. While this condition is mostly satisfied by social interactions models and some network models, the condition excludes the case where one observes many small private information games as in Aradillas-Lopez (2010) and de Paula and Tang (2011).

This paper's framework is relevant to many models of social interactions and a certain class of social network models. See Brock and Durlauf (2001a) and Ioannides (2013) for a review of the literature of social interactions. This paper's approach is related to the structural approach of Brock and Durlauf (2001b, 2007) who developed discrete choice-based social interactions models. See also Krauth (2006) for a simulation-based method of inference for such models, Li and Lee (2009) for investigating the role of subjective expectations in social interactions models based on Brock and Durlauf (2001b). Blume, Brock, Durlauf and Jayraman (2011) explored the issue of identification in linear social network models. While the perspective of a large game in this literature has already appeared in various forms in Brock and Durlauf (2001, 2007), recent researches by Xu (2012), Bisin, Moro, and Topa (2011), and Menzel (2012) are more explicit about the inferential issues in a large game model. The inferential procedure of Xu (2012) employs conditions that yield uniqueness of the equilibrium. On the other hand, Bisin, Moro, and Topa (2011) admit multiple equilibria, but their method is built on the characterization of an equilibrium through aggregate quantities of outcomes, where the asymptotic stability of the aggregate quantities (as the number of the players increases) is part of their equilibrium concept. In contrast, this paper defines the set of equilibria within a finite player game, without prescribing a particular asymptotic behavior of the equilibria as the number of the players grows to infinity. Menzel (2012) recently developed asymptotic theory for inference based on large games where type-action profiles are exchangeable sequences.

This research was originally motivated as an exploration of a way to elucidate the structural source of cross-sectional dependence among the outcomes observed by the econometrician. The literature of cross-sectional dependence assumes a cross-sectional dependence structure directly on the observed outcomes either through spatial version of weak dependence or using factor models. For example, see Conley (1999) for an introduction of cross-sectionally stationary series, and see Jia (2008) for its application in the model of entry decisions by Wal Mart. Also see Bai and Ng (2002) and Phillips and Sul (2003) for models of strong cross-sectional dependence through factors, among many others. The structure of cross-sectional dependence arises from the informational assumptions about the players in the game, and are mainly due to two sources: the public signal commonly observed among the players in each information group, and the types of the players shared among each other within each acquaintance group.

This paper's approach for inference is inspired by the work of Andrews (2005) who explored how the presence of unobserved public signals affect the asymptotic properties of estimators from linear factor models with short panel data. Kuersteiner and Prutcha (2012) provides asymptotic theory for a wide class of dynamic panel models with regressors that are not strictly exogenous and may even exhibit strong cross-sectional dependence of unknown form.

As in Andrews (2005) and Kuersteiner and Prutcha (2012), the test statistic we are using has a functional of a mixture normal distribution as its limiting distribution, but we cannot use the random norming to pivotize the test statistic, because the restrictions here are inequalities rather than equalities. Instead, this paper develops a modified form of a wild bootstrap procedure that is asymptotically valid uniformly over a large class of objective probabilities that the Nature adopts.

This paper is organized as follows. The first section formally introduces a large Bayesian game. The section also defines the notion of hindsight regret, and offers its bound. Section 3 develops a general inference method and establishes its asymptotic validity. The proofs of the results in this paper appear in the Appendix.

2. A Large Bayesian Game with Acquaintance Groups

2.1. **Acquaintance Groups, Information Groups and Bayesian Nash Equilibria.**

In this section, we introduce a general form of a Bayesian game that essentially defines the scope of the paper. The game is played by N players, where each player $i \in \mathbb{N} \equiv \{1, 2, \cdots, N\}$ chooses an action from a common finite action set $\mathbb{A} \equiv \{\bar{a}_1, \cdots, \bar{a}_K\}$, after observing his own payoff types and some other signals.

More specifically, the nature draws an outcome ω from a sample space Ω which realizes the (payoff) type profile

$$T(\omega) = (T_1(\omega), \cdots, T_N(\omega)) \in \mathbb{T}^N,$$

for the N players, where $T_i(\omega) \in \mathbb{T} \subset \mathbf{R}^t$ represents a type vector for player i. Let the distribution of T adopted by the Nature be denoted by P.

To specify the cross-sectional correlation structure of the types, we suppose that there exists a partition of $\{1, \cdots, N\} = \cup_{s=1}^S \mathbb{N}_s$, where for each $s = 1, \cdots, S$, every player $i \in \mathbb{N}_s$ observes a common signal vector C_s, but does not observe C_t with $t \neq s$. For each $s \in \mathbb{S} \equiv \{1, \cdots, S\}$, we call an *information group* s the set of players who commonly observe the signal vector C_s. Hence there are S number of information groups, and each information group s contains $N_s \equiv |\mathbb{N}_s|$ number of players.

Each player i observes the payoff types of some other players in a set denoted by $I(i)$. We call $I(i)$ *player i's acquaintance group*, and each member of $I(i)$ player i's *acquaintance*. Hence the type vector that player i observes is $T_{I(i)} \equiv (T_j)_{j \in I(i)}$. We assume for the sake of analytical simplicity that we have $j \in I(i)$ if and only if $i \in I(j)$, and the acquaintance groups form a finer partition of the players than that of information groups so that for all

$i \in \mathbb{N}$, $I(i) \subset \mathbb{N}_s$ for some $s \in \mathbb{S}$. Acquaintance groups are useful for representing situations with complete information games and some network games, as we shall show later.

Once the Nature draws a type vector T that realizes to be $t = (t_j)_{j \in \mathbb{N}} \in \mathbb{T}^N$, each player i observes $t_{I(i)} \equiv (t_j)_{j \in I(i)}$ and forms a belief on T. The belief is denoted by $Q_i(\cdot|t_{I(i)})$, a probability measure on $\mathbb{T}^N$ for each $t_{I(i)}$. The beliefs Q_i are allowed to be heterogeneous across players.

The probability P is the *objective probability* that the econometrician uses to express the validity of his inference method. On the other hand, the probability Q_i for each player i is a *subjective probability* formed according to player i's prior and possibly through higher order beliefs about other players' beliefs. When the Nature's P belongs to common knowledge, we have $Q_i(\cdot|t_{I(i)}) = P(\cdot|t_{I(i)})$ for all $i \in \mathbb{N}$ and $t_{I(i)} \in \mathbb{T}^{|I(i)|}$, so that the distinction between the objective probability that the econometrician's inference centers on and the subjective probabilities that the players take as their beliefs is not necessary. Here it is, as we are not making such an assumption.

We introduce a conditional independence assumption for P and Q_i's. For this, we define the following notion: for each $s \in \mathbb{S}$, let $\mathbb{M}_s \subset \mathbb{N}_s$ be such that for $i, j \in \mathbb{M}_s$, $i \neq j$ if and only if $I(i) \cap I(j) = \varnothing$. Hence players from $\mathbb{M}_s$ have no common acquaintance. We call such a set $\mathbb{M}_s$ a *disjoint selection* from $\mathbb{N}_s$.

Assumption 1: For each $s \in \mathbb{S}$ and any disjoint selection $\mathbb{M}_s$ from $\mathbb{N}_s$, $\{T_{I(i)} : i \in \mathbb{M}_s\}$ is the set of random vectors that are conditionally independent given C_s both under P and under $Q_i(\cdot|t_{I(i)})$ for all $i \in \mathbb{N}$ and all $t_{I(i)} \in \mathbb{T}^{|I(i)|}$.

Assumption 1 assumes that the types of players from disjoint selections are conditionally independent among each other given the information group's public signal C_s according to the Nature's probability, and all the players know this fact. However, we do not require that this conditional independence be common knowledge among the players.

Assumption 2: There exists a small $\rho > 0$ such that for any event $B \subset \mathbb{T}^N$ such that $Q_i(B|t_{I(i)}) \geq 1 - \rho$ for all $i \in \mathbb{N}$ and all $t_{I(i)} \in \mathbb{T}^{|I(i)|}$, it is satisfied that $P\{B|T_{I(i)} = t_{I(i)}\} \geq 1 - \rho$ for all $i \in \mathbb{N}$ and all $t_{I(i)} \in \mathbb{T}^{|I(i)|}$.

Assumption 2 says that any event that all the players in the game believe strongly to occur does occur with high probability according to the Nature's conditional probability. This assumption imposes a limited (one-sided) version of rational expectations on the players' beliefs on events that are commonly believed to be highly likely among the players. The version is one-sided in the sense that a high probability event (according to the Nature's experiment) does not need to be viewed as a high probability event by each player. Assumption 2 is used later to translate the results about belief free bounds for hindsight regrets into moment inequalities that the econometrician can use.

Without loss of generality, we assume that for each $s \in \mathbb{S}$, and each $i \in \mathbb{N}_s$, the public signal C_s is a deterministic function of T_i so that C_s is measurable with respect to the σ-field of T_i for all $i \in \mathbb{N}_s$.

Once the Nature draws $T = t$ with distribution P, each player i, facing the other players choosing $a_{-i} \in \mathbb{A}^{N-1}$, receives payoff $u_i(a_i, a_{-i}; t_{I(i)})$ from choosing $a_i \in \mathbb{A}$, so that the payoff of player i is not directly affected by the types of other players outside his acquaintance group, although it is affected by their actions.

It is assumed that the information group structure and the payoff functions are common knowledge. A *pure strategy y_i for player i* is an $\mathbb{A}$-valued map on $\mathbb{T}$, and a *pure strategy profile* is a profile $y = (y_1, \cdots, y_N)$ of such maps y_i. Given a strategy profile y, the expected payoff for player i is given by

$$U_i(y|t_{I(i)}) = \int_{\mathbb{T}^N} \int_{\mathbb{A}^N} u_i(y(t); t_{I(i)}) Q_i(dt|t_{I(i)}),$$

where $y(t) = (y_1(t_{I(1)}), y_2(t_{I(2)}), \cdots, y_N(t_{I(N)}))$. Then we say that a strategy profile y is a *pure strategy Bayesian Nash equilibrium*, if for each $i \in \mathbb{N}$, $t_{I(i)} \in \mathbb{T}^{|I(i)|}$, and any pure strategy y_i' for i,

$$U_i(y|t_{I(i)}) \geq U_i(y_i', y_{-i}|t_{I(i)}). \tag{2.1}$$

This paper does not place restrictions on Q_i other than that Assumptions 1 and 2 hold.[4]

Instead of directly observing a strategy profile, the econometrician typically observes its realized action profile. Given a pure strategy equilibrium $y = (y_1, \cdots, y_N)$, define

$$Y_i \equiv y_i(T_{I(i)}),\ i \in \mathbb{N}, \tag{2.2}$$

and let $Y \equiv (Y_1, \cdots, Y_N) \in \mathbb{A}^N$. The econometrician observes Y_i's and part of T_i. (We will specify the econometrician's observations in detail later.) The equation (2.2) is a *reduced form* for Y_i. When the game has multiple equilibria, this reduced form is not uniquely determined by the game.

Given a pure strategy equilibrium y, let P^y be the joint distribution of $(y(T), T)$, with $y(T) = (y_1(T_{I(1)}), \cdots, y_N(T_{I(N)}))$, when its marginal distribution of T is equal to P. Also

[4]Existence of a Nash equilibrium in this general set-up is not always ensured. See Stinchcombe (2010) for an example of a Bayesian game that does not have an equilibrium, failing the absolute continuity condition of Milgrom and Weber (1985). However, existence of an equilibrium can be established by invoking more special structure of the game in application. For example, see Milgrom and Weber (1985), Balder (1988), Athey (2001), McAdams (2003) and Reny (2011) and references therein for general results.

given y, let $Q^y = (Q_1^y, \cdots, Q_N^y)$, where Q_i^y is the joint distribution of $(y(T), T)$ representing the player i's beliefs about the actions and the types (in equilibrium y).

2.2. **Belief-Free Hindsight Regrets.** The equilibrium constraints in (2.1) can be rewritten as follows: for all $i \in \mathbb{N}$ and all $\bar{a} \in \mathbb{A}$,

$$\mathbf{E}_i^y \left[u_i(Y; T_{I(i)}) - u_i(\bar{a}, Y_{-i}; T_{I(i)}) | T_{I(i)} = t_{I(i)} \right] \geq 0, \tag{2.3}$$

where $\mathbf{E}_i^y[\cdot | T_{I(i)} = t_{I(i)}]$ denotes conditional expectation (under Q_i^y) given $T_{I(i)} = t_{I(i)}$. The constraints are generally useful for the econometrician to generate moment inequalities for observations. However, they cannot be directly used in our context, because there is no general way for the econometrician to recover the subjective beliefs.

The hindsight regret approach in this paper replaces the inequality in (2.3) by the following *ex post* version: for all $i \in \mathbb{N}$ and all $\bar{a} \in \mathbb{A}$,

$$u_i(Y; T_{I(i)}) - u_i(\bar{a}, Y_{-i}; T_{I(i)}) \geq -\lambda_{\bar{a}}, \tag{2.4}$$

with large probability according to his belief Q_i^y, where $\lambda_{\bar{a}} \geq 0$ is a compensation for player i needed to prevent her from switching from her action Y_i in equilibrium to action $\bar{a}$ (with large probability) after the types of all the players are revealed to her.

We seek to formulate the inequality (2.4) in a way that have three main features. First, we would like the inequalities to hold jointly for all the players whose equilibrium actions can be determined by player i, i.e., the players in player i's acquaintance group. Behind the inequality (2.4) is the assumption that other players also maintain their equilibrium strategy after the types are revealed. But from the perspective of player i, it is hard to determine the minimal compensation level for those players whose equilibrium action player i cannot determine from his type information $t_{I(i)}$. Hence we envisage a scheme in which the compensation level for such players is chosen so that the violation of the inequality in (2.4) due to the deviation by such players' actions becomes highly unlikely from player i's perspective. Second, for simplicity, we do not allow the compensation scheme to rely on any signal about the beliefs or higher order beliefs of the players. Otherwise there may arise a complex mechanism design issue regarding the compensation. We call this belief-free compensation a *belief-free hindsight regret*, because the compensation represents a potential payoff loss to the players in hindsight due to not being able to observe the types of other players outside his acquaintance group. Third, we seek to formulate a minimal amount of belief-free hindsight regret. This is because the quality of prediction from the incentive constraints in (2.4) becomes better with tighter compensation scheme. As we shall see later, the econometrician makes inference about the payoff parameter by checking whether the equilibrium constraints in (2.4) are plausible with each parameter value given the observed equilibrium outcomes. Hence using a tighter hindsight regret tends to yield a sharper empirical result for the econometrician.

To define a belief-free hindsight regret formally, we introduce some notation. Given $i \in \mathbb{N}$ and any $\lambda_i = (\lambda_{i,\bar{a}})_{\bar{a}\in\mathbb{A}}$, $\lambda_{i,\bar{a}} \geq 0$, let

$$b_i(\lambda_i) \equiv \left\{a \in \mathbb{A} : u_i(a, Y_{-i}; T_{I(i)}) - u_i(\bar{a}, Y_{-i}; T_{I(i)}) \geq -\lambda_{i,\bar{a}},\ \forall \bar{a} \in \mathbb{A}\backslash\{a\}\right\}.$$

When $Y_i \in b_i(\lambda)$, it means that Y_i remains a best response for player i to $(Y_{-i}, T_{I(i)})$ even after all the types of the players are revealed, as long as player i is compensated by the compensation scheme λ_i for adopting the response Y_i.

Given small $\rho > 0$, we say that the vector $\lambda_i = (\lambda_{ij})_{j\in I(i)}$ (with $\lambda_{ij} = (\lambda_{ij,\bar{a}})_{\bar{a}\in\mathbb{A}}$ and $\lambda_{ij,\bar{a}} \geq 0$) is a *$\rho$-hindsight regret* for player i, if

$$Q_i^y\left\{Y_j \in b_j(\lambda_{ij}),\ \forall j \in I(i)|T_{I(i)} = t_{I(i)}\right\} \geq 1 - \rho.$$

A hindsight regret λ_i for player i lists the amount of compensations for players in $I(i)$ to induce them to maintain their strategies in equilibrium y with high probability.

Hindsight regrets are closely related to strategic interdependence among the players: the regrets tend to be high for a player who does not observe the types of those players whose action can have a large impact on his payoff. To formally introduce this interdependence, we define a maximal variation of a real function. Suppose that $f(x_1, \cdots, x_N)$ is a real-valued function on a set $\mathcal{X}^N \subset \mathbf{R}^N$. Then, we write

$$V_j(f) = \max\left|f(x) - f(x_j(x))\right|,$$

where the maximum is over all x's in $\mathcal{X}^N$ and over all $x_j(x)$'s in $\mathcal{X}^N$ such that $x_j(x)$ is x except for its j-th entry. We call $V_j(f)$ a *maximal variation of f at the j-th coordinate.*

For $i \in \mathbb{N}$, and $\bar{a} \in \mathbb{A}$, let

$$u_{i,m}^{\Delta}\left(a_{-i}; t_{I(i)}, \bar{a}\right) \equiv u_i(\bar{a}_m, a_{-i}; t_{I(i)}) - u_i(\bar{a}, a_{-i}; t_{I(i)}),$$

and let for $i \in \mathbb{N}$, $j \in \mathbb{N}$, and $t_{I(i)} \in \mathbb{T}^{|I(i)|}$,

$$\Delta_{ij}(t_{I(i)}, \bar{a}) \equiv \max_{m=1,\cdots,K} V_j(u_{i,m}^{\Delta}\left(\cdot; t_{I(i)}, \bar{a}\right)). \tag{2.5}$$

The quantity $\Delta_{ij}(t_{I(i)}, \bar{a})$ measures the worst payoff loss for player i from his benchmark payoff action $\bar{a}$ that can be inflicted by player j's choice of an action. Hence $\Delta_{ij}(t_{I(i)}, \bar{a})$ measures strategic relevance of player j to player i. When $\Delta_{ij}(t_{I(i)}, \bar{a}) = 0$ for all $t_{I(i)}$ and $\bar{a} \in \mathbb{A}$, there is no way player j can affect the optimal choice of player i. We say that players

i and j are *payoff disjoint,* if $\Delta_{ij}(t_{I(i)}, \bar{a}) = \Delta_{ji}(t_{I(j)}, \bar{a}) = 0$ for all $t_{I(i)} \in \mathbb{T}^{|I(i)|}$, $t_{I(j)} \in \mathbb{T}^{|I(j)|}$, and all $\bar{a} \in \mathbb{A}$. [5]

Assumption 3: Any two players from different information groups $\mathbb{N}_s$ and $\mathbb{N}_t$ with $s \neq t$ are payoff disjoint.

Assumption 3 simplifies the framework by allowing that a player in one information group $\mathbb{N}_s$ does not need to form beliefs about the types of the players from a different information group $\mathbb{N}_t$. In many empirical examples, the different information groups $\mathbb{N}_s$ can be thought of as constituting separate games observed by the econometrician.

To characterize the belief-free hindsight regrets, for each $i \in \mathbb{N}_s$, $j \in I(i)$, $\bar{a} \in \mathbb{A}$, and $\rho \in (0,1)$, we let

$$B_{ij,\rho}(t_{I(j)}, \bar{a}) \equiv \sqrt{-\frac{1}{2}\psi_{ij}(t_{I(j)}, \bar{a}) \cdot \log\left(\frac{\rho}{(K-1)|I(i)|}\right)}, \tag{2.6}$$

where

$$\psi_{ij}(t_{I(j)}, \bar{a}) \equiv \sum_{l \in \mathbb{N}_s \setminus I(i)} \left(\sum_{k \in I(l)} \Delta_{jk}(t_{I(j)}, \bar{a}) \right)^2 . \tag{2.7}$$

The following theorem shows that $B_{ij,\rho}$ is a belief-free ρ-hindsight regret for player j (from the perspective of player i.)

Theorem 1: *Suppose that Assumptions 1-3 hold. Then for each pure strategy equilibrium y, each $\rho \in (0,1)$, for all $i \in \mathbb{N}$ and all $t_{I(i)} \in \mathbb{T}_{I(i)}$,*

$$Q_i^y \left\{ Y_j \in b_j(B_{ij,\rho}(T_{I(i)})), \ \forall j \in I(i) | T_{I(i)} = t_{I(i)} \right\} \geq 1 - \rho,$$

where $B_{ij,\rho}(T_{I(i)}) = (B_{ij,\rho}(T_{I(i)}, \bar{a}))_{\bar{a} \in \mathbb{A}}$.

The hindsight regret $B_{ij,\rho}(\cdot)$ is fully determined only by the payoff functions and the acquaintance group structure I of the game, and does not depend on Q_i's. Hence it is *belief free.* The bound is a consequence of McDiarmid's inequality (McDiarmid (1989)) and the Nash equilibrium constraints in (2.3).

Recall that $B_{ij,\rho}$ is the compensation for player j from player i's perspective. Player i knows all the types of the acquaintances of the players in $I(i)$, and hence he can fully determine their equilibrium actions, and the compensation level for them. The uncertainty

[5] When two players are payoff disjoint, this does not necessarily mean that one player's change of action has no way of having an impact on the other player's payoff. It only means that one player's change of action has no way of having an impact on an optimal decision made by the other player.

faced by player i is due the actions of those players outside $I(i)$. To maintain the *ex post* incentive constraint with high probability, the compensation level has to take these players into account, i.e. the impact the change of their types may have on player j's payoff. Perturbing the types of those players outside $I(i)$ can affect player j's payoff by affecting the actions of those players who observe the perturbed types. Thus the component $\psi_{ij}(t_{I(j)}, \bar{a})$ in the definition of $B_{ij,\rho}(t_{I(j)}, \bar{a})$ now measures the overall strategic relevance (to player j) of those players who observe the types of the players outside $I(i)$.

In the case of private information game, we have $I(i) = \{i\}$. Hence in this case, Theorem 1 is reduced to the following simple form: for each pure strategy equilibrium y, each $\rho \in (0,1)$, for all $i \in \mathbb{N}$ and all $t \in \mathbb{T}$,

$$Q_i^y \{Y_i \in b_i(B_{i,\rho}(T_i)) | T_i = t\} \geq 1 - \rho,$$

where $B_{i,\rho}(T_i) = (B_{i,\rho}(T_i, \bar{a}))_{\bar{a} \in \mathbb{A}}$,

$$B_{i,\rho}(t_i, \bar{a}) \equiv \sqrt{-\frac{1}{2}\psi_i(t_i, \bar{a}) \cdot \log \rho},$$

and $\psi_i(t_i, \bar{a}) \equiv \sum_{l \in \mathbb{N}_s \setminus \{i\}} \Delta_{il}^2(t_i, \bar{a})$. The component ψ_i now summarizes the strategic relevance to player i of all the other players in player i's information group. The larger ψ_i is the higher the hindsight regret $B_{i,\rho}$.

2.3. **Examples.**

2.3.1. *Many Small Entry Games with Complete Information.* We consider a large game constituted by M markets where each market m forms an acquaintance group. The acquaintance group $I(i)$ of i represents the group of players in the same market that player i belongs to. The action space is $\mathbb{A} = \{0, 1\}$, with 1 representing entry into the market, and 0 non-entry. For each player i in market s, the payoff differential is specified as

$$u_i(1, a_{-i}; T_i) - u_i(0, a_{-i}; T_i) = \alpha^\top T_{i,1} + \theta \sum_{j \in I(i) \setminus \{i\}} a_j + T_{i,2},$$

where $T_i = (T_{i,1}, T_{i,2})$. The payoff differential shows that the players from different acquaintance groups are payoff disjoint. Each market constitutes a complete information game.

This type of payoff specification was used by many researchers (Bresnahan and Reiss (1991), Tamer (2003), Ciliberto and Tamer (2008).) If we focus only on pure strategies, it turns out that $\psi_{ij}(T_{I(j)}, \bar{a}) = 0$ and hence $B_{ij,\rho}(T_{I(j)}, \bar{a}) = 0$. This is not surprising; there is no hindsight regret, because each player observes the types of all the players that are strategically relevant to her (that is, her opponents in her market).

2.3.2. *Large Games with Social Interactions.* Suppose that we have S groups of players where each group s has N_s number of players. Each group constitutes a large game with private information. For player i in group s, we consider either of the following two specifications of payoff functions:

$$u_i(a_i, a_{-i}; t_i) = v(a_i; t_i) + \frac{\theta \cdot a_i}{N_s - 1} \sum_{j \neq i}^{N_s} a_j,$$

or

$$u_i(a_i, a_{-i}; t_i) = v(a_i; t_i) - \frac{\theta}{2} \cdot \left(a_i - \frac{1}{N_s - 1} \sum_{j \neq i}^{N_s} a_j \right)^2,$$

where $a_{-i} = (a_j)_{j=1, j \neq i}^{N_s}$ and $v(a_i; t_i)$ is a component depending only on (a_i, t_i). The parameter θ captures strategic interaction of player i with the other players within his group s. The first specification expresses interaction between player i's action (a_i) and the average actions of the other players. The second specification captures preference for conformity to the average actions of the other players. (See Bernheim (1994).) If we replace the other players' actions a_j by $\mathbf{E}Y_j$, i.e., the expected value of the actions, both specifications coincide with those considered in the payoff specification of Brock and Durlauf (2001).

In both cases,

$$u_i^{\Delta}(a_{-i}; t_i, \bar{a}) = \min_{m=1,\cdots,K} \left\{ v_m(t_i, \bar{a}) + \frac{\theta \cdot (\bar{a}_m - \bar{a})}{N_s - 1} \sum_{j \neq i}^{N_s} a_j \right\},$$

where $v_m(t_i, \bar{a}) = v(\bar{a}_m; t_i) - v(\bar{a}; t_i)$ in the first specification, and $v_m(t_i, \bar{a}) = v(\bar{a}_m; t_i) - v(\bar{a}; t_i) - \theta \cdot (\bar{a}_m^2 - \bar{a}^2)/2$ in the second specification.

Also, in both cases, $\psi_{ij}(t_i, \bar{a}) = \sum_{j \in \mathbb{N}_{-i}} \Delta_{ij}(t_i, \bar{a})^2$ with $\Delta_{ij}(t_i, \bar{a}) = |\theta| \cdot \max_m (\bar{a}_m - \bar{a})/(N_s - 1)$ and the belief-free hindsight regret in Theorem 1 becomes:

$$B_{i,\rho}(t_i, \bar{a}) = \frac{|\theta| \cdot \max_m (\bar{a}_m - \bar{a})}{\sqrt{N_s - 1}} \cdot \sqrt{-\frac{1}{2} \log \left(\frac{\rho}{K - 1} \right)}.$$

Hence the hindsight regret becomes asymptotically negligible, as long as the number of the players in each information group $\mathbb{N}_s$ increases.

3. Econometric Inference

3.1. **The Econometrician's Observations and Parametrization.**

Suppose that the econometrician observes (Y, X), where $Y \in \mathbb{A}^N$ is an N-dimensional vector of actions by N players and X is an $N \times d_X$ matrix whose i-th row is $X_i^\top$, with X_i being a subvector of T_i, represents a covariate vector of player i. As for (Y, X), we make the following assumptions.

Assumption 4 (The Econometrician's Observation): (i) The distribution of (Y, T) is equal to P^y associated with a pure strategy equilibrium y.

(ii) For each $i \in \mathbb{N}$, $T_i = (\eta_i, X_i)$, where $X_i \in \mathbf{R}^{d_X}$ is observed but $\eta_i \in \mathbb{H} \subset \mathbf{R}^{d_\eta}$ is not observed by the econometrician.

The distribution of (Y,T) that the econometrician focuses on is originated from the Nature's objective probability P and a pure strategy equilibrium y. The econometrician does not know which equilibrium the vector of observed outcomes Y is associated with. The players' subjective beliefs affect the distribution of (Y,T) through their impact on the associated equilibrium y.

Assumption 4(ii) specifies that T_i involves components η_i and X_i which are unobserved and observed by the econometrician respectively. Thus the econometrician may not observe part of the type information each player has.

We first introduce parametrization of unobserved heterogeneity η_i and payoffs.

Assumption 5 (Parametrization of Unobserved Heterogeneity and Payoffs): For all $s \in \mathbb{S}$, $i \in \mathbb{N}_s$, $t_{I(i)} \in \mathbb{T}_{I(i)}$, and $a \in \mathbb{A}^{|I(i)|}$,

$$P\left\{\eta_{I(i)} \leq t_{I(i)} | X_{I(i)}\right\} = G_{s,\theta_0}\left(t_{I(i)} | X_{I(i)}\right) \text{ and } u_i(a; t_{I(i)}) = u_{i,\theta_0}(a; t_{I(i)}),$$

where $\theta_0 \in \Theta \subset \mathbf{R}^d$, $G_{s,\theta}(\cdot | X_{I(i)})$ and $u_{i,\theta}(\cdot; t_{I(i)})$ are parametric functions parametrized by $\theta \in \Theta$.

Assumption 5 assumes that the conditional CDF of $\eta_{I(i)}$ given $X_{I(i)}$ and the payoff function are parametrized by a finite dimensional vector $\theta \in \Theta$. Assumption 5 does not require that the payoff function be additive in unobserved heterogeneity $\eta_{I(i)}$. Therefore, the model accommodates specifications where the observed covariates $X_{I(i)}$ are intertwined with $\eta_{I(i)}$ in a complicated manner as in random coefficient specifications.

Assumption 6 (Conditional Independence): (i) For any $s \in \mathbb{S}$ and disjoint selection $\mathbb{M}_s$ from $\mathbb{N}$, the set $\{T_{I(i)} : i \in \mathbb{M}_s\}$ is the set of conditionally independent random vectors given $C = (C_s)_{s=1}^S$ under P.
(ii) $\{\eta_1, \cdots, \eta_N, C\}$ is conditionally independent given $X = (X_1, \cdots, X_N)$ under P.

Assumption 6(i) requires that the type vectors across different acquaintance groups are conditionally independent given C. This follows by Assumption 1 and the additional assumption that $(T_{I(i)}, C_s)_{i \in \mathbb{N}_s}$'s are independent across $s = 1, \cdots, S$. Assumption 6(ii) requires that the unobserved components $\eta_1, \cdots, \eta_N$ and C are conditionally independent of X. This condition is satisfied, for example, if $\eta_1, \cdots, \eta_N$ are conditionally independent given X and for each $i \in \mathbb{N}$, $X_i = (Z_s, W_i)$, where W_i is an idiosyncratic component and Z_s is a component common in information group s, and C_s is a nonstochastic function of Z_s. Note that the econometric may observe X_i, but may or may not observe C_s, although the players observe C_s.

Assumption 6 is also concerned only with the primitives of the game. It does not impose restrictions on the equilibrium y or the way beliefs of the agents are formed in equilibrium. It is only concerned with the objective probability P that the Nature uses to draw types for the players.

3.2. **Moment Inequalities.** In this subsection, we derive moment inequalities from the Nash equilibrium constraints. For simplicity, we write $u_i(a_{I(i)}) = u_{i,\theta}(a_{I(i)}, Y_{-I(i)}; T_{I(i)})$, suppressing $Y_{-I(i)}$ and $T_{I(i)}$ and θ from the notation.

Let

$$\gamma_i^*\left(a_{I(i)}, T_{I(i)}\right) \equiv 1\left\{\mathbf{E}_j^y\left[u_j(a_{I(j)})|T_{I(j)}\right] \geq \max_{c\in\mathbb{A}} \mathbf{E}_j^y\left[u_j(c, a_{I(j)\backslash\{j\}})|T_{I(j)}\right],\ \forall j \in I(i)\right\}.$$

Recall that the conditional expectation $\mathbf{E}_j^y\left[\cdot|T_{I(j)}\right]$ originates from player j's subjective beliefs. Since $Y = y(T)$ for some pure strategy Nash equilibrium y, for any $a_{I(i)} \in \mathbb{A}^{|I(i)|}$ and for all values of $T_{I(i)}$ such that $Y_{I(i)} = a_{I(i)}$, we have $\gamma_i^*\left(a_{I(i)}, T_{I(i)}\right) = 1$, i.e.,

$$1\left\{Y_{I(i)} = a_{I(i)}\right\} \leq \gamma_i^*\left(a_{I(i)}, T_{I(i)}\right). \tag{3.1}$$

Similarly, for all values of $T_{I(i)}$ such that $Y_{I(i)} \neq a_{I(i)}$, we have $Y_{I(i)} = c_{I(i)}$ for some $c_{I(i)} \in \mathbb{A}^{|I(i)|}\backslash\{a_{I(i)}\}$ so that we have $\gamma_i^*\left(c_{I(i)}, T_{I(i)}\right) = 1$ for some $c_{I(i)} \in \mathbb{A}^{|I(i)|}\backslash\{a_{I(i)}\}$. In other words, we have

$$1\left\{Y_{I(i)} \neq a_{I(i)}\right\} \leq 1\left\{\exists\ c_{I(i)} \in \mathbb{A}^{|I(i)|}\backslash\{a_{I(i)}\} \text{ s.t. } \gamma_i^*\left(c_{I(i)}, T_{I(i)}\right) = 1\right\}. \tag{3.2}$$

We take conditional expectations (given $X_{I(i)}$) of both sides in (3.1) and (3.2), and deduce that for each $i \in \mathbb{N}$, and $a_{I(i)} \in \mathbb{A}^{|I(i)|}$,

$$1 - \pi_{i,L}^*(a_{I(i)}) \leq P\left\{Y_{I(i)} = a_{I(i)}|X_{I(i)}\right\} \leq \pi_{i,U}^*(a_{I(i)}). \tag{3.3}$$

where

$$\begin{aligned}
\pi_{i,U}^*(a_{I(i)}) &\equiv P\left\{\gamma_i^*\left(a_{I(i)}, T_{I(i)}\right) = 1|X_{I(i)}\right\} \text{ and} \\
\pi_{i,L}^*(a_{I(i)}) &\equiv P\left\{\exists\ c_{I(i)} \in \mathbb{A}^{|I(i)|}\backslash\{a_{I(i)}\} \text{ s.t. } \gamma_i^*\left(c_{I(i)}, T_{I(i)}\right) = 1|X_{I(i)}\right\}.
\end{aligned}$$

In fact the bounds in (3.3) can be viewed as a Bayesian-game version of what Ciliberto and Tamer (2009) derived from a complete information game. As Tamer (2003) and Ciliberto and Tamer (2009) illustrated well, when there are multiple equilibria and acquaintance groups are not singletons, we may have

$$1 < \pi_{i,U}^*(a_{I(i)}) + \pi_{i,L}^*(a_{I(i)}).$$

This is because we may have both $\gamma_i^*(a_{I(i)},\eta_{I(i)}) = 1$ and $\gamma_i^*(a'_{I(i)},\eta_{I(i)}) = 1$ with different $a_{I(i)}$ and $a'_{I(i)}$, when there are multiple equilibria. The inequalities in (3.3) become equalities when there is a unique equilibria (with zero probability of a tie occurring) or acquaintance groups are singletons as in the private information set-up. In the case of private information (so that $I(i) = I(i) = \{i\}$), we have for different actions a_i and a'_i, $\gamma_i^*(a_i,\eta_i) = 1$ if and only if $\gamma_i^*(a'_i,\eta_i) = 0$ for almost all η_i. Hence in this case, the inequalities in (3.3) become equalities even when there are multiple equilibria

Unfortunately, the inequalities in (3.3) cannot be directly used in our set-up for inference for two reasons. First, the sets $H_{i,L}^*\left(a_{I(i)}\right)$ and $H_{i,U}^*\left(a_{I(i)}\right)$ involve heterogeneous subjective beliefs which the econometrician has difficulty recovering from the observations. Second, the probabilities in both bounds of (3.3) cannot be simulated, because the bounds depend on the unknown distribution of $Y_{-I(i)}$.

We formulate the moment inequalities through probabilities that do not involve subjective beliefs of players and can be simulated. First, we define an *ex post* version of γ_i^* :

$$\gamma_i\left(a_{I(i)},T_{I(i)}\right) \equiv 1\left\{u_j(a_{I(j)}) \geq \max_{c\in\mathbb{A}} u_j(c,a_{I(j)\backslash\{j\}}) - B_{ij,\rho}(T_{I(j)}),\ \forall j\in I(i)\right\}. \tag{3.4}$$

Now we construct probabilities that can be simulated:

$$\begin{aligned}\pi_{i,U}(a_{I(i)}) &\equiv P\left\{\gamma_i\left(a_{I(i)},\eta_{I(i)}\right) = 1|\mathcal{W}_i\right\} \text{ and}\\ \pi_{i,L}(a_{I(i)}) &\equiv P\left\{\exists\ c_{I(j)}\in\mathbb{A}^{|I(j)|}\backslash\{a_{I(j)}\} \text{ s.t. } \gamma_i\left(c_{I(i)},\eta_{I(i)}\right) = 1|\mathcal{W}_i\right\},\end{aligned}$$

where $\mathcal{W}_i = (X_{I(i)},Y_{-I(i)},C)$. To see that $\pi_{i,U}(a_{I(i)})$ and $\pi_{i,L}(a_{I(i)})$ can be simulated, note that the randomness of $\gamma_{i,U}\left(a_{I(i)},\eta_{I(i)}\right)$ and $\gamma_{i,L}\left(a_{I(i)},\eta_{I(i)}\right)$ is due to $(Y_{-I(i)},\eta_{I(i)},X_{I(i)})$, and that $\eta_{I(i)}$ is conditionally independent of $(Y_{-I(i)},X_{I(i)},C_s)$ given $X_{I(i)}$ by Assumption 5(ii). Hence we can write

$$\begin{aligned}\pi_{i,U}(a_{I(i)}) &= \int_{H_{i,U}(a_{I(i)};X_{I(i)})} dG_{s,\theta}(\eta_{I(i)}|X_{I(i)}) \text{ and}\\ \pi_{i,L}(a_{I(i)}) &= \int_{H_{i,L}(a_{I(i)};X_{I(i)})} dG_{s,\theta}(\eta_{I(i)}|X_{I(i)}),\end{aligned} \tag{3.5}$$

where

$$\begin{aligned}H_{i,U}\left(a_{I(i)};X_{I(i)}\right) &\equiv \left\{\eta_{I(i)}\in\mathbb{H}^{d_\eta|I(i)|} : \gamma_i\left(a_{I(i)},\eta_{I(i)},X_{I(i)}\right) = 1\right\} \text{ and}\\ H_{i,L}\left(a_{I(i)};X_{I(i)}\right) &\equiv \left\{\eta_{I(i)}\in\mathbb{H}^{d_\eta|I(i)|} : \exists\ c_{I(j)}\in\mathbb{A}^{|I(i)|}\backslash\{a_{I(i)}\} \text{ s.t. } \gamma_i\left(c_{I(i)},\eta_{I(i)},X_{I(i)}\right) = 1\right\}.\end{aligned}$$

Thus the quantities $\pi_{i,U}(a_{I(i)})$ and $\pi_{i,L}(a_{I(i)})$ can be simulated using the parametric specification of the conditional distribution of $\eta_{I(i)}$ given $X_{I(i)}$ in Assumption 4.

To construct moment inequalities, we assume that $|I(i)| \leq q$ for all $i\in\mathbb{N}$, for some integer $q > 0$. Given $\mathbf{a} = (a_1,\cdot\cdot\cdot,a_J)\in\mathbb{A}^q$, we write $\mathbf{a}_i = (a_1,\cdot\cdot\cdot,a_{I(i)})\in\mathbb{A}^{|I(i)|}$ for each $i\in\mathbb{N}$.

Then, define

$$\begin{aligned} e_{i,L}(\mathbf{a}_i) &\equiv P\left\{Y_{I(i)} = \mathbf{a}_i|\mathcal{W}_i\right\} - \left(1 - \frac{1}{1-\rho_i}\cdot\pi_{i,L}(\mathbf{a}_i)\right) \text{ and} \\ e_{i,U}(\mathbf{a}_i) &\equiv P\left\{Y_{I(i)} = \mathbf{a}_i|\mathcal{W}_i\right\} - \frac{1}{1-\rho_i}\cdot\pi_{i,U}(\mathbf{a}_i), \end{aligned}$$

where $\rho_i \equiv \rho\cdot(c\bar{\psi}_i/(1+c\bar{\psi}_i))$ with some constant $c > 0$, and $\bar{\psi}_i = \max_{j\in I(i)} \max_{t_{I(j)},\bar{a}} \psi_{ij}(t_{I(j)}, \bar{a})$. The choice of ρ_i is made such that when there is no hindsight regret, i.e., $\bar{\psi}_i = 0$, ρ_i is taken to be zero. (In simulation studies, it works fine to choose its extreme version with $c = \infty$, i.e., $\rho_i = \rho 1\{\bar{\psi}_i > 0\}$.)

We choose nonnegative functions g_{ij} on $\mathbf{R}^{|I(i)|}$ to construct moment inequalities in a spirit similar to Andrews and Shi (2012). For this, first, define the event: for $w = (w_{j,U}(\mathbf{a}), w_{j,L}(\mathbf{a}) : \mathbf{a} \in \mathbb{A}^q)_{j=1}^J$, $w_{j,U}(\mathbf{a}) \geq 0$, $w_{j,L}(\mathbf{a}) \geq 0$ and $\mathbb{D} \subset \mathbb{A}^q$,

$$\mathcal{M}(w;\mathbb{D}) = \left\{\forall j = 1,\cdots,J,\ \forall \mathbf{a} \in \mathbb{D} : \begin{array}{l} \frac{1}{N}\sum_{i=1}^N e_{i,L}(\mathbf{a}_i) g_{ij}(X_{I(i)}) + w_{j,L}(\mathbf{a}) \geq 0 \\ \frac{1}{N}\sum_{i=1}^N e_{i,U}(\mathbf{a}_i) g_{ij}(X_{I(i)}) - w_{j,U}(\mathbf{a}) \leq 0 \end{array}\right\}.$$

The event $\mathcal{M}(w;\mathbb{D})$ expresses the event that certain moment inequality restrictions hold. We aim to find a good bound w such that the probability of $\mathcal{M}(w;\mathbb{D})$ becomes sufficiently large.

To define the bound, we first note that $\pi_{i,L}(\mathbf{a}_i)$ and $\pi_{i,U}(\mathbf{a}_i)$ are nonstochastic functions of (Y, X) from (3.5) and write

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N \left(1 - \frac{1}{1-\rho_i}\cdot\pi_{i,L}(\mathbf{a}_i)\right) g_{ij}(X_i) &= f_{j,L}(Y, X; \mathbf{a}) \text{ and} \\ \frac{1}{N}\sum_{i=1}^N \left(\frac{1}{1-\rho_i}\cdot\pi_{i,U}(\mathbf{a}_i)\right) g_{ij}(X_i) &= f_{j,L}(Y, X; \mathbf{a}), \end{aligned}$$

for some functions $f_{j,L}$ and $f_{j,L}$. Then we let $\delta_{ij,L}(X, \mathbf{a})$ and $\delta_{ij,U}(X, \mathbf{a})$ be bounds for maximal variations of $f_{j,L}(\cdot, X; \mathbf{a})$ and $f_{j,U}(\cdot, X; \mathbf{a})$ at the i-th coordinate, i.e.,

$$\max_{k\in I(i)} V_k\left(f_{j,L}(\cdot, X; \mathbf{a})\right) \leq \delta_{ij,L}(X, \mathbf{a}) \text{ and } \max_{k\in I(i)} V_k\left(f_{j,U}(\cdot, X; \mathbf{a})\right) \leq \delta_{ij,U}(X, \mathbf{a}).$$

Since the payoffs and the conditional distribution of η_i given X_i are fully parametrized, we can usually obtain reasonable bounds $\delta_{ij,L}(X, \mathbf{a})$ and $\delta_{ij,U}(X, \mathbf{a})$ as we illustrate later. Now we are ready to state the main result of moment inequality restrictions.

THEOREM 2: *Suppose that Assumptions 1-6 hold. Then it is satisfied that for any small* $\tau \in (0, 1)$, $\bar{\rho}$, *and any* $\mathbb{D} \subset \mathbb{A}^q$,

$$P\left[\mathcal{M}(w(\tau, X);\mathbb{D})|X, C\right] \geq 1 - \frac{\tau}{2}\left(1\{|\bar{\delta}_U| > 0\} + 1\{|\bar{\delta}_L| > 0\}\right),\ \textit{everywhere,}$$

where $w(\tau, X) = (w_{j,U}(\tau;\mathbf{a},X), w_{j,L}(\tau;\mathbf{a},X) : \mathbf{a} \in \mathbb{A}^q)_{j=1}^J$,

$$\begin{aligned} w_{j,L}(\tau;\mathbf{a},X) &\equiv \sqrt{-\frac{1}{2}\left(\sum_{i=1}^N \delta_{ij,L}^2(X,\mathbf{a})\right)\log\left(\frac{\tau}{2J|\mathbb{D}|}\right)}, \\ w_{j,U}(\tau;\mathbf{a},X) &\equiv \sqrt{-\frac{1}{2}\left(\sum_{i=1}^N \delta_{ij,U}^2(X,\mathbf{a})\right)\log\left(\frac{\tau}{2J|\mathbb{D}|}\right)}, \end{aligned}$$

$\bar{\delta}_U = \max_{\mathbf{a}\in\mathbb{A}^q, j=1,\cdots,J}\sum_{i=1}^N \delta_{ij,U}^2(X,\mathbf{a})$ *and* $\bar{\delta}_L = \max_{\mathbf{a}\in\mathbb{A}^q, j=1,\cdots,J}\sum_{i=1}^N \delta_{ij,L}^2(X,\mathbf{a})$.

Theorem 2 offers finite sample moment inequalities in a general form. When the game is private information and $\psi_i = 0$ for each $i \in \mathbb{N}$, so that there is no strategic interaction between players, all the players have zero hindsight regret and thus we have $B_{j,\rho}(T_i) = 0$, $\bar{\rho}_i = 0$, $\delta_{ij}(\mathbf{a}) = 0$ and $e_{i,L}(\mathbf{a}_i) = e_{i,U}(\mathbf{a}_i)$. We obtain the following result:

$$P\left\{\frac{1}{N}\sum_{i=1}^N e_{i,L}(\mathbf{a}_i)g(X_i) = 0, \forall \mathbf{a} \in \mathbb{A}^q\right\} = 1.$$

One can use this restriction to construct the usual moment restrictions for discrete choice models without strategic interactions.

On the other hand, when ψ_i is large so that the strategic relevance of the players among each other is strong, we have larger $B_{i,\rho}(T_i)$ and $\delta_{ij}^2(\mathbf{a})$, so that the testable implications in Theorem 2 become weaker. This can be viewed as a cost to the econometrician for not being able to recover fully the beliefs of individual players despite their strategic interactions.

To appreciate the components of the bounds in Theorem 2 more concretely, let us consider a private information large game where each player chooses an action from $\{0, 1\}$ so that we only consider the action 1. Furthermore, we specify the payoff function as follows:

$$u_i(1, a_{-i}; T_i) - u_i(0, a_{-i}; T_i) = \theta_1^\top X_i + \frac{\theta_2}{N_s - 1}\sum_{j=1, j\neq i}^{N_s} a_j - \eta_i,$$

where we assume that η_i is unobserved and is independent of the observed covariate X_i and follows a standard normal distribution. As we saw in Section 2.3.2, we have in this case

$$B_{i,\rho}(t_i) = \frac{|\theta_2|}{\sqrt{N_s - 1}}\sqrt{-\frac{1}{2}\log\rho} \equiv B_\rho, \text{ say.}$$

Then, with $\bar{Y}_{-i,s} = \frac{1}{N_s - 1}\sum_{j=1, j\neq i}^{N_s} Y_j$,

$$\begin{aligned} \pi_{i,U}(1) &= \Phi\left(\theta_1^\top X_i + \theta_2\bar{Y}_{-i,s} + B_\rho\right) \text{ and} \\ \pi_{i,L}(1) &= 1 - \Phi\left(\theta_1^\top X_i + \theta_2\bar{Y}_{-i,s} - B_\rho\right). \end{aligned}$$

For each $i \in \mathbb{N}_s$, let us find bounds $\delta_{ij,U}(1)$ and $\delta_{ij,L}(1)$. For each $a < b$ and $\Delta \in \mathbf{R}$, define

$$\begin{aligned} z(\Delta; a, b) &\equiv \text{argmax}_{z \in [a,b]} \left|\Phi\left(z+\Delta\right) - \Phi\left(z\right)\right| \\ &= 1\{-\Delta < a\}a + 1\{a \leq -2\Delta \leq b\}(-2\Delta) + 1\{b \leq -2\Delta\}b. \end{aligned}$$

Also, we define if $\theta_2 \geq 0$,

$$z_i = z\left(\frac{\theta_2}{N_s - 1}; \theta_1^\top X_i + B_\rho + \frac{\theta_2}{N_s - 1}, \theta_1^\top X_i + B_\rho\right)$$

and if $\theta_2 < 0$,

$$z_i = z\left(\frac{\theta_2}{N_s - 1}; \theta_1^\top X_i + B_\rho, \theta_1^\top X_i + B_\rho + \frac{\theta_2}{N_s - 1}\right).$$

Then we obtain that $\delta_{ij,U}(1) = \delta_{ij,L}(1) = \delta_{ij}$, where

$$\delta_{ij} = \frac{1}{(1-\bar{\rho})^2 S} \cdot \left(\Phi\left(z_i + \frac{\theta_2}{N_s - 1}\right) - \Phi\left(z_i\right)\right)^2 g_{ij}(X_{I(i)})^2, \tag{3.6}$$

where $\bar{\rho} = \rho 1\{|\theta_2| > 0\}$. Therefore, the inequality in Theorem 2 holds with

$$w_{j,U}(\tau; 1) = w_{j,L}(\tau; 1) = \sqrt{-\frac{1}{2}\sum_{i=1}^{N} \delta_{ij}^2 \log\left(\frac{\tau}{2J}\right)}.$$

When $N_s = n$ for all $s = 1, \cdots, S$ (so that all the information groups are of same size n) we can check that $B_\rho = O(n^{-1/2})$ and $w(\tau) = O((nS)^{-1/2}) = O(N^{-1/2})$. As $n \to \infty$, B_ρ and $w(\tau)$ become asymptotically negligible. But when $S \to \infty$ only with n fixed, only $w(\tau)$ becomes asymptotically negligible.

When $\theta_2 = 0$, so that there is no strategic interaction among the players. Then, we obtain that $\pi_{i,U}(1) = \pi_{i,U}(1) = \Phi\left(\theta_1^\top X_i\right)$ and with $P = 1$ and $g_{ij}(x) = g_j(x)$ for all i for some nonnegative real function g_j,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\Phi\left(\theta_1^\top X_i\right) - P\left\{Y_i = 1|X_i\right\}\right) g_j(X_i) = 0.$$

Thus as for the sample version of the moment restrictions, we use this restriction to derive that

$$\frac{1}{N}\sum_{i=1}^{N}\{Y_i - \Phi\left(\theta_1^\top X_i\right)\}g_j(X_i) = \frac{1}{N}\sum_{i=1}^{N}\{Y_i - P\left\{Y_i = 1|X_i\right\}\}g_j(X_i).$$

The asymptotic analysis of the sample moments now can be done using the standard arguments. The tuning parameters τ and ρ do not have any role here.

However, when θ_2 is away from zero, B_ρ and $w(\tau)$ increase, and the testable implications from Theorem 2 become weaker, yielding wider confidence intervals for the structural parameters. This issue becomes alleviated when the sample size (i.e., the number of the players) is large.

3.3. **Bootstrap Inference.** For inference, we compare the actual actions of all the players and their predicted actions conditional on X. Let $\mathcal{P}_0$ be the collection of distributions P that the Nature uses for drawing types for individual players. Given each pure strategy equilibrium y, let $\mathcal{P}^y \equiv \{P^y : P \in \mathcal{P}_0\}$, i.e., the collection of the joint distributions of $(y(T), T)$, as the distribution of T runs in $\mathcal{P}_0$. To accommodate multiple equilibria, we let $\mathcal{P}$ be the collection of the joint distributions of $(y(T), T)$, as the distribution of T runs in $\mathcal{P}_0$ and y runs in the set of pure strategy Bayesian Nash equilibria (according to some heterogeneous beliefs $Q = (Q_1, \cdots, Q_N)$). We search for an inference procedure that is *robust* against any choice of distributions in $\mathcal{P}$, that is, against any choice of the probability by the Nature, any equilibrium among the multiple equilibria, and any beliefs formed by the individual players (within the boundary set by Assumptions 1 and 2). From now on, we do not distinguish between P^y and P and simply write P to denote a generic joint distribution of $(y(T), T)$ (or the corresponding marginal distribution of T) from the collection $\mathcal{P}$.

First, we define

$$r_{i,U}(\mathbf{a}_i;\theta) \equiv 1\{Y_{I(i)} = \mathbf{a}_i\} - \left(1 - \frac{1}{1-\rho_i} \cdot \pi_{i,L}(\mathbf{a}_i)\right) \text{ and} \tag{3.7}$$

$$r_{i,L}(\mathbf{a}_i;\theta) \equiv 1\{Y_{I(i)} = \mathbf{a}_i\} - \frac{1}{1-\rho_i} \cdot \pi_{i,U}(\mathbf{a}_i).$$

Let, for $\mathbf{a} \in \mathbb{D}$,

$$L_{j,U}(\mathbf{a};\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} r_{i,U}(\mathbf{a}_i;\theta) g_j(X_{I(i)}) \text{ and} \tag{3.8}$$

$$L_{j,L}(\mathbf{a};\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} r_{i,L}(\mathbf{a}_i;\theta) g_j(X_{I(i)}).$$

Using the sums $L_{j,U}(\mathbf{a};\theta)$ and $L_{j,L}(\mathbf{a};\theta)$, we take the following as our test statistic:

$$T(\theta) = \frac{1}{J|\mathbb{D}|}\sum_{\mathbf{a}\in\mathbb{D}}\sum_{j=1}^{J}\left([L_{j,U}(\mathbf{a};\theta) - w_U(\tau)]_+ + [L_{j,L}(\mathbf{a};\theta) + w_L(\tau)]_-\right)^2, \tag{3.9}$$

where $[x]_+ = \max\{x, 0\}$ and $[x]_- = \max\{-x, 0\}$, $x \in \mathbf{R}$.

Although the test statistic takes a similar form as many researches in the literature of moment inequalities (e.g. Rosen (2008), Andrews and Soares (2010), and Andrews and Shi (2010) among others), the sample moments $L_{j,U}(\mathbf{a};\theta)$ and $L_{j,L}(\mathbf{a};\theta)$ here are not necessarily the sum of independent or conditionally independent random variables as typically assumed in the literature. The summands $r_{i,U}(\mathbf{a};\theta)g_j(X_{I(i)})$ and $r_{i,L}(\mathbf{a};\theta)g_j(X_{I(i)})$ involve $\mathcal{W}_i$ whose subvector is Y_{-i} for each i. Since Y_{-i} are obviously dependent across i's, the summands $r_{i,U}(\mathbf{a};\theta)g_j(X_{I(i)})$ and $r_{i,L}(\mathbf{a};\theta)g_j(X_{I(i)})$ are dependent across i's in a complicated manner.

We use the testable implications in Theorem 2 to deal with this issue. First, we write

$$L_{j,U}(\mathbf{a};\theta) = \zeta_j(\mathbf{a}) + \frac{1}{N}\sum_{i=1}^{N} e_{i,U}(\mathbf{a})g_j(X_{I(i)}). \tag{3.10}$$

where $\zeta_j(\mathbf{a}) \equiv \sum_{i=1}^{N} r_i^*(\mathbf{a})g_j(X_{I(i)})$ and

$$r_i^*(\mathbf{a}) \equiv 1\{Y_{I(i)} = \mathbf{a}_i\} - P\{Y_{I(i)} = \mathbf{a}_i|X_{I(i)}\}.$$

The first sum $\zeta_j(\mathbf{a})$ is the sum of the mean zero random variables, and the second sum in (3.10) is the sum of the terms that depend on $\mathcal{W}_i$ for each i, rendering its asymptotic analysis complex. Using the similar inequality for $L_j^L(\mathbf{a};\theta)$ and applying Theorem 2, we deduce that with probability at least greater than $1-\tau$,

$$\begin{aligned} &[L_{j,U}(\mathbf{a};\theta) - w_U(\tau)]_+ + [L_{j,L}(\mathbf{a};\theta) + w_L(\tau)]_- \\ \leq\ &[\zeta_j(\mathbf{a})]_+ + [\zeta_j(\mathbf{a})]_- = |\zeta_j(\mathbf{a})|. \end{aligned} \tag{3.11}$$

We base the inference on the asymptotic distribution of $\zeta_j(\mathbf{a})$. By Assumption 6(i) and the condition that the acquaintance groups are not overlapping, we can easily show that $\zeta_j(\mathbf{a})$ is a sum of martingale difference arrays. Under some regularity conditions, the martingale central limit theorem gives us the following: for $\mathbf{a} \in \mathbb{A}^q$ (as $N \to \infty$)

$$r_N \zeta_j(\mathbf{a}) \stackrel{D}{\to} \sigma_j(\mathbf{a})\mathbb{Z}_j(\mathbf{a}), \tag{3.12}$$

jointly over j's and $\mathbf{a}$'s, where $r_N \to \infty$ is an appropriate normalizing sequence, $\mathbb{Z}_j(\mathbf{a})$ is a random variable distributed as $N(0,1)$ and $\sigma_j^2(\mathbf{a})$ is a nonnegative random variable independent of $\mathbb{Z}_j(\mathbf{a})$.

It remains to obtain an approximate distribution of $\sigma_j(\mathbf{a})\mathbb{Z}_j(\mathbf{a})$ in large samples. Andrews (2005) and Kuersteiner and Prucha (2012) used random norming to pivotize the test statistic. However, this is not possible in our case for two reasons. First, the quantities $\sigma_j(a)$ involve $P\{Y_{I(i)} = \mathbf{a}_i|X_{I(i)}\}$ and the nonparametric functions $P\{Y_{I(i)} = \mathbf{a}_i|X_{I(i)} = \cdot\}$ are heterogeneous across i's unless we assume a symmetric equilibrium where $y_1 = \cdots = y_N$. Second, the test is on inequality restrictions rather than equality restrictions. Thus, we cannot pivotize the test, for example, by using an inverse covariance matrix.

To deal with this situation, we first propose a benchmark method of constructing bootstrap critical values that are shown later to be asymptotically valid, and then develop a way to improve the inference when all the players have hindsight regrets that are asymptotically negligible.

3.3.1. *A Benchmark Method.* We adapt the cluster wild bootstrap method proposed by Cameron, Gelbach, and Miller (2008) to our inference environment. We first relabel the acquaintance groups as

$$\{I_m : m = 1, \cdots, N_0\} = \{I(i) : i \in \mathbb{N}\}. \tag{3.13}$$

Then we first take sufficiently large integer B and draw $\varepsilon_{m,b}$'s with $m = 1, \cdots, N_0$, and $b = 1, \cdots, B$, from a common distribution with mean zero and variance one (independently and identically distributed across m's, and b's.) (For each (m, b), the random numbers $\varepsilon_{m,b}$ are taken to be common across $j = 1, \cdots, J$.) We consider the following bootstrap test statistic:

$$T_b^* \equiv \sum_{\mathbf{a} \in \mathbb{A}^q} \sum_{j=1}^{J} \left(\frac{1}{N} \sum_{i=1}^{N} \hat{Z}_{ij}(\mathbf{a}_i) \varepsilon_{m(i),b} \right)^2, \; b = 1, \cdots, B,$$

where $\varepsilon_{m(i),b} = \varepsilon_{m,b}$ if and only if $I(i) = I_m$, and

$$\hat{Z}_{ij}(\mathbf{a}) \equiv 1\{Y_{I(i)} = \mathbf{a}_i\} g_j(X_{I(i)}) - \frac{1}{N} \sum_{k=1}^{N} 1\{Y_{I(k)} = \mathbf{a}_k\} g_j(X_{I(k)}).$$

Let $c_{1-\alpha,B}^*$ be the $(1-\alpha+\tau(\theta))$-th percentile of the bootstrap test statistics T_b^*, $b = 1, 2, \cdots, B$, where

$$\tau(\theta) = \frac{\tau}{2} \left(1\left\{ |\bar{\delta}_U| > 0 \right\} + 1\left\{ |\bar{\delta}_L| > 0 \right\} \right). \tag{3.14}$$

The confidence set for $\theta \in \Theta$ is defined to be

$$\mathcal{C}_B = \left\{ \theta \in \Theta : T(\theta) \le \max\{c_{1-\alpha,B}^*, \varepsilon\} \right\}, \tag{3.15}$$

where $\varepsilon > 0$ is a fixed small number such as 0.001. The maximum in the critical value in $\mathcal{C}_B$ is introduced to ensure the uniform validity of the bootstrap confidence set even when the test statistic becomes degenerate.

Conveniently, the critical value $c_{1-\alpha,B}^*$ depends on $\theta \in \Theta$ only through $\tau(\theta)$, not through the bootstrap test statistic T_b^*. This expedites the computation of the confidence set substantially. This confidence set is asymptotically valid as shown in Theorem 3 below.

THEOREM 3: *Suppose that Assumptions 1-6 hold and that there exist $c > 0$ and $\nu > 0$ such that for all $s = 1, \cdots, S$, and $i \in \mathbb{N}_s$,*

$$\inf_{P \in \mathcal{P}} P\left\{ \mathbf{E}\left[||g_j(X_{I(i)})||^{4+\nu} | C \right] < c \right\} = 1. \tag{3.16}$$

Then

$$\liminf_{N_0 \to \infty} \inf_{P \in \mathcal{P}} P\left\{\theta_0 \in \mathcal{C}_\infty\right\} \geq 1 - \alpha.$$

The condition (3.16) is the moment condition for $g_j(X_{I(i)})$. This condition is satisfied if one take indicator functions or some polynomials of $X_{I(i)}$ up to r degree, and $\mathbf{E}[||X_{I(i)}||^{r(4+v)}|C]$ is finite almost everywhere.

To see the intuition of why the bootstrap method works, first note that for each $\theta \in \Theta$, the distribution of $T(\theta)$ is first order stochastically dominated by that of

$$\sum_{\mathbf{a} \in \mathbb{A}^q} \sum_{j=1}^{J} \left([\zeta_j(\mathbf{a})]_+ + [\zeta_j(\mathbf{a})]_-\right)^2 = \sum_{\mathbf{a} \in \mathbb{A}^q} \sum_{j=1}^{J} \zeta_j^2(\mathbf{a}), \tag{3.17}$$

by (3.11). By comparing the variances, one can show that the asymptotic distribution of the last quantity is again first order stochastically dominated by the asymptotic distribution of

$$\sum_{\mathbf{a} \in \mathbb{A}^q} \sum_{j=1}^{J} \left(\frac{1}{N} \sum_{i=1}^{N} Z_{ij}(\mathbf{a}_i)\right)^2, \tag{3.18}$$

where

$$Z_{ij}(\mathbf{a}_i) \equiv 1\{Y_{I(i)} = \mathbf{a}_i\} g_j(X_{I(i)}) - \mathbf{E}\left[1\{Y_{I(i)} = \mathbf{a}_i\} g_j(X_{I(i)}) | \mathcal{F}\right],$$

and $\mathcal{F}$ is a certain sigma field. The last sum is approximated by the conditional distribution of T_b^* given (Y, X) when N and B are sufficiently large. Since

$$P\{T_b^* > \max\{c_{1-\alpha,B}^*, \varepsilon\} | Y, X\} \leq \alpha,$$

by the definition of $c_{1-\alpha,B}^*$, we also have

$$P\{T(\theta) > \max\{c_{1-\alpha,B}^*, \varepsilon\} | \mathcal{F}\} \precsim \alpha,$$

where $\precsim$ denotes inequality that holds in the limit. Hence taking the expectation on both sides, we find that the bootstrap test is still asymptotically valid. In the proof of Theorem 3 in the Appendix, we make this approximation rigorous using a Berry-Esseen bound for martingale difference arrays, to ensure validity that is uniform over $P \in \mathcal{P}$.

Since we allow for the types within each acquaintance group to be correlated in an arbitrary way, the effective sample size among the N players is in fact the number of acquaintance groups N_0. It is worth noting that the asymptotic validity of the confidence sets in Theorem 3 does not require any particular way the ratio N_0/N behaves asymptotically as $N \to \infty$. See Hansen (2007) for a similar observation in the context of estimating a robust covariance matrix in linear panel models. More recently Ibragimov and Müller (2010) suggested a method of very general robust inference based on t-statistic. Their method is directly applicable here, unless the dimension J is one and the game is a binary action game.

3.3.2. *Modified Bootstrap with Asymptotically Negligible Hindsight Regrets.* We assume that the hindsight regrets are asymptotically negligible as $N_s \to \infty$ for each $s \in \mathbb{S}$. Many models of large games with semianonymous payoffs exhibit belief-free hindsight regrets that are asymptotically negligible. Typically when the belief-free hindsight regrets are asymptotically negligible as the number of the players goes to infinity, we can show that $\bar{\delta}_U$ and $\bar{\delta}_L$ are asymptotically negligible, so that the bounds $w_U(\tau)$ and $w_L(\tau)$ become also asymptotically negligible. In this section, we propose an alternative inference method that utilizes this fact.

We consider the following bootstrap test statistic:

$$T_b^{Mod*} \equiv \frac{1}{J|\mathbb{D}|} \sum_{\mathbf{a} \in \mathbb{D}} \sum_{j=1}^{J} \left(\left[L_{j,U,b}^*(\mathbf{a};\theta) \right]_+ + \left[L_{j,L,b}^*(\mathbf{a};\theta) \right]_- \right)^2, \ b = 1, \cdots, B,$$

where

$$\begin{aligned} L_{j,U,b}^*(\mathbf{a};\theta) &= \frac{1}{N} \sum_{i=1}^{N} r_{i,U}(\mathbf{a}_i;\theta) g_j(X_{I(i)}) \varepsilon_{m(i),b} \text{ and} \\ L_{j,L,b}^*(\mathbf{a};\theta) &= \frac{1}{N} \sum_{i=1}^{N} r_{i,L}(\mathbf{a}_i;\theta) g_j(X_{I(i)}) \varepsilon_{m(i),b}. \end{aligned}$$

We take $c_{1-\alpha,B}^{Mod*}(\theta)$ to be the $(1-\alpha+\tau(\theta))$-th percentile of the bootstrap test statistics T_b^{Mod*}, $b = 1, 2, \cdots, B$, and define the confidence set for $\theta \in \Theta$ to be

$$\mathcal{C}_B^{Mod} = \left\{ \theta \in \Theta : T(\theta) \le \max\{ c_{1-\alpha,B}^{Mod*}(\theta), \varepsilon \} \right\}.$$

To see how this method achieves validity, we note that

$$\begin{aligned} L_{j,U,b}^*(\mathbf{a};\theta) &= \frac{1}{N} \sum_{i=1}^{N} r_{i,U}(\mathbf{a}_i;\theta) g_j(X_{I(i)}) \varepsilon_{m(i),b} \\ &= \frac{1}{N} \sum_{i=1}^{N} r_i^*(\mathbf{a}_i) g_j(X_{I(i)}) \varepsilon_{m(i),b} + \frac{1}{N} \sum_{i=1}^{N} e_{i,U}(\mathbf{a}_i) g_j(X_{I(i)}) \varepsilon_{m(i),b}. \end{aligned}$$

The conditional variance of the last term given (Y, X) is equal to

$$\left(\frac{N_0}{N^2} \right) \cdot \frac{1}{N_0} \sum_{m=1}^{N_0} \left(\sum_{i \in I_m} e_{i,U}(\mathbf{a}_i) g_j(X_{I(i)}) \right)^2.$$

We can show that when the hindsight regrets are asymptotically negligible (or more specifically, $\bar{\psi}_{ij}$ in Theorem 2 is asymptotically negligible uniformly over $i \in \mathbb{N}$), we have the last term also asymptotically negligible. This means that we have

$$L^*_{j,U,b}(\mathbf{a};\theta) \approx \frac{1}{N}\sum_{i=1}^{N} r^*_i(\mathbf{a}_i) g_j(X_{I(i)})\varepsilon_{m(i),b}, \tag{3.19}$$

and similarly, using the fact that $\bar{\delta}_L$ is also asymptotically negligible,

$$L^*_{j,L,b}(\mathbf{a};\theta) \approx \frac{1}{N}\sum_{i=1}^{N} r^*_i(\mathbf{a}_i) g_j(X_{I(i)})\varepsilon_{m(i),b}. \tag{3.20}$$

The bootstrap distribution of the last sum approximates the distribution of $\zeta_j(\mathbf{a})$. From (3.11), we conclude that this modified bootstrap is asymptotically valid. We formally state the result as follows.

Theorem 4: *Suppose that the conditions of Theorem 3 hold. Furthermore assume that as* $N \to \infty$, *we have* $\max_{i\in\mathbb{N}} \bar{\psi}_{ij} \to 0$ *for all* $j = 1, \cdots, J$. *Then*

$$\liminf_{N\to\infty} \inf_{P\in\mathcal{P}} P\left\{\theta_0 \in \mathcal{C}^{Mod}_{\infty}\right\} \geq 1-\alpha.$$

While the confidence set $\mathcal{C}^{Mod}_B$ is less conservative than $\mathcal{C}_B$, it is asymptotically valid only when the hindsight regrets of all the players are asymptotically negligible.

It is worth noting that the wild bootstrap procedure preserves both the cross-sectional dependence due to the acquaintance group structure and the cross-sectional dependence due to the public signals. Cross-sectional dependence due to the acquaintance group structure is captured by using the same $\varepsilon_{m,b}$ for individuals in the same acquaintance group. Cross-sectional dependence due to the public signals is preserved by the wild bootstrap procedure. To see this latter fact, assume for simplicity a private information game with a single information group with a public signal C, and take $J = 1$ and $\mathbb{A} = \{0, 1\}$ so that we have in this case

$$N \cdot T^{Mod*}_b \approx \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} r^*_i(1) g_j(X_i)\varepsilon_{i,b}\right)^2, \; b = 1, \cdots, B.$$

The conditional variance of the sum is equal to

$$\frac{1}{N}\sum_{i=1}^{N} r^{*2}_i(1) g^2_j(X_i).$$

By the law of the large numbers applied to conditionally independent sums, the last sum is approximated by

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{E}\left[r^{*2}_i(1) g^2_j(X_i)|C\right],$$

which is also the conditional variance of $\zeta_1(1)$ given C. Hence the cross-sectional dependence through the presence of public signal is captured by the wild bootstrap in this case.

As mentioned before, the cluster bootstrap method that this paper relies on was originally proposed by Cameron, Gelbach, and Miller (2008) for linear regression models to deal with within-cluster dependence. Their wild bootstrap procedure draws regression residuals multiplied by random variables with a two-point distribution. The latter random variables are chosen to be cluster-specific, to preserve the cross-sectional dependence structure. The main difference here is that we have two layers of cross-sectional dependence one with acquaintance groups and the other with information groups. While both groups can be viewed as clusters, our bootstrap method controls only the acquaintance group structure. The main reason is that the cross-sectional dependence within information group is due to some public signals, and can be captured by the wild bootstrap procedure as we saw before. Note that this cross-sectional dependence will not be captured by simply resampling (with replacement) from acquaintance groups in general.

3.4. **Partial Observation of the Players.** When there are many players in a game, it is not rare that the econometrician does not observe all the players. The issue that the econometrician faces in the case of partial observation of the players is distinct from the well-known issue of missing values in the econometrics literature, because the payoff function that the econometrician uses should depend on the actions of all the players (including those unobserved by the econometrician). Under certain conditions, the hindsight regret approach delivers an asymptotically valid inference method even when the econometrician observes only part of the players.

Condition 1: (i) The game is a private information game consisting of several or many information groups indexed by $s = 1, \cdots, S$, and players from different information groups are payoff disjoint (Assumption 3).
(ii) In each large game, the payoff differential is additive in the proportion of the actions of the other players in his payoff group, i.e.,

$$\Delta u_{mk}(a_{-i}; t_i) = v_{mk}(t_i) + \theta \cdot (\bar{a}_m - \bar{a}_k) \cdot \frac{1}{N_s - 1} \sum_{j \neq i}^{N_s} a_j.$$

Condition 2: The outcomes observed by the econometrician are realizations from a pure strategy Bayesian Nash equilibrium, where within each information group, the strategies are identical among the players, and the types are drawn from the same distribution by the Nature.

Condition 1(ii) tells us that the expected payoff differential does not depend on the number or identity of the other players. Condition 2 has the consequence that the observed outcomes have the identical distribution across the players within each information group.

When Conditions 1-2 are satisfied, we have for all $m, k = 1, \cdots, K$,

$$\mathbf{E}\left[u_{mk}^{\Delta}(Y_{-i}; T_i)|T_i\right] = v_{mk}(T_i) + \theta \cdot (\bar{a}_m - \bar{a}_k)\mathbf{E}[Y_j|T_i],$$

where player j belongs to player i's information group. The last conditional expectation $\mathbf{E}[Y_j|T_i]$ does not depend on the number of the players in the large game, so one can proceed with only a subsample of the players. More specifically, suppose that the set of players in information group s is a proper subset $\mathbb{N}_s^1$ of $\mathbb{N}_s$. Then the *ex post* version of the equilibrium constraints can be formulated using the players in $\mathbb{N}_s^1$, where one uses $(N_s^1 - 1)^{-1} \sum_{j \in \mathbb{N}_s^1 \setminus \{i\}} Y_j$, instead of $\mathbf{E}[Y_j|T_i]$. To address the bias caused by this replacement, one uses the hindsight regret bound in Theorem 1 which becomes in this case $B_{i,\rho}(t_i)$ as in Section 2.3.1, except that we use N_s^1 now instead of N_s. Using N_s^1 instead of N_s makes the hindsight regret bound larger. Hence the moment inequalities in Theorem 2 still hold for the players in the subsample $\mathbb{N}_s^1$. Therefore, the asymptotic validity of the inference procedure is not affected, when one uses only part of the moment inequalities, as long as the number of the players is still large.

4. Monte Carlo Simulation Studies

4.1. **Basic Data Generating Processes.** We consider S number of private information Bayesian games, where each game s is populated by N_s number of players. The action space for each player is $\{0, 1\}$. The i-th player in game $s = 1, \cdots, S$ has the following form of a payoff differential:

$$u_i(1, a_{-i}; T_i) - u_i(0, a_{-i}; T_i) = X_{i,s}\beta_0 + \phi_0 \left(\frac{1}{N_s - 1} \sum_{j=1, j \neq i}^{N_s} a_j \right) + \eta_{i,s},$$

where $X_{i,s}$ and $W_{i,s}$ are observable characteristics of individual i, and η_{is} is an unobservable characteristic. We have specified

$$X_{i,s} = Z_{i,s}/3 + \gamma_{1,0} C_{1,s} - 0.2. \tag{4.1}$$

where $Z_{i,s}$ is an idiosyncratic component and $C_{1,s}$ is a public signal that is specific to group s. The random variables $\eta_{i,s}$, $Z_{1,i,s}$, $Z_{2,i,s}$, $C_{1,s}$ and $C_{2,s}$ are drawn independently from $N(0, 1)$. The variables, $Z_{1,i,s}$'s and $Z_{2,i,s}$'s are independent across i's and s's and $C_{1,s}$'s and $C_{2,s}$'s are independent across s's.

The utility specification is similar to the one that is often used in the literature of social interactions. The parameter ϕ_0 measures the presence of social interactions. Throughout the exercise, we set β_0 to be 1.

To generate observed outcomes, we assume private information game, i.e., $I(i) = \{i\}$, and equilibrium strategies are identical within each subgroups and find any p_s for each game s such that

$$p_s = \mathbf{E}\left[\Phi\left(X_{i,s}\beta_0 - 0.2 + \phi_0 \cdot p_s\right)\right].$$

Then, we generate

$$Y_{i,s} = 1\left\{X_{i,s}\beta_0 + \phi_0 \cdot p_s + \eta_{i,s}\right\}.$$

The parameter β_0 is set to be 1. For the simulations, we assume that each game has the same number of players equal to $N_s = n$. Then, the belief-free hindsight regret for player i in game s (denoted by $B_{i,s,\rho}(T_i)$) is simplified as

$$B_{i,s,\rho}(T_i) = \sqrt{-\frac{\phi_0^2}{N_s - 1}\frac{\log(\rho)}{2}}. \tag{4.2}$$

Throughout the simulation study, we have set $\rho = 0.01$. For the inference, we set the Monte Carlo simulation number to be 1,000, and the bootstrap Monte Carlo simulation number (B) to be 1,000.[6] For the construction of moment inequalities, we used the following:[7]

$$g_1(X_{i,s}) = 1, \; g_2(X_{i,s}) = |X_{i,s}|, \text{ and } g_3(X_{i,s}) = 2 \cdot 1\{X_{i,s} \geq 0\}.$$

Throughout the simulation studies, we have chosen $\rho = 0.01$ and $\tau = 0.01$.

4.2. **Finite Sample Coverage Probabilities of Bootstrap Tests.** We first investigate the finite sample validity of the confidence intervals. For this study, we have fixed $\beta_0 = 1$ and vary ϕ_0 in $\{0, 0.5\}$. The main focus is on the role of belief-free hindsight regret. Since the belief-free hindsight regret in (4.2) is increasing in ϕ_0, we expect that the confidence set will become more conservative as ϕ_0 moves away from zero. This is analogous to a situation where all the moment inequalities are away from being binding.

The first simulation study investigate the finite sample coverage probabilities. The results are shown in Table 1. In Table 1, we observe that the benchmark bootstrap confidence set gives conservative finite sample coverage probabilities even when $\phi_0 = 0$, i.e. there is no hindsight regret. This is because the term $\sum_{\mathbf{a} \in \mathbb{A}^q} \sum_{j=1}^J \zeta_j^2(\mathbf{a})$ in (3.17) is stochastically dominated by the term $\sum_{\mathbf{a} \in \mathbb{A}^q} \sum_{j=1}^J (\frac{1}{N} \sum_{i=1}^N Z_{ij}(\mathbf{a}_i))^2$ in (3.18). Recall that the bootstrap test

[6]The simulation study in the current version used $\rho_i = 1\{\bar{\psi}_i > 0\}$ instead of using $\rho_i = \rho \cdot (c\bar{\psi}_{ii}/(1 + c\bar{\psi}_{ii}))$ for the test statistic, and the modified bootstrap procedure.

[7]An alternative choice of $g_3(X_{i,s}) = X_{i,s}^2$ was tried initially. This choice leads to a very stable coverage properties, but the power of the test for $\phi_0 = \phi$ and $\beta_0 = \beta$ was very weak. Hence the choice of the nonnegative map $g_j(\cdot)$ is crucial for sharp inference results.

relies on the latter quantity, because there is no reasonable way to simulate the distribution of the former quantity, when there exists nonnegligible hindsight regrets and the equilibrium is asymmetric with heterogeneous beliefs. This stochastic dominance is not removed even when there is no hindsight regret.

Table 1: Finite Sample Coverage Probabilities at 95%

		Benchmark		Modified	
		ϕ_0= 0.0	ϕ_0= 0.5	ϕ_0= 0.0	ϕ_0= 0.5
	N_s= 100	0.983	1.000	0.942	1.000
$S=3$	N_s= 200	0.979	1.000	0.941	1.000
	N_s= 500	0.991	1.000	0.942	1.000
	N_s= 100	0.988	1.000	0.956	1.000
$S=10$	N_s= 200	0.990	1.000	0.946	1.000
	N_s= 500	0.991	1.000	0.942	1.000
	N_s= 100	0.986	1.000	0.935	1.000
$S=50$	N_s= 200	0.991	1.000	0.939	1.000
	N_s= 500	0.992	1.000	0.949	1.000

The confidence sets from the modified bootstrap assuming asymptotically negligible hindsight regrets are shown to produce nonconservative coverage probabilities, when there is no hindsight regret ($\phi_0 = 0$). This is because when the hindsight regrets are asymptotically negligible, one can simulate the distribution $\sum_{\mathbf{a}\in\mathbb{A}^q}\sum_{j=1}^J \zeta_j^2(\mathbf{a})$ directly as we saw previously. Hence as compared to the benchmark method, the modified method is much less conservative. However, this modified method becomes conservative as ϕ_0 moves away from zero. This is the situation where we are further into the interior of the moment inequality restrictions, and hence the coverage probabilities show conservativeness.

One might think that when the hindsight regrets are asymptotically negligible, we may ignore the hindsight regrets altogether and consider a test statistic of the following:

$$\frac{1}{J|\mathbb{D}|}\sum_{\mathbf{a}\in\mathbb{D}}\sum_{j=1}^J L_j^2(\mathbf{a};\theta),$$

where $L_j(\mathbf{a};\theta)$ is equal to $L_{j,U}(\mathbf{a};\theta) = L_{j,L}(\mathbf{a};\theta)$ with $\rho_i = 0$, and $B_{ij,\rho}$ is replaced by zero, and one may use the bootstrap critical values using

$$\frac{1}{J|\mathbb{D}|}\sum_{\mathbf{a}\in\mathbb{D}}\sum_{j=1}^J L_j^{*2}(\mathbf{a};\theta),$$

where $L_j^*(\mathbf{a};\theta)$ is the same as $L_{j,U}^*(\mathbf{a};\theta)$ except that $B_{ij,\rho}$ is set to be zero. This inference is tantamount to assuming that the Bayesian Nash equilibrium that the econometrician observes is in fact its refinement as an *ex post Nash equilibrium*, where the equilibrium strategies remain an equilibrium even after all the types are revealed to the players. However, as we will see in Table 2 below, this leads to invalid inference in general.

Table 2: Finite Sample Coverage Probabilities at 90% When the Hindsight Regrets are Ignored

		Benchmark		Modified	
		$\phi_0 = -1.0$	$\phi_0 = -3.0$	$\phi_0 = -1.0$	$\phi_0 = -3.0$
	$N_s = 100$	0.894	0.757	0.837	0.723
$S = 3$	$N_s = 200$	0.899	0.741	0.828	0.699
	$N_s = 500$	0.909	0.757	0.831	0.702
	$N_s = 100$	0.895	0.752	0.830	0.731
$S = 10$	$N_s = 200$	0.887	0.739	0.814	0.694
	$N_s = 500$	0.877	0.712	0.797	0.654
	$N_s = 100$	0.905	0.739	0.840	0.705
$S = 50$	$N_s = 200$	0.900	0.706	0.839	0.662
	$N_s = 500$	0.872	0.576	0.784	0.509

Table 2 reports the finite sample coverage probabilities of the bootstrap tests when the hindsight regrets are entirely ignored in the construction of the test statistic. Here we set $\gamma_{0,1} = 0$ and $\beta_0 = 1$. The table shows that the finite sample coverage probabilities deteriorate when the hindsight regrets are ignored. (The benchmark bootstrap procedure and the modified bootstrap procedure that properly take into account hindsight regrets as proposed in this paper produced the coverage probability of 1 in this set-up.) The deterioration becomes severe as the sample size becomes larger. Therefore, when the hindsight regrets are asymptotically negligible, one cannot simply eliminate the hindsight regret component $B_{ij,\rho}$ and ρ_i in the test statistic.[8]

4.3. **Finite Sample Power of the Bootstrap Tests for Parameter Values.** From the coverage probability results that are conservative, especially when there is nonzero hindsight regret, one may be concerned about the small sample power properties of the bootstrap method. Hence we have investigated the small sample power properties of the bootstrap method.

[8]When we took ϕ_0 to be a positive number, the coverage probability deterioration did not arise.

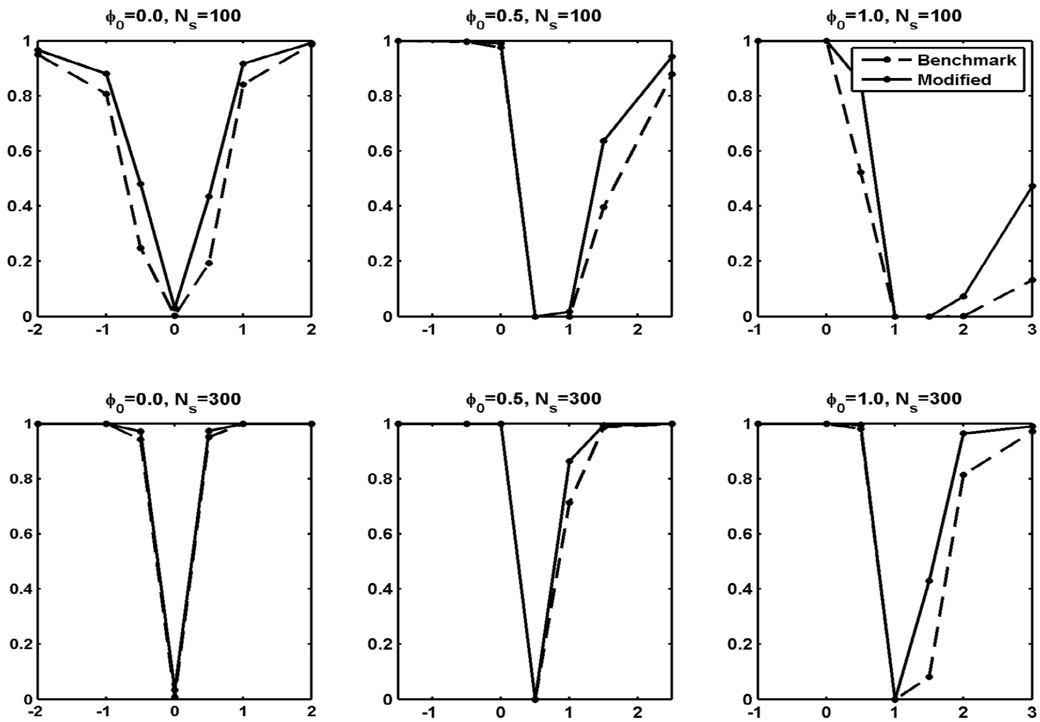

FIGURE 1. The Empirical Power of the Test: $H_0 : \phi_0 = \phi$ vs. $H_1 : \phi_0 \neq \phi$ at 5%.with $S = 10$. The x-axis represents the true value of ϕ_0. The test is constructed using the Bonferroni approach to construct a confidence set for ϕ_0. As expected, as ϕ_0 is away from zero, the hindsight regret becomes larger, and hence it becomes harder to distinguish the null hypothesis from the alternatives. Note the conspicuous improvement in power by the modified bootstrap.

We first consider the test for ϕ_0. The null hypothesis and the alternative hypothesis is given by

$$H_0 : \phi_0 = \phi \quad vs. \quad H_1 : \phi_0 \neq \phi.$$

The nominal level of the test is set to be at 5%, and the number of the information groups are set to be 10. The test is performed by constructing a Bonferroni bootstrap confidence set for ϕ_0 and checking whether the hypothesized value ϕ belongs to the the confidence set.

The results are shown in Figure 1. The x-axis represents the hypothesized value of ϕ under the null hypothesis. The upper row panels use the true value of $\phi_0 = 0.0$, 0.5, and 1.0, with the group size $N_s = 100$. And the lower row panels use the same values of ϕ_0, but with a larger group size $N_s = 300$.

From Figure 1, the bootstrap test has a higher power when $\phi_0 = 0$, but the power becomes lower as ϕ_0 is away from zero. However, it is clear that even when the empirical coverage probability for the bootstrap is highly conservative (i.e., when $\phi_0 = 0.5$ and $\phi_0 = 1.0$),

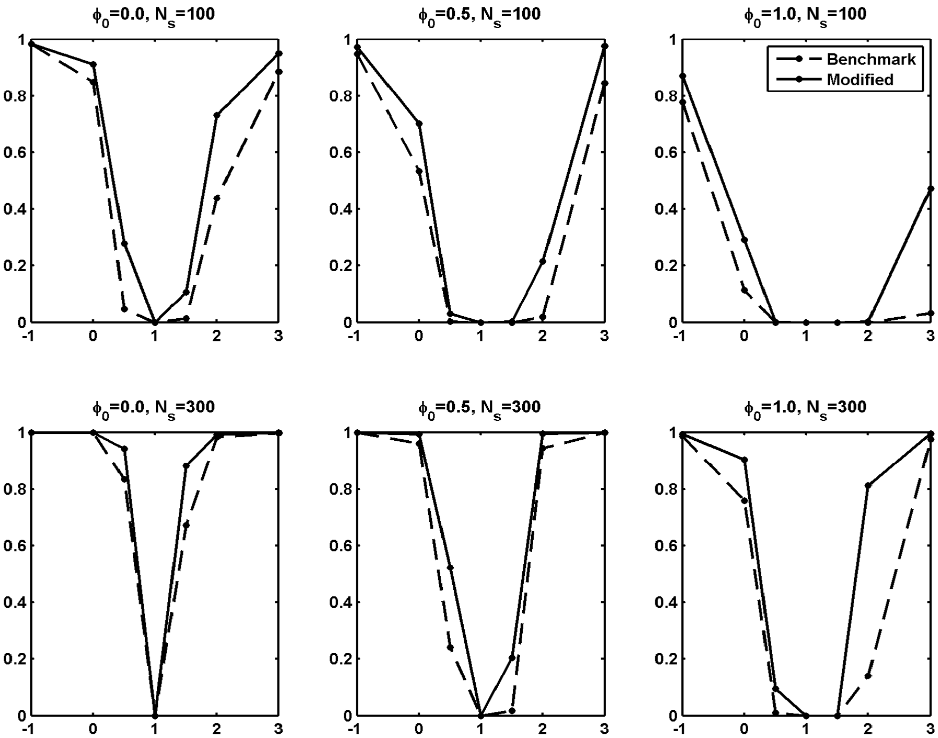

FIGURE 2. The Empirical Power of the Test: $H_0 : \beta_0 = \beta$ vs. $H_1 : \beta_0 \neq \beta$ at 5%. .with $S = 10$. The x-axis represents the value of ϕ_0.under the alternative hypothesis. The test is constructed using the Bonferroni approach to construct a confidence set for β_0. As expected, as ϕ_0 is away from zero, the hindsight regret becomes larger, and hence it becomes harder to distinguish the null hypothesis from the alternatives. Note the conspicuous improvement in power by the modified bootstrap.

this does not necessarily mean that the test has negligible power for the range of values ϕ considered here.

The power improvement by the modified bootstrap procedure appears conspicuous. Again, even when the coverage probabilities are both 1 almost always for the case of $\phi_0 = 0.5$ and $\phi_0 = 1$, the power improvement achieved by the modified bootstrap looks outstanding. Also, as expected, the increased information group size increases the power of the test, apparently by reducing the size of hindsight regrets.

One can expect that the conservativeness of the bootstrap test also affects the estimation of β_0, depending on the size of the hindsight regrets. To investigate this, we consider the test for β_0 now:

$$H_0 : \beta_0 = \beta \quad vs. \quad H_1 : \beta_0 \neq \beta$$

with different values of ϕ_0. Again, the test is performed by constructing a Bonferroni bootstrap confidence set for β_0 and checking whether the hypothesized value β belongs to the the confidence set.

Similarly as in the case of testing $\phi_0 = \phi$, the test shows decrease in power as we move ϕ_0 away from zero, due to higher hindsight regrets. Also, note that the power improvement by the modified bootstrap method appears conspicuous throughout all the cases.

One might be concerned about the lower power properties when ϕ_0 is away from zero. While one may investigate further and produce a method with better power properties, the low power with larger hindsight regrets is general a phenomenon not with the method, but inherent in the inference problem, because inference in the presence of belief heterogeneity and its larger role in the players' strategic decisions will obviously make the inference problem harder.

Lastly, the bootstrap test performs very well in testing the null hypothesis of no strategic interaction $H_0 : \phi_0 = 0$ vs. $H_1 : \phi_0 \neq 0$. The test performs quite well because the null hypothesis corresponds to the case with no hindsight regret.

4.4. **Bootstrap Validity in the Presence of Public Signals.** Previously, we emphasized that the modified bootstrap test is designed to capture both layers of cross-sectional dependence, one due to the acquaintance groups and the other due to the information groups. Since we are focusing on the private information game here, the focus now is on the cross-sectional dependence due to the public signals.

The magnitude of the public signals is determined by the parameter $\gamma_{0,1}$. When $\gamma_{0,1} = 0$, there is no public signal, but when $\gamma_{0,1}$ is large, there is a large role for the public signal in the game.

We investigate the finite sample coverage probabilities by using different values of $\gamma_{0,1}$. The results are shown in Table 3. The number of the information groups (S) was set to be 10. We focus on the case of no hindsight regret ($\phi_0 = 0$). The parameter $\gamma_{0,1}$ was chosen from $\{0, 1/3, 1\}$.

Table 3 shows that the benchmark bootstrap method becomes more and more conservative as the public signals become larger. However, the modified bootstrap method gives stable coverage probabilities over different values of $\gamma_{0,1}$. This supports our observation that the modified bootstrap captures the cross-sectional dependence structure well.

Here are a summary of the results from the Monte Carlo study. First, as expected, the inference becomes more conservative, as the strategic interaction (among unacquainted players) becomes larger (i.e. ϕ_0 becomes away from zero.) In particular, the coverage probabilities becomes 1 mostly when $\phi_0 = 0.5$ or larger.

Table 3: Coverage Probabilities with Various Magnitudes of Public Signals: $\phi_0 = 0$, $\beta_0 = 1$, and $S = 10$

		90%			95%		
		$\gamma_{1,0}= 0$	$\gamma_{1,0}= 1/3$	$\gamma_{1,0}= 1$	$\gamma_{1,0}= 0$	$\gamma_{1,0}= 1/3$	$\gamma_{1,0}= 1$
	$N_s= 100$	0.957	0.972	0.988	0.989	0.988	0.995
Benchmark	$N_s= 200$	0.968	0.969	0.985	0.987	0.990	0.995
	$N_s= 500$	0.967	0.965	0.992	0.987	0.991	0.997
	$N_s= 100$	0.885	0.909	0.912	0.940	0.956	0.949
Modified	$N_s= 200$	0.902	0.904	0.893	0.944	0.946	0.958
	$N_s= 500$	0.906	0.897	0.903	0.956	0.942	0.946

Second, the power properties reflect the same phenomenon. As ϕ_0 becomes away from zero, the empirical power of the tests for ϕ_0 or for β_0 becomes lower, as expected. However, it should be noted that even when the coverage probability becomes 1, the test still shows reasonable power properties. In particular, the performance of modified bootstrap method shows outstanding improvement over the benchmark method, and hence this method is recommendable, when all the players are thought to have asymptotically negligible hindsight regrets. It is also shown that even when the hindsight regrets are thought to be asymptotically negligible, one should not remove them from the construction of the test statistic, as it invalidates the inference procedure.

Finally, the presence of public signals tend to make the benchmark bootstrap test more conservative. In contrast, the modified bootstrap method is not affected by the presence the public signals. This reflects that the modified bootstrap captures the cross-sectional dependence characterized by the public signals.

5. Conclusion

This paper proposes a large game perspective for empirical researches on many agents interacting with each other. For this, the paper develops a general inference framework for a large Bayesian game where the econometrician observes a Nash equilibrium profile. The development proceeds by fully utilizing the structure of cross-sectional dependence that is derived from the structural primitives of the game through the notion of hindsight regret. The hindsight regret approach has the merit of not relying on the assumptions about the way individual players form their beliefs about other players' types. Hence the approach immediately accommodates the situation where players have heterogeneous beliefs. In this framework, nontrivial inference is still possible when the players' payoffs are impacted asymptotically negligibly by a deviation by another player outside his acquaintance group.

6. Appendix

The following is a conditional version of McDiarmid's inequality (McDiarmid (1989)) which is used in the proof of Theorem 1.

Lemma A1 (McDiarmid's Inequality): *Let $X = (X_1, \cdots, X_m)$ be a vector of random vectors taking values in a subset $\mathcal{X}$ of a Euclidean space and X_i's are conditionally independent given a random vector Z. Then for any $f : \mathcal{X}^m \to \mathbf{R}$, and all $\varepsilon > 0$,*

$$P\left\{f(X) - \mathbf{E}\left[f(X)|Z\right] \geq \varepsilon|Z\right\} \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^m V_i(f)^2}\right),$$

where $V_i(f)$ is the maximal variation of f along the i-th coordinate.

Proof of Theorem 1: Given a pure strategy equilibrium y, let $u_i(a) \equiv u_i(a, Y_{-i}; T_{I(i)})$. For each $s = 1, \cdots, S$, let $\mathbb{N}_s$ be the set of players in information group s. The finite collection $\{T_{I(l)} : l \in \mathbb{N}_s\}$ is conditionally independent given C_s by Assumption 1(i). (Also recall that T_i for each $i \in \mathbb{N}_s$ already contains C_s as its subvector.)

For any $\lambda > 0$, $a \in \mathbb{A}\backslash\{Y_j\}$, $i \in \mathbb{M}_s$, and $j \in I(i)$, we write

$$\begin{aligned} (6.1) \qquad & Q_i^y\left\{u_j(Y_j) - u_j(a) \leq -\lambda|T_{I(i)}\right\} \\ = \; & Q_i^y\left\{\begin{array}{c} u_j(Y_j) - u_j(a) - \mathbf{E}_i^y\left[u_j(Y_j) - u_j(a)|T_{I(i)}\right] \\ \leq -\lambda - \mathbf{E}_i^y\left[u_j(Y_j) - u_j(a)|T_{I(i)}\right] \end{array} |T_{I(i)}\right\} \\ \leq \; & Q_i^y\left\{u_j(Y_j) - u_j(a) - \mathbf{E}_i^y\left[u_j(Y_j) - u_j(a)|T_{I(i)}\right] \leq -\lambda|T_{I(i)}\right\}, \end{aligned}$$

where the last inequality holds everywhere. We write

$$\begin{aligned} & u_j(Y_j) - u_j(a) - \mathbf{E}_i^y\left[u_j(Y_j) - u_i(a)|T_{I(i)}\right] \\ = \; & v_j(a, T) - \mathbf{E}_i^y\left[v_j(a, T)|T_{I(i)}\right], \end{aligned}$$

where $v_j(a, t) \equiv u_j(y_j(t_j), y_{-j}(t); t_{I(j)}) - u_j(a, y_{-j}(t); t_{I(j)})$ and $y_{-j}(t) = (y_k(t_{I(k)}))_{k=1,k\neq j}^N$. When we choose $l \in \mathbb{N}_s/I(i)$ and perturb player l's type t_l, this alters $v_j(a, t)$ through its effect on the actions of all players k who observe the type of player l, i.e., for all $k \in I(l)$. Hence this perturbation affects $v_j(a, t)$ by not more than

$$\sum_{k \in I(l)} \Delta_{jk}(t_{I(j)}, a).$$

(We do not need to consider players outside $\mathbb{N}_s$ due to the assumption of payoff disjointedness across the information groups.) We replace λ by $B_{ij,\rho}(T_{I(j)}, a)$ in (6.1) and use Lemma A1 to bound the last probability in (6.1) by

$$\exp\left(-\frac{2B_{ij,\rho}^2(T_{I(j)},a)}{\sum_{l\in\mathbb{M}_s\backslash I(i)}\left(\sum_{k\in I(l)}\Delta_{jk}(T_{I(j)},a)\right)^2}\right)=\exp\left(-\frac{2B_{ij,\rho}^2(T_{I(j)},a)}{\psi_{ij}(T_{I(j)},a)}\right), \tag{6.2}$$

where the inequality follows by the definition of $\psi_j(T_j)$. Now

$$\begin{aligned}&P\left\{u_j(Y_j)-u_j(a)\leq -B_{ij,\rho}(T_{I(j)},a)\text{ for some }(a,j)\in(\mathbb{A}\backslash\{Y_j\})\times I(i)|T_{I(i)}\right\}\\ \leq\ &(K-1)\sum_{j\in I(i)}\exp\left(-\frac{2B_{ij,\rho}^2(T_{I(j)},a)}{\psi_{ij}(T_{I(j)},a)}\right)=\rho,\end{aligned}$$

where we used the definition of $B_{ij,\rho}(T_{I(j)})$ and the fact that $|\mathbb{A}\backslash\{Y_j\}|=K-1$. We conclude that

$$Q_i^y\left\{u_j(Y_j)-u_j(a)>-B_{ij,\rho}(T_{I(j)},a),\ \forall a\in\mathbb{A}\backslash\{Y_j\}\text{ and }\forall j\in I(i)|T_{I(i)}\right\}>1-\rho,$$

which is the desired result. *Q.E.D.*

Proof of Theorem 2: Given a pure strategy equilibrium y, let $u_i(a_{I(i)})\equiv u_i(a_{I(i)},Y_{-I(i)};T_{I(i)})$ as before. Define the events:

$$S_{i,U}(a_{I(i)})\equiv\left\{\begin{array}{c}u_j(a_{I(j)})-u_j(c,a_{I(j)\backslash\{j\}})\geq -B_{ij,\rho}(T_{I(j)},c)\\ \text{for all }c\in\mathbb{A}\text{ and all }j\in I(i)\end{array}\right\}.$$

By the definition of $B_{ij,\rho}(T_{I(j)})$, Assumption 1(ii), and Theorem 1, we have (everywhere)

$$P(S_{i,U}(Y_{I(i)})|T_{I(i)})\geq 1-\rho. \tag{6.3}$$

Now, observe that

$$P(S_{i,U}(Y_{I(i)})|T_{I(i)})=\sum_{a_{I(i)}\in\mathbb{A}_{I(i)}}P(S_{i,U}(a_{I(i)})|T_{I(i)})1\left\{Y_{I(i)}=a_{I(i)}\right\}\geq 1-\rho. \tag{6.4}$$

From this, we deduce that

$$1\left\{Y_{I(i)}=a_{I(i)}\right\}\leq 1\left\{P(\tilde{S}_{i,U}(a_{I(i)})|T_{I(i)})\geq 1-\rho\right\}, \tag{6.5}$$

where $\tilde{S}_{i,U}(a_{I(i)})\equiv S_{i,U}(a_{I(i)})\cap\left\{Y_{I(i)}=a_{I(i)}\right\}$. Similarly also from (6.4), we have

$$\begin{aligned}1\left\{Y_{I(i)}\neq a_{I(i)}\right\}\ &\leq\ 1\left\{\sum_{d_{I(i)}\in\mathbb{A}_{I(i)}\backslash\{a_{I(i)}\}}P(S_{i,U}(d_{I(i)})|T_{I(i)})1\left\{Y_{I(i)}=d_{I(i)}\right\}\geq 1-\rho\right\}\\ &=\ 1\left\{\sum_{d_{I(i)}\in\mathbb{A}_{I(i)}\backslash\{a_{I(i)}\}}P(S_{i,U}(d_{I(i)})\cap\left\{Y_{I(i)}=d_{I(i)}\right\}|T_{I(i)})\geq 1-\rho\right\}.\end{aligned}$$

Since $S_{i,U}(d_{I(i)}) \cap \left\{Y_{I(i)} = d_{I(i)}\right\}$ is disjoint across $d_{I(i)}$'s, we conclude that

$$1\left\{Y_{I(i)} \neq a_{I(i)}\right\} \leq 1\left\{P\left(\tilde{S}_{i,L}(a_{I(i)})|T_{I(i)}\right) \geq 1-\rho\right\},$$

where

$$\tilde{S}_{i,L}(a_{I(i)}) \equiv \bigcup_{d_{I(i)} \in \mathbb{A}_{I(i)} \backslash \{a_{I(i)}\}} S_{i,U}(d_{I(i)}) \cap \left\{Y_{I(i)} = d_{I(i)}\right\}.$$

Recall the definition $\mathcal{W}_i = (X_{I(i)}, Y_{-I(i)}, C)$. Taking conditional expectation on both sides on (6.5) and using Markov's inequality, we find that

$$\begin{aligned} P\left\{Y_{I(i)} = a_{I(i)}|X_{I(i)}\right\} &\leq \frac{1}{1-\rho}\mathbf{E}\left[P(\tilde{S}_{i,U}(a_{I(i)})|T_{I(i)})|X_{I(i)}\right] \\ &= \frac{1}{1-\rho}\mathbf{E}\left[P(\tilde{S}_{i,U}(a_{I(i)})|\mathcal{W}_i)|X_{I(i)}\right] \\ &= \frac{1}{1-\rho}P(\tilde{S}_{i,U}(a_{I(i)})|\mathcal{W}_i) + \frac{1}{1-\rho}R_{i,U}(a_{I(i)}), \end{aligned} \tag{6.6}$$

where $R_{i,U}(a_{I(i)}) \equiv P(\tilde{S}_{i,U}(a_{I(i)})|\mathcal{W}_i) - P(\tilde{S}_{i,U}(a_{I(i)})|X_{I(i)})$. Similarly,

$$P\left\{Y_{I(i)} \neq a_{I(i)}|X_{I(i)}\right\} \leq \frac{1}{1-\rho}P(\tilde{S}_{i,L}(a_{I(i)})|\mathcal{W}_i) + \frac{1}{1-\rho}R_{i,L}(a_{I(i)}), \tag{6.7}$$

where $R_{i,L}(a_{I(i)}) \equiv P(\tilde{S}_{i,L}(a_{I(i)})|\mathcal{W}_i) - P(\tilde{S}_{i,L}(a_{I(i)})|X_{I(i)})$.

Given $\mathbf{a} = (a_1, \cdots, a_J) \in \mathbb{A}^q$, let $\mathbf{a}_i = (a_1, \cdots, a_{I(i)}) \in \mathbb{A}^{|I(i)|}$. Since $g_j(X_{I(i)}) \geq 0$, we multiply both ends of (6.6) by $g_j(X_{I(i)})$ and summing them up over $i = 1, \cdots, N$, we find that for $\mathbf{a} = (a_1, \cdots, a_J) \in \mathbb{A}^q$,

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} P\left\{Y_{I(i)} = \mathbf{a}_i|X_{I(i)}\right\} g_j(X_{I(i)}) \\ &\leq \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1-\rho_i}P(\tilde{S}_{i,U}(\mathbf{a}_i)|\mathcal{W}_i)g_j(X_{I(i)}) + v_{j,U}(\mathbf{a}), \end{aligned} \tag{6.8}$$

where $v_{j,U}(\mathbf{a}) \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1-\rho_i}R_{i,U}(\mathbf{a}_i)g_j(X_{I(i)})$. Similarly, from (6.7), we also find that

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} P\left\{Y_{I(i)} = \mathbf{a}_i|X_{I(i)}\right\} g_j(X_{I(i)}) \\ &\geq \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{1}{1-\rho}P(\tilde{S}_{i,L}(\mathbf{a}_i)|\mathcal{W}_i)\right) g_j(X_{I(i)}) - v_{j,L}(\mathbf{a}), \end{aligned} \tag{6.9}$$

where $v_{j,L}(\mathbf{a}) \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1-\rho_i}R_{i,L}(\mathbf{a})g_j(X_{I(i)})$.

Now it suffices to control $v_{j,U}(\mathbf{a})$ and $v_{j,L}(\mathbf{a})$. We define

$$e_{i,U}(\mathbf{a}_i) \equiv P\left\{Y_{I(i)} = \mathbf{a}_i | X_{I(i)}\right\} - \frac{1}{1-\rho} P(\tilde{S}_{i,U}(\mathbf{a}_i)|\mathcal{W}_i) g_j(X_{I(i)})$$

and

$$e_{i,L}(\mathbf{a}_i) \equiv P\left\{Y_{I(i)} = \mathbf{a}_i | X_{I(i)}\right\} - \left(1 - \frac{1}{1-\rho} P(\tilde{S}_{i,L}(\mathbf{a}_i)|\mathcal{W}_i)\right) g_j(X_{I(i)}).$$

We write for a given vector of nonnegative constants $w_U = (w_{j,U}(\mathbf{a}))_{j=1,\mathbf{a}\in\mathbb{A}^q}^J$ and $w_L = (w_{j,L}(\mathbf{a}))_{j=1,\mathbf{a}\in\mathbb{A}^q}^J$,

$$\mathcal{M}(w) = \mathcal{M}_L(w_L) \cap \mathcal{M}_U(w_U), \tag{6.10}$$

where $w = (w_U, w_L)$,

$$\begin{aligned} \mathcal{M}_L(w_L) &= \left\{ \sum_{\mathbf{a}\in\mathbb{A}^q} \sum_{j=1}^J \left[\frac{1}{N} \sum_{i=1}^N e_{i,L}(\mathbf{a}_i) g_j(X_{I(i)}) + w_{j,L}(\mathbf{a}) \right]_- = 0 \right\} \text{ and} \\ \mathcal{M}_U(w_U) &= \left\{ \sum_{\mathbf{a}\in\mathbb{A}^q} \sum_{j=1}^J \left[\frac{1}{N} \sum_{i=1}^N e_{i,U}(\mathbf{a}_i) g_j(X_{I(i)}) - w_{j,U}(\mathbf{a}) \right]_+ = 0 \right\}. \end{aligned}$$

By (6.8) and (6.9),

$$P\left(\mathcal{M}_L(v_L)|X, C\right) = 1 \text{ and } P\left(\mathcal{M}_U\left(v_U\right)|X, C\right) = 1, \tag{6.11}$$

where $v_L = (v_{j,L}(\mathbf{a}))_{j=1,\mathbf{a}\in\mathbb{A}^q}^J$ and $v_U = (v_{j,U}(\mathbf{a}))_{j=1,\mathbf{a}\in\mathbb{A}^q}^J$.

Fix nonnegative constants w_U and w_L and define the event

$$\mathcal{M}_A = \left\{v_{j,U}(\mathbf{a}) \le w_U \text{ and } v_{j,L}(\mathbf{a}) \le w_L,\ \forall \mathbf{a} \in \mathbb{A}^q \text{ and } j = 1, \cdots, J\right\}.$$

We write

$$P(\mathcal{M}\left(w\right)|X, C) = P\left(\mathcal{M}\left(w\right) \cap \mathcal{M}_A | X, C\right) + P\left(\mathcal{M}\left(w\right) \cap \mathcal{M}_A^c | X, C\right). \tag{6.12}$$

Note that first probability is increasing in w. Hence by the definition of $\mathcal{M}_A$, the first probability is bounded from below by

$$P\left(\mathcal{M}_L\left(v_L\right) \cap \mathcal{M}_U\left(v_U\right) \cap \mathcal{M}_A | X, C\right) = P\left(\mathcal{M}_A | X, C\right),$$

where the last equality follows by (6.11). Also, we write the last probability in (6.12) as

$$P\left(\mathcal{M}_A^c | X, C\right) - P\left(\mathcal{M}\left(w\right)^c \cap \mathcal{M}_A^c | X, C\right).$$

From (6.12), we conclude that

$$P\left(\mathcal{M}\left(w\right)|X, C\right) \ge 1 - P\left(\mathcal{M}_A^c | X, C\right). \tag{6.13}$$

Now, it suffices to obtain a bound for the last probability. For this, first note that

$$\begin{aligned} P\left[\mathcal{M}_A^c | X, C\right] &\leq P\left\{\max_{\mathbf{a}\in\mathbb{A}^q, j=1,\cdots,J} v_{j,U}(\mathbf{a}) > w_U \text{ or } \max_{\mathbf{a}\in\mathbb{A}^q, j=1,\cdots,J} v_{j,L}(\mathbf{a}) > w_L | X, C\right\} \\ &\leq \sum_{\mathbf{a}\in\mathbb{A}^q}\sum_{j=1}^{J} P\left\{v_{j,U}(\mathbf{a}) > w_U | X, C\right\} + \sum_{\mathbf{a}\in\mathbb{A}^q}\sum_{j=1}^{J} P\left\{v_{j,L}(\mathbf{a}) > w_L | X, C\right\}, \end{aligned}$$

because $w_L, w_U > 0$. We analyze the first probability only. The second probability can be analyzed similarly. Note that essentially $v_{j,U}(\mathbf{a})$ is a random variable as a real valued function of Y_{-i}. Note that $v_{j,U}(\mathbf{a})$ and $v_{j,L}(\mathbf{a})$ are nonstochastic functions of (η, X), where $\eta = (\eta_i)_{i\in\mathbb{N}}$ and $X = (X_i)_{i\in\mathbb{N}}$. By Assumption 6(ii), η_i's are conditionally independent given $X = (X_i)_{i\in\mathbb{N}}$ and C.

We write

$$\begin{aligned} v_{j,U}(\mathbf{a}) &= f_{j,U}(Y, X; \mathbf{a}) = \tilde{f}_{j,U}(\eta, X; \mathbf{a}) \text{ and} \\ v_{j,L}(\mathbf{a}) &= f_{j,L}(Y, X; \mathbf{a}) = \tilde{f}_{j,L}(\eta, X; \mathbf{a}), \end{aligned}$$

for some functions $f_{j,U}(\cdot; \mathbf{a})$ and $f_{j,L}(\cdot; \mathbf{a})$, where

$$\tilde{f}_{j,U}(\eta, X; \mathbf{a}) = f_{j,U}(y(\eta, X), X; \mathbf{a}) \text{ and } \tilde{f}_{j,L}(\eta, X; \mathbf{a}) = f_{j,L}(y(\eta, X), X; \mathbf{a}),$$

using the fact that $Y = y(T)$ and $T = (\eta, X)$, with $y(T) = (y_1(T_{I(1)}), \cdots, y_N(T_{I(N)}))$.

We use McDiarmid's inequality to deduce that

$$\begin{aligned} P\left\{v_{j,U}(\mathbf{a}) > w_U | X, C\right\} &\leq \exp\left(-\frac{2w_U^2}{\sum_{i=1}^N V_i^2(\tilde{f}_{j,U}(\cdot, X; \mathbf{a}))}\right) \\ &\leq \exp\left(-\frac{2w_U^2}{\sum_{i=1}^N \max_{k\in H(i)} V_k^2(f_{j,U}(\cdot, X; \mathbf{a}))}\right) \\ &\leq \exp\left(-\frac{2w_U^2}{\sum_{i=1}^N \delta_{ij,U}^2(X; \mathbf{a})}\right), \end{aligned}$$

where the second inequality follows because the variations in η_i affects $f_{j,U}$ only through the pure strategies $y_k(\eta_{I(k)}, X_{I(k)})$ of players k who observe η_i. Similarly, we also obtain

$$P\left\{v_{j,L}(\mathbf{a}) > w_L | X, C\right\} \leq \exp\left(-\frac{2w_L^2}{\sum_{i=1}^N \delta_{ij,L}^2(X; \mathbf{a})}\right).$$

Using the definitions of $w_U(\tau)$ and $w_L(\tau)$ in Theorem 2, we have

$$\begin{aligned} P\left\{v_{j,U}(\mathbf{a}) > w_U(\tau) | X, C\right\} &\leq \frac{\tau}{2} \text{ and} \\ P\left\{v_{j,L}(\mathbf{a}) > w_L(\tau) | X, C\right\} &\leq \frac{\tau}{2}. \end{aligned}$$

By applying this to (6.13), we obtain the desired inequality:

$$P\left(\mathcal{M}\left(w(\tau)\right)|X,C\right) \geq 1-\tau.$$

Q.E.D.

Proof of Theorem 3: We consider $\mathcal{C}_\infty$ first. Since $\eta_{I(i)}$ is conditionally independent of $(Y_{-I(i)},C)$ given $X_{I(i)}$ for each $i \in \mathbb{N}$ by Assumption 6(ii), we can write

$$\int_{H_{i,U}(a_{I_m};X_{I_m})} dG_{s,\theta}(\eta_{I_m}|X_{I_m}) = P\left\{\eta_{I_m} \in H_{i,U}(a_{I_m};X_{I_m})|\mathcal{W}_i\right\}.$$

On the other hand, because Y is from a pure strategy equilibrium y (Assumption 4(i)), we write $P\left\{Y_{I_m} = \mathbf{a}_{[m]}|\mathcal{W}_i\right\}$ as

$$\begin{aligned} P\left\{y_{I_m}(T_{I_m}) = \mathbf{a}_{[m]}|X_{I_m},Y_{-I_m},C\right\} &= P\left\{y_{I_m}(T_{I_m}) = \mathbf{a}_{[m]}|X_{I_m},C\right\} \\ &= P\left\{y_{I_m}(T_{I_m}) = \mathbf{a}_{[m]}|X_{I_m}\right\} = P\left\{Y_{I_m} = \mathbf{a}_{[m]}|X_{I_m}\right\}, \end{aligned}$$

where $y_{I_m}(T_{I_m}) = (y_j(T_{I(j)}))_{j\in I_m}$ and the equalities are due to Assumptions 5(ii) and 4(ii).

Similarly as in (3.10), we decompose the last sum and apply the same decomposition to $L_{j,L}(\mathbf{a};\theta)$ to obtain that

$$\begin{aligned} L_{j,U}(\mathbf{a};\theta) &= \frac{1}{N}\sum_{i=1}^N \zeta_{ij}(\mathbf{a}_i) + \frac{1}{N}\sum_{i=1}^N e_{i,U}(\mathbf{a}_i)g_j(X_{I(i)}) \text{ and} \\ L_{j,L}(\mathbf{a};\theta) &= \frac{1}{N}\sum_{i=1}^N \zeta_{ij}(\mathbf{a}_i) + \frac{1}{N}\sum_{i=1}^N e_{i,L}(\mathbf{a}_i)g_j(X_{I(i)}), \end{aligned}$$

where $\zeta_{ij}(\mathbf{a}_i) = r_i^*(\mathbf{a}_i)g_j(X_{I(i)})$, $r_i^*(\mathbf{a}_i) = 1\{Y_{I(i)} = \mathbf{a}_i\} - P\{Y_{I(i)} = \mathbf{a}_i|\mathcal{F}_{m(i)-1}\}$, and

$$\begin{aligned} \tilde{e}_{i,U}(\mathbf{a}_i) &= P\{Y_{I(i)} = \mathbf{a}_i|\mathcal{F}_{m(i)-1}\} - \frac{1}{1-\rho_i}\cdot\pi_{i,U}\left(\mathbf{a}_i\right) \text{ and} \\ \tilde{e}_{i,L}(\mathbf{a}_i) &= P\{Y_{I(i)} = \mathbf{a}_i|\mathcal{F}_{m(i)-1}\} - \left(1 - \frac{1}{1-\rho_i}\cdot\pi_{i,L}\left(\mathbf{a}_i\right)\right), \end{aligned}$$

and $m(i) = m$ if and only if $i \in I_m$.

By Theorem 2, we have for all $\bar{a} \in \mathbb{A}$,

$$L_{j,U}(\mathbf{a};\theta) - w_U(\tau) \leq \frac{1}{N}\sum_{i=1}^N \zeta_{ij}(\mathbf{a}_i) \leq L_{j,L}(\mathbf{a};\theta) + w_L(\tau), \tag{6.14}$$

with probability at least $1-\tau$.

Recall the definition

$$Z_{ij}(\mathbf{a}_i) \equiv 1\{Y_{I(i)} = \mathbf{a}_i\}g_j(X_{I(i)}) - \frac{1}{N}P\{Y_{I(i)} = \mathbf{a}_i|\mathcal{F}_{m(i)-1}\}.$$

It is convenient if we stack up $\zeta_{ij}(\mathbf{a})$, $Z_{ij}(\mathbf{a})$, and $\hat{Z}_{ij}(\mathbf{a})$ as vectors with $j = 1, \cdots, J$ and $\mathbf{a}$ running in $\mathbb{A}^q$. To do this, we define $\zeta_i(\mathbf{a})$ $Z_i(\mathbf{a})$, and $\hat{Z}_i(\mathbf{a})$ to be the J-dimensional column vector whose j-th entry is given by $\zeta_{ij}(\mathbf{a})$, $Z_{ij}(\mathbf{a})$, and $\hat{Z}_{ij}(\mathbf{a})$. Then let ζ_i, Z_i, and $\hat{Z}_i$ be $J|\mathbb{A}^q|$-dimensional vector constructed by stacking up $\zeta_i(\mathbf{a})$, $Z_i(\mathbf{a})$, and $\hat{Z}_i(\mathbf{a})$ starting from the first member of $\mathbb{A}^q$ to the last (e.g. by a lexicographic order).

Therefore from (6.14), we can bound

$$\sum_{\mathbf{a}\in\mathbb{A}^q}\sum_{j=1}^{J}\left(\begin{array}{c}[L_{j,U}(\mathbf{a};\theta) - w_{j,U}(\tau,\mathbf{a})]_+ \\ +[L_{j,L}(\mathbf{a};\theta) - w_{j,L}(\tau,\mathbf{a})]_-\end{array}\right)^2 \leq \left\|\frac{1}{N}\sum_{i=1}^{N}\zeta_i\right\|^2, \tag{6.15}$$

with probability at least $1-\tau$. Now, our asymptotic analysis focuses on $||\frac{1}{N}\sum_{i=1}^{N}\zeta_i||^2$.

We define $F_m \equiv \cup_{k=1}^{m} I_k$ and write

$$\frac{1}{N}\sum_{i=1}^{N}\zeta_i = \frac{1}{N}\sum_{m=1}^{N_0}\sum_{i\in I_m}\zeta_i = \frac{N_0}{N}\frac{1}{N_0}\sum_{m=1}^{N_0}\zeta_{m,S},$$

where $\zeta_{m,S} = \sum_{i\in I_m}\zeta_i$. Similarly we define $\hat{Z}^{\varepsilon}_{m,S} = \sum_{i\in I_m}\hat{Z}_i\varepsilon_{m,b}$.

To deal with the case of asymptotically degenerate distribution in a way that is uniform over $P \in \mathcal{P}$, we use arguments similar to those in the proof of Theorem 2 of Lee, Song and Whang (2013). First, fix a small number $0 < \lambda < 1/2$ and let $\{\nu_{i,\lambda} : i \in \mathbb{N}\}$ be a sequence of i.i.d. $J|\mathbb{A}^q|$-dimensional random vectors such that the entries are distributed i.i.d. as uniform $[-\lambda, \lambda]$ and $\{\nu_{i,\lambda}\}_{i=1}^{\infty}$ is independent of $\{(Y_i, X_i, C)\}_{i=1}^{\infty}$. We define

$$\begin{aligned}\zeta_{m,S,\lambda} &\equiv \sum_{i\in I_m}\left(\zeta_i + \frac{\sqrt{3}\nu_{i,\lambda}}{\sqrt{|I_m|}}\right) \text{ and} \\ \hat{Z}^{\varepsilon}_{m,S,\lambda} &\equiv \sum_{i\in I_m}\left(\hat{Z}_i + \frac{\sqrt{3}\nu_{i,\lambda}}{\sqrt{|I_m|}}\right)\varepsilon_{m,b}.\end{aligned}$$

Let $C \equiv (C_s)_{s=1}^{S}$ and for each $j = 1, 2, \cdots$, let

$$\mathcal{F}_{m,\lambda} \equiv \sigma\left(\{(Y_{I_k}, X_{I_k})_{k=1}^{m}, X_{I_{m+1}}, (v_{I_k,\lambda})_{k=1}^{m}, C\right),$$

where $v_{I_k,\lambda} = (v_{i,\lambda})_{i\in I_k}$. Define

$$\mathcal{F}_{\lambda} \equiv \bigcap_{m=1}^{\infty}\mathcal{F}_{m,\lambda}.$$

And

$$F_{\zeta}^{[\lambda]}(t|\mathcal{F}_{\lambda}) \equiv P\left\{\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\zeta_{m,S,\lambda}\right\| \leq t|\mathcal{F}_{\lambda}\right\} \text{ and}$$
$$F_{Z}^{[\lambda]}(t|Y,X) \equiv P\left\{\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\hat{Z}_{m,S,\lambda}^{\varepsilon}\right\| \leq t|Y,X\right\}.$$

Let $F_{\zeta}(t|\mathcal{F}_{\lambda})$ and $F_Z(t|Y,X)$ be $F_{\zeta}^{[\lambda]}(t|\mathcal{F}_{\lambda})$ and $F_{Z}^{[\lambda]}(t|Y,X)$ with $\zeta_{m,S,\lambda}$ and $\hat{Z}_{m,S,\lambda}^{\varepsilon}$ replaced by $\zeta_{m,S}$ and $\hat{Z}_{m,S}^{\varepsilon}$, respectively. Define for a large number $N_0 > 0$ and let a_{N_0} be a positive sequence such that $a_{N_0} \to \infty$ and $a_{N_0}/\sqrt{N_0} \to 0$ as $N_0 \to \infty$. For $\lambda \in (0, 1/2)$ and $\varepsilon > \sqrt{\lambda}$, let

$$h_{N_0}(\lambda) \equiv \frac{a_{N_0}\left|\log a_{N_0} - \log N_0\right|^{12K(\lambda)+3}}{\sqrt{N_0}},$$

where $K(\lambda) \equiv -\log(\lambda^2)/\log(J|\mathbb{A}|^q)$. Note that $h_{N_0}(\lambda)$ does not depend on $P \in \mathcal{P}$.

Recall the sequence $a_{N_0} \to \infty$ in the definition of $h_{N_0}(\lambda, \varepsilon)$. We introduce truncated versions of $\zeta_{m,S,\lambda}$ and $\hat{Z}_{m,S,\lambda}^{\varepsilon}$:

$$\bar{\zeta}_{m,S,\lambda} \equiv \sum_{i \in I_m}\left(\zeta_i 1\left\{\left\|g(X_{I(i)})\right\| \leq a_{N_0}\right\} + \frac{\sqrt{3}\nu_{i,\lambda}}{\sqrt{|I_m|}}\right) \text{ and}$$
$$\bar{Z}_{m,S,\lambda}^{\varepsilon} \equiv \sum_{i \in I_m}\left(\hat{Z}_i 1\left\{\left\|g(X_{I(i)})\right\| \leq a_{N_0}\right\} + \frac{\sqrt{3}\nu_{i,\lambda}}{\sqrt{|I_m|}}\right)\varepsilon_{m,b}.$$

Let

$$\bar{F}_{\zeta}^{[\lambda]}(t|\mathcal{F}_{\lambda}) \equiv P\left\{\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda}\right\| \leq t|\mathcal{F}_{\lambda}\right\} \text{ and}$$
$$\bar{F}_{Z}^{[\lambda]}(t|Y,X) \equiv P\left\{\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{Z}_{m,S,\lambda}^{\varepsilon}\right\| \leq t|Y,X\right\}.$$

Also, we define

$$V_{\lambda}^2 \equiv \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\bar{\zeta}_{m,S,\lambda}^2|\mathcal{F}_{m-1,\lambda}\right] \text{ and}$$
$$W_{\lambda}^2 \equiv \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\bar{Z}_{m,S,\lambda}^2|\mathcal{G}_{m-1,\lambda}\right],$$

where $\mathcal{G}_{m,\lambda} = \sigma\left((\varepsilon_{m(i),b})_{i \in F_m}, (\nu_{i,\lambda})_{i \in F_m}, Y, X\right)$.

First, we show the following claims.

Claim 1: There exists $c_1 > 0$ such that for each $\lambda \in (0, 1/2)$ and $\varepsilon > 0$,

$$P\left\{\sup_{t\geq 0}\left|\bar{F}_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda) - \bar{F}_{\zeta,\infty}^{[\lambda]}(t|\mathcal{F}_\lambda)\right| \geq c_1 h_{N_0}(\lambda)\right\} = 1,$$

where $\bar{F}_{\zeta,\infty}^{[\lambda]}(t|\mathcal{F}_\lambda) = P\{V_\lambda \mathbb{Z} \leq t|\mathcal{F}_\lambda\}$.

Claim 2: There exists $c_2 > 0$ such that for each $\lambda \in (0, 1/2)$ and $\varepsilon > 0$,

$$P\left\{\sup_{t\geq \varepsilon}\left|\bar{F}_Z^{[\lambda]}(t|Y,X) - \bar{F}_{\infty}^{[\lambda]}(t|Y,X)\right| \geq c_2 h_{N_0}(\lambda)\right\} = 1,$$

where $F_{Z,\infty}^{[\lambda]}(t|Y,X) = P\{W_\lambda \mathbb{Z} \leq t|Y,X\}$.

Claim 3: For any choice of $c_{N_0} > 0$ and $\kappa_{N_0} > 0$, we have

$$P\left\{\bar{F}_{\zeta,\infty}^{[\lambda]}(t|\mathcal{F}_\lambda) - \bar{F}_{Z,\infty}^{[\lambda]}(t|Y,X) \geq -c_{N_0}\right\} \geq 1 - \left(\frac{3J|\mathbb{A}^q|\bar{M}^2 a_{N_0}^2}{\kappa_{N_0} c_{N_0}\sqrt{N_0}} + \frac{C_1 \kappa_{N_0}}{c_{N_0}\lambda^2}\right),$$

where C_1 is the maximum value of χ^2 density with degree of freedom equal to $J|\mathbb{A}^q|$.

Claim 4: For $C = \max_{i\in\mathbb{N}} \sup_{P\in\mathcal{P}} \mathbf{E}_P\left[||g(X_{I(i)})||^4\right]$, we have (as $N_0 \to \infty$)

$$\begin{aligned} P\left\{\left|\bar{F}_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda) - F_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda)\right| \leq C\bar{M}a_{N_0}^{-\frac{2}{3}} + \varphi_{N_0}^{[1]}(\lambda,\varepsilon) + c_1 h_{N_0}(\lambda)\right\} &= 1 \text{ and} \\ P\left\{\left|\bar{F}_Z^{[\lambda]}(t|\mathcal{F}_\lambda) - F_Z^{[\lambda]}(t|\mathcal{F}_\lambda)\right| \leq C\bar{M}a_{N_0}^{-\frac{2}{3}} + \varphi_{N_0}^{[1]}(\lambda,\varepsilon) + \tilde{h}_{N_0}(\lambda)\right\} &\to 1, \end{aligned}$$

where $\varphi_{N_0}^{[1]}(\lambda,\varepsilon)$ is a nonstochastic map that does not depend on P such that $\lim_{N_0\to\infty}\varphi_{N_0}^{[1]}(\lambda,\varepsilon)$ for all $\lambda > 0, \varepsilon > 0$, and $\tilde{h}_{N_0}(\lambda) \equiv C\bar{M}a_{N_0}^{-2/3}/(N_0 \log N_0) + c_2 h_{N_0}(\lambda)$.

Claim 5: For each $\lambda > 0$ and $\varepsilon > 0$, there exists a positive number $\varphi_{N_0}^{[2]}(\lambda,\varepsilon) > 0$ such that $\lim_{\lambda\to 0}\lim_{N_0\to\infty}\varphi_{N_0}^{[2]}(\lambda,\varepsilon) = 0$ for each $\varepsilon > 0$ and

$$\sup_{P\in\mathcal{P}}\sup_{t\geq\varepsilon}\left|F_\zeta(t) - F_\zeta^{[\lambda]}(t)\right| \leq \varphi_{N_0}^{[2]}(\lambda,\varepsilon) + \lambda \text{ and}$$

$$P\left\{\sup_{t\geq\varepsilon}\left|F_Z(t|Y,X) - F_Z^{[\lambda]}(t|Y,X)\right| \leq \varphi_{N_0}^{[2]}(\lambda,\varepsilon) + \lambda\right\} = 1.$$

By chaining Claims 1-5. we conclude that

$$\inf_{P\in\mathcal{P}} P\left\{\sup_{t\geq\varepsilon}|F_\zeta(t) - F_Z(t|Y,X)| \leq (c_1 + c_2)\, h_{N_0}(\lambda,\varepsilon)\right\} \to 0, \tag{6.16}$$

where

$$h_{N_0}(\lambda,\varepsilon) = 2(\varphi_{N_0}^{[2]}(\lambda,\varepsilon) + \lambda) + 2\{C\bar{M}a_{N_0}^{-\frac{2}{3}} + \varphi_{N_0}^{[1]}(\lambda,\varepsilon) + 2c_1 h_{N_0}(\lambda) + 2\tilde{h}_{N_0}(\lambda)\}.$$

Note that as $N_0 \to \infty$, and then $\lambda \to \infty$, we have $h_{N_0}(\lambda,\varepsilon) \to 0$.

By the definition of $c^*_{1-\alpha,\infty}$

$$\begin{aligned}1-\alpha+\tau(\theta) &\leq P\left\{\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{Z}_i\varepsilon_{i,b}\right\|^2 \leq c^*_{1-\alpha,\infty}\vee\varepsilon|Y,X\right\}\\ &= P\left\{\left\|\frac{1}{\sqrt{N_0}}\sum_{i=1}^{N}\hat{Z}_i\varepsilon_{i,b}\right\|^2 \leq \frac{N^2(c^*_{1-\alpha,\infty}\vee\varepsilon)}{N_0}|Y,X\right\}.\end{aligned}$$

The last term is bounded by

$$P\left\{\left\|\frac{1}{\sqrt{N_0}}\sum_{i=1}^{N}\zeta_i\right\|^2 \leq \frac{N^2(c^*_{1-\alpha,\infty}\vee\varepsilon)}{N_0}|\mathcal{F}_\lambda\right\}+o_P(1),$$

where $o_P(1)$ is obtained by applying (6.16) and sending $N_0 \to 0$ and then $\lambda \to 0$, uniformly over $P \in \mathcal{P}$. The leading probability is bounded by

$$P\left\{T(\theta) \leq c^*_{1-\alpha,\infty}\vee\varepsilon|\mathcal{F}_\lambda\right\}+o_P(1),$$

uniformly over $P \in \mathcal{P}$. Hence, we obtain the desired result.

Proof of Claim 1: We note that $\{\bar{\zeta}_{m,S,\lambda} : m = 1, \cdots, N_0\}$ is a martingale difference array with respect to the filtration $\{\mathcal{F}_{m,\lambda}\}_{m=1}^{\infty}$. First, define

$$V_\lambda \equiv \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\bar{\zeta}_{m,S,\lambda}\bar{\zeta}^\top_{m,S,\lambda}|\mathcal{F}_{m-1,\lambda}\right] = \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\bar{\zeta}_{m,S,\lambda}\bar{\zeta}^\top_{m,S,\lambda}|C\right], \tag{6.17}$$

because $\bar{\zeta}_{m,S,\lambda}$'s across m's are conditionally independent given C. By the definition of $\nu_{i,\lambda}$, for any N_0 and $\omega \in \Omega$,

$$V_\lambda(\omega) = \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\zeta_{m,S}\zeta^\top_{m,S}|\mathcal{F}_{m-1,\lambda}\right](\omega)+\lambda^2 I \geq \lambda^2 I. \tag{6.18}$$

Now we apply the Berry-Esseen bound for Hilbert-valued martingale difference arrays in Rackauskas (1991) (Theorem on page 345). (Instead of the unconditional probability there, we use the conditional probability $P\{\cdot|\mathcal{F}_\lambda\}$.) For this we prove the conditions in the theorem. We focus on the following sum:

$$\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda}.$$

By (6.17), Conditions (i) and (ii) in Theorem of Rackauskas (1991) are satisfied trivially with $\sigma^2_{ii} = 1$ there. As for Condition (iii) in the theorem, note that

$$\frac{\max_{1\leq m\leq N_0}\left\|\bar{\zeta}_{m,S,\lambda}\right\|}{\sqrt{N_0}} \leq \frac{a_{N_0}}{\sqrt{N_0}}.$$

As for Condition (iv), the eigen values of $V_\lambda(\omega)$ are bounded from below by λ^2 from (6.18). Hence Condition (iv) in Theorem of Rackauskas (1991) is satisfied with constant K in the theorem equals 1. Thus, we apply Theorem of Rackauskas (1991) (p.345) to deduce that for some nonstochastic $c_1 > 0$,

$$\sup_{t\geq 0}\left|\bar{F}_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda) - \bar{F}_{\zeta,\infty}^{[\lambda]}(t|\mathcal{F}_\lambda)\right| \leq \frac{c_1 a_{N_0}\left|\log a_{N_0} - \log N_0\right|^{12K(\lambda)+3}}{\sqrt{N_0}}.$$

Proof of Claim 2: We find that $\mathbf{E}\left[\bar{Z}_{m,S,\lambda}^\varepsilon|\mathcal{G}_{m-1,\lambda}\right]$ is equal to

$$\mathbf{E}\left[\sum_{i\in I_m}\left(\hat{Z}_i 1\left\{\left\|g(X_{I(i)})\right\| \leq a_{N_0}\right\} + \frac{\sqrt{3}\nu_{i,\lambda}}{\sqrt{|I_m|}}\right)\varepsilon_{m,b}|\mathcal{G}_{m-1,\lambda}\right] = 0.$$

Furthermore, $\bar{Z}_{m,S,\lambda}^\varepsilon$ is $\mathcal{G}_{m,\lambda}$-measurable for each $m = 1,\cdots,N_0$. Hence $\{\bar{Z}_{m,S,\lambda}^\varepsilon : m = 1,\cdots,N_0\}$ is a martingale difference array with respect to $\mathcal{G}_{m-1,\lambda}$. We first define

$$W_\lambda \equiv \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\bar{Z}_{m,S,\lambda}\bar{Z}_{m,S,\lambda}^\top|\mathcal{G}_{m-1,\lambda}\right] = \frac{1}{N_0}\sum_{m=1}^{N_0}\hat{Z}_{m,S}\hat{Z}_{m,S}^\top + \lambda^2 I.$$

Observe that

$$\max_{1\leq m\leq N_0}\left|\frac{\bar{Z}_{m,S,\lambda}}{\sqrt{N_0}}\right| \leq \frac{a_{N_0}}{\sqrt{N_0}}.$$

Similarly as before, we use Theorem of Rackauskas (1991) to obtain the desired result.

Proof of Claim 3: Let $R_{m,S} = \sum_{i\in I_m} R_{i,S}$ and $R_{i,S}$ is a $J|\mathbb{A}|^q$-vector with

$$1\{Y_{I(i)} = \mathbf{a}_i\}g_j(X_{I(i)})1\left\{\max_{i\in I_m}\left\|g(X_{I(i)})\right\| \leq a_{N_0}\right\}$$

stacked up with j's running $j = 1,\cdots,J$ and $\mathbf{a}$ running in $|\mathbb{A}|$. Note that

$$\bar{\zeta}_{m,S} = R_{m,S} - \mathbf{E}\left[R_{m,S}|\mathcal{F}_{m-1}\right].$$

As for V_λ and W_λ, note that

$$\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[\bar{\zeta}_{m,S}\bar{\zeta}_{m,S}^\top|\mathcal{F}_{m-1}\right] = \frac{1}{N_0}\sum_{m=1}^{N_0}\left(\mathbf{E}\left[R_{m,S}R_{m,S}^\top|\mathcal{F}_{m-1}\right] - \left(\mathbf{E}\left[R_{m,S}|\mathcal{F}_{m-1}\right]\mathbf{E}\left[R_{m,S}^\top|\mathcal{F}_{m-1}\right]\right)\right),$$

and

$$\frac{1}{N_0}\sum_{m=1}^{N_0}\hat{Z}_{m,S}\hat{Z}_{m,S}^\top = \frac{1}{N_0}\sum_{m=1}^{N_0}R_{m,S}R_{m,S}^\top - \left(\frac{1}{N_0}\sum_{m=1}^{N_0}R_{m,S}\right)\left(\frac{1}{N_0}\sum_{m=1}^{N_0}R_{m,S}^\top\right).$$

Using the similar arguments as before, we can show that

$$\frac{1}{N_0}\sum_{m=1}^{N_0}\hat{Z}_{m,S}\hat{Z}_{m,S}^\top = W + r_{N_0},$$

where $r_{N_0} = \xi_{1,N_0} + \xi_{2,N_0}$,

$$\begin{aligned}\xi_{1,N_0} &= \frac{1}{N_0}\sum_{m=1}^{N_0}\left(R_{m,S}R_{m,S}^\top - \mathbf{E}\left[R_{m,S}R_{m,S}^\top|\mathcal{F}_{m-1}\right]\right)\\ \xi_{2,N_0} &= \left(\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}|\mathcal{F}_{m-1}\right]\right)\left(\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}^\top|\mathcal{F}_{m-1}\right]\right)\\ &\quad -\left(\frac{1}{N_0}\sum_{m=1}^{N_0}R_{m,S}\right)\left(\frac{1}{N_0}\sum_{m=1}^{N_0}R_{m,S}^\top\right),\end{aligned}$$

and

$$\tilde{W} = \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}R_{m,S}^\top|\mathcal{F}_{m-1}\right] - \left(\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}|\mathcal{F}_{m-1}\right]\right)\left(\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}^\top|\mathcal{F}_{m-1}\right]\right).$$

Let $\tilde{W}_\lambda = \tilde{W} + \lambda^2 I$. Hence we have (everywhere)

$$\begin{aligned}\tilde{W}_\lambda - V_\lambda &= \frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}|\mathcal{F}_{m-1}\right]\mathbf{E}\left[R_{m,S}^\top|\mathcal{F}_{m-1}\right]\\ &\quad -\left(\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}|\mathcal{F}_{m-1}\right]\right)\left(\frac{1}{N_0}\sum_{m=1}^{N_0}\mathbf{E}\left[R_{m,S}^\top|\mathcal{F}_{m-1}\right]\right).\end{aligned}$$

We conclude that $\tilde{W}_\lambda \geq V_\lambda$ everywhere.

We consider

$$\begin{aligned}&\left|P\left\{\left\|\tilde{W}_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\} - P\left\{\left\|W_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\}\right|\\ \leq\ & P\left\{t^2 - \Delta_\lambda \leq \left\|W_\lambda^{1/2}\mathbb{Z}\right\|^2 \leq t^2 + \Delta_\lambda|Y,X\right\},\end{aligned} \tag{6.19}$$

where $\Delta_\lambda = |||\tilde{W}_\lambda^{1/2}\mathbb{Z}||^2 - ||W_\lambda^{1/2}\mathbb{Z}||^2|$. Note that

$$\Delta_\lambda = \left|tr\left(\tilde{W}_\lambda\mathbb{Z}\mathbb{Z}^\top\right) - tr\left(W_\lambda\mathbb{Z}\mathbb{Z}^\top\right)\right| = \left|tr\left((\tilde{W}_\lambda - W_\lambda)\mathbb{Z}\mathbb{Z}^\top\right)\right|.$$

Choose $\kappa > 0$ and bound the last probability in (6.19) from below by

$$\begin{aligned}& P\left\{t^2/\lambda^2 - \Delta_\lambda/\lambda^2 \leq \|\mathbb{Z}\|^2 \leq t^2/\lambda^2 + \Delta_\lambda/\lambda^2|Y,X\right\}\\ \leq\ & P\left\{t^2/\lambda^2 - \kappa/\lambda^2 \leq \|\mathbb{Z}\|^2 \leq t^2/\lambda^2 + \kappa/\lambda^2|Y,X\right\} + P\left\{\Delta_\lambda > \kappa|Y,X\right\}.\end{aligned}$$

The first probability is bounded by $C_1\left(\kappa/\lambda^2\right)$, where C_1 is the maximum value of χ^2 distribution with degree of freedom equal to $J|\mathbb{A}^q|$. The second probability is bounded by

$$\begin{aligned}\frac{1}{\kappa}\mathbf{E}\left[\left|tr\left((\tilde{W}_\lambda - W_\lambda)\mathbb{Z}\mathbb{Z}^\top\right)\right||Y,X\right] &= \frac{1}{\kappa}\mathbf{E}\left[tr\left((W_\lambda - \tilde{W}_\lambda)\mathbf{E}\left[\mathbb{Z}\mathbb{Z}^\top\right]\right)|Y,X\right]\\ &\leq \frac{J|\mathbb{A}^q|}{\kappa}\mathbf{E}\left[tr(W_\lambda - \tilde{W}_\lambda)|Y,X\right]\\ &= \frac{J|\mathbb{A}^q|\,\|r_{N_0}\|}{\kappa}.\end{aligned}$$

First, note that ξ_{1,N_0} is a sum of uncorrelated matrices, and that $||R_{m,S}|| \leq \bar{M}a_{N_0}$. hence

$$\mathbf{E}\left[\left\|\xi_{1,N_0}\right\|^2|\mathcal{F}_\lambda\right] \leq \frac{\bar{M}^4a_{N_0}^4}{N_0}$$

and similarly, $\mathbf{E}\left[\left\|\xi_{2,N_0}\right\|^2|\mathcal{F}_\lambda\right] \leq 2\bar{M}^4a_{N_0}^4/N_0$ after some algebra. We conclude that

$$\mathbf{E}\left[|r_{N_0}|\,|\mathcal{F}_\lambda\right] \leq \frac{3\bar{M}^2a_{N_0}^2}{\sqrt{N_0}}.$$

From (6.19), we find that

$$\begin{aligned}&\mathbf{E}\left[\left|P\left\{\left\|\tilde{W}_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\} - P\left\{\left\|W_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\}\right||\mathcal{F}_\lambda\right]\\ \leq\ & \frac{3J|\mathbb{A}^q|\bar{M}^2a_{N_0}^2}{\kappa\sqrt{N_0}} + \frac{C_1\kappa}{\lambda^2}.\end{aligned} \tag{6.20}$$

Now, we take $c_{N_0} > 0$ and write

$$\begin{aligned}&P\left\{\bar{F}_{\zeta,\infty}^{[\lambda]}(t|\mathcal{F}_\lambda) - \bar{F}_{Z,\infty}^{[\lambda]}(t|Y,X) \geq -c_{N_0}\right\}\\ \geq\ & P\left\{P\left\{\left\|V_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\} - P\left\{\left\|\tilde{W}_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\} \geq -c_{N_0} + D_{N_0}\right\},\end{aligned}$$

where

$$D_{N_0} = \left|P\left\{\left\|\tilde{W}_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\} - P\left\{\left\|W_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\}\right|.$$

We bound the last probability from below by

$$\begin{aligned}&P\left\{P\left\{\left\|V_\lambda^{1/2}\mathbb{Z}\right\| \leq t|\mathcal{F}_\lambda\right\} - P\left\{\left\|\tilde{W}_\lambda^{1/2}\mathbb{Z}\right\| \leq t|Y,X\right\} \geq 0\right\} - P\left\{D_{N_0} > c_{N_0}\right\}\\ =\ & 1 - P\left\{D_{N_0} > c_{N_0}\right\}.\end{aligned}$$

By Markov's inequality and (6.20), the last probability is bounded by

$$\frac{3J|\mathbb{A}^q|\bar{M}^2a_{N_0}^2}{\kappa c_{N_0}\sqrt{N_0}} + \frac{C_1\kappa}{c_{N_0}\lambda^2}.$$

Hence we obtain the desired result.

Proof of Claim 4: Define

$$\begin{aligned} D_\zeta^{[\lambda]} &= \frac{1}{\sqrt{N_0}} \sum_{m=1}^{N_0} (\bar{\zeta}_{m,S,\lambda} - \zeta_{m,S,\lambda}) \text{ and} \\ D_Z^{[\lambda]} &= \frac{1}{\sqrt{N_0}} \sum_{m=1}^{N_0} (\bar{Z}_{m,S,\lambda} - \hat{Z}_{m,S,\lambda})\varepsilon_{m,b}. \end{aligned}$$

Note that $\{\bar{\zeta}_{m,S,\lambda} - \zeta_{m,S,\lambda}\}_{m=1}^{N_0}$ is a (multi-dimensional) martingale difference array with respect to $\{\mathcal{F}_{m,\lambda}\}_{m=1}^{N_0}$ and $\{\bar{Z}_{m,S,\lambda} - \hat{Z}_{m,S,\lambda}\}_{m=1}^{N_0}$ is also a martingale difference array with respect to $\{\mathcal{G}_{m,\lambda}\}_{m=1}^{N_0}$. Hence

$$\begin{aligned} \mathbf{E}\left[\left\|D_\zeta^{[\lambda]}\right\|^2 |\mathcal{F}_\lambda\right] &\leq \frac{1}{N_0}\sum_{m=1}^{N_0} \mathbf{E}\left[\left\|\bar{\zeta}_{m,S,\lambda} - \zeta_{m,S,\lambda}\right\|^2 |\mathcal{F}_\lambda\right] \\ &\leq \frac{1}{N_0}\sum_{m=1}^{N_0} P\left\{\sum_{i\in I_m} ||g(X_{I(i)})|| > a_{N_0}|\mathcal{F}_\lambda\right\} \\ &\leq \frac{1}{a_{N_0}^2 N_0}\sum_{m=1}^{N_0} \mathbf{E}\left[\left(\sum_{i\in I_m} ||g(X_{I(i)})||\right)^2\right] \leq \frac{C\bar{M}}{a_{N_0}^2}. \end{aligned} \tag{6.21}$$

Similarly, we also deduce that The last term vanishes uniformly in $P \in \mathcal{P}$, as $N \to \infty$. This means that $|\bar{F}_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda) - F_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda)|$ is bounded by

$$\begin{aligned} & P\left\{\left|\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda}\right\| - t\right| \leq \left\|D_\zeta^{[\lambda]}\right\| |\mathcal{F}_\lambda\right\} \\ = & P\left\{\left|\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda}\right\| - t\right| \leq \left\|D_\zeta^{[\lambda]}\right\| \text{ and } \left\|D_\zeta^{[\lambda]}\right\| \leq a_{N_0}^{-\frac{2}{3}}|\mathcal{F}_\lambda\right\} \\ & +P\left\{\left|\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda}\right\| - t\right| \leq \left\|D_\zeta^{[\lambda]}\right\| \text{ and } \left\|D_\zeta^{[\lambda]}\right\| > a_{N_0}^{-\frac{2}{3}}|\mathcal{F}_\lambda\right\}. \end{aligned} \tag{6.22}$$

The first probability on the right hand side of (6.22) is bounded by

$$P\left\{\left|\left\|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda}\right\| - t\right| \leq a_{N_0}^{-\frac{2}{3}}|\mathcal{F}_\lambda\right\} \leq P\left\{\left|\left\|V_\lambda^{1/2}\mathbb{Z}\right\| - t\right| \leq a_{N_0}^{-\frac{2}{3}}|\mathcal{F}_\lambda\right\} + 2\Delta_1(\mathcal{F}_\lambda),$$

where

$$\Delta_1(\mathcal{F}_\lambda) \equiv \sup_{t\geq\varepsilon}\left|\bar{F}_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda) - F_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda)\right|.$$

As for the leading probability, note that it is written as

$$P\left\{t - a_{N_0}^{-\frac{2}{3}} \leq \left\|V_\lambda^{1/2}\mathbb{Z}\right\| \leq t + a_{N_0}^{-\frac{2}{3}}|\mathcal{F}_\lambda\right\}.$$

Since $t \geq \varepsilon$, $a_{N_0} \to \infty$ and $||V_\lambda^{1/2}|| \leq C\bar{M}$, it is bounded by some function $\varphi_{N_0}^{[1]}(\lambda, \varepsilon)$ that does not depend on P such that

$$\lim_{N_0 \to 0} \varphi_{N_0}^{[1]}(\lambda, \varepsilon) = 0.$$

On the other hand, the last probability in (6.22) is bounded by

$$a_{N_0}^{\frac{4}{3}} \mathbf{E}\left[\left(D_\zeta^{[\lambda]}\right)^2 |\mathcal{F}_\lambda\right] \leq \frac{a_{N_0}^{\frac{4}{3}} C\bar{M}}{a_{N_0}^2} = C\bar{M} a_{N_0}^{-\frac{2}{3}}.$$

Thus we conclude that

$$\left|\bar{F}_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda) - F_\zeta^{[\lambda]}(t|\mathcal{F}_\lambda)\right| \leq C\bar{M} a_{N_0}^{-\frac{2}{3}} + \varphi_{N_0}^{[1]}(\lambda, \varepsilon) + 2\Delta_1(\mathcal{F}_\lambda). \tag{6.23}$$

As for the second statement of Claim 4, we note that

$$\begin{aligned}
\mathbf{E}\left[\left\|D_Z^{[\lambda]}\right\|^2 |Y, X\right] &\leq \frac{1}{N_0} \sum_{m=1}^{N_0} 1\left\{\sum_{i \in I_m} ||g(X_{I(i)})|| > a_{N_0}\right\} \\
&= \frac{1}{N_0} \sum_{m=1}^{N_0} \left(1\left\{\sum_{i \in I_m} ||g(X_{I(i)})|| > a_{N_0}\right\} - P\left\{\sum_{i \in I_m} ||g(X_{I(i)})|| > a_{N_0} |\mathcal{F}_\lambda\right\}\right) \\
&\quad + \frac{1}{N_0} \sum_{m=1}^{N_0} P\left\{\sum_{i \in I_m} ||g(X_{I(i)})|| > a_{N_0} |\mathcal{F}_\lambda\right\},
\end{aligned}$$

where the last sum is bounded by

$$\frac{1}{a_{N_0}^2 N_0} \sum_{m=1}^{N_0} \mathbf{E}\left[\left(\sum_{i \in I_m} ||g(X_{I(i)})||\right)^2 |\mathcal{F}_\lambda\right] \leq \frac{C\bar{M}}{a_{N_0}^2},$$

similarly as in (6.21). Hence using the same arguments that led to (6.23), we obtain the following inequality:

$$\begin{aligned}
\left|\bar{F}_Z^{[\lambda]}(t|Y, X) - F_Z^{[\lambda]}(t|Y, X)\right| &\leq (C\bar{M} + 8\lambda^{-1}) a_{N_0}^{-\frac{2}{3}} + 2\Delta_2(Y, X) \\
&\quad + \Delta_3(Y, X) a_{N_0}^{\frac{4}{3}},
\end{aligned}$$

where $\Delta_2(Y, X) \equiv \sup_{t \geq \varepsilon} |\bar{F}_\zeta^{[\lambda]}(t|Y, X) - F_Z^{[\lambda]}(t|Y, X)|$ and

$$\Delta_3(Y, X) \equiv \frac{1}{N_0} \sum_{m=1}^{N_0} \left(1\left\{\sum_{i \in I_m} ||g(X_{I(i)})|| > a_{N_0}\right\} - P\left\{\sum_{i \in I_m} ||g(X_{I(i)})|| > a_{N_0} |\mathcal{F}_\lambda\right\}\right).$$

Note that $\mathbf{E}\left[\Delta_3^2(Y,X)|\mathcal{F}_\lambda\right] \leq C\bar{M}/(N_0 a_{N_0}^2)$. Hence

$$\begin{aligned}
& P\left\{\left|\bar{F}_Z^{[\lambda]}(t|Y,X) - F_Z^{[\lambda]}(t|Y,X)\right| \leq C\bar{M}a_{N_0}^{-\frac{2}{3}} + \varphi_{N_0}^{[1]}(\lambda,\varepsilon) + 2\Delta_2(Y,X) + \frac{C\bar{M}a_{N_0}^{-\frac{2}{3}}}{N_0 \log N_0}\right\} \\
\geq \; & P\left\{\begin{array}{c} \left|\bar{F}_Z^{[\lambda]}(t|Y,X) - F_Z^{[\lambda]}(t|Y,X)\right| \leq C\bar{M}a_{N_0}^{-\frac{2}{3}} + \varphi_{N_0}^{[1]}(\lambda,\varepsilon) + 2\Delta_2(Y,X) + \mathbf{\Delta}_3(Y,X)a_{N_0}^{\frac{4}{3}} \\ \text{and } \mathbf{\Delta}_3(Y,X) \leq \frac{C\bar{M}}{N_0 \log N_0 a_{N_0}^2} \end{array}\right\}.
\end{aligned}$$

The last probability is bounded from below by

$$1 - P\left\{\mathbf{\Delta}_3(Y,X) > \frac{C\bar{M}}{N_0 \log N_0 a_{N_0}^2}\right\} \geq 1 - \frac{C\bar{M}}{N_0 \log N_0 a_{N_0}^2} \to 1,$$

as $N_0 \to \infty$. Using Claims 1 and 2, we obtain the desired result.

Proof of Claim 5: First, we use Claim 4 to find that

$$\begin{aligned}
\left|F_\zeta(t) - F_\zeta^{[\lambda]}(t)\right| &\leq P\left\{\left|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\zeta_{m,S,\lambda} - t\right| \leq \Delta_\eta\right\} \\
&\leq P\left\{\left|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda} - t\right| \leq \Delta_\eta\right\} + (C\bar{M} + \varphi_{N_0}^{[1]}(\lambda,\varepsilon))a_{N_0}^{-\frac{2}{3}} + h_{N_0}(\lambda),
\end{aligned}$$

where

$$\Delta_\eta = \left|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\frac{\sqrt{3}}{\sqrt{|I_m|}}\sum_{i\in I_m}\nu_{i,\lambda}\right|.$$

Using the previous arguments,

$$P\left\{\left|\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\bar{\zeta}_{m,S,\lambda} - t\right| \leq \Delta_\eta\right\} \leq P\left\{|V_\lambda \mathbb{Z} - t| \leq \Delta_\eta\right\} + h_{N_0}(\lambda).$$

We write the last probability as

$$\begin{aligned}
& P\left\{|V_\lambda \mathbb{Z} - t| \leq \Delta_\eta \text{ and } \Delta_\eta \leq \sqrt{\lambda}\right\} \\
& + P\left\{|V_\lambda \mathbb{Z} - t| \leq \Delta_\eta \text{ and } \Delta_\eta > \sqrt{\lambda}\right\}.
\end{aligned} \tag{6.24}$$

The probability is written as

$$\begin{aligned}
& P\left\{t - \Delta_\eta \leq V_\lambda \mathbb{Z} \leq \Delta_\eta + t \text{ and } \Delta_\eta \leq \sqrt{\lambda}\right\} \\
\leq \; & P\left\{t - \sqrt{\lambda} \leq V_\lambda \mathbb{Z} \leq \sqrt{\lambda} + t\right\}.
\end{aligned}$$

Since $V_\lambda \in [\lambda, C\bar{M}]$ and $t \geq \varepsilon$, there exists a positive function $\varphi_{N_0}^{[2]}(\lambda,\varepsilon)$ that does not depend on P such that the last probability is bounded by $\varphi_{N_0}^{[2]}(\lambda,\varepsilon)$ and $\lim_{\lambda\to 0}\lim_{N_0\to\infty}\varphi_{N_0}^{[2]}(\lambda,\varepsilon) = 0$.

The second probability in (6.24) is bounded by (for some $c_1 > 0$)

$$\frac{1}{\lambda}\mathbf{E}\left[\left(\frac{1}{\sqrt{N_0}}\sum_{m=1}^{N_0}\frac{\sqrt{3}}{\sqrt{|I_m|}}\sum_{i\in I_m}\nu_{i,\lambda}\right)^2\right] \leq \lambda.$$

We conclude that

$$\left|F_\zeta(t) - F_\zeta^{[\lambda]}(t)\right| \leq \varphi_{N_0}^{[2]}(\lambda,\varepsilon) + \lambda.$$

Using the same arguments as previously, we also conclude that with $P = 1$,

$$\left|F_Z(t|Y,X) - F_Z^{[\lambda]}(t|Y,X)\right| \leq \varphi_{N_0}^{[2]}(\lambda,\varepsilon) + \lambda.$$

Thus we obtain the desired result. *Q.E.D.*

Proof of Theorem 4: It suffices to show that

$$\mathbf{E}\left[\left|\frac{1}{\sqrt{N_0}}\sum_{i=1}^{N} e_{i,U}(\mathbf{a})g_j(X_{I(i)})\varepsilon_{m(i),b}\right||Y,X\right] \text{ and}$$
$$\mathbf{E}\left[\left|\frac{1}{\sqrt{N_0}}\sum_{i=1}^{N} e_{i,L}(\mathbf{a})g_j(X_{I(i)})\varepsilon_{m(i),b}\right||Y,X\right]$$

are asymptotically negligible conditional on (Y,X). We only deal with the first conditional expectation. Note that it is bounded by

$$C\sqrt{\frac{1}{N_0}\sum_{m=1}^{N_0}\left(\sum_{i\in I_m} e_{i,U}(\mathbf{a})g_j(X_{I(i)})\right)^2}.$$

Since all the players are asymptotically negligible, we have

$$\max_{i\in\mathbb{N}} e_{i,U}(\mathbf{a}) \to 0 \text{ and } \max_{i\in\mathbb{N}} e_{i,L}(\mathbf{a}) \to 0$$

as $N \to \infty$. Hence we obtain the desired result. *Q.E.D.*

References

[1] Andreoni, J., Y.-K. Che, and J. Kim (2007): "Asymmetric Information about Rivals' Types in Standard Auctions: An Experiment," *Games and Economic Behavior* 59, 240-259.

[2] Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica* 59, 307-345.

[3] Andrews, D. W. K. (2005): "Cross-section Regression with Common Shocks," *Econometrica* 73, 1551-1585.

[4] Andrews, D. W. K., X. Cheng, and P. Guggenberger (2011), "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Paper 1813.

[5] Andrews, D. W. K. and X. Shi (2011): "Inference Based on Conditional Moment Inequalities," *Econometrica* 81, 609-666.

[6] Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica* 78, 119–157.

[7] Aradillas-Lopez, A. (2010): "Semiparametric estimation of a simultaneous game with incomplete information," *Journal of Econometrics* 157, 409-431.

[8] Aradillas-Lopez, A. and E. Tamer (2011): "The Identification Power of Equilibrium in Simple Games," *Journal of Business and Economic Statistics* 26, 261-283.

[9] Athey, S. (2001): "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information," *Econometrica* 69, 861-889.

[10] Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70, 191-221.

[11] Balder, E. J. (1988): "Generalized Equilibrium Results for Games with Incomplete Information," *Mathematics of Operations Research* 13, 265-276.

[12] Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica* 79, 1785-1821.

[13] Bergemann, D. and S. Morris (2011): "Robust Mechanism Design," *Econometrica* 73 1771-1813.

[14] Bisin, A., A. Moro, and G. Topa (2011): "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions," NBER Working Paper 17196.

[15] Blume, L. E., W. A. Brock, S. N. Durlauf and R. Jayaraman (2011): "Linear Social Network Models," Working Paper.

[16] Bresnahan, T. F. and P. C. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics* 48, 57-81.

[17] Brock, W. A. and S. N. Durlauf (2001a): "Interaction-Based Models," *Handbook of Econometrics,* Vol. 5. Edited by J.J. Heckman, and E. Leamer. Elsevier Science.

[18] Brock, W. A. and S. N. Durlauf (2001b): "Discrete Choice with Social Interactions," *Review of Economic Studies* 68, 235-260.

[19] Brock, W. A. and S. N. Durlauf (2007): "Identification of Binary Choice Models with Social Interactions," *Journal of Econometrics* 140, 52-75.

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

[21] Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer Effects and Social Networks in Education," *Review of Economic Studies* 76, 1239-1267.

[22] Cameron, A. C., J. B. Gelbach, D. L. Miller (2008): "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics* 90, 414-427.

[23] Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica* 75, 1243-1284.

[24] Chernozhukov, V. S. Lee, and A. Rosen (2008): "Intersection Bounds: Estimation and Inference," *Econometrica* 81, 667-737.

[25] Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791-1828.

[26] Conley, T. (1999): "GMM Estimation with Cross-Sectional Dependence," *Journal of Econometrics* 92, 1-45.

[27] Dawid, A. P. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B.* 41, 1-31.

[28] de Paula, A. and X. Tang (2011): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica* 80, 143-172.

[29] Deb, J. and E. Kalai (2010): "Stability in Large Bayesian Games with Heterogeneous Players," Working Paper.

[30] Hansen, C. B. (2007): "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T is Large," *Journal of Econometrics* 141, 597-620.

[31] Ibragimov, R. and U. K. Müller (2010: "t-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics* 28, 453-468.

[32] Ioannides, Y. M. (2013): *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton University Press, New Jersey.

[33] Kalai, E. (2004): "Large Robust Games," *Econometrica* 72, 1631-1665.

[34] Kim, J. and Y.-K. Che (2004): "Asymmetric Information about Rivals' Types in Standard Auctions," *Games and Economic Behavior* 46, 383-397.

[35] Krauth, B. (2006): "Simulation-Based Estimation of Peer Effects," *Journal of Econometrics*, 133, 243-271.

[36] Kuersteiner, G. M. and I. R. Prucha (2012): "Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity," Forthcoming in *Journal of Econometrics*.

[37] McAdams, D. (2003): "Isotone Equilibrium in Games of Incomplete Information," *Econometrica* 71, 1191–1214.

[38] McDiarmid, C. (1989): "On the Method of Bounded Differences," In *Surveys in Combinatorics*, Cambridge University Press, Cambridge.

[39] Menzel, K. (2012): "Inference for Large Games with Exchangeable Players," Working Paper.

[40] Milgrom, P. R. and R. J. Weber (1985): "Distributional Strategies for Games with Incomplete Information," *Mathematics of Operations Research*, 10, 619-632.

[41] Phillips, P. C. B . and D. Sul (2003): "Dynamic Panel Estimation and Homogeneity Testing under Cross-Section Dependence," *Econometrics Journal* 6, 217-259.

[42] Reny, P. (2011): "On the Existence of Monotone Pure-Strategy Equilibria in Bayesian Games," *Econometrica* 79, 499-553.

[43] Rosen, A. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics* 146, 107-117.

[44] Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies* 70, 147-165.

[45] Wilson, R. (1985): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley. Cambridge, UK: Cambridge University Press, Chap. 2, 33-70.

[46] Xu, H. (2012): "Social Interactions: A Game-Theoretic Approach," Working Paper.

Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada

E-mail address: kysong@mail.ubc.ca